MongoDB Documentation

Release 3.2.5

MongoDB, Inc.

April 25, 2016

Contents

1 Introduction to MongoDB **3**
- 1.1 Document Database . . . 3
- 1.2 Key Features . . . 4

2 Install MongoDB **21**
- 2.1 Supported Platforms . . . 21
- 2.2 Deprecation of 32-bit Versions . . . 21
- 2.3 Tutorials . . . 22
- 2.4 Additional Resources . . . 75

3 The mongo Shell **77**
- 3.1 Introduction . . . 77
- 3.2 Start the mongo Shell . . . 77
- 3.3 Working with the mongo Shell . . . 78
- 3.4 Tab Completion and Other Keyboard Shortcuts . . . 79
- 3.5 Exit the Shell . . . 80

4 MongoDB CRUD Operations **97**
- 4.1 MongoDB CRUD Introduction . . . 97
- 4.2 MongoDB CRUD Concepts . . . 99
- 4.3 MongoDB CRUD Tutorials . . . 136
- 4.4 MongoDB CRUD Reference . . . 178

5 Aggregation **195**
- 5.1 Aggregation Pipeline . . . 195
- 5.2 Map-Reduce . . . 197
- 5.3 Single Purpose Aggregation Operations . . . 198
- 5.4 Additional Features and Behaviors . . . 198
- 5.5 Additional Resources . . . 237

6 Text Search **239**
- 6.1 Overview . . . 239
- 6.2 Example . . . 239
- 6.3 Language Support . . . 240

7 Data Models **247**
- 7.1 Data Modeling Introduction . . . 247
- 7.2 Document Validation . . . 250
- 7.3 Data Modeling Concepts . . . 252

7.4 Data Model Examples and Patterns . . . 258
7.5 Data Model Reference . . . 276

8 Administration **281**
8.1 Administration Concepts . . . 281
8.2 Administration Tutorials . . . 318
8.3 Administration Reference . . . 372
8.4 Production Checklist . . . 386

9 Security **391**
9.1 Security Checklist . . . 391
9.2 Authentication . . . 393
9.3 Role-Based Access Control . . . 433
9.4 Encryption . . . 449
9.5 Auditing . . . 466
9.6 Security Hardening . . . 472
9.7 Implement Field Level Redaction . . . 482
9.8 Security Reference . . . 484
9.9 Create a Vulnerability Report . . . 512
9.10 Additional Resources . . . 514

10 Indexes **515**
10.1 Default `_id` Index . . . 515
10.2 Create an Index . . . 516
10.3 Index Types . . . 516
10.4 Index Properties . . . 518
10.5 Index Use . . . 519
10.6 Covered Queries . . . 519
10.7 Index Intersection . . . 520
10.8 Restrictions . . . 520
10.9 Additional Considerations . . . 520
10.10 Additional Resources . . . 593

11 Storage **595**
11.1 Storage Engines . . . 595
11.2 Journaling . . . 606
11.3 GridFS . . . 611
11.4 FAQ: MongoDB Storage . . . 615

12 Replication **623**
12.1 Replication Introduction . . . 623
12.2 Replication Concepts . . . 627
12.3 Replica Set Tutorials . . . 665
12.4 Replication Reference . . . 716

13 Sharding **733**
13.1 Sharding Introduction . . . 733
13.2 Sharding Concepts . . . 739
13.3 Sharded Cluster Tutorials . . . 764
13.4 Sharding Reference . . . 822

14 Frequently Asked Questions **831**
14.1 FAQ: MongoDB Fundamentals . . . 831
14.2 FAQ: Indexes . . . 833
14.3 FAQ: Concurrency . . . 835

14.4 FAQ: Sharding with MongoDB . . . 841
14.5 FAQ: Replication and Replica Sets . . . 846
14.6 FAQ: MongoDB Storage . . . 850
14.7 FAQ: MongoDB Diagnostics . . . 856

15 Release Notes **865**
15.1 Current Stable Release . . . 865
15.2 Previous Stable Releases . . . 911
15.3 MongoDB Version Numbers . . . 1070

16 About MongoDB Documentation **1073**
16.1 License . . . 1073
16.2 Editions . . . 1073
16.3 Version and Revisions . . . 1074
16.4 Report an Issue or Make a Change Request . . . 1074
16.5 Contribute to the Documentation . . . 1075

Note: This version of the PDF does *not* include the reference section, see MongoDB Reference Manual[1] for a PDF edition of all MongoDB Reference Material.

[1] http://docs.mongodb.org/master/MongoDB-reference-manual.pdf

CHAPTER 1

Introduction to MongoDB

On this page

- Document Database (page 3)
- Key Features (page 4)

MongoDB is an open-source document database that provides high performance, high availability, and automatic scaling.

1.1 Document Database

A record in MongoDB is a document, which is a data structure composed of field and value pairs. MongoDB documents are similar to JSON objects. The values of fields may include other documents, arrays, and arrays of documents.

The advantages of using documents are:

- Documents (i.e. objects) correspond to native data types in many programming languages.
- Embedded documents and arrays reduce need for expensive joins.
- Dynamic schema supports fluent polymorphism.

1.2 Key Features

1.2.1 High Performance

MongoDB provides high performance data persistence. In particular,

- Support for embedded data models reduces I/O activity on database system.
- Indexes support faster queries and can include keys from embedded documents and arrays.

1.2.2 Rich Query Language

MongoDB supports a rich query language to support *read and write operations* (page 97) as well as:

- *data aggregation* (page 199)
- `Text Search` and *Geospatial Queries* (page 545).

1.2.3 High Availability

MongoDB's replication facility, called *replica set* (page 623), provides:

- *automatic* failover and
- data redundancy.

A *replica set* (page 623) is a group of MongoDB servers that maintain the same data set, providing redundancy and increasing data availability.

1.2.4 Horizontal Scalability

MongoDB provides horizontal scalability as part of its *core* functionality:

- *Sharding* (page 733) distributes data across a cluster of machines.
- Tag aware sharding allows for directing data to specific shards, such as to take into consideration geographic distribution of the shards.

1.2.5 Support for Multiple Storage Engines

MongoDB supports *multiple storage engines* (page 595), such as:

- *WiredTiger Storage Engine* (page 595) and
- *MMAPv1 Storage Engine* (page 603).

In addition, MongoDB provides pluggable storage engine API that allows third parties to develop storage engines for MongoDB.

Databases and Collections

On this page

- Databases (page 5)
- Collections (page 5)

MongoDB stores *BSON documents* (page 8), i.e. data records, in *collections*; the collections in databases.

Databases

In MongoDB, databases hold collections of documents.

To select a database to use, in the `mongo` shell, issue the `use <db>` statement, as in the following example:

```
use myDB
```

Create a Database If a database does not exist, MongoDB creates the database when you first store data for that database. As such, you can switch to a non-existent database and perform the following operation in the `mongo` shell:

```
use myNewDB

db.myNewCollection1.insert( { x: 1 } )
```

The `insert()` operation creates both the database `myNewDB` and the collection `myNewCollection1` if they do not already exist.

For a list of restrictions on database names, see *restrictions-on-db-names*.

Collections

MongoDB stores documents in collections. Collections are analogous to tables in relational databases.

Create a Collection If a collection does not exist, MongoDB creates the collection when you first store data for that collection.

```
db.myNewCollection2.insert( { x: 1 } )
db.myNewCollection3.createIndex( { y: 1 } )
```

Both the `insert()` and the `createIndex()` operations create their respective collection if they do not already exist.

For a list of restrictions on database names, see *restrictions-on-collection-names*.

Explicit Creation MongoDB provides the `db.createCollection()` method to explicitly create a collection with various options, such as setting the maximum size or the documentation validation rules. If you are not specifying these options, you do not need to explicitly create the collection since MongoDB creates new collections when you first store data for the collections.

To modify these collection options, see `collMod`.

Document Validation New in version 3.2.

By default, a collection does not require its documents to have the same schema; i.e. the documents in a single collection do not need to have the same set of fields and the data type for a field can differ across documents within a collection.

Starting in MongoDB 3.2, however, you can enforce *document validation rules* (page 250) for a collection during update and insert operations. See *Document Validation* (page 250) for details.

Modifying Document Structure To change the structure of the documents in a collection, such as add new fields, remove existing fields, or change the field values to a new type, update the documents to the new structure.

Capped Collections

On this page

- Overview (page 6)
- Behavior (page 6)
- Restrictions and Recommendations (page 6)
- Procedures (page 7)

Overview *Capped collections* are fixed-size collections that support high-throughput operations that insert and retrieve documents based on insertion order. Capped collections work in a way similar to circular buffers: once a collection fills its allocated space, it makes room for new documents by overwriting the oldest documents in the collection.

See `createCollection()` or `create` for more information on creating capped collections.

Behavior

Insertion Order Capped collections guarantee preservation of the insertion order. As a result, queries do not need an index to return documents in insertion order. Without this indexing overhead, capped collections can support higher insertion throughput.

Automatic Removal of Oldest Documents To make room for new documents, capped collections automatically remove the oldest documents in the collection without requiring scripts or explicit remove operations.

For example, the *oplog.rs* collection that stores a log of the operations in a *replica set* uses a capped collection. Consider the following potential use cases for capped collections:

- Store log information generated by high-volume systems. Inserting documents in a capped collection without an index is close to the speed of writing log information directly to a file system. Furthermore, the built-in *first-in-first-out* property maintains the order of events, while managing storage use.
- Cache small amounts of data in a capped collections. Since caches are read rather than write heavy, you would either need to ensure that this collection *always* remains in the working set (i.e. in RAM) *or* accept some write penalty for the required index or indexes.

`_id` Index Capped collections have an `_id` field and an index on the `_id` field by default.

Restrictions and Recommendations

Updates If you plan to update documents in a capped collection, create an index so that these update operations do not require a collection scan.

Document Size Changed in version 3.2.

If an update or a replacement operation changes the document size, the operation will fail.

Document Deletion You cannot delete documents from a capped collection. To remove all documents from a collection, use the `drop()` method to drop the collection and recreate the capped collection.

Sharding You cannot shard a capped collection.

Query Efficiency Use natural ordering to retrieve the most recently inserted elements from the collection efficiently. This is (somewhat) analogous to tail on a log file.

Aggregation `$out` The aggregation pipeline operator `$out` cannot write results to a capped collection.

Procedures

Create a Capped Collection You must create capped collections explicitly using the `db.createCollection()` method, which is a helper in the `mongo` shell for the `create` command. When creating a capped collection you must specify the maximum size of the collection in bytes, which MongoDB will pre-allocate for the collection. The size of the capped collection includes a small amount of space for internal overhead.

```
db.createCollection( "log", { capped: true, size: 100000 } )
```

If the `size` field is less than or equal to 4096, then the collection will have a cap of 4096 bytes. Otherwise, MongoDB will raise the provided size to make it an integer multiple of 256.

Additionally, you may also specify a maximum number of documents for the collection using the `max` field as in the following document:

```
db.createCollection("log", { capped : true, size : 5242880, max : 5000 } )
```

Important: The `size` argument is *always* required, even when you specify `max` number of documents. MongoDB will remove older documents if a collection reaches the maximum size limit before it reaches the maximum document count.

See

`db.createCollection()` and `create`.

Query a Capped Collection If you perform a `find()` on a capped collection with no ordering specified, MongoDB guarantees that the ordering of results is the same as the insertion order.

To retrieve documents in reverse insertion order, issue `find()` along with the `sort()` method with the `$natural` parameter set to `-1`, as shown in the following example:

```
db.cappedCollection.find().sort( { $natural: -1 } )
```

Check if a Collection is Capped Use the `isCapped()` method to determine if a collection is capped, as follows:

```
db.collection.isCapped()
```

Convert a Collection to Capped You can convert a non-capped collection to a capped collection with the `convertToCapped` command:

```
db.runCommand({"convertToCapped": "mycoll", size: 100000});
```

The `size` parameter specifies the size of the capped collection in bytes.

> **Warning:** This command obtains a global write lock and will block other operations until it has completed.

Automatically Remove Data After a Specified Period of Time For additional flexibility when expiring data, consider MongoDB's *TTL* indexes, as described in *Expire Data from Collections by Setting TTL* (page 567). These indexes allow you to expire and remove data from normal collections using a special type, based on the value of a date-typed field and a TTL value for the index.

TTL Collections (page 567) are not compatible with capped collections.

Tailable Cursor You can use a *tailable cursor* with capped collections. Similar to the Unix `tail -f` command, the tailable cursor "tails" the end of a capped collection. As new documents are inserted into the capped collection, you can use the tailable cursor to continue retrieving documents.

See *Create Tailable Cursor* (page 172) for information on creating a tailable cursor.

Documents

On this page

- Document Structure (page 8)
- Dot Notation (page 9)
- Document Limitations (page 10)
- Other Uses of the Document Structure (page 11)
- Additional Resources (page 12)

MongoDB stores data records as BSON documents. BSON is a binary representation of *JSON* documents, though it contains more data types than JSON. For the BSON spec, see bsonspec.org[1]. See also *BSON Types* (page 12).

Document Structure

MongoDB documents are composed of field-and-value pairs and have the following structure:

[1] http://bsonspec.org/

```
{
   field1: value1,
   field2: value2,
   field3: value3,
   ...
   fieldN: valueN
}
```

The value of a field can be any of the BSON *data types* (page 12), including other documents, arrays, and arrays of documents. For example, the following document contains values of varying types:

```
var mydoc = {
               _id: ObjectId("5099803df3f4948bd2f98391"),
               name: { first: "Alan", last: "Turing" },
               birth: new Date('Jun 23, 1912'),
               death: new Date('Jun 07, 1954'),
               contribs: [ "Turing machine", "Turing test", "Turingery" ],
               views : NumberLong(1250000)
             }
```

The above fields have the following data types:

- `_id` holds an *ObjectId* (page 14).
- `name` holds an *embedded document* that contains the fields `first` and `last`.
- `birth` and `death` hold values of the *Date* type.
- `contribs` holds an *array of strings*.
- `views` holds a value of the *NumberLong* type.

Field Names Field names are strings.

Documents (page 8) have the following restrictions on field names:

- The field name `_id` is reserved for use as a primary key; its value must be unique in the collection, is immutable, and may be of any type other than an array.
- The field names **cannot** start with the dollar sign (`$`) character.
- The field names **cannot** contain the dot (`.`) character.
- The field names **cannot** contain the `null` character.

BSON documents may have more than one field with the same name. Most `MongoDB interfaces`, however, represent MongoDB with a structure (e.g. a hash table) that does not support duplicate field names. If you need to manipulate documents that have more than one field with the same name, see the `driver documentation` for your driver.

Some documents created by internal MongoDB processes may have duplicate fields, but *no* MongoDB process will *ever* add duplicate fields to an existing user document.

Field Value Limit For *indexed collections* (page 515), the values for the indexed fields have a `Maximum Index Key Length` limit. See `Maximum Index Key Length` for details.

Dot Notation

MongoDB uses the *dot notation* to access the elements of an array and to access the fields of an embedded document.

Arrays To specify or access an element of an array by the zero-based index position, concatenate the array name with the dot (`.`) and zero-based index position, and enclose in quotes:

```
"<array>.<index>"
```

For example, given the following field in a document:

```
{
   ...
   contribs: [ "Turing machine", "Turing test", "Turingery" ],
   ...
}
```

To specify the third element in the `contribs` array, use the dot notation `"contribs.2"`.

See also:

- `$` positional operator for update operations,
- `$` projection operator when array index position is unknown
- *Arrays* (page 143) for dot notation examples with arrays.

Embedded Documents To specify or access a field of an embedded document with dot notation, concatenate the embedded document name with the dot (`.`) and the field name, and enclose in quotes:

```
"<embedded document>.<field>"
```

For example, given the following field in a document:

```
{
   ...
   name: { first: "Alan", last: "Turing" },
   ...
}
```

To specify the field named `last` in the `name` field, use the dot notation `"name.last"`.

See also:

Embedded Documents (page 142) for dot notation examples with embedded documents.

Document Limitations

Documents have the following attributes:

Document Size Limit The maximum BSON document size is 16 megabytes.

The maximum document size helps ensure that a single document cannot use excessive amount of RAM or, during transmission, excessive amount of bandwidth. To store documents larger than the maximum size, MongoDB provides the GridFS API. See `mongofiles` and the documentation for your `driver` for more information about GridFS.

Document Field Order MongoDB preserves the order of the document fields following write operations *except* for the following cases:

- The `_id` field is always the first field in the document.
- Updates that include `renaming` of field names may result in the reordering of fields in the document.

Changed in version 2.6: Starting in version 2.6, MongoDB actively attempts to preserve the field order in a document. Before version 2.6, MongoDB did not actively preserve the order of the fields in a document.

The `_id` Field The `_id` field has the following behavior and constraints:

- By default, MongoDB creates a unique index on the `_id` field during the creation of a collection.
- The `_id` field is always the first field in the documents. If the server receives a document that does not have the `_id` field first, then the server will move the field to the beginning.
- The `_id` field may contain values of any *BSON data type* (page 12), other than an array.

> **Warning:** To ensure functioning replication, do not store values that are of the BSON regular expression type in the `_id` field.

The following are common options for storing values for `_id`:

- Use an *ObjectId* (page 14).
- Use a natural unique identifier, if available. This saves space and avoids an additional index.
- Generate an auto-incrementing number. See *Create an Auto-Incrementing Sequence Field* (page 173).
- Generate a UUID in your application code. For a more efficient storage of the UUID values in the collection and in the `_id` index, store the UUID as a value of the BSON `BinData` type.

  Index keys that are of the `BinData` type are more efficiently stored in the index if:

  - the binary subtype value is in the range of 0-7 or 128-135, and
  - the length of the byte array is: 0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 20, 24, or 32.
- Use your driver's BSON UUID facility to generate UUIDs. Be aware that driver implementations may implement UUID serialization and deserialization logic differently, which may not be fully compatible with other drivers. See your driver documentation[2] for information concerning UUID interoperability.

Note: Most MongoDB driver clients will include the `_id` field and generate an `ObjectId` before sending the insert operation to MongoDB; however, if the client sends a document without an `_id` field, the `mongod` will add the `_id` field and generate the `ObjectId`.

Other Uses of the Document Structure

In addition to defining data records, MongoDB uses the document structure throughout, including but not limited to:

- *Query filters or specifications* (page 140). Query filter documents specify the conditions that determine which records to select for read, update, and delete operations.

  You can use `<field>:<value>` expressions to specify the equality condition and `query operator` expressions.

  ```
  {
     <field1>: <value1>,
     <field2>: { <operator>: <value> },
     ...
  }
  ```

  For examples, see *Query filters or specifications* (page 140).

[2] https://api.mongodb.org/

- *Update specifications* (page 148). Update specifications documents use *update operators* to specify the data modifications to perform on specific fields during an `db.collection.update()` operation.

```
{
   <operator1>: { <field1>: <value1>, ... },
   <operator2>: { <field2>: <value2>, ... },
   ...
}
```

- *Index specifications* (page 515). Index specifications document define the field to index and the index type:

```
{ <field1>: <type1>, <field2>: <type2>, ...  }
```

Additional Resources

- Thinking in Documents Part 1 (Blog Post)[3]

BSON Types

On this page

- Comparison/Sort Order (page 13)
- ObjectId (page 14)
- String (page 14)
- Timestamps (page 15)
- Date (page 15)

BSON is a binary serialization format used to store documents and make remote procedure calls in MongoDB. The BSON specification is located at bsonspec.org[4].

BSON supports the following data types as values in documents. Each data type has a corresponding number and string alias that can be used with the `$type` operator to query documents by BSON type.

[3] https://www.mongodb.com/blog/post/thinking-documents-part-1?jmp=docs
[4] http://bsonspec.org/

Type	Number	Alias	Notes
Double	1	“double”	
String	2	“string”	
Object	3	“object”	
Array	4	“array”	
Binary data	5	“binData”	
Undefined	6	“undefined”	Deprecated.
ObjectId	7	“objectId”	
Boolean	8	“bool”	
Date	9	“date”	
Null	10	“null”	
Regular Expression	11	“regex”	
DBPointer	12	“dbPointer”	
JavaScript	13	“javascript”	
Symbol	14	“symbol”	
JavaScript (with scope)	15	“javascriptWithScope”	
32-bit integer	16	“int”	
Timestamp	17	“timestamp”	
64-bit integer	18	“long”	
Min key	-1	“minKey”	
Max key	127	“maxKey”	

To determine a field’s type, see *Check Types in the mongo Shell* (page 89).

If you convert BSON to JSON, see the *Extended JSON* (page 16) reference.

Comparison/Sort Order

When comparing values of different *BSON* types, MongoDB uses the following comparison order, from lowest to highest:

1. MinKey (internal type)
2. Null
3. Numbers (ints, longs, doubles)
4. Symbol, String
5. Object
6. Array
7. BinData
8. ObjectId
9. Boolean
10. Date
11. Timestamp
12. Regular Expression
13. MaxKey (internal type)

MongoDB treats some types as equivalent for comparison purposes. For instance, numeric types undergo conversion before comparison.

Changed in version 3.0.0: Date objects sort before Timestamp objects. Previously Date and Timestamp objects sorted together.

The comparison treats a non-existent field as it would an empty BSON Object. As such, a sort on the `a` field in documents `{ }` and `{ a: null }` would treat the documents as equivalent in sort order.

With arrays, a less-than comparison or an ascending sort compares the smallest element of arrays, and a greater-than comparison or a descending sort compares the largest element of the arrays. As such, when comparing a field whose value is a single-element array (e.g. `[ 1 ]`) with non-array fields (e.g. `2`), the comparison is between `1` and `2`. A comparison of an empty array (e.g. `[ ]`) treats the empty array as less than `null` or a missing field.

MongoDB sorts `BinData` in the following order:

1. First, the length or size of the data.
2. Then, by the BSON one-byte subtype.
3. Finally, by the data, performing a byte-by-byte comparison.

The following sections describe special considerations for particular BSON types.

ObjectId

ObjectIds are small, likely unique, fast to generate, and ordered. ObjectId values consists of 12-bytes, where the first four bytes are a timestamp that reflect the ObjectId's creation, specifically:

- a 4-byte value representing the seconds since the Unix epoch,
- a 3-byte machine identifier,
- a 2-byte process id, and
- a 3-byte counter, starting with a random value.

In MongoDB, documents stored in a collection require a unique *_id* field that acts as a *primary key*. If the `_id` field is unspecified in documents, MongoDB uses ObjectIds as the default value for the `_id` field; i.e. if a document does not contain a top-level `_id` field, the MongoDB driver adds the `_id` field that holds an ObjectId.

In addition, if the `mongod` receives a document to insert that does not contain an `_id` field, `mongod` will add the `_id` field that holds an ObjectId.

MongoDB clients should add an `_id` field with a unique ObjectId. Using ObjectIds for the `_id` field provides the following additional benefits:

- in the `mongo` shell, you can access the creation time of the `ObjectId`, using the `ObjectId.getTimestamp()` method.
- sorting on an `_id` field that stores `ObjectId` values is roughly equivalent to sorting by creation time.

> **Important:** The relationship between the order of `ObjectId` values and generation time is not strict within a single second. If multiple systems, or multiple processes or threads on a single system generate values, within a single second; `ObjectId` values do not represent a strict insertion order. Clock skew between clients can also result in non-strict ordering even for values because client drivers generate `ObjectId` values.

See also:

`ObjectId()`

String

BSON strings are UTF-8. In general, drivers for each programming language convert from the language's string format to UTF-8 when serializing and deserializing BSON. This makes it possible to store most international characters in

BSON strings with ease. [5] In addition, MongoDB `$regex` queries support UTF-8 in the regex string.

Timestamps

BSON has a special timestamp type for *internal* MongoDB use and is **not** associated with the regular *Date* (page 15) type. Timestamp values are a 64 bit value where:

- the first 32 bits are a `time_t` value (seconds since the Unix epoch)
- the second 32 bits are an incrementing `ordinal` for operations within a given second.

Within a single `mongod` instance, timestamp values are always unique.

In replication, the *oplog* has a `ts` field. The values in this field reflect the operation time, which uses a BSON timestamp value.

Note: The BSON timestamp type is for *internal* MongoDB use. For most cases, in application development, you will want to use the BSON date type. See *Date* (page 15) for more information.

If you insert a document containing an empty BSON timestamp in a top-level field, the MongoDB server will replace that empty timestamp with the current timestamp value. For example, if you create an insert a document with a timestamp value, as in the following operation:

```
var a = new Timestamp();

db.test.insert( { ts: a } );
```

Then, the `db.test.find()` operation will return a document that resembles the following:

```
{ "_id" : ObjectId("542c2b97bac0595474108b48"), "ts" : Timestamp(1412180887, 1) }
```

If `ts` were a field in an embedded document, the server would have left it as an empty timestamp value.

Changed in version 2.6: Previously, the server would only replace empty timestamp values in the first two fields, including `_id`, of an inserted document. Now MongoDB will replace any top-level field.

Date

BSON Date is a 64-bit integer that represents the number of milliseconds since the Unix epoch (Jan 1, 1970). This results in a representable date range of about 290 million years into the past and future.

The official BSON specification[6] refers to the BSON Date type as the *UTC datetime*.

BSON Date type is signed. [7] Negative values represent dates before 1970.

Example

Construct a Date using the `new Date()` constructor in the `mongo` shell:

```
var mydate1 = new Date()
```

[5] Given strings using UTF-8 character sets, using `sort()` on strings will be reasonably correct. However, because internally `sort()` uses the C++ `strcmp` api, the sort order may handle some characters incorrectly.

[6] http://bsonspec.org/#/specification

[7] Prior to version 2.0, `Date` values were incorrectly interpreted as *unsigned* integers, which affected sorts, range queries, and indexes on `Date` fields. Because indexes are not recreated when upgrading, please re-index if you created an index on `Date` values with an earlier version, and dates before 1970 are relevant to your application.

Example

Construct a Date using the `ISODate()` constructor in the `mongo` shell:

```
var mydate2 = ISODate()
```

Example

Return the `Date` value as string:

```
mydate1.toString()
```

Example

Return the month portion of the Date value; months are zero-indexed, so that January is month `0`:

```
mydate1.getMonth()
```

MongoDB Extended JSON

On this page

- Parsers and Supported Format (page 16)
- BSON Data Types and Associated Representations (page 17)

JSON can only represent a subset of the types supported by *BSON*. To preserve type information, MongoDB adds the following extensions to the JSON format:

- *Strict mode*. Strict mode representations of BSON types conform to the JSON RFC[8]. Any JSON parser can parse these strict mode representations as key/value pairs; however, only the MongoDB internal JSON parser recognizes the type information conveyed by the format.
- `mongo` *Shell mode*. The MongoDB internal JSON parser and the `mongo` shell can parse this mode.

The representation used for the various data types depends on the context in which the JSON is parsed.

Parsers and Supported Format

Input in Strict Mode The following can parse representations in strict mode *with* recognition of the type information.

- REST Interfaces[9]
- `mongoimport`
- `--query` option of various MongoDB tools

Other JSON parsers, including `mongo` shell and `db.eval()`, can parse strict mode representations as key/value pairs, but *without* recognition of the type information.

[8] http://www.json.org
[9] https://docs.mongodb.org/ecosystem/tools/http-interfaces

Input in `mongo` Shell Mode The following can parse representations in `mongo` shell mode *with* recognition of the type information.

- REST Interfaces[10]
- `mongoimport`
- `--query` option of various MongoDB tools
- `mongo` shell

Output in Strict mode `mongoexport` and REST and HTTP Interfaces[11] output data in *Strict mode*.

Output in `mongo` Shell Mode `bsondump` outputs in `mongo` *Shell mode*.

BSON Data Types and Associated Representations

The following presents the BSON data types and the associated representations in *Strict mode* and `mongo` *Shell mode*.

Binary
`data_binary`

Strict Mode		mongo Shell Mode
`{ "$binary": "<bindata>", "$type": "<t>" }`		`BinData ( <t>, <bindata> )`

- `<bindata>` is the base64 representation of a binary string.
- `<t>` is a representation of a single byte indicating the data type. In *Strict mode* it is a hexadecimal string, and in *Shell mode* it is an integer. See the extended bson documentation. http://bsonspec.org/spec.html

Date
`data_date`

Strict Mode		mongo Shell Mode
`{ "$date": "<date>" }`		`new Date ( <date> )`

In *Strict mode*, `<date>` is an ISO-8601 date format with a mandatory time zone field following the template `YYYY-MM-DDTHH:mm:ss.mmm<+/-Offset>`.

The MongoDB JSON parser currently does not support loading ISO-8601 strings representing dates prior to the *Unix epoch*. When formatting pre-epoch dates and dates past what your system's `time_t` type can hold, the following format is used:

```
{ "$date" : { "$numberLong" : "<dateAsMilliseconds>" } }
```

In *Shell mode*, `<date>` is the JSON representation of a 64-bit signed integer giving the number of milliseconds since epoch UTC.

[10] https://docs.mongodb.org/ecosystem/tools/http-interfaces
[11] https://docs.mongodb.org/ecosystem/tools/http-interfaces

Timestamp

data_timestamp

Strict Mode		mongo Shell Mode
`{ "$timestamp": { "t": <t>, "i": <i> } }`		`Timestamp( <t>, <i> )`

- `<t>` is the JSON representation of a 32-bit unsigned integer for seconds since epoch.
- `<i>` is a 32-bit unsigned integer for the increment.

Regular Expression

data_regex

Strict Mode		mongo Shell Mode
`{ "$regex": "<sRegex>", "$options": "<sOptions>" }`		`/<jRegex>/<jOptions>`

- `<sRegex>` is a string of valid JSON characters.
- `<jRegex>` is a string that may contain valid JSON characters and unescaped double quote (`"`) characters, but may not contain unescaped forward slash (`https://docs.mongodb.org/manual/`) characters.
- `<sOptions>` is a string containing the regex options represented by the letters of the alphabet.
- `<jOptions>` is a string that may contain only the characters 'g', 'i', 'm' and 's' (added in v1.9). Because the `JavaScript` and `mongo Shell` representations support a limited range of options, any nonconforming options will be dropped when converting to this representation.

OID

data_oid

Strict Mode		mongo Shell Mode
`{ "$oid": "<id>" }`		`ObjectId( "<id>" )`

`<id>` is a 24-character hexadecimal string.

DB Reference

data_ref

Strict Mode		mongo Shell Mode
`{ "$ref": "<name>", "$id": "<id>" }`		`DBRef("<name>", "<id>")`

- `<name>` is a string of valid JSON characters.
- `<id>` is any valid extended JSON type.

Undefined Type

data_undefined

Strict Mode		mongo Shell Mode
`{ "$undefined": true }`		`undefined`

The representation for the JavaScript/BSON undefined type.

You *cannot* use `undefined` in query documents. Consider the following document inserted into the `people` collection:

```
db.people.insert( { name : "Sally", age : undefined } )
```

The following queries return an error:

```
db.people.find( { age : undefined } )
db.people.find( { age : { $gte : undefined } } )
```

However, you can query for undefined values using `$type`, as in:

```
db.people.find( { age : { $type : 6 } } )
```

This query returns all documents for which the `age` field has value `undefined`.

MinKey

`data_minkey`

Strict Mode		mongo Shell Mode
`{ "$minKey": 1 }`		`MinKey`

The representation of the MinKey BSON data type that compares lower than all other types. See *Comparison/Sort Order* (page 13) for more information on comparison order for BSON types.

MaxKey

`data_maxkey`

Strict Mode		mongo Shell Mode
`{ "$maxKey": 1 }`		`MaxKey`

The representation of the MaxKey BSON data type that compares higher than all other types. See *Comparison/Sort Order* (page 13) for more information on comparison order for BSON types.

NumberLong New in version 2.6.

`data_numberlong`

Strict Mode		mongo Shell Mode
`{ "$numberLong": "<number>" }`		`NumberLong( "<number>" )`

`NumberLong` is a 64 bit signed integer. You must include quotation marks or it will be interpreted as a floating point number, resulting in a loss of accuracy.

For example, the following commands insert `9223372036854775807` as a `NumberLong` with and without quotation marks around the integer value:

```
db.json.insert( { longQuoted : NumberLong("9223372036854775807") } )
db.json.insert( { longUnQuoted : NumberLong(9223372036854775807) } )
```

When you retrieve the documents, the value of `longUnquoted` has changed, while `longQuoted` retains its accuracy:

```
db.json.find()
{ "_id" : ObjectId("54ee1f2d33335326d70987df"), "longQuoted" : NumberLong("9223372036854775807")
{ "_id" : ObjectId("54ee1f7433335326d70987e0"), "longUnquoted" : NumberLong("-92233720368547758
```

CHAPTER 2

Install MongoDB

On this page

- Supported Platforms (page 21)
- Deprecation of 32-bit Versions (page 21)
- Tutorials (page 22)
- Additional Resources (page 75)

This section of the manual contains tutorials on installation of MongoDB.

2.1 Supported Platforms

Platform	3.2	3.0	2.6	2.4	2.2
Amazon Linux	Y	Y	Y	Y	Y
Debian 7	Y	Y	Y	Y	Y
Fedora 8+			Y	Y	Y
RHEL/CentOS 6.2+	Y	Y	Y	Y	Y
RHEL/CentOS 7.0+	Y	Y	Y		
SLES 11	Y	Y	Y	Y	Y
SLES 12	Y				
Solaris 64-bit	Y	Y	Y	Y	Y
Ubuntu 12.04	Y	Y	Y	Y	Y
Ubuntu 14.04	Y	Y	Y		
Microsoft Azure	Y	Y	Y	Y	Y
Windows Vista/Server 2008R2/2012+	Y	Y	Y	Y	Y
OSX 10.7+	Y	Y	Y	Y	

2.2 Deprecation of 32-bit Versions

Changed in version 3.2: Starting in MongoDB 3.2, 32-bit binaries are deprecated and will be unavailable in future releases.

Changed in version 3.0: Commercial support is no longer provided for MongoDB on 32-bit platforms (Linux and Windows). See *Platform Support* (page 952).

In addition, the 32-bit versions of MongoDB have the following limitations:

- 32-bit versions of MongoDB do **not** support the WiredTiger storage engine.
- 32-bit builds disable *journaling* by default because journaling further limits the maximum amount of data that the database can store.
- When running a 32-bit build of MongoDB, the total storage size for the server, including data and indexes, is 2 gigabytes. For this reason, do not deploy MongoDB to production on 32-bit machines.

If you're running a 64-bit build of MongoDB, there's virtually no limit to storage size. For production deployments, 64-bit builds and operating systems are strongly recommended.

See also:

Blog Post: 32-bit Limitations[1]

2.3 Tutorials

2.3.1 MongoDB Community Edition

Install on Linux **(page 22)** Install MongoDB Community Edition and required dependencies on Linux.

Install on OS X **(page 42)** Install MongoDB Community Edition on OS X systems from Homebrew packages or from MongoDB archives.

Install on Windows **(page 44)** Install MongoDB Community Edition on Windows systems and optionally start MongoDB as a Windows service.

2.3.2 MongoDB Enterprise

Install on Linux **(page 50)** Install the official builds of MongoDB Enterprise on Linux-based systems.

Install on OS X **(page 68)** Install the official build of MongoDB Enterprise on OS X

Install on Windows **(page 69)** Install MongoDB Enterprise on Windows using the `.msi` installer.

Install MongoDB Community Edition

These documents provide instructions to install MongoDB Community Edition.

Install on Linux **(page 22)** Install MongoDB Community Edition and required dependencies on Linux.

Install on OS X **(page 42)** Install MongoDB Community Edition on OS X systems from Homebrew packages or from MongoDB archives.

Install on Windows **(page 44)** Install MongoDB Community Edition on Windows systems and optionally start MongoDB as a Windows service.

Install MongoDB Community Edition on Linux

On this page

- Recommended (page 23)
- Manual Installation (page 23)

[1] http://blog.mongodb.org/post/137788967/32-bit-limitations

These documents provide instructions to install MongoDB Community Edition for various Linux systems.

Note: Starting in MongoDB 3.2, 32-bit binaries are deprecated and will be unavailable in future releases.

Recommended For the best installation experience, MongoDB provides packages for popular Linux distributions. These packages, which support specific platforms and provide improved performance and TLS/SSL support, are the preferred way to run MongoDB. The following guides detail the installation process for these systems:

Install on Red Hat **(page 23)** Install MongoDB Community Edition on Red Hat Enterprise and related Linux systems using `.rpm` packages.

Install on SUSE **(page 27)** Install MongoDB Community Edition on SUSE Linux systems using `.rpm` packages.

Install on Amazon **(page 30)** Install MongoDB Community Edition on Amazon Linux AMI systems using `.rpm` packages.

Install on Ubuntu **(page 33)** Install MongoDB Community Edition on Ubuntu Linux systems using `.deb` packages.

Install on Debian **(page 36)** Install MongoDB Community Edition on Debian systems using `.deb` packages.

For systems without supported packages, refer to the Manual Installation tutorial.

Manual Installation For Linux systems without supported packages, MongoDB provides a generic Linux release. These versions of MongoDB don't include TLS/SSL, and may not perform as well as the targeted packages, but are compatible on most contemporary Linux systems. See the following guides for installation:

Install From Tarball **(page 39)** Install MongoDB Community Edition on other Linux systems from MongoDB archives.

Install MongoDB Community Edition on Red Hat Enterprise or CentOS Linux
On this page

- Overview (page 23)
- Packages (page 23)
- Init Scripts (page 24)
- Install MongoDB Community Edition (page 24)
- Run MongoDB Community Edition (page 25)
- Uninstall MongoDB Community Edition (page 27)

Overview Use this tutorial to install MongoDB Community Edition on Red Hat Enterprise Linux or CentOS Linux versions 6 and 7 using `.rpm` packages. While some of these distributions include their own MongoDB packages, the official MongoDB Community Edition packages are generally more up to date.

Platform Support

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

MongoDB 3.2 deprecates support for Red Hat Enterprise Linux 5.

Packages MongoDB provides officially supported packages in their own repository. This repository contains the following packages:

mongodb-org	A metapackage that will automatically install the four component packages listed below.
mongodb-org-server	Contains the mongod daemon and associated configuration and init scripts.
mongodb-org-mongos	Contains the mongos daemon.
mongodb-org-shell	Contains the mongo shell.
mongodb-org-tools	Contains the following MongoDB tools: mongoimport bsondump, mongodump, mongoexport, mongofiles, mongooplog, mongoperf, mongorestore, mongostat, and mongotop.

The default /etc/mongod.conf configuration file supplied by the packages have bind_ip set to 127.0.0.1 by default. Modify this setting as needed for your environment before initializing a *replica set*.

Init Scripts The mongodb-org package includes various *init scripts*, including the init script /etc/rc.d/init.d/mongod. You can use these scripts to stop, start, and restart daemon processes.

The package configures MongoDB using the /etc/mongod.conf file in conjunction with the init scripts. See the Configuration File reference for documentation of settings available in the configuration file.

As of version 3.2.5, there are no init scripts for mongos. The mongos process is used only in *sharding* (page 739). You can use the mongod init script to derive your own mongos init script for use in such environments. See the mongos reference for configuration details.

The default /etc/mongod.conf configuration file supplied by the packages have bind_ip set to 127.0.0.1 by default. Modify this setting as needed for your environment before initializing a *replica set*.

Install MongoDB Community Edition

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[2].

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

Step 1: Configure the package management system (yum). Create a /etc/yum.repos.d/mongodb-org-3.2.repo file so that you can install MongoDB directly, using yum.

Changed in version 3.0: MongoDB Linux packages are in a new repository beginning with 3.0.

For the *latest* stable release of MongoDB Use the following repository file:

```
[mongodb-org-3.2]
name=MongoDB Repository
baseurl=https://repo.mongodb.org/yum/redhat/$releasever/mongodb-org/3.2/x86_64/
gpgcheck=1
enabled=1
gpgkey=https://www.mongodb.org/static/pgp/server-3.2.asc
```

For versions of MongoDB *earlier* than 3.0 To install the packages from an earlier *release series* (page 1070), such as 2.4 or 2.6, you can specify the release series in the repository configuration. For example, to restrict your system to the 2.6 release series, create a /etc/yum.repos.d/mongodb-org-2.6.repo file to hold the following configuration information for the MongoDB 2.6 repository:

```
[mongodb-org-2.6]
name=MongoDB 2.6 Repository
baseurl=http://downloads-distro.mongodb.org/repo/redhat/os/x86_64/
```

[2] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-red-hat/

```
gpgcheck=0
enabled=1
```

You can find `.repo` files for each release in the repository itself[3]. Remember that odd-numbered minor release versions (e.g. 2.5) are development versions and are unsuitable for production use.

Step 2: Install the MongoDB packages and associated tools. When you install the packages, you choose whether to install the current release or a previous one. This step provides the commands for both.

To install the latest stable version of MongoDB, issue the following command:

```
sudo yum install -y mongodb-org
```

To install a specific release of MongoDB, specify each component package individually and append the version number to the package name, as in the following example:

```
sudo yum install -y mongodb-org-3.2.5 mongodb-org-server-3.2.5 mongodb-org-shell-3.2.5 mongodb-org-mo
```

You can specify any available version of MongoDB. However `yum` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package. To pin a package, add the following `exclude` directive to your `/etc/yum.conf` file:

```
exclude=mongodb-org,mongodb-org-server,mongodb-org-shell,mongodb-org-mongos,mongodb-org-tools
```

Run MongoDB Community Edition

Prerequisites

Configure SELinux

Important: You must configure SELinux to allow MongoDB to start on Red Hat Linux-based systems (Red Hat Enterprise Linux or CentOS Linux).

To configure SELinux, administrators have three options:

Note: All three options require `root` privileges. The first two options each requires a system reboot and may have larger implications for your deployment.

- Disable SELinux entirely by changing the `SELINUX` setting to `disabled` in `/etc/selinux/config`.

  ```
  SELINUX=disabled
  ```

- Set SELinux to `permissive` mode in `/etc/selinux/config` by changing the `SELINUX` setting to `permissive`.

  ```
  SELINUX=permissive
  ```

  Note: You can use `setenforce` to change to permissive mode; this method does not require a reboot but is **not** persistent.

- Enable access to the relevant ports (e.g. 27017) for SELinux if in `enforcing` mode. See `https://docs.mongodb.org/manual/reference/default-mongodb-port` for more information on MongoDB's default ports. For default settings, this can be accomplished by running

[3] https://repo.mongodb.org/yum/redhat/

```
semanage port -a -t mongod_port_t -p tcp 27017
```

> **Warning:** On RHEL 7.0, if you change the data path, the *default* SELinux policies will prevent `mongod` from having write access on the new data path if you do not change the security context.

You may alternatively choose not to install the SELinux packages when you are installing your Linux operating system, or choose to remove the relevant packages. This option is the most invasive and is not recommended.

Data Directories and Permissions

> **Warning:** On RHEL 7.0, if you change the data path, the *default* SELinux policies will preve
> having write access on the new data path if you do not change the security context.

The MongoDB instance stores its data files in `/var/lib/mongo` and its log files in `/var/log/mongodb` by default, and runs using the `mongod` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongo` and `/var/log/mongodb` directories to give this user access to these directories.

Procedure

Step 1: Start MongoDB. You can start the `mongod` process by issuing the following command:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully You can verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

You can optionally ensure that MongoDB will start following a system reboot by issuing the following command:

```
sudo chkconfig mongod on
```

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. You can restart the `mongod` process by issuing the following command:

```
sudo service mongod restart
```

You can follow the state of the process for errors or important messages by watching the output in the `/var/log/mongodb/mongod.log` file.

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB Community Edition To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo yum erase $(rpm -qa | grep mongodb-org)
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongo
```

Install MongoDB Community Edition on SUSE

On this page

- Overview (page 27)
- Packages (page 27)
- Init Scripts (page 28)
- Install MongoDB Community Edition (page 28)
- Run MongoDB Community Edition (page 29)
- Uninstall MongoDB Community Edition (page 30)

Overview Use this tutorial to install MongoDB Community Edition on SUSE Linux from `.rpm` packages. While SUSE distributions include their own MongoDB Community Edition packages, the official MongoDB Community Edition packages are generally more up to date.

Platform Support

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

Packages MongoDB provides officially supported packages in their own repository. This repository contains the following packages:

`mongodb-org`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-org-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-org-mongos`	Contains the `mongos` daemon.
`mongodb-org-shell`	Contains the `mongo` shell.
`mongodb-org-tools`	Contains the following MongoDB tools: `mongoimport` `bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

These packages conflict with the `mongodb`, `mongodb-server`, and `mongodb-clients` packages provided by Ubuntu.

The default `/etc/mongod.conf` configuration file supplied by the packages have `bind_ip` set to `127.0.0.1` by default. Modify this setting as needed for your environment before initializing a *replica set*.

Init Scripts The `mongodb-org` package includes various *init scripts*, including the init script `/etc/rc.d/init.d/mongod`. You can use these scripts to stop, start, and restart daemon processes.

The package configures MongoDB using the `/etc/mongod.conf` file in conjunction with the init scripts. See the `Configuration File` reference for documentation of settings available in the configuration file.

As of version 3.2.5, there are no init scripts for `mongos`. The `mongos` process is used only in *sharding* (page 739). You can use the `mongod` init script to derive your own `mongos` init script for use in such environments. See the `mongos` reference for configuration details.

Note: SUSE Linux Enterprise Server and potentially other SUSE distributions ship with virtual memory address space limited to 8 GB by default. You *must* adjust this in order to prevent virtual memory allocation failures as the database grows.

The SLES packages for MongoDB adjust these limits in the default scripts, but you will need to make this change manually if you are using custom scripts and/or the tarball release rather than the SLES packages.

Install MongoDB Community Edition

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[4].

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

Step 1: Configure the package management system (`zypper`). Add the repository so that you can install MongoDB using `zypper`.

Changed in version 3.0: MongoDB Linux packages are in a new repository beginning with 3.0.

For the *latest* stable release of MongoDB Use the following command:

```
sudo zypper addrepo --no-gpgcheck https://repo.mongodb.org/zypper/suse/$(sed -rn 's/VERSION=.*([0-9]
```

For versions of MongoDB *earlier* than 3.2 To install MongoDB packages from a previous *release series* (page 1070), such as 3.0, you can specify the release series in the repository configuration. For example, to restrict your SUSE 11 system to the 3.0 release series, use the following command:

```
sudo zypper addrepo --no-gpgcheck https://repo.mongodb.org/zypper/suse/11/mongodb-org/3.0/x86_64/ mon
```

Step 2: Install the MongoDB packages and associated tools. When you install the packages, you choose whether to install the current release or a previous one. This step provides the commands for both.

To install the latest stable version of MongoDB, issue the following command:

```
sudo zypper -n install mongodb-org
```

[4] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-suse/

To install a specific release of MongoDB, specify each component package individually and append the version number to the package name, as in the following example:

```
sudo zypper install mongodb-org-3.2.5 mongodb-org-server-3.2.5 mongodb-org-shell-3.2.5 mongodb-org-mo
```

You can specify any available version of MongoDB. However `zypper` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the packages by running the following command:

```
sudo zypper addlock mongodb-org-3.2.5 mongodb-org-server-3.2.5 mongodb-org-shell-3.2.5 mongodb-org-mo
```

Previous versions of MongoDB packages use a different repository location. Refer to the version of the documentation appropriate for your MongoDB version.

Run MongoDB Community Edition

Prerequisites The MongoDB instance stores its data files in `/var/lib/mongo` and its log files in `/var/log/mongodb` by default, and runs using the `mongod` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongo` and `/var/log/mongodb` directories to give this user access to these directories.

Procedure

Step 1: Start MongoDB. You can start the `mongod` process by issuing the following command:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully You can verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

You can optionally ensure that MongoDB will start following a system reboot by issuing the following command:

```
sudo chkconfig mongod on
```

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. You can restart the `mongod` process by issuing the following command:

```
sudo service mongod restart
```

You can follow the state of the process for errors or important messages by watching the output in the `/var/log/mongodb/mongod.log` file.

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB Community Edition To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo zypper remove $(rpm -qa | grep mongodb-org)
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongo
```

Install MongoDB Community Edition on Amazon Linux

On this page

- Overview (page 30)
- Packages (page 30)
- Init Scripts (page 31)
- Install MongoDB Community Edition (page 31)
- Run MongoDB Community Edition (page 32)
- Uninstall MongoDB Community Edition (page 33)

Overview Use this tutorial to install MongoDB Community Edition on Amazon Linux from `.rpm` packages.

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

Packages MongoDB provides officially supported packages in their own repository. This repository contains the following packages:

`mongodb-org`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-org-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-org-mongos`	Contains the `mongos` daemon.
`mongodb-org-shell`	Contains the `mongo` shell.
`mongodb-org-tools`	Contains the following MongoDB tools: `mongoimport` `bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

The default `/etc/mongod.conf` configuration file supplied by the packages have `bind_ip` set to `127.0.0.1` by default. Modify this setting as needed for your environment before initializing a *replica set*.

Init Scripts The `mongodb-org` package includes various *init scripts*, including the init script `/etc/rc.d/init.d/mongod`. You can use these scripts to stop, start, and restart daemon processes.

The package configures MongoDB using the `/etc/mongod.conf` file in conjunction with the init scripts. See the `Configuration File` reference for documentation of settings available in the configuration file.

As of version 3.2.5, there are no init scripts for `mongos`. The `mongos` process is used only in *sharding* (page 739). You can use the `mongod` init script to derive your own `mongos` init script for use in such environments. See the `mongos` reference for configuration details.

Install MongoDB Community Edition

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[5].

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

Step 1: Configure the package management system (`yum`). Create a `/etc/yum.repos.d/mongodb-org-3.2.repo` file so that you can install MongoDB directly, using `yum`.

Changed in version 3.0: MongoDB Linux packages are in a new repository beginning with 3.0.

For the *latest* stable release of MongoDB Use the following repository file:

```
[mongodb-org-3.2]
name=MongoDB Repository
baseurl=https://repo.mongodb.org/yum/amazon/2013.03/mongodb-org/3.2/x86_64/
gpgcheck=1
enabled=1
gpgkey=https://www.mongodb.org/static/pgp/server-3.2.asc
```

For versions of MongoDB *earlier* than 3.0 To install the packages from an earlier *release series* (page 1070), such as 2.4 or 2.6, you can specify the release series in the repository configuration. For example, to restrict your system to the 2.6 release series, create a `/etc/yum.repos.d/mongodb-org-2.6.repo` file to hold the following configuration information for the MongoDB 2.6 repository:

```
[mongodb-org-2.6]
name=MongoDB 2.6 Repository
baseurl=http://downloads-distro.mongodb.org/repo/redhat/os/x86_64/
gpgcheck=0
enabled=1
```

You can find `.repo` files for each release in the repository itself[6]. Remember that odd-numbered minor release versions (e.g. 2.5) are development versions and are unsuitable for production use.

[5] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-amazon/
[6] https://repo.mongodb.org/yum/amazon/

Step 2: Install the MongoDB packages and associated tools. When you install the packages, you choose whether to install the current release or a previous one. This step provides the commands for both.

To install the latest stable version of MongoDB, issue the following command:

```
sudo yum install -y mongodb-org
```

To install a specific release of MongoDB, specify each component package individually and append the version number to the package name, as in the following example:

```
sudo yum install -y mongodb-org-3.2.5 mongodb-org-server-3.2.5 mongodb-org-shell-3.2.5 mongodb-org-mo
```

You can specify any available version of MongoDB. However `yum` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package. To pin a package, add the following `exclude` directive to your `/etc/yum.conf` file:

```
exclude=mongodb-org,mongodb-org-server,mongodb-org-shell,mongodb-org-mongos,mongodb-org-tools
```

Run MongoDB Community Edition The MongoDB instance stores its data files in `/var/lib/mongo` and its log files in `/var/log/mongodb` by default, and runs using the `mongod` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongo` and `/var/log/mongodb` directories to give this user access to these directories.

Step 1: Start MongoDB. You can start the `mongod` process by issuing the following command:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully You can verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

You can optionally ensure that MongoDB will start following a system reboot by issuing the following command:

```
sudo chkconfig mongod on
```

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. You can restart the `mongod` process by issuing the following command:

```
sudo service mongod restart
```

You can follow the state of the process for errors or important messages by watching the output in the `/var/log/mongodb/mongod.log` file.

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB Community Edition To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo yum erase $(rpm -qa | grep mongodb-org)
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongo
```

Install MongoDB Community Edition on Ubuntu

On this page

- Overview (page 33)
- Packages (page 34)
- Init Scripts (page 34)
- Install MongoDB Community Edition (page 34)
- Run MongoDB Community Edition (page 35)
- Uninstall MongoDB Community Edition (page 36)

Overview Use this tutorial to install MongoDB Community Edition on LTS Ubuntu Linux systems from `.deb` packages. While Ubuntu includes its own MongoDB packages, the official MongoDB Community Edition packages are generally more up-to-date.

Platform Support

MongoDB only provides packages for 64-bit long-term support Ubuntu releases. Currently, this means 12.04 LTS (Precise Pangolin) and 14.04 LTS (Trusty Tahr). While the packages may work with other Ubuntu releases, this is not a supported configuration.

Packages MongoDB provides officially supported packages in their own repository. This repository contains the following packages:

`mongodb-org`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-org-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-org-mongos`	Contains the `mongos` daemon.
`mongodb-org-shell`	Contains the `mongo` shell.
`mongodb-org-tools`	Contains the following MongoDB tools: `mongoimport` `bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

These packages conflict with the `mongodb`, `mongodb-server`, and `mongodb-clients` packages provided by Ubuntu.

The default `/etc/mongod.conf` configuration file supplied by the packages have `bind_ip` set to `127.0.0.1` by default. Modify this setting as needed for your environment before initializing a *replica set*.

Init Scripts The `mongodb-org` package includes various *init scripts*, including the init script `/etc/init.d/mongod`. You can use these scripts to stop, start, and restart daemon processes.

The package configures MongoDB using the `/etc/mongod.conf` file in conjunction with the init scripts. See the `Configuration File` reference for documentation of settings available in the configuration file.

As of version 3.2.5, there are no init scripts for `mongos`. The `mongos` process is used only in *sharding* (page 739). You can use the `mongod` init script to derive your own `mongos` init script for use in such environments. See the `mongos` reference for configuration details.

Install MongoDB Community Edition

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[7].

MongoDB only provides packages for 64-bit long-term support Ubuntu releases. Currently, this means 12.04 LTS (Precise Pangolin) and 14.04 LTS (Trusty Tahr). While the packages may work with other Ubuntu releases, this is not a supported configuration.

Step 1: Import the public key used by the package management system. The Ubuntu package management tools (i.e. `dpkg` and `apt`) ensure package consistency and authenticity by requiring that distributors sign packages with GPG keys. Issue the following command to import the MongoDB public GPG Key[8]:

```
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv EA312927
```

Step 2: Create a list file for MongoDB. Create the `/etc/apt/sources.list.d/mongodb-org-3.2.list` list file using the command appropriate for your version of Ubuntu:

Ubuntu 12.04

```
echo "deb http://repo.mongodb.org/apt/ubuntu precise/mongodb-org/3.2 multiverse" | sudo tee /etc/apt/
```

Ubuntu 14.04

```
echo "deb http://repo.mongodb.org/apt/ubuntu trusty/mongodb-org/3.2 multiverse" | sudo tee /etc/apt/s
```

[7] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-ubuntu/
[8] https://www.mongodb.org/static/pgp/server-3.2.asc

Step 3: Reload local package database. Issue the following command to reload the local package database:

```
sudo apt-get update
```

Step 4: Install the MongoDB packages. You can install either the latest stable version of MongoDB or a specific version of MongoDB.

Install the latest stable version of MongoDB. Issue the following command:

```
sudo apt-get install -y mongodb-org
```

Install a specific release of MongoDB. To install a specific release, you must specify each component package individually along with the version number, as in the following example:

```
sudo apt-get install -y mongodb-org=3.2.5 mongodb-org-server=3.2.5 mongodb-org-shell=3.2.5 mongodb-or
```

If you only install `mongodb-org=3.2.5` and do not include the component packages, the latest version of each MongoDB package will be installed regardless of what version you specified.

Pin a specific version of MongoDB. Although you can specify any available version of MongoDB, `apt-get` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package. To pin the version of MongoDB at the currently installed version, issue the following command sequence:

```
echo "mongodb-org hold" | sudo dpkg --set-selections
echo "mongodb-org-server hold" | sudo dpkg --set-selections
echo "mongodb-org-shell hold" | sudo dpkg --set-selections
echo "mongodb-org-mongos hold" | sudo dpkg --set-selections
echo "mongodb-org-tools hold" | sudo dpkg --set-selections
```

Run MongoDB Community Edition The MongoDB instance stores its data files in `/var/lib/mongodb` and its log files in `/var/log/mongodb` by default, and runs using the `mongodb` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongodb` and `/var/log/mongodb` directories to give this user access to these directories.

Step 1: Start MongoDB. Issue the following command to start `mongod`:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully Verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. Issue the following command to restart `mongod`:

```
sudo service mongod restart
```

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB Community Edition To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo apt-get purge mongodb-org*
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongodb
```

Install MongoDB Community Edition on Debian

On this page

- Overview (page 36)
- Packages (page 37)
- Init Scripts (page 37)
- Install MongoDB Community Edition (page 37)
- Run MongoDB Community Edition (page 38)
- Uninstall MongoDB Community Edition (page 39)

Overview Use this tutorial to install MongoDB Community Edition from `.deb` packages on Debian 7 "Wheezy". While Debian includes its own MongoDB packages, the official MongoDB Community Edition packages are more up to date.

MongoDB only provides packages for 64-bit Debian "Wheezy". These packages may work with other Debian releases, but this is not a supported configuration.

Packages MongoDB provides officially supported packages in their own repository. This repository contains the following packages:

`mongodb-org`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-org-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-org-mongos`	Contains the `mongos` daemon.
`mongodb-org-shell`	Contains the `mongo` shell.
`mongodb-org-tools`	Contains the following MongoDB tools: `mongoimport` `bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

These packages conflict with the `mongodb`, `mongodb-server`, and `mongodb-clients` packages provided by Debian.

The default `/etc/mongod.conf` configuration file supplied by the packages have `bind_ip` set to `127.0.0.1` by default. Modify this setting as needed for your environment before initializing a *replica set*.

Init Scripts The `mongodb-org` package includes various *init scripts*, including the init script `/etc/init.d/mongod`. You can use these scripts to stop, start, and restart daemon processes.

The package configures MongoDB using the `/etc/mongod.conf` file in conjunction with the init scripts. See the `Configuration File` reference for documentation of settings available in the configuration file.

As of version 3.2.5, there are no init scripts for `mongos`. The `mongos` process is used only in *sharding* (page 739). You can use the `mongod` init script to derive your own `mongos` init script for use in such environments. See the `mongos` reference for configuration details.

Install MongoDB Community Edition

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[9].

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

The Debian package management tools (i.e. `dpkg` and `apt`) ensure package consistency and authenticity by requiring that distributors sign packages with GPG keys.

Step 1: Import the public key used by the package management system. The Ubuntu package management tools (i.e. `dpkg` and `apt`) ensure package consistency and authenticity by requiring that distributors sign packages with GPG keys. Issue the following command to import the MongoDB public GPG Key[10]:

```
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv EA312927
```

Step 2: Create a `/etc/apt/sources.list.d/mongodb-org-3.2.list` file for MongoDB. Create the list file using the following command:

```
echo "deb http://repo.mongodb.org/apt/debian wheezy/mongodb-org/3.2 main" | sudo tee /etc/apt/sources
```

Currently packages are only available for Debian 7 (Wheezy).

Step 3: Reload local package database. Issue the following command to reload the local package database:

[9] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-debian/
[10] https://www.mongodb.org/static/pgp/server-3.2.asc

```
sudo apt-get update
```

Step 4: Install the MongoDB packages. You can install either the latest stable version of MongoDB or a specific version of MongoDB.

Install the latest stable version of MongoDB. Issue the following command:

```
sudo apt-get install -y mongodb-org
```

Install a specific release of MongoDB. To install a specific release, you must specify each component package individually along with the version number, as in the following example:

```
sudo apt-get install -y mongodb-org=3.2.5 mongodb-org-server=3.2.5 mongodb-org-shell=3.2.5 mongodb-or
```

If you only install `mongodb-org=3.2.5` and do not include the component packages, the latest version of each MongoDB package will be installed regardless of what version you specified.

Pin a specific version of MongoDB. Although you can specify any available version of MongoDB, `apt-get` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package. To pin the version of MongoDB at the currently installed version, issue the following command sequence:

```
echo "mongodb-org hold" | sudo dpkg --set-selections
echo "mongodb-org-server hold" | sudo dpkg --set-selections
echo "mongodb-org-shell hold" | sudo dpkg --set-selections
echo "mongodb-org-mongos hold" | sudo dpkg --set-selections
echo "mongodb-org-tools hold" | sudo dpkg --set-selections
```

Run MongoDB Community Edition The MongoDB instance stores its data files in `/var/lib/mongodb` and its log files in `/var/log/mongodb` by default, and runs using the `mongodb` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongodb` and `/var/log/mongodb` directories to give this user access to these directories.

Step 1: Start MongoDB. Issue the following command to start `mongod`:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully Verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. Issue the following command to restart `mongod`:

```
sudo service mongod restart
```

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB Community Edition To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo apt-get purge mongodb-org*
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongodb
```

Install MongoDB Community Edition From Tarball

On this page

- Overview (page 39)
- Install MongoDB Community Edition (page 40)
- Run MongoDB Community Edition (page 41)

Overview Compiled versions of MongoDB Community Edition for Linux provide a simple option for installing MongoDB Community Edition for other Linux systems without supported packages.

Note: Do not use this installation method unless you have a specific need that the available *Linux Packages* (page 23) do not address.

Install MongoDB Community Edition MongoDB provides archives for both 64-bit and 32-bit (deprecated) builds of Linux. Follow the installation procedure appropriate for your system.

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[11].

Install for 64-bit Linux

Step 1: Download the binary files for the desired release of MongoDB. Download the binaries from `https://www.mongodb.org/downloads`.

For example, to download the latest release through the shell, issue the following:

```
curl -O https://fastdl.mongodb.org/linux/mongodb-linux-x86_64-3.2.5.tgz
```

Step 2: Extract the files from the downloaded archive. For example, from a system shell, you can extract through the `tar` command:

```
tar -zxvf mongodb-linux-x86_64-3.2.5.tgz
```

Step 3: Copy the extracted archive to the target directory. Copy the extracted folder to the location from which MongoDB will run.

```
mkdir -p mongodb
cp -R -n mongodb-linux-x86_64-3.2.5/ mongodb
```

Step 4: Ensure the location of the binaries is in the `PATH` variable. The MongoDB binaries are in the `bin/` directory of the archive. To ensure that the binaries are in your `PATH`, you can modify your `PATH`.

For example, you can add the following line to your shell's `rc` file (e.g. `~/.bashrc`):

```
export PATH=<mongodb-install-directory>/bin:$PATH
```

Replace `<mongodb-install-directory>` with the path to the extracted MongoDB archive.

Install for 32-bit Linux

Note: Starting in MongoDB 3.2, 32-bit binaries are deprecated and will be unavailable in future releases.

Step 1: Download the binary files for the desired release of MongoDB. Download the binaries from `https://www.mongodb.org/downloads`.

For example, to download the latest release through the shell, issue the following:

```
curl -O https://fastdl.mongodb.org/linux/mongodb-linux-i686-3.2.5.tgz
```

Step 2: Extract the files from the downloaded archive. For example, from a system shell, you can extract through the `tar` command:

[11] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-linux/

```
tar -zxvf mongodb-linux-i686-3.2.5.tgz
```

Step 3: Copy the extracted archive to the target directory. Copy the extracted folder to the location from which MongoDB will run.

```
mkdir -p mongodb
cp -R -n mongodb-linux-i686-3.2.5/ mongodb
```

Step 4: Ensure the location of the binaries is in the `PATH` variable. The MongoDB binaries are in the `bin/` directory of the archive. To ensure that the binaries are in your `PATH`, you can modify your `PATH`.

For example, you can add the following line to your shell's `rc` file (e.g. `~/.bashrc`):

```
export PATH=<mongodb-install-directory>/bin:$PATH
```

Replace `<mongodb-install-directory>` with the path to the extracted MongoDB archive.

Run MongoDB Community Edition

Step 1: Create the data directory. Before you start MongoDB for the first time, create the directory to which the `mongod` process will write data. By default, the `mongod` process uses the `/data/db` directory. If you create a directory other than this one, you must specify that directory in the `dbpath` option when starting the `mongod` process later in this procedure.

The following example command creates the default `/data/db` directory:

```
mkdir -p /data/db
```

Step 2: Set permissions for the data directory. Before running `mongod` for the first time, ensure that the user account running `mongod` has read and write permissions for the directory.

Step 3: Run MongoDB. To run MongoDB, run the `mongod` process at the system prompt. If necessary, specify the path of the `mongod` or the data directory. See the following examples.

Run without specifying paths If your system `PATH` variable includes the location of the `mongod` binary and if you use the default data directory (i.e., `/data/db`), simply enter `mongod` at the system prompt:

```
mongod
```

Specify the path of the `mongod` If your `PATH` does not include the location of the `mongod` binary, enter the full path to the `mongod` binary at the system prompt:

```
<path to binary>/mongod
```

Specify the path of the data directory If you do not use the default data directory (i.e., `/data/db`), specify the path to the data directory using the *`--dbpath`* option:

```
mongod --dbpath <path to data directory>
```

Step 4: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Install MongoDB Community Edition on OS X

On this page

- Overview (page 42)
- Install MongoDB Community Edition (page 42)
- Run MongoDB (page 43)

Overview Use this tutorial to install MongoDB Community Edition on OS X systems.

Platform Support

Starting in version 3.0, MongoDB only supports OS X versions 10.7 (Lion) and later on Intel x86-64.

MongoDB Community Edition is available through the popular OS X package manager Homebrew[12] or through the MongoDB Download site[13].

Install MongoDB Community Edition

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[14].

You can install MongoDB Community Edition with Homebrew[15] or manually. This section describes both methods.

Install MongoDB Community Edition with Homebrew Homebrew[16] installs binary packages based on published "formulae." This section describes how to update `brew` to the latest packages and install MongoDB Community Edition. Homebrew requires some initial setup and configuration, which is beyond the scope of this document.

Step 1: Update Homebrew's package database. In a system shell, issue the following command:

```
brew update
```

Step 2: Install MongoDB. You can install MongoDB via `brew` with several different options. Use one of the following operations:

Install the MongoDB Binaries To install the MongoDB binaries, issue the following command in a system shell:

```
brew install mongodb
```

[12] http://brew.sh/
[13] http://www.mongodb.org/downloads
[14] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-os-x/
[15] http://brew.sh/
[16] http://brew.sh/

Build MongoDB from Source with TLS/SSL Support To build MongoDB from the source files and include TLS/SSL support, issue the following from a system shell:

```
brew install mongodb --with-openssl
```

Install the Latest Development Release of MongoDB To install the latest development release for use in testing and development, issue the following command in a system shell:

```
brew install mongodb --devel
```

Install MongoDB Community Edition Manually Only install MongoDB Community Edition using this procedure if you cannot use *homebrew* (page 42).

Step 1: Download the binary files for the desired release of MongoDB. Download the binaries from `https://www.mongodb.org/downloads`.

For example, to download the latest release through the shell, issue the following:

```
curl -O https://fastdl.mongodb.org/osx/mongodb-osx-x86_64-3.2.5.tgz
```

Step 2: Extract the files from the downloaded archive. For example, from a system shell, you can extract through the `tar` command:

```
tar -zxvf mongodb-osx-x86_64-3.2.5.tgz
```

Step 3: Copy the extracted archive to the target directory. Copy the extracted folder to the location from which MongoDB will run.

```
mkdir -p mongodb
cp -R -n mongodb-osx-x86_64-3.2.5/ mongodb
```

Step 4: Ensure the location of the binaries is in the `PATH` variable. The MongoDB binaries are in the `bin/` directory of the archive. To ensure that the binaries are in your `PATH`, you can modify your `PATH`.

For example, you can add the following line to your shell's `rc` file (e.g. `~/.bashrc`):

```
export PATH=<mongodb-install-directory>/bin:$PATH
```

Replace `<mongodb-install-directory>` with the path to the extracted MongoDB archive.

Run MongoDB

Step 1: Create the data directory. Before you start MongoDB for the first time, create the directory to which the `mongod` process will write data. By default, the `mongod` process uses the `/data/db` directory. If you create a directory other than this one, you must specify that directory in the `dbpath` option when starting the `mongod` process later in this procedure.

The following example command creates the default `/data/db` directory:

```
mkdir -p /data/db
```

Step 2: Set permissions for the data directory. Before running `mongod` for the first time, ensure that the user account running `mongod` has read and write permissions for the directory.

Step 3: Run MongoDB. To run MongoDB, run the `mongod` process at the system prompt. If necessary, specify the path of the `mongod` or the data directory. See the following examples.

Run without specifying paths If your system `PATH` variable includes the location of the `mongod` binary and if you use the default data directory (i.e., `/data/db`), simply enter `mongod` at the system prompt:

```
mongod
```

Specify the path of the `mongod` If your `PATH` does not include the location of the `mongod` binary, enter the full path to the `mongod` binary at the system prompt:

```
<path to binary>/mongod
```

Specify the path of the data directory If you do not use the default data directory (i.e., `/data/db`), specify the path to the data directory using the *`--dbpath`* option:

```
mongod --dbpath <path to data directory>
```

Step 4: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Install MongoDB Community Edition on Windows

On this page

- Overview (page 44)
- Requirements (page 45)
- Get MongoDB Community Edition (page 45)
- Install MongoDB Community Edition (page 45)
- Run MongoDB Community Edition (page 46)
- Configure a Windows Service for MongoDB Community Edition (page 47)
- Manually Create a Windows Service for MongoDB Community Edition (page 48)
- Additional Resources (page 49)

Overview Use this tutorial to install MongoDB Community Edition on Windows systems.

Platform Support

Starting in version 2.2, MongoDB does not support Windows XP. Please use a more recent version of Windows to use more recent releases of MongoDB.

Important: If you are running any edition of Windows Server 2008 R2 or Windows 7, please install a hotfix to

resolve an issue with memory mapped files on Windows[17].

Requirements MongoDB Community Edition requires Windows Server 2008 R2, Windows Vista, or later. The `.msi` installer includes all other software dependencies and will automatically upgrade any older version of MongoDB installed using an `.msi` file.

Get MongoDB Community Edition

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[18].

Step 1: Determine which MongoDB build you need. The following MongoDB builds are available for Windows:

MongoDB for Windows 64-bit runs only on Windows Server 2008 R2, Windows 7 64-bit, and newer versions of Windows. This build takes advantage of recent enhancements to the Windows Platform and cannot operate on older versions of Windows.

MongoDB for Windows 64-bit Legacy runs on Windows Vista, and Windows Server 2008 and does not include recent performance enhancements.

To find which version of Windows you are running, enter the following commands in the *Command Prompt* or *Powershell*:

```
wmic os get caption
wmic os get osarchitecture
```

Step 2: Download MongoDB for Windows. Download the latest production release of MongoDB from the MongoDB downloads page[19]. Ensure you download the correct version of MongoDB for your Windows system. The 64-bit versions of MongoDB do not work with 32-bit Windows.

Install MongoDB Community Edition

Interactive Installation

Step 1: Install MongoDB for Windows. In Windows Explorer, locate the downloaded MongoDB `.msi` file, which typically is located in the default `Downloads` folder. Double-click the `.msi` file. A set of screens will appear to guide you through the installation process.

You may specify an installation directory if you choose the "Custom" installation option.

Note: These instructions assume that you have installed MongoDB to `C:\mongodb`.

MongoDB is self-contained and does not have any other system dependencies. You can run MongoDB from any folder you choose. You may install MongoDB in any folder (e.g. `D:\test\mongodb`).

Unattended Installation You may install MongoDB Community unattended on Windows from the command line using `msiexec.exe`.

[17] http://support.microsoft.com/kb/2731284
[18] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-on-windows/
[19] http://www.mongodb.org/downloads

Step 1: Open an Administrator command prompt. Press the `Win` key, type `cmd.exe`, and press `Ctrl + Shift + Enter` to run the *Command Prompt* as Administrator.

Execute the remaining steps from the Administrator command prompt.

Step 2: Install MongoDB for Windows. Change to the directory containing the `.msi` installation binary of your choice and invoke:

```
msiexec.exe /q /i mongodb-win32-x86_64-2008plus-ssl-3.2.5-signed.msi ^
            INSTALLLOCATION="C:\mongodb" ^
            ADDLOCAL="all"
```

You can specify the installation location for the executable by modifying the `INSTALLLOCATION` value.

By default, this method installs all MongoDB binaries. To install specific MongoDB component sets, you can specify them in the `ADDLOCAL` argument using a comma-separated list including one or more of the following component sets:

Component Set	Binaries
`Server`	`mongod.exe`
`Router`	`mongos.exe`
`Client`	`mongo.exe`
`MonitoringTools`	`mongostat.exe`, `mongotop.exe`
`ImportExportTools`	`mongodump.exe`, `mongorestore.exe`, `mongoexport.exe`, `mongoimport.exe`
`MiscellaneousTools`	`bsondump.exe`, `mongofiles.exe`, `mongooplog.exe`, `mongoperf.exe`

For instance, to install *only* the MongoDB utilities, invoke:

```
msiexec.exe /q /i mongodb-win32-x86_64-2008plus-ssl-3.2.5-signed.msi ^
            INSTALLLOCATION="C:\mongodb" ^
            ADDLOCAL="MonitoringTools,ImportExportTools,MiscellaneousTools"
```

Run MongoDB Community Edition

Warning: Do not make `mongod.exe` visible on public networks without running in "Sec
`auth` setting. MongoDB is designed to be run in trusted environments, and the database doe
Mode" by default.

Step 1: Set up the MongoDB environment. MongoDB requires a *data directory* to store all data. MongoDB's default data directory path is `\data\db`. Create this folder using the following commands from a *Command Prompt*:

```
md \data\db
```

You can specify an alternate path for data files using the *`--dbpath`* option to `mongod.exe`, for example:

```
C:\mongodb\bin\mongod.exe --dbpath d:\test\mongodb\data
```

If your path includes spaces, enclose the entire path in double quotes, for example:

```
C:\mongodb\bin\mongod.exe --dbpath "d:\test\mongo db data"
```

You may also specify the `dbpath` in a `configuration file`.

Step 2: Start MongoDB. To start MongoDB, run `mongod.exe`. For example, from the *Command Prompt*:

```
C:\mongodb\bin\mongod.exe
```

This starts the main MongoDB database process. The `waiting for connections` message in the console output indicates that the `mongod.exe` process is running successfully.

Depending on the security level of your system, Windows may pop up a *Security Alert* dialog box about blocking "some features" of `C:\mongodb\bin\mongod.exe` from communicating on networks. All users should select `Private Networks, such as my home or work network` and click `Allow access`. For additional information on security and MongoDB, please see the *Security Documentation* (page 391).

Step 3: Connect to MongoDB. To connect to MongoDB through the `mongo.exe` shell, open another *Command Prompt*.

```
C:\mongodb\bin\mongo.exe
```

If you want to develop applications using .NET, see the documentation of C# and MongoDB[20] for more information.

Step 4: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Configure a Windows Service for MongoDB Community Edition

Step 1: Open an Administrator command prompt. Press the `Win` key, type `cmd.exe`, and press `Ctrl + Shift + Enter` to run the *Command Prompt* as Administrator.

Execute the remaining steps from the Administrator command prompt.

Step 2: Create directories. Create directories for your database and log files:

```
mkdir c:\data\db
mkdir c:\data\log
```

Step 3: Create a configuration file. Create a configuration file. The file **must** set `systemLog.path`. Include additional `configuration options` as appropriate.

For example, create a file at `C:\mongodb\mongod.cfg` that specifies both `systemLog.path` and `storage.dbPath`:

```
systemLog:
    destination: file
    path: c:\data\log\mongod.log
storage:
    dbPath: c:\data\db
```

[20] https://docs.mongodb.org/ecosystem/drivers/csharp

Step 4: Install the MongoDB service.

Important: Run all of the following commands in *Command Prompt* with "Administrative Privileges".

Install the MongoDB service by starting `mongod.exe` with the `--install` option and the `-config` option to specify the previously created configuration file.

```
"C:\mongodb\bin\mongod.exe" --config "C:\mongodb\mongod.cfg" --install
```

To use an alternate `dbpath`, specify the path in the configuration file (e.g. `C:\mongodb\mongod.cfg`) or on the command line with the *`--dbpath`* option.

If needed, you can install services for multiple instances of `mongod.exe` or `mongos.exe`. Install each service with a unique *`--serviceName`* and *`--serviceDisplayName`*. Use multiple instances only when sufficient system resources exist and your system design requires it.

Step 5: Start the MongoDB service.

```
net start MongoDB
```

Step 6: Stop or remove the MongoDB service as needed. To stop the MongoDB service use the following command:

```
net stop MongoDB
```

To remove the MongoDB service use the following command:

```
"C:\mongodb\bin\mongod.exe" --remove
```

Manually Create a Windows Service for MongoDB Community Edition You can set up the MongoDB server as a *Windows Service* that starts automatically at boot time.

The following procedure assumes you have installed MongoDB Community using the `.msi` installer with the path `C:\mongodb\`.

If you have installed in an alternative directory, you will need to adjust the paths as appropriate.

Step 1: Open an Administrator command prompt. Press the `Win` key, type `cmd.exe`, and press `Ctrl + Shift + Enter` to run the *Command Prompt* as Administrator.

Execute the remaining steps from the Administrator command prompt.

Step 2: Create directories. Create directories for your database and log files:

```
mkdir c:\data\db
mkdir c:\data\log
```

Step 3: Create a configuration file. Create a configuration file. The file **must** set `systemLog.path`. Include additional `configuration options` as appropriate.

For example, create a file at `C:\mongodb\mongod.cfg` that specifies both `systemLog.path` and `storage.dbPath`:

```
systemLog:
    destination: file
    path: c:\data\log\mongod.log
storage:
    dbPath: c:\data\db
```

Step 4: Create the MongoDB service. Create the MongoDB service.

```
sc.exe create MongoDB binPath= "C:\mongodb\bin\mongod.exe --service --config=\"C:\mongodb\mongod.cfg
```

`sc.exe` requires a space between "=" and the configuration values (eg "binPath= "), and a "\" to escape double quotes.

If successfully created, the following log message will display:

```
[SC] CreateService SUCCESS
```

Step 5: Start the MongoDB service.

```
net start MongoDB
```

Step 6: Stop or remove the MongoDB service as needed. To stop the MongoDB service, use the following command:

```
net stop MongoDB
```

To remove the MongoDB service, first stop the service and then run the following command:

```
sc.exe delete MongoDB
```

Additional Resources

- MongoDB for Developers Free Course[21]
- MongoDB for .NET Developers Free Online Course[22]
- MongoDB Architecture Guide[23]

Install MongoDB Enterprise

These documents provide instructions to install MongoDB Enterprise.

MongoDB Enterprise is available for MongoDB Enterprise subscribers and includes several additional features including support for SNMP monitoring, LDAP authentication, Kerberos authentication, and System Event Auditing.

Install on Linux **(page 50)** Install the official builds of MongoDB Enterprise on Linux-based systems.

Install on OS X **(page 68)** Install the official build of MongoDB Enterprise on OS X

Install on Windows **(page 69)** Install MongoDB Enterprise on Windows using the `.msi` installer.

[21] https://university.mongodb.com/courses/M101P/about?jmp=docs
[22] https://university.mongodb.com/courses/M101N/about?jmp=docs
[23] https://www.mongodb.com/lp/white-paper/architecture-guide?jmp=docs

Install MongoDB Enterprise on Linux

Install on Red Hat **(page 50)** Install MongoDB Enterprise and required dependencies on Red Hat Enterprise or CentOS Systems using packages.

Install on Ubuntu **(page 54)** Install MongoDB Enterprise and required dependencies on Ubuntu Linux Systems using packages.

Install on Debian **(page 57)** Install MongoDB Enterprise and required dependencies on Debian Linux Systems using packages.

Install on SUSE **(page 60)** Install MongoDB Enterprise and required dependencies on SUSE Enterprise Linux.

Install on Amazon **(page 63)** Install MongoDB Enterprise and required dependencies on Amazon Linux AMI.

Install From Tarball **(page 66)** Install MongoDB Enterprise from a tarball.

Install MongoDB Enterprise on Red Hat Enterprise or CentOS

On this page

- Overview (page 50)
- Install MongoDB Enterprise (page 50)
- Install MongoDB Enterprise From Tarball (page 51)
- Run MongoDB Enterprise (page 52)
- Uninstall MongoDB (page 53)

Overview Use this tutorial to install MongoDB Enterprise[24] on Red Hat Enterprise Linux or CentOS Linux versions 6 and 7 from `.rpm` packages.

Platform Support

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

MongoDB 3.2 deprecates support for Red Hat Enterprise Linux 5.

MongoDB provides officially supported Enterprise packages in their own repository. This repository contains the following packages:

`mongodb-enterprise`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-enterprise-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-enterprise-mongos`	Contains the `mongos` daemon.
`mongodb-enterprise-shell`	Contains the `mongo` shell.
`mongodb-enterprise-tools`	Contains the following MongoDB tools: `mongoimport` `bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

The default `/etc/mongod.conf` configuration file supplied by the packages have `bind_ip` set to `127.0.0.1` by default. Modify this setting as needed for your environment before initializing a *replica set*.

Install MongoDB Enterprise

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[25].

[24] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[25] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-enterprise-on-red-hat/

Use the provided distribution packages as described in this page if possible. These packages will automatically install all of MongoDB's dependencies, and are the recommended installation method.

Step 1: Configure repository. Create an `/etc/yum.repos.d/mongodb-enterprise.repo` file so that you can install MongoDB enterprise directly, using `yum`.

For the *latest* stable release of MongoDB Enterprise Use the following repository file:

```
[mongodb-enterprise]
name=MongoDB Enterprise Repository
baseurl=https://repo.mongodb.com/yum/redhat/$releasever/mongodb-enterprise/stable/$basearch/
gpgcheck=1
enabled=1
gpgkey=https://www.mongodb.org/static/pgp/server-3.2.asc
```

`.repo` files for each release can also be found in the repository itself[26]. Remember that odd-numbered minor release versions (e.g. 2.5) are development versions and are unsuitable for production deployment.

Step 2: Install the MongoDB Enterprise packages and associated tools. You can install either the latest stable version of MongoDB Enterprise or a specific version of MongoDB Enterprise.

To install the latest stable version of MongoDB Enterprise, issue the following command:

```
sudo yum install -y mongodb-enterprise
```

Step 3: Optional: Manage Installed Version

Install a specific release of MongoDB Enterprise. Specify each component package individually and append the version number to the package name, as in the following example that installs the `3.2.1` release of MongoDB:

```
sudo yum install -y mongodb-enterprise-3.2.1 mongodb-enterprise-server-3.2.1 mongodb-enterprise-shell
```

Pin a specific version of MongoDB Enterprise. Although you can specify any available version of MongoDB Enterprise, `yum` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package by adding the following `exclude` directive to your `/etc/yum.conf` file:

```
exclude=mongodb-enterprise,mongodb-enterprise-server,mongodb-enterprise-shell,mongodb-enterprise-mong
```

Previous versions of MongoDB packages use different naming conventions. See the 2.4 version of documentation for more information[27].

Step 4: When the install completes, you can run MongoDB.

Install MongoDB Enterprise From Tarball While you should use the `.rpm` packages as previously described, you may also manually install MongoDB using the tarballs.

First you must install any dependencies as appropriate:

Version 5

[26] https://repo.mongodb.com/yum/redhat/
[27] https://docs.mongodb.org/v2.4/tutorial/install-mongodb-on-linux

```
yum install perl cyrus-sasl cyrus-sasl-plain cyrus-sasl-gssapi krb5-libs \
            lm_sensors net-snmp openssl popt rpm-libs tcp_wrappers zlib
```

Version 6

```
yum install cyrus-sasl cyrus-sasl-plain cyrus-sasl-gssapi krb5-libs \
            net-snmp openssl
```

Version 7

```
yum install cyrus-sasl cyrus-sasl-plain cyrus-sasl-gssapi krb5-libs \
            lm_sensors-libs net-snmp-agent-libs net-snmp openssl rpm-libs \
            tcp_wrappers-libs
```

To perform the installation, see *Install MongoDB Enterprise From Tarball* (page 66).

Run MongoDB Enterprise

Prerequisites

Configure SELinux

Important: You must configure SELinux to allow MongoDB to start on Red Hat Linux-based systems (Red Hat Enterprise Linux or CentOS Linux).

To configure SELinux, administrators have three options:

Note: All three options require `root` privileges. The first two options each requires a system reboot and may have larger implications for your deployment.

- Disable SELinux entirely by changing the `SELINUX` setting to `disabled` in `/etc/selinux/config`.

  ```
  SELINUX=disabled
  ```

- Set SELinux to `permissive` mode in `/etc/selinux/config` by changing the `SELINUX` setting to `permissive`.

  ```
  SELINUX=permissive
  ```

  Note: You can use `setenforce` to change to permissive mode; this method does not require a reboot but is **not** persistent.

- Enable access to the relevant ports (e.g. 27017) for SELinux if in `enforcing` mode. See `https://docs.mongodb.org/manual/reference/default-mongodb-port` for more information on MongoDB's default ports. For default settings, this can be accomplished by running

  ```
  semanage port -a -t mongod_port_t -p tcp 27017
  ```

  Warning: On RHEL 7.0, if you change the data path, the *default* SELinux policies will prevent `mongod` from having write access on the new data path if you do not change the security context.

You may alternatively choose not to install the SELinux packages when you are installing your Linux operating system, or choose to remove the relevant packages. This option is the most invasive and is not recommended.

Data Directories and Permissions

Warning: On RHEL 7.0, if you change the data path, the *default* SELinux policies will preve having write access on the new data path if you do not change the security context.

The MongoDB instance stores its data files in `/var/lib/mongo` and its log files in `/var/log/mongodb` by default, and runs using the `mongod` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongo` and `/var/log/mongodb` directories to give this user access to these directories.

Procedure

Step 1: Start MongoDB. You can start the `mongod` process by issuing the following command:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully You can verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

You can optionally ensure that MongoDB will start following a system reboot by issuing the following command:

```
sudo chkconfig mongod on
```

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. You can restart the `mongod` process by issuing the following command:

```
sudo service mongod restart
```

You can follow the state of the process for errors or important messages by watching the output in the `/var/log/mongodb/mongod.log` file.

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

Warning: This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo yum erase $(rpm -qa | grep mongodb-enterprise)
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongo
```

Install MongoDB Enterprise on Ubuntu

On this page

- Overview (page 54)
- Install MongoDB Enterprise (page 54)
- Install MongoDB Enterprise From Tarball (page 56)
- Run MongoDB Enterprise (page 56)
- Uninstall MongoDB (page 57)

Overview Use this tutorial to install MongoDB Enterprise[28] on LTS Ubuntu Linux systems from `.deb` packages.

Platform Support

MongoDB only provides packages for 64-bit long-term support Ubuntu releases. Currently, this means 12.04 LTS (Precise Pangolin) and 14.04 LTS (Trusty Tahr). While the packages may work with other Ubuntu releases, this is not a supported configuration.

MongoDB provides officially supported Enterprise packages in their own repository. This repository contains the following packages:

`mongodb-enterprise`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-enterprise-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-enterprise-mongos`	Contains the `mongos` daemon.
`mongodb-enterprise-shell`	Contains the `mongo` shell.
`mongodb-enterprise-tools`	Contains the following MongoDB tools: `mongoimport bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

Install MongoDB Enterprise

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[29].

MongoDB only provides packages for 64-bit long-term support Ubuntu releases. Currently, this means 12.04 LTS (Precise Pangolin) and 14.04 LTS (Trusty Tahr). While the packages may work with other Ubuntu releases, this is not a supported configuration.

[28] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[29] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-enterprise-on-ubuntu/

Use the provided distribution packages as described in this page if possible. These packages will automatically install all of MongoDB's dependencies, and are the recommended installation method.

Step 1: Import the public key used by the package management system. The Ubuntu package management tools (i.e. `dpkg` and `apt`) ensure package consistency and authenticity by requiring that distributors sign packages with GPG keys. Issue the following command to import the MongoDB public GPG Key[30]:

```
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv EA312927
```

Step 2: Create a `/etc/apt/sources.list.d/mongodb-enterprise.list` file for MongoDB. Create the list file using the command appropriate for your version of Ubuntu:

Ubuntu 12.04

```
echo "deb http://repo.mongodb.com/apt/ubuntu precise/mongodb-enterprise/stable multiverse" | sudo tee
```

Ubuntu 14.04

```
echo "deb http://repo.mongodb.com/apt/ubuntu trusty/mongodb-enterprise/stable multiverse" | sudo tee
```

If you'd like to install MongoDB Enterprise packages from a particular *release series* (page 1070), such as 2.4 or 2.6, you can specify the release series in the repository configuration. For example, to restrict your system to the 2.6 release series, add the following repository:

```
echo "deb http://repo.mongodb.com/apt/ubuntu "$(lsb_release -sc)"/mongodb-enterprise/2.6 multiverse"
```

Step 3: Reload local package database. Issue the following command to reload the local package database:

```
sudo apt-get update
```

Step 4: Install the MongoDB Enterprise packages. You can install either the latest stable version of MongoDB or a specific version of MongoDB.

Install the latest stable version of MongoDB Enterprise. Issue the following command:

```
sudo apt-get install -y mongodb-enterprise
```

Install a specific release of MongoDB Enterprise. To install a specific release, you must specify each component package individually along with the version number, as in the following example:

```
sudo apt-get install -y mongodb-enterprise=3.2.5 mongodb-enterprise-server=3.2.5 mongodb-enterprise-s
```

If you only install `mongodb-enterprise=3.2.5` and do not include the component packages, the latest version of each MongoDB package will be installed regardless of what version you specified.

Pin a specific version of MongoDB Enterprise. Although you can specify any available version of MongoDB, `apt-get` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package. To pin the version of MongoDB at the currently installed version, issue the following command sequence:

[30] https://www.mongodb.org/static/pgp/server-3.2.asc

```
echo "mongodb-enterprise hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-server hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-shell hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-mongos hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-tools hold" | sudo dpkg --set-selections
```

Versions of the MongoDB packages before 2.6 use a different repository location. Refer to the version of the documentation appropriate for your MongoDB version.

Install MongoDB Enterprise From Tarball While you should use the `.deb` packages as previously described, you may also manually install MongoDB using the tarballs.

First you must install any dependencies as appropriate:

```
sudo apt-get install libgssapi-krb5-2 libsasl2-2 libssl1.0.0 libstdc++6 snmp
```

To perform the installation, see *Install MongoDB Enterprise From Tarball* (page 66).

Run MongoDB Enterprise The MongoDB instance stores its data files in `/var/lib/mongodb` and its log files in `/var/log/mongodb` by default, and runs using the `mongodb` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongodb` and `/var/log/mongodb` directories to give this user access to these directories.

Step 1: Start MongoDB. Issue the following command to start `mongod`:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully Verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. Issue the following command to restart `mongod`:

```
sudo service mongod restart
```

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo apt-get purge mongodb-enterprise*
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongodb
```

Install MongoDB Enterprise on Debian

On this page

- Overview (page 57)
- Install MongoDB Enterprise (page 58)
- Install MongoDB Enterprise From Tarball (page 59)
- Run MongoDB Enterprise (page 59)
- Uninstall MongoDB (page 60)

Overview Use this tutorial to install MongoDB Enterprise[31] from `.deb` packages on Debian 7 "Wheezy".

Platform Support

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

MongoDB provides officially supported Enterprise packages in their own repository. This repository contains the following packages:

`mongodb-enterprise`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-enterprise-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-enterprise-mongos`	Contains the `mongos` daemon.
`mongodb-enterprise-shell`	Contains the `mongo` shell.
`mongodb-enterprise-tools`	Contains the following MongoDB tools: `mongoimport` `bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

[31] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

Install MongoDB Enterprise

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[32].

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

Use the provided distribution packages as described in this page if possible. These packages will automatically install all of MongoDB's dependencies, and are the recommended installation method.

Step 1: Import the public key used by the package management system. The Ubuntu package management tools (i.e. `dpkg` and `apt`) ensure package consistency and authenticity by requiring that distributors sign packages with GPG keys. Issue the following command to import the MongoDB public GPG Key[33]:

```
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv EA312927
```

Step 2: Create a `/etc/apt/sources.list.d/mongodb-enterprise.list` file for MongoDB. Create the list file using the following command:

```
echo "deb http://repo.mongodb.com/apt/debian wheezy/mongodb-enterprise/stable main" | sudo tee /etc/a
```

If you'd like to install MongoDB Enterprise packages from a particular *release series* (page 1070), such as 2.6, you can specify the release series in the repository configuration. For example, to restrict your system to the 2.6 release series, add the following repository:

```
echo "deb http://repo.mongodb.com/apt/debian wheezy/mongodb-enterprise/2.6 main" | sudo tee /etc/apt/
```

Currently packages are only available for Debian 7 (Wheezy).

Step 3: Reload local package database. Issue the following command to reload the local package database:

```
sudo apt-get update
```

Step 4: Install the MongoDB Enterprise packages. You can install either the latest stable version of MongoDB or a specific version of MongoDB.

Install the latest stable version of MongoDB Enterprise. Issue the following command:

```
sudo apt-get install -y mongodb-enterprise
```

Install a specific release of MongoDB Enterprise. To install a specific release, you must specify each component package individually along with the version number, as in the following example:

```
sudo apt-get install -y mongodb-enterprise=3.2.5 mongodb-enterprise-server=3.2.5 mongodb-enterprise-s
```

If you only install `mongodb-enterprise=3.2.5` and do not include the component packages, the latest version of each MongoDB package will be installed regardless of what version you specified.

[32] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-enterprise-on-debian/
[33] https://www.mongodb.org/static/pgp/server-3.2.asc

Pin a specific version of MongoDB Enterprise. Although you can specify any available version of MongoDB, `apt-get` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package. To pin the version of MongoDB at the currently installed version, issue the following command sequence:

```
echo "mongodb-enterprise hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-server hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-shell hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-mongos hold" | sudo dpkg --set-selections
echo "mongodb-enterprise-tools hold" | sudo dpkg --set-selections
```

Versions of the MongoDB packages before 2.6 use a different repository location. Refer to the version of the documentation appropriate for your MongoDB version.

Install MongoDB Enterprise From Tarball While you should use the `.deb` packages as previously described, you may also manually install MongoDB using the tarballs.

First you must install any dependencies as appropriate:

```
sudo apt-get install libgssapi-krb5-2 libsasl2-2 libssl1.0.0 libstdc++6 snmp
```

To perform the installation, see *Install MongoDB Enterprise From Tarball* (page 66).

Run MongoDB Enterprise The MongoDB instance stores its data files in `/var/lib/mongodb` and its log files in `/var/log/mongodb` by default, and runs using the `mongodb` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongodb` and `/var/log/mongodb` directories to give this user access to these directories.

Step 1: Start MongoDB. Issue the following command to start `mongod`:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully Verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. Issue the following command to restart `mongod`:

```
sudo service mongod restart
```

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo apt-get purge mongodb-enterprise*
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongodb
```

Install MongoDB Enterprise on SUSE

On this page

- Overview (page 60)
- Considerations (page 61)
- Install MongoDB Enterprise (page 61)
- Install MongoDB Enterprise From Tarball (page 62)
- Run MongoDB Enterprise (page 62)
- Uninstall MongoDB (page 63)

Overview Use this tutorial to install MongoDB Enterprise[34] on SUSE Linux. MongoDB Enterprise is available on select platforms and contains support for several features related to security and monitoring.

Platform Support

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

MongoDB provides officially supported Enterprise packages in their own repository. This repository contains the following packages:

[34] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

`mongodb-enterprise`	A `metapackage` that will automatically install the four component packages listed below.
`mongodb-enterprise-server`	Contains the `mongod` daemon and associated configuration and init scripts.
`mongodb-enterprise-mongos`	Contains the `mongos` daemon.
`mongodb-enterprise-shell`	Contains the `mongo` shell.
`mongodb-enterprise-tools`	Contains the following MongoDB tools: `mongoimport` `bsondump`, `mongodump`, `mongoexport`, `mongofiles`, `mongooplog`, `mongoperf`, `mongorestore`, `mongostat`, and `mongotop`.

Considerations MongoDB only provides Enterprise packages for 64-bit builds of SUSE Enterprise Linux versions 11 and 12.

Use the provided distribution packages as described in this page if possible. These packages will automatically install all of MongoDB's dependencies, and are the recommended installation method.

Note: SUSE Linux Enterprise Server and potentially other SUSE distributions ship with virtual memory address space limited to 8 GB by default. You *must* adjust this in order to prevent virtual memory allocation failures as the database grows.

The SLES packages for MongoDB adjust these limits in the default scripts, but you will need to make this change manually if you are using custom scripts and/or the tarball release rather than the SLES packages.

Install MongoDB Enterprise

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[35].

Step 1: Configure the package management system (`zypper`). Add the repository so that you can install MongoDB using `zypper`.

Specify the latest stable release of MongoDB using the command appropriate for your version of SUSE:

SUSE 11

```
sudo zypper addrepo --no-gpgcheck "https://repo.mongodb.com/zypper/suse/11/mongodb-enterprise/stable/
```

SUSE 12

```
sudo zypper addrepo --no-gpgcheck "https://repo.mongodb.com/zypper/suse/12/mongodb-enterprise/stable/
```

If you'd like to install MongoDB packages from a previous *release series* (page 1070), such as 2.6, you can specify the release series in the repository configuration. For example, to restrict your SUSE 11 system to the 2.6 release series, use the following command:

```
sudo zypper addrepo --no-gpgcheck https://repo.mongodb.com/zypper/suse/11/mongodb-enterprise/2.6/x86_
```

Step 2: Install the MongoDB packages and associated tools. When you install the packages, you choose whether to install the current release or a previous one. This step provides the commands for both.

To install the latest stable version of MongoDB, issue the following command:

```
sudo zypper -n install mongodb-enterprise
```

[35] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-enterprise-on-suse/

To install a specific release of MongoDB, specify each component package individually and append the version number to the package name, as in the following example:

```
sudo zypper install mongodb-enterprise-3.2.5 mongodb-enterprise-server-3.2.5 mongodb-enterprise-shell
```

You can specify any available version of MongoDB. However `zypper` will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the packages by running the following command:

```
sudo zypper addlock mongodb-enterprise-3.2.5 mongodb-enterprise-server-3.2.5 mongodb-enterprise-shell
```

Previous versions of MongoDB packages use a different repository location. Refer to the version of the documentation appropriate for your MongoDB version.

Install MongoDB Enterprise From Tarball While you should use the `.rpm` packages as previously described, you may also manually install MongoDB using the tarballs.

First you must install any dependencies as appropriate:

```
zypper install cyrus-sasl cyrus-sasl-plain cyrus-sasl-gssapi krb5 \
           libopenssl0_9_8 net-snmp libstdc++46 zlib
```

To perform the installation, see *Install MongoDB Enterprise From Tarball* (page 66).

Run MongoDB Enterprise

Prerequisites The MongoDB instance stores its data files in `/var/lib/mongo` and its log files in `/var/log/mongodb` by default, and runs using the `mongod` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongo` and `/var/log/mongodb` directories to give this user access to these directories.

Procedure

Step 1: Start MongoDB. You can start the `mongod` process by issuing the following command:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully You can verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

You can optionally ensure that MongoDB will start following a system reboot by issuing the following command:

```
sudo chkconfig mongod on
```

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. You can restart the `mongod` process by issuing the following command:

```
sudo service mongod restart
```

You can follow the state of the process for errors or important messages by watching the output in the `/var/log/mongodb/mongod.log` file.

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

Warning: This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo zypper remove $(rpm -qa | grep mongodb-enterprise)
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongo
```

Install MongoDB Enterprise on Amazon Linux

On this page

- Overview (page 63)
- Install MongoDB Enterprise (page 64)
- Install MongoDB Enterprise From Tarball (page 65)
- Run MongoDB Enterprise (page 65)
- Uninstall MongoDB (page 66)

Overview Use this tutorial to install MongoDB Enterprise[36] on Amazon Linux AMI. MongoDB Enterprise is available on select platforms and contains support for several features related to security and monitoring.

This installation guide only supports 64-bit systems. See *Platform Support* (page 952) for details.

MongoDB provides officially supported Enterprise packages in their own repository. This repository contains the following packages:

[36] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

mongodb-enterprise	A metapackage that will automatically install the four component packages listed below.
mongodb-enterprise-server	Contains the mongod daemon and associated configuration and init scripts.
mongodb-enterprise-mongos	Contains the mongos daemon.
mongodb-enterprise-shell	Contains the mongo shell.
mongodb-enterprise-tools	Contains the following MongoDB tools: mongoimport bsondump, mongodump, mongoexport, mongofiles, mongooplog, mongoperf, mongorestore, mongostat, and mongotop.

Install MongoDB Enterprise

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[37].

Step 1: Configure repository. Create an /etc/yum.repos.d/mongodb-enterprise.repo file so that you can install MongoDB enterprise directly, using yum.

For the *latest* stable release of MongoDB Enterprise Use the following repository file:

```
[mongodb-enterprise]
name=MongoDB Enterprise Repository
baseurl=https://repo.mongodb.com/yum/amazon/2013.03/mongodb-enterprise/stable/$basearch/
gpgcheck=1
enabled=1
gpgkey=https://www.mongodb.org/static/pgp/server-3.2.asc
```

.repo files for each release can also be found in the repository itself[38]. Remember that odd-numbered minor release versions (e.g. 2.5) are development versions and are unsuitable for production deployment.

Step 2: Install the MongoDB Enterprise packages and associated tools. You can install either the latest stable version of MongoDB Enterprise or a specific version of MongoDB Enterprise.

To install the latest stable version of MongoDB Enterprise, issue the following command:

```
sudo yum install -y mongodb-enterprise
```

Step 3: Optional: Manage Installed Version

Install a specific release of MongoDB Enterprise. Specify each component package individually and append the version number to the package name, as in the following example that installs the 3.2.1 release of MongoDB:

```
sudo yum install -y mongodb-enterprise-3.2.1 mongodb-enterprise-server-3.2.1 mongodb-enterprise-shell
```

Pin a specific version of MongoDB Enterprise. Although you can specify any available version of MongoDB Enterprise, yum will upgrade the packages when a newer version becomes available. To prevent unintended upgrades, pin the package by adding the following exclude directive to your /etc/yum.conf file:

```
exclude=mongodb-enterprise,mongodb-enterprise-server,mongodb-enterprise-shell,mongodb-enterprise-mong
```

[37] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-enterprise-on-amazon/
[38] https://repo.mongodb.com/yum/amazon/

Previous versions of MongoDB packages use different naming conventions. See the 2.4 version of documentation for more information[39].

Step 4: When the install completes, you can run MongoDB.

Install MongoDB Enterprise From Tarball While you should use the `.rpm` packages as previously described, you may also manually install MongoDB using the tarballs.

First you must install any dependencies as appropriate:

```
yum install cyrus-sasl cyrus-sasl-plain cyrus-sasl-gssapi krb5-libs \
            lm_sensors-libs net-snmp-agent-libs net-snmp openssl rpm-libs \
            tcp_wrappers-libs
```

To perform the installation, see *Install MongoDB Enterprise From Tarball* (page 66).

Run MongoDB Enterprise The MongoDB instance stores its data files in `/var/lib/mongo` and its log files in `/var/log/mongodb` by default, and runs using the `mongod` user account. You can specify alternate log and data file directories in `/etc/mongod.conf`. See `systemLog.path` and `storage.dbPath` for additional information.

If you change the user that runs the MongoDB process, you **must** modify the access control rights to the `/var/lib/mongo` and `/var/log/mongodb` directories to give this user access to these directories.

Step 1: Start MongoDB. You can start the `mongod` process by issuing the following command:

```
sudo service mongod start
```

Step 2: Verify that MongoDB has started successfully You can verify that the `mongod` process has started successfully by checking the contents of the log file at `/var/log/mongodb/mongod.log` for a line reading

```
[initandlisten] waiting for connections on port <port>
```

where `<port>` is the port configured in `/etc/mongod.conf`, `27017` by default.

You can optionally ensure that MongoDB will start following a system reboot by issuing the following command:

```
sudo chkconfig mongod on
```

Step 3: Stop MongoDB. As needed, you can stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 4: Restart MongoDB. You can restart the `mongod` process by issuing the following command:

```
sudo service mongod restart
```

You can follow the state of the process for errors or important messages by watching the output in the `/var/log/mongodb/mongod.log` file.

[39] https://docs.mongodb.org/v2.4/tutorial/install-mongodb-on-linux

Step 5: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Uninstall MongoDB To completely remove MongoDB from a system, you must remove the MongoDB applications themselves, the configuration files, and any directories containing data and logs. The following section guides you through the necessary steps.

> **Warning:** This process will *completely* remove MongoDB, its configuration, and *all* databases. This process is not reversible, so ensure that all of your configuration and data is backed up before proceeding.

Step 1: Stop MongoDB. Stop the `mongod` process by issuing the following command:

```
sudo service mongod stop
```

Step 2: Remove Packages. Remove any MongoDB packages that you had previously installed.

```
sudo yum erase $(rpm -qa | grep mongodb-enterprise)
```

Step 3: Remove Data Directories. Remove MongoDB databases and log files.

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongo
```

On this page

Install MongoDB Enterprise From Tarball

- Overview (page 66)
- Install MongoDB (page 66)
- Run MongoDB (page 67)

Overview Compiled versions of MongoDB Enterprise for Linux provide a simple option for installing MongoDB for other Linux systems without supported packages.

Install MongoDB

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[40].

Step 1: Install any missing dependencies. To manually install MongoDB Enterprise, first install any dependencies as appropriate.

[40] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-enterprise-on-linux/

Step 2: Download and install the MongoDB Enterprise packages. After you have installed the required pre-requisite packages, download and install the MongoDB Enterprise packages from https://mongodb.com/download/. The MongoDB binaries are located in the `bin/` directory of the archive. To download and install, use the following sequence of commands.

Step 3: Ensure the location of the MongoDB binaries is included in the `PATH` variable. Once you have copied the MongoDB binaries to their target location, ensure that the location is included in your `PATH` variable. If it is not, either include it or create symbolic links from the binaries to a directory that is included.

Run MongoDB

Step 1: Create the data directory. Before you start MongoDB for the first time, create the directory to which the `mongod` process will write data. By default, the `mongod` process uses the `/data/db` directory. If you create a directory other than this one, you must specify that directory in the `dbpath` option when starting the `mongod` process later in this procedure.

The following example command creates the default `/data/db` directory:

```
mkdir -p /data/db
```

Step 2: Set permissions for the data directory. Before running `mongod` for the first time, ensure that the user account running `mongod` has read and write permissions for the directory.

Step 3: Run MongoDB. To run MongoDB, run the `mongod` process at the system prompt. If necessary, specify the path of the `mongod` or the data directory. See the following examples.

Run without specifying paths If your system `PATH` variable includes the location of the `mongod` binary and if you use the default data directory (i.e., `/data/db`), simply enter `mongod` at the system prompt:

```
mongod
```

Specify the path of the `mongod` If your `PATH` does not include the location of the `mongod` binary, enter the full path to the `mongod` binary at the system prompt:

```
<path to binary>/mongod
```

Specify the path of the data directory If you do not use the default data directory (i.e., `/data/db`), specify the path to the data directory using the *`--dbpath`* option:

```
mongod --dbpath <path to data directory>
```

Step 4: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Install MongoDB Enterprise on OS X

Overview Use this tutorial to install MongoDB Enterprise[41] on OS X systems. MongoDB Enterprise is available on select platforms and contains support for several features related to security and monitoring.

Platform Support

MongoDB only supports OS X versions 10.7 (Lion) and later on Intel x86-64. Versions of MongoDB Enterprise prior to 3.2 did not support OS X.

Install MongoDB Enterprise

Step 1 Download the latest production release of MongoDB Enterprise[42].

Step 2: Extract the files from the downloaded archive. For example, from a system shell, you can extract through the `tar` command:

```
tar -zxvf mongodb-osx-x86_64-enterprise-3.2.5.tgz
```

Step 3: Copy the extracted archive to the target directory. Copy the extracted folder to the location from which MongoDB will run.

```
mkdir -p mongodb
cp -R -n mongodb-osx-x86_64-enterprise-3.2.5/ mongodb
```

Step 4: Ensure the location of the binaries is in the `PATH` variable. The MongoDB binaries are in the `bin/` directory of the archive. To ensure that the binaries are in your `PATH`, you can modify your `PATH`.

For example, you can add the following line to your shell's `rc` file (e.g. `~/.bashrc`):

```
export PATH=<mongodb-install-directory>/bin:$PATH
```

Replace `<mongodb-install-directory>` with the path to the extracted MongoDB archive.

Run MongoDB Enterprise

Step 1: Create the data directory. Before you start MongoDB for the first time, create the directory to which the `mongod` process will write data. By default, the `mongod` process uses the `/data/db` directory. If you create a directory other than this one, you must specify that directory in the `dbpath` option when starting the `mongod` process later in this procedure.

The following example command creates the default `/data/db` directory:

```
mkdir -p /data/db
```

Step 2: Set permissions for the data directory. Before running `mongod` for the first time, ensure that the user account running `mongod` has read and write permissions for the directory.

[41] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[42] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Step 3: Run MongoDB. To run MongoDB, run the `mongod` process at the system prompt. If necessary, specify the path of the `mongod` or the data directory. See the following examples.

Run without specifying paths If your system `PATH` variable includes the location of the `mongod` binary and if you use the default data directory (i.e., `/data/db`), simply enter `mongod` at the system prompt:

```
mongod
```

Specify the path of the `mongod` If your `PATH` does not include the location of the `mongod` binary, enter the full path to the `mongod` binary at the system prompt:

```
<path to binary>/mongod
```

Specify the path of the data directory If you do not use the default data directory (i.e., `/data/db`), specify the path to the data directory using the *`--dbpath`* option:

```
mongod --dbpath <path to data directory>
```

Step 4: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Install MongoDB Enterprise on Windows

On this page

- Overview (page 69)
- Prerequisites (page 69)
- Get MongoDB Enterprise (page 70)
- Install MongoDB Enterprise (page 70)
- Run MongoDB Enterprise (page 71)
- Configure a Windows Service for MongoDB Enterprise (page 72)
- Manually Create a Windows Service for MongoDB Enterprise (page 73)

New in version 2.6.

Overview Use this tutorial to install MongoDB Enterprise[43] on Windows systems. MongoDB Enterprise is available on select platforms and contains support for several features related to security and monitoring.

Prerequisites MongoDB Enterprise Server for Windows requires Windows Server 2008 R2 or later. The `.msi` installer includes all other software dependencies and will automatically upgrade any older version of MongoDB installed using an `.msi` file.

[43] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

Get MongoDB Enterprise

Note: To install a version of MongoDB prior to 3.2, please refer to that version's documentation. For example, see version 3.0[44].

Step 1: Download MongoDB Enterprise for Windows. Download the latest production release of MongoDB Enterprise[45].

To find which version of Windows you are running, enter the following commands in the *Command Prompt* or *Powershell*:

```
wmic os get caption
wmic os get osarchitecture
```

Install MongoDB Enterprise

Interactive Installation

Step 1: Install MongoDB Enterprise for Windows. In Windows Explorer, locate the downloaded MongoDB `.msi` file, which typically is located in the default `Downloads` folder. Double-click the `.msi` file. A set of screens will appear to guide you through the installation process.

You may specify an installation directory if you choose the "Custom" installation option.

Note: These instructions assume that you have installed MongoDB to `C:\mongodb`.

MongoDB is self-contained and does not have any other system dependencies. You can run MongoDB from any folder you choose. You may install MongoDB in any folder (e.g. `D:\test\mongodb`).

Unattended Installation You may install MongoDB unattended on Windows from the command line using `msiexec.exe`.

Step 1: Install MongoDB Enterprise for Windows. Change to the directory containing the `.msi` installation binary of your choice and invoke:

```
msiexec.exe /q /i mongodb-win32-x86_64-2008plus-ssl-3.2.5-signed.msi ^
            INSTALLLOCATION="C:\mongodb" ^
            ADDLOCAL="all"
```

You can specify the installation location for the executable by modifying the `INSTALLLOCATION` value.

By default, this method installs all MongoDB binaries. To install specific MongoDB component sets, you can specify them in the `ADDLOCAL` argument using a comma-separated list including one or more of the following component sets:

[44] https://docs.mongodb.org/v3.0/tutorial/install-mongodb-enterprise-on-windows/
[45] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Component Set	Binaries
Server	mongod.exe
Router	mongos.exe
Client	mongo.exe
MonitoringTools	mongostat.exe, mongotop.exe
ImportExportTools	mongodump.exe, mongorestore.exe, mongoexport.exe, mongoimport.exe
MiscellaneousTools	bsondump.exe, mongofiles.exe, mongooplog.exe, mongoperf.exe

For instance, to install *only* the MongoDB utilities, invoke:

```
msiexec.exe /q /i mongodb-win32-x86_64-2008plus-ssl-3.2.5-signed.msi ^
            INSTALLLOCATION="C:\mongodb" ^
            ADDLOCAL="MonitoringTools,ImportExportTools,MiscellaneousTools"
```

Run MongoDB Enterprise

Warning: Do not make `mongod.exe` visible on public networks without running in "Secure Mode
`auth` setting. MongoDB is designed to be run in trusted environments, and the database does not enab
Mode" by default.

Step 1: Set up the MongoDB environment. MongoDB requires a *data directory* to store all data. MongoDB's default data directory path is `\data\db`. Create this folder using the following commands from a *Command Prompt*:

```
md \data\db
```

You can specify an alternate path for data files using the *`--dbpath`* option to `mongod.exe`, for example:

```
C:\mongodb\bin\mongod.exe --dbpath d:\test\mongodb\data
```

If your path includes spaces, enclose the entire path in double quotes, for example:

```
C:\mongodb\bin\mongod.exe --dbpath "d:\test\mongo db data"
```

You may also specify the `dbpath` in a `configuration file`.

Step 2: Start MongoDB. To start MongoDB, run `mongod.exe`. For example, from the *Command Prompt*:

```
C:\mongodb\bin\mongod.exe
```

This starts the main MongoDB database process. The `waiting for connections` message in the console output indicates that the `mongod.exe` process is running successfully.

Depending on the security level of your system, Windows may pop up a *Security Alert* dialog box about blocking "some features" of `C:\mongodb\bin\mongod.exe` from communicating on networks. All users should select `Private Networks, such as my home or work network` and click `Allow access`. For additional information on security and MongoDB, please see the *Security Documentation* (page 391).

Step 3: Connect to MongoDB. To connect to MongoDB through the `mongo.exe` shell, open another *Command Prompt*.

```
C:\mongodb\bin\mongo.exe
```

If you want to develop applications using .NET, see the documentation of C# and MongoDB[46] for more information.

[46] https://docs.mongodb.org/ecosystem/drivers/csharp

Step 4: Begin using MongoDB. To help you start using MongoDB, MongoDB provides *Getting Started Guides* in various driver editions. See *getting-started* for the available editions.

Before deploying MongoDB in a production environment, consider the *Production Notes* (page 296) document.

Later, to stop MongoDB, press `Control+C` in the terminal where the `mongod` instance is running.

Configure a Windows Service for MongoDB Enterprise

Step 1: Open an Administrator command prompt. Press the `Win` key, type `cmd.exe`, and press `Ctrl + Shift + Enter` to run the *Command Prompt* as Administrator.

Execute the remaining steps from the Administrator command prompt.

Step 2: Create directories. Create directories for your database and log files:

```
mkdir c:\data\db
mkdir c:\data\log
```

Step 3: Create a configuration file. Create a configuration file. The file **must** set `systemLog.path`. Include additional `configuration options` as appropriate.

For example, create a file at `C:\mongodb\mongod.cfg` that specifies both `systemLog.path` and `storage.dbPath`:

```
systemLog:
    destination: file
    path: c:\data\log\mongod.log
storage:
    dbPath: c:\data\db
```

Step 4: Install the MongoDB service.

Important: Run all of the following commands in *Command Prompt* with "Administrative Privileges".

Install the MongoDB service by starting `mongod.exe` with the `--install` option and the `-config` option to specify the previously created configuration file.

```
"C:\mongodb\bin\mongod.exe" --config "C:\mongodb\mongod.cfg" --install
```

To use an alternate `dbpath`, specify the path in the configuration file (e.g. `C:\mongodb\mongod.cfg`) or on the command line with the *`--dbpath`* option.

If needed, you can install services for multiple instances of `mongod.exe` or `mongos.exe`. Install each service with a unique *`--serviceName`* and *`--serviceDisplayName`*. Use multiple instances only when sufficient system resources exist and your system design requires it.

Step 5: Start the MongoDB service.

```
net start MongoDB
```

Step 6: Stop or remove the MongoDB service as needed. To stop the MongoDB service use the following command:

```
net stop MongoDB
```

To remove the MongoDB service use the following command:

```
"C:\mongodb\bin\mongod.exe" --remove
```

Manually Create a Windows Service for MongoDB Enterprise You can set up the MongoDB server as a *Windows Service* that starts automatically at boot time.

The following procedure assumes you have installed MongoDB using the `.msi` installer with the path `C:\mongodb\`.

If you have installed in an alternative directory, you will need to adjust the paths as appropriate.

Step 1: Open an Administrator command prompt. Press the `Win` key, type `cmd.exe`, and press `Ctrl + Shift + Enter` to run the *Command Prompt* as Administrator.

Execute the remaining steps from the Administrator command prompt.

Step 2: Create directories. Create directories for your database and log files:

```
mkdir c:\data\db
mkdir c:\data\log
```

Step 3: Create a configuration file. Create a configuration file. The file **must** set `systemLog.path`. Include additional `configuration options` as appropriate.

For example, create a file at `C:\mongodb\mongod.cfg` that specifies both `systemLog.path` and `storage.dbPath`:

```
systemLog:
    destination: file
    path: c:\data\log\mongod.log
storage:
    dbPath: c:\data\db
```

Step 4: Create the MongoDB service. Create the MongoDB service.

```
sc.exe create MongoDB binPath= "C:\mongodb\bin\mongod.exe --service --config=\"C:\mongodb\mongod.cfg\
```

`sc.exe` requires a space between "=" and the configuration values (eg "binPath= "), and a "\" to escape double quotes.

If successfully created, the following log message will display:

```
[SC] CreateService SUCCESS
```

Step 5: Start the MongoDB service.

```
net start MongoDB
```

Step 6: Stop or remove the MongoDB service as needed. To stop the MongoDB service, use the following command:

```
net stop MongoDB
```

To remove the MongoDB service, first stop the service and then run the following command:

```
sc.exe delete MongoDB
```

Verify Integrity of MongoDB Packages

On this page

- Overview (page 74)
- Procedures (page 74)

Overview

The MongoDB release team digitally signs all software packages to certify that a particular MongoDB package is a valid and unaltered MongoDB release. Before installing MongoDB, you should validate the package using either the provided PGP signature or SHA-256 checksum.

PGP signatures provide the strongest guarantees by checking both the authenticity and integrity of a file to prevent tampering.

Cryptographic checksums only validate file integrity to prevent network transmission errors.

Procedures

Use PGP/GPG MongoDB signs each release branch with a different PGP key. The public key files for each release branch since MongoDB 2.2 are available for download from the key server[47] in both textual `.asc` and binary `.pub` formats.

Step 1: Download the MongoDB installation file. Download the binaries from `https://www.mongodb.org/downloads` based on your environment.

For example, to download the `3.0.5` release for OS X through the shell, type this command:

```
curl -LO https://fastdl.mongodb.org/osx/mongodb-osx-x86_64-3.0.5.tgz
```

Step 2: Download the public signature file.

```
curl -LO https://fastdl.mongodb.org/osx/mongodb-osx-x86_64-3.0.5.tgz.sig
```

Step 3: Download then import the key file. If you have not downloaded and imported the MongoDB 3.0 public key, enter these commands:

```
curl -LO https://www.mongodb.org/static/pgp/server-3.0.asc
gpg --import server-3.0.asc
```

[47] https://www.mongodb.org/static/pgp/

You should receive this message:

```
gpg: key 24F3C978: public key "MongoDB 3.0 Release Signing Key <packaging@mongodb.com>" imported
gpg: Total number processed: 1
gpg:               imported: 1
```

Step 4: Verify the MongoDB installation file. Type this command:

```
gpg --verify mongodb-osx-x86_64-3.0.5.tgz.sig mongodb-osx-x86_64-3.0.5.tgz
```

You should receive this message:

```
gpg: Signature made Mon 27 Jul 2015 07:51:53 PM EDT using RSA key ID 24F3C978
gpg: Good signature from "MongoDB 3.0 Release Signing Key <packaging@mongodb.com>" [unknown]
```

If you receive a message such as the following, confirm that you imported the correct public key:

```
gpg: Signature made Mon 27 Jul 2015 07:51:53 PM EDT using RSA key ID 24F3C978
gpg: Can't check signature: public key not found
```

`gpg` will return the following message if the package is properly signed, but you do not currently trust the signing key in your local `trustdb`.

```
gpg: WARNING: This key is not certified with a trusted signature!
gpg:          There is no indication that the signature belongs to the owner.
Primary key fingerprint: 89AE C6ED 5423 0831 793F  1384 BE0E B6AA 24F3 C978
```

Use SHA-256

Step 1: Download the MongoDB installation file. Download the binaries from `https://www.mongodb.org/downloads` based on your environment.

For example, to download the `3.0.5` release for OS X through the shell, type this command:

```
curl -LO https://fastdl.mongodb.org/osx/mongodb-osx-x86_64-3.0.5.tgz
```

Step 2: Download the SHA256 file.

```
curl -LO https://fastdl.mongodb.org/osx/mongodb-osx-x86_64-3.0.5.tgz.sha256
```

Step 3: Use the SHA-256 checksum to verify the MongoDB package file. Compute the checksum of the package file:

```
shasum -c mongodb-osx-x86_64-3.0.5.tgz.sha256
```

which should return the following if the checksum matched the downloaded package:

```
mongodb-osx-x86_64-3.0.5.tgz: OK
```

2.4 Additional Resources

- Install MongoDB using MongoDB Cloud Manager[48]

[48] https://docs.cloud.mongodb.com/tutorial/getting-started?jmp=docs

- Create a New MongoDB Deployment with Ops Manager[49]: Ops Manager is an on-premise solution available in MongoDB Enterprise Advanced[50].
- *MongoDB CRUD Concepts* (page 99)
- *Data Models* (page 247)

[49] https://docs.opsmanager.mongodb.com/current/tutorial/nav/management
[50] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

CHAPTER 3

The mongo Shell

On this page

- Introduction (page 77)
- Start the mongo Shell (page 77)
- Working with the mongo Shell (page 78)
- Tab Completion and Other Keyboard Shortcuts (page 79)
- Exit the Shell (page 80)

3.1 Introduction

The mongo shell is an interactive JavaScript interface to MongoDB. You can use the mongo shell to query and update data as well as perform administrative operations.

The mongo shell is a component of the MongoDB distributions[1]. Once you have *installed and have started MongoDB* (page 21), connect the mongo shell to your running MongoDB instance.

Most examples in the MongoDB Manual use the mongo shell; however, many drivers provide similar interfaces to MongoDB.

3.2 Start the mongo Shell

Important: Ensure that MongoDB is running before attempting to start the mongo shell.

To start the mongo shell and connect to your MongoDB instance running on **localhost** with **default port**:

1. At a prompt in a terminal window (or a command prompt for Windows), go to your <mongodb installation dir>:

   ```
   cd <mongodb installation dir>
   ```

2. Type ./bin/mongo to start mongo:

   ```
   ./bin/mongo
   ```

[1] http://www.mongodb.org/downloads

If you have added the <mongodb installation dir>/bin to the PATH environment variable, you can just type mongo instead of ./bin/mongo.

3.2.1 Options

When you run mongo without any arguments, the mongo shell will attempt to connect to the MongoDB instance running on the localhost interface on port 27017. To specify a different host or port number, as well as other options, see *examples of starting up mongo* and mongo reference which provides details on the available options.

3.2.2 .mongorc.js File

When starting, mongo checks the user's HOME directory for a JavaScript file named *.mongorc.js*. If found, mongo interprets the content of .mongorc.js before displaying the prompt for the first time. If you use the shell to evaluate a JavaScript file or expression, either by using the *--eval* option on the command line or by specifying *a .js file to mongo*, mongo will read the .mongorc.js file *after* the JavaScript has finished processing. You can prevent .mongorc.js from being loaded by using the *--norc* option.

3.3 Working with the mongo Shell

To display the database you are using, type db:

```
db
```

The operation should return test, which is the default database. To switch databases, issue the use <db> helper, as in the following example:

```
use <database>
```

To list the available databases, use the helper show dbs. See also db.getSiblingDB() method to access a different database from the current database without switching your current database context (i.e. db).

You can switch to non-existing databases. When you first store data in the database, such as by creating a collection, MongoDB creates the database. For example, the following creates both the database myNewDatabase and the *collection* myCollection during the insert() operation:

```
use myNewDatabase
db.myCollection.insert( { x: 1 } );
```

The db.myCollection.insert() is one of the methods available in the mongo shell

- db refers to the current database.
- myCollection is the name of the collection.

If the mongo shell does not accept the name of the collection, for instance if the name contains a space, hyphen, or starts with a number, you can use an alternate syntax to refer to the collection, as in the following:

```
db["3test"].find()

db.getCollection("3test").find()
```

For more documentation of basic MongoDB operations in the mongo shell, see:

- Getting Started Guide[2]

[2] https://docs.mongodb.org/getting-started/shell

- *Insert Documents* (page 137)
- *Query Documents* (page 140)
- *Modify Documents* (page 148)
- *Remove Documents* (page 152)
- `https://docs.mongodb.org/manual/reference/method`

3.3.1 Format Printed Results

The `db.collection.find()` method returns a *cursor* to the results; however, in the `mongo` shell, if the returned cursor is not assigned to a variable using the `var` keyword, then the cursor is automatically iterated up to 20 times to print up to the first 20 documents that match the query. The `mongo` shell will prompt `Type it` to iterate another 20 times.

To format the printed result, you can add the `.pretty()` to the operation, as in the following:

```
db.myCollection.find().pretty()
```

In addition, you can use the following explicit print methods in the `mongo` shell:

- `print()` to print without formatting
- `print(tojson(<obj>))` to print with *JSON* formatting and equivalent to `printjson()`
- `printjson()` to print with *JSON* formatting and equivalent to `print(tojson(<obj>))`

For more information and examples on cursor handling in the `mongo` shell, see *Cursors* (page 103). See also *Cursor Help* (page 83) for list of cursor help in the `mongo` shell.

3.3.2 Multi-line Operations in the `mongo` Shell

If you end a line with an open parenthesis (`'('`), an open brace (`'{'`), or an open bracket (`'['`), then the subsequent lines start with ellipsis (`"..."`) until you enter the corresponding closing parenthesis (`')'`), the closing brace (`'}'`) or the closing bracket (`']'`). The `mongo` shell waits for the closing parenthesis, closing brace, or the closing bracket before evaluating the code, as in the following example:

```
> if ( x > 0 ) {
... count++;
... print (x);
... }
```

You can exit the line continuation mode if you enter two blank lines, as in the following example:

```
> if (x > 0
...
...
>
```

3.4 Tab Completion and Other Keyboard Shortcuts

The `mongo` shell supports keyboard shortcuts. For example,

- Use the up/down arrow keys to scroll through command history. See *.dbshell* documentation for more information on the `.dbshell` file.

- Use <Tab> to autocomplete or to list the completion possibilities, as in the following example which uses <Tab> to complete the method name starting with the letter 'c':

  ```
  db.myCollection.c<Tab>
  ```

  Because there are many collection methods starting with the letter 'c', the <Tab> will list the various methods that start with 'c'.

For a full list of the shortcuts, see *Shell Keyboard Shortcuts*

3.5 Exit the Shell

To exit the shell, type `quit()` or use the <Ctrl-c> shortcut.

See also:

- Getting Started Guide[3]
- `mongo` Reference Page

3.5.1 Configure the `mongo` Shell

On this page

- Customize the Prompt (page 80)
- Use an External Editor in the `mongo` Shell (page 81)
- Change the `mongo` Shell Batch Size (page 82)

Customize the Prompt

You may modify the content of the prompt by setting the variable `prompt` in the `mongo` shell. The `prompt` variable can hold strings as well as JavaScript code. If `prompt` holds a function that returns a string, `mongo` can display dynamic information in each prompt.

You can add the logic for the prompt in the *.mongorc.js* file to set the prompt each time you start up the `mongo` shell.

Customize Prompt to Display Number of Operations

For example,to create a `mongo` shell prompt with the number of operations issued in the current session, define the following variables in the `mongo` shell:

```
cmdCount = 1;
prompt = function() {
             return (cmdCount++) + "> ";
         }
```

The prompt would then resemble the following:

```
1>
2>
3>
```

[3] https://docs.mongodb.org/getting-started/shell

Customize Prompt to Display Database and Hostname

To create a `mongo` shell prompt in the form of `<database>@<hostname>$`, define the following variables:

```
host = db.serverStatus().host;

prompt = function() {
             return db+"@"+host+"$ ";
         }
```

The prompt would then resemble the following:

```
test@myHost1$
```

Customize Prompt to Display Up Time and Document Count

To create a `mongo` shell prompt that contains the system up time *and* the number of documents in the current database, define the following `prompt` variable in the `mongo` shell:

```
prompt = function() {
           return "Uptime:"+db.serverStatus().uptime+" Documents:"+db.stats().objects+" > ";
         }
```

The prompt would then resemble the following:

```
Uptime:5897 Documents:6 >
```

Use an External Editor in the `mongo` Shell

You can use your own editor in the `mongo` shell by setting the `EDITOR` environment variable *before* starting the `mongo` shell.

```
export EDITOR=vim
mongo
```

Once in the `mongo` shell, you can edit with the specified editor by typing `edit <variable>` or `edit <function>`, as in the following example:

1. Define a function `myFunction`:

   ```
   function myFunction () { }
   ```

2. Edit the function using your editor:

   ```
   edit myFunction
   ```

   The command should open the `vim` edit session. When finished with the edits, save and exit `vim` edit session.

3. In the `mongo` shell, type `myFunction` to see the function definition:

   ```
   myFunction
   ```

   The result should be the changes from your saved edit:

   ```
   function myFunction() {
       print("This was edited");
   }
   ```

Note: As mongo shell interprets code edited in an external editor, it may modify code in functions, depending on the JavaScript compiler. For mongo may convert 1+1 to 2 or remove comments. The actual changes affect only the appearance of the code and will vary based on the version of JavaScript used but will not affect the semantics of the code.

Change the mongo Shell Batch Size

The db.collection.find() method is the JavaScript method to retrieve documents from a *collection*. The db.collection.find() method returns a *cursor* to the results; however, in the mongo shell, if the returned cursor is not assigned to a variable using the var keyword, then the cursor is automatically iterated up to 20 times to print up to the first 20 documents that match the query. The mongo shell will prompt Type it to iterate another 20 times.

You can set the DBQuery.shellBatchSize attribute to change the number of documents from the default value of 20, as in the following example which sets it to 10:

```
DBQuery.shellBatchSize = 10;
```

3.5.2 Access the mongo Shell Help

On this page

- Command Line Help (page 82)
- Shell Help (page 82)
- Database Help (page 82)
- Collection Help (page 83)
- Cursor Help (page 83)
- Wrapper Object Help (page 84)

In addition to the documentation in the MongoDB Manual, the mongo shell provides some additional information in its "online" help system. This document provides an overview of accessing this help information.

Command Line Help

To see the list of options and help for starting the mongo shell, use the *--help* option from the command line:

```
mongo --help
```

Shell Help

To see the list of help, in the mongo shell, type help:

```
help
```

Database Help

In the mongo shell:

- To see the list of databases on the server, use the show dbs command:

```
show dbs
```

New in version 2.4: `show databases` is now an alias for `show dbs`

- To see the list of help for methods you can use on the `db` object, call the `db.help()` method:

```
db.help()
```

- To see the implementation of a method in the shell, type the `db.<method name>` without the parenthesis (`()`), as in the following example which will return the implementation of the method `db.updateUser()`:

```
db.updateUser
```

Collection Help

In the `mongo` shell:

- To see the list of collections in the current database, use the `show collections` command:

```
show collections
```

- To see the help for methods available on the collection objects (e.g. `db.<collection>`), use the `db.<collection>.help()` method:

```
db.collection.help()
```

`<collection>` can be the name of a collection that exists, although you may specify a collection that doesn't exist.

- To see the collection method implementation, type the `db.<collection>.<method>` name without the parenthesis (`()`), as in the following example which will return the implementation of the `save()` method:

```
db.collection.save
```

Cursor Help

When you perform *read operations* (page 100) with the `find()` method in the `mongo` shell, you can use various cursor methods to modify the `find()` behavior and various JavaScript methods to handle the cursor returned from the `find()` method.

- To list the available modifier and cursor handling methods, use the `db.collection.find().help()` command:

```
db.collection.find().help()
```

`<collection>` can be the name of a collection that exists, although you may specify a collection that doesn't exist.

- To see the implementation of the cursor method, type the `db.<collection>.find().<method>` name without the parenthesis (`()`), as in the following example which will return the implementation of the `toArray()` method:

```
db.collection.find().toArray
```

Some useful methods for handling cursors are:

- `hasNext()` which checks whether the cursor has more documents to return.
- `next()` which returns the next document and advances the cursor position forward by one.

- `forEach(<function>)` which iterates the whole cursor and applies the `<function>` to each document returned by the cursor. The `<function>` expects a single argument which corresponds to the document from each iteration.

For examples on iterating a cursor and retrieving the documents from the cursor, see *cursor handling* (page 103). See also *js-query-cursor-methods* for all available cursor methods.

Wrapper Object Help

To get a list of the wrapper classes available in the `mongo` shell, such as `BinData()`, type `help misc` in the `mongo` shell:

```
help misc
```

See also:

`https://docs.mongodb.org/manual/reference/method`

3.5.3 Write Scripts for the `mongo` Shell

On this page

- Opening New Connections (page 84)
- Differences Between Interactive and Scripted `mongo` (page 85)
- Scripting (page 86)

You can write scripts for the `mongo` shell in JavaScript that manipulate data in MongoDB or perform administrative operation. For more information about the `mongo` shell, see the *Running .js files via a mongo shell Instance on the Server* (page 384) section for more information about using these `mongo` script.

This tutorial provides an introduction to writing JavaScript that uses the `mongo` shell to access MongoDB.

Opening New Connections

From the `mongo` shell or from a JavaScript file, you can instantiate database connections using the `Mongo()` constructor:

```
new Mongo()
new Mongo(<host>)
new Mongo(<host:port>)
```

Consider the following example that instantiates a new connection to the MongoDB instance running on localhost on the default port and sets the global `db` variable to `myDatabase` using the `getDB()` method:

```
conn = new Mongo();
db = conn.getDB("myDatabase");
```

If connecting to a MongoDB instance that has enforces access control, you can use the `db.auth()` method to authenticate.

Additionally, you can use the `connect()` method to connect to the MongoDB instance. The following example connects to the MongoDB instance that is running on `localhost` with the non-default port `27020` and set the global `db` variable:

```
db = connect("localhost:27020/myDatabase");
```

See also:

https://docs.mongodb.org/manual/reference/method/

Differences Between Interactive and Scripted mongo

When writing scripts for the mongo shell, consider the following:

- To set the db global variable, use the getDB() method or the connect() method. You can assign the database reference to a variable other than db.
- Write operations in the mongo shell use a write concern of *{ w: 1 }* (page 180) by default. If performing bulk operations, use the Bulk() methods. See *Write Method Acknowledgements* (page 1002) for more information.

  Changed in version 2.6: Before MongoDB 2.6, call db.getLastError() explicitly to wait for the result of *write operations* (page 114).
- You **cannot** use any shell helper (e.g. use <dbname>, show dbs, etc.) inside the JavaScript file because they are not valid JavaScript.

  The following table maps the most common mongo shell helpers to their JavaScript equivalents.

Shell Helpers	JavaScript Equivalents
show dbs, show databases	db.adminCommand('listDatabases')
use <db>	db = db.getSiblingDB('<db>')
show collections	db.getCollectionNames()
show users	db.getUsers()
show roles	db.getRoles({showBuiltinRoles: true})
show log <logname>	db.adminCommand({ 'getLog' : '<logname>' })
show logs	db.adminCommand({ 'getLog' : '*' })

it	cursor = db.collection.find()
	if (cursor.hasNext()){
	  cursor.next();
	}

- In interactive mode, mongo prints the results of operations including the content of all cursors. In scripts, either use the JavaScript print() function or the mongo specific printjson() function which returns formatted JSON.

Example

To print all items in a result cursor in mongo shell scripts, use the following idiom:

```
cursor = db.collection.find();
while ( cursor.hasNext() ) {
   printjson( cursor.next() );
}
```

Scripting

From the system prompt, use `mongo` to evaluate JavaScript.

`--eval` option

Use the *`--eval`* option to `mongo` to pass the shell a JavaScript fragment, as in the following:

```
mongo test --eval "printjson(db.getCollectionNames())"
```

This returns the output of `db.getCollectionNames()` using the `mongo` shell connected to the `mongod` or `mongos` instance running on port `27017` on the `localhost` interface.

Execute a JavaScript file

You can specify a `.js` file to the `mongo` shell, and `mongo` will execute the JavaScript directly. Consider the following example:

```
mongo localhost:27017/test myjsfile.js
```

This operation executes the `myjsfile.js` script in a `mongo` shell that connects to the `test` *database* on the `mongod` instance accessible via the `localhost` interface on port `27017`.

Alternately, you can specify the mongodb connection parameters inside of the javascript file using the `Mongo()` constructor. See *Opening New Connections* (page 84) for more information.

You can execute a `.js` file from within the `mongo` shell, using the `load()` function, as in the following:

```
load("myjstest.js")
```

This function loads and executes the `myjstest.js` file.

The `load()` method accepts relative and absolute paths. If the current working directory of the `mongo` shell is `/data/db`, and the `myjstest.js` resides in the `/data/db/scripts` directory, then the following calls within the `mongo` shell would be equivalent:

```
load("scripts/myjstest.js")
load("/data/db/scripts/myjstest.js")
```

Note: There is no search path for the `load()` function. If the desired script is not in the current working directory or the full specified path, `mongo` will not be able to access the file.

3.5.4 Data Types in the `mongo` Shell

On this page

- Types (page 87)
- Check Types in the mongo Shell (page 89)

MongoDB *BSON* provides support for additional data types than *JSON*. Drivers provide native support for these data types in host languages and the mongo shell also provides several helper classes to support the use of these data types in the mongo JavaScript shell. See the *Extended JSON* (page 16) reference for additional information.

Types

Date

The mongo shell provides various methods to return the date, either as a string or as a Date object:

- Date() method which returns the current date as a string.
- new Date() constructor which returns a Date object using the ISODate() wrapper.
- ISODate() constructor which returns a Date object using the ISODate() wrapper.

Internally, *Date* (page 15) objects are stored as a 64 bit integer representing the number of milliseconds since the Unix epoch (Jan 1, 1970), which results in a representable date range of about 290 millions years into the past and future.

Return Date as a String To return the date as a string, use the Date() method, as in the following example:

```
var myDateString = Date();
```

To print the value of the variable, type the variable name in the shell, as in the following:

```
myDateString
```

The result is the value of myDateString:

```
Wed Dec 19 2012 01:03:25 GMT-0500 (EST)
```

To verify the type, use the typeof operator, as in the following:

```
typeof myDateString
```

The operation returns string.

Return Date The mongo shell wraps objects of Date type with the ISODate helper; however, the objects remain of type Date.

The following example uses both the new Date() constructor and the ISODate() constructor to return Date objects.

```
var myDate = new Date();
var myDateInitUsingISODateWrapper = ISODate();
```

You can use the new operator with the ISODate() constructor as well.

To print the value of the variable, type the variable name in the shell, as in the following:

```
myDate
```

The result is the `Date` value of `myDate` wrapped in the `ISODate()` helper:

```
ISODate("2012-12-19T06:01:17.171Z")
```

To verify the type, use the `instanceof` operator, as in the following:

```
myDate instanceof Date
myDateInitUsingISODateWrapper instanceof Date
```

The operation returns `true` for both.

ObjectId

The `mongo` shell provides the `ObjectId()` wrapper class around the *ObjectId* (page 14) data type. To generate a new ObjectId, use the following operation in the `mongo` shell:

```
new ObjectId
```

See

`ObjectId`

NumberLong

By default, the `mongo` shell treats all numbers as floating-point values. The `mongo` shell provides the `NumberLong()` wrapper to handle 64-bit integers.

The `NumberLong()` wrapper accepts the long as a string:

```
NumberLong("2090845886852")
```

The following examples use the `NumberLong()` wrapper to write to the collection:

```
db.collection.insert( { _id: 10, calc: NumberLong("2090845886852") } )
db.collection.update( { _id: 10 },
                      { $set:  { calc: NumberLong("2555555000000") } } )
db.collection.update( { _id: 10 },
                      { $inc: { calc: NumberLong(5) } } )
```

Retrieve the document to verify:

```
db.collection.findOne( { _id: 10 } )
```

In the returned document, the `calc` field contains a `NumberLong` object:

```
{ "_id" : 10, "calc" : NumberLong("2555555000005") }
```

If you use the `$inc` to increment the value of a field that contains a `NumberLong` object by a **float**, the data type changes to a floating point value, as in the following example:

1. Use `$inc` to increment the `calc` field by `5`, which the `mongo` shell treats as a float:

   ```
   db.collection.update( { _id: 10 },
                         { $inc: { calc: 5 } } )
   ```

2. Retrieve the updated document:

```
db.collection.findOne( { _id: 10 } )
```

In the updated document, the `calc` field contains a floating point value:

```
{ "_id" : 10, "calc" : 2555555000010 }
```

NumberInt

By default, the `mongo` shell treats all numbers as floating-point values. The `mongo` shell provides the `NumberInt()` constructor to explicitly specify 32-bit integers.

Check Types in the `mongo` Shell

To determine the type of fields, the `mongo` shell provides the `instanceof` and `typeof` operators.

`instanceof`

`instanceof` returns a boolean to test if a value is an instance of some type.

For example, the following operation tests whether the `_id` field is an instance of type `ObjectId`:

```
mydoc._id instanceof ObjectId
```

The operation returns `true`.

`typeof`

`typeof` returns the type of a field.

For example, the following operation returns the type of the `_id` field:

```
typeof mydoc._id
```

In this case `typeof` will return the more generic `object` type rather than `ObjectId` type.

3.5.5 `mongo` Shell Quick Reference

On this page

- `mongo` Shell Command History (page 90)
- Command Line Options (page 90)
- Command Helpers (page 90)
- Basic Shell JavaScript Operations (page 90)
- Keyboard Shortcuts (page 91)
- Queries (page 92)
- Error Checking Methods (page 94)
- Administrative Command Helpers (page 94)
- Opening Additional Connections (page 94)
- Miscellaneous (page 94)
- Additional Resources (page 95)

mongo Shell Command History

You can retrieve previous commands issued in the mongo shell with the up and down arrow keys. Command history is stored in ~/.dbshell file. See *.dbshell* for more information.

Command Line Options

The mongo shell can be started with numerous options. See mongo shell page for details on all available options.

The following table displays some common options for mongo:

Option	Description
--help	Show command line options
--nodb	Start mongo shell without connecting to a database. To connect later, see *Opening New Connections* (page 84).
--shell	Used in conjunction with a JavaScript file (i.e. <*file.js*>) to continue in the mongo shell after running the JavaScript file. See *JavaScript file* (page 86) for an example.

Command Helpers

The mongo shell provides various help. The following table displays some common help methods and commands:

Help Methods and Commands	Description
help	Show help.
db.help()	Show help for database methods.
db.<collection>.help()	Show help on collection methods. The <collection> can be the name of an existing collection or a non-existing collection.
show dbs	Print a list of all databases on the server.
use <db>	Switch current database to <db>. The mongo shell variable db is set to the current database.
show collections	Print a list of all collections for current database
show users	Print a list of users for current database.
show roles	Print a list of all roles, both user-defined and built-in, for the current database.
show profile	Print the five most recent operations that took 1 millisecond or more. See documentation on the *database profiler* (page 326) for more information.
show databases	Print a list of all available databases.
load()	Execute a JavaScript file. See *Write Scripts for the mongo Shell* (page 84) for more information.

Basic Shell JavaScript Operations

The mongo shell provides a JavaScript API for database operations.

In the mongo shell, db is the variable that references the current database. The variable is automatically set to the default database test or is set when you use the use <db> to switch current database.

The following table displays some common JavaScript operations:

JavaScript Database Operations	Description
`db.auth()`	If running in secure mode, authenticate the user.

`coll = db.<collection>`	Set a specific collection in the current database to a variable `coll`, as in the following example:
	`coll = db.myCollection;`
	You can perform operations on the `myCollection` using the variable, as in the following example:
	`coll.find();`

`db.collection.find()`	Find all documents in the collection and returns a cursor. See the `db.collection.find()` and *Query Documents* (page 140) for more information and examples. See *Cursors* (page 103) for additional information on cursor handling in the `mongo` shell.
`db.collection.insert()`	Insert a new document into the collection.
`db.collection.update()`	Update an existing document in the collection. See *Write Operations* (page 114) for more information.
`db.collection.save()`	Insert either a new document or update an existing document in the collection. See *Write Operations* (page 114) for more information.
`db.collection.remove()`	Delete documents from the collection. See *Write Operations* (page 114) for more information.
`db.collection.drop()`	Drops or removes completely the collection.
`db.collection.createIndex()`	Create a new index on the collection if the index does not exist; otherwise, the operation has no effect.
`db.getSiblingDB()`	Return a reference to another database using this same connection without explicitly switching the current database. This allows for cross database queries.

For more information on performing operations in the shell, see:

- *MongoDB CRUD Concepts* (page 99)
- *Read Operations* (page 100)
- *Write Operations* (page 114)
- *js-administrative-methods*

Keyboard Shortcuts

The `mongo` shell provides most keyboard shortcuts similar to those found in the `bash` shell or in Emacs. For some functions `mongo` provides multiple key bindings, to accommodate several familiar paradigms.

The following table enumerates the keystrokes supported by the `mongo` shell:

Keystroke	Function
Up-arrow	previous-history
Down-arrow	next-history
Home	beginning-of-line
End	end-of-line
Tab	autocomplete
Left-arrow	backward-character
Right-arrow	forward-character
Ctrl-left-arrow	backward-word

Continued on next page

Table 3.1 – continued from previous page

Keystroke	Function
Ctrl-right-arrow	forward-word
Meta-left-arrow	backward-word
Meta-right-arrow	forward-word
Ctrl-A	beginning-of-line
Ctrl-B	backward-char
Ctrl-C	exit-shell
Ctrl-D	delete-char (or exit shell)
Ctrl-E	end-of-line
Ctrl-F	forward-char
Ctrl-G	abort
Ctrl-J	accept-line
Ctrl-K	kill-line
Ctrl-L	clear-screen
Ctrl-M	accept-line
Ctrl-N	next-history
Ctrl-P	previous-history
Ctrl-R	reverse-search-history
Ctrl-S	forward-search-history
Ctrl-T	transpose-chars
Ctrl-U	unix-line-discard
Ctrl-W	unix-word-rubout
Ctrl-Y	yank
Ctrl-Z	Suspend (job control works in linux)
Ctrl-H (i.e. Backspace)	backward-delete-char
Ctrl-I (i.e. Tab)	complete
Meta-B	backward-word
Meta-C	capitalize-word
Meta-D	kill-word
Meta-F	forward-word
Meta-L	downcase-word
Meta-U	upcase-word
Meta-Y	yank-pop
Meta-[Backspace]	backward-kill-word
Meta-<	beginning-of-history
Meta->	end-of-history

Queries

In the `mongo` shell, perform read operations using the `find()` and `findOne()` methods.

The `find()` method returns a cursor object which the `mongo` shell iterates to print documents on screen. By default, `mongo` prints the first 20. The `mongo` shell will prompt the user to "`Type it`" to continue iterating the next 20 results.

The following table provides some common read operations in the `mongo` shell:

Read Operations	Description

`db.collection.find(<query>)`	Find the documents matching the `<query>` criteria in the collection. If the `<query>` criteria is not specified or is empty (i.e `{}`), the read operation selects all documents in the collection.
	The following example selects the documents in the `users` collection with the `name` field equal to `"Joe"`:
	`coll = db.users;`
	`coll.find( { name: "Joe" } );`
	For more information on specifying the `<query>` criteria, see *Query Documents* (page 140).

`db.collection.find(<query>, <projection>)`	Find documents matching the `<query>` criteria and return just specific fields in the `<projection>`.
	The following example selects all documents from the collection but returns only the `name` field and the `_id` field. The `_id` is always returned unless explicitly specified to not return.
	`coll = db.users;`
	`coll.find( { }, { name: true } );`
	For more information on specifying the `<projection>`, see *Limit Fields to Return from a Query* (page 153).

`db.collection.find().sort(<sort order>)`	Return results in the specified `<sort order>`.
	The following example selects all documents from the collection and returns the results sorted by the `name` field in ascending order (`1`). Use `-1` for descending order:
	`coll = db.users;`
	`coll.find().sort( { name: 1 } );`

`db.collection.find(<query>).sort(<sort order>)`	Return the documents matching the `<query>` criteria in the specified `<sort order>`.
`db.collection.find( ... ).limit( <n> )`	Limit result to `<n>` rows. Highly recommended if you need only a certain number of rows for best performance.
`db.collection.find( ... ).skip( <n> )`	Skip `<n>` results.
`db.collection.count()`	Returns total number of documents in the collection.
`db.collection.find(<query>).count()`	Returns the total number of documents that match the query. The `count()` ignores `limit()` and `skip()`. For example, if 100 records match but the limit is 10, `count()` will return 100. This will be faster than iterating yourself, but still take time.

`db.collection.findOne(<query>)`	Find and return a single document. Returns null if not found.
	The following example selects a single document in the `users` collection with the `name` field matches to `"Joe"`:
	`coll = db.users;`
	`coll.findOne( { name: "Joe" } );`
	Internally, the `findOne()` method is the `find()` method with a `limit(1)`.

See *Query Documents* (page 140) and *Read Operations* (page 100) documentation for more information and examples. See `https://docs.mongodb.org/manual/reference/operator/query` to specify other query operators.

Error Checking Methods

Changed in version 2.6.

The `mongo` shell write methods now integrates the *Write Concern* (page 179) directly into the method execution rather than with a separate `db.getLastError()` method. As such, the write methods now return a `WriteResult()` object that contains the results of the operation, including any write errors and write concern errors.

Previous versions used `db.getLastError()` and `db.getLastErrorObj()` methods to return error information.

Administrative Command Helpers

The following table lists some common methods to support database administration:

JavaScript Database Administration Methods	Description
`db.cloneDatabase(<host>)`	Clone the current database from the `<host>` specified. The `<host>` database instance must be in noauth mode.
`db.copyDatabase(<from>, <to>, <host>)`	Copy the `<from>` database from the `<host>` to the `<to>` database on the current server. The `<host>` database instance must be in `noauth` mode.
`db.fromColl.renameCollection(<toColl>)`	Rename collection from `fromColl` to `<toColl>`.
`db.repairDatabase()`	Repair and compact the current database. This operation can be very slow on large databases.
`db.getCollectionNames()`	Get the list of all collections in the current database.
`db.dropDatabase()`	Drops the current database.

See also *administrative database methods* for a full list of methods.

Opening Additional Connections

You can create new connections within the `mongo` shell.

The following table displays the methods to create the connections:

JavaScript Connection Create Methods	Description
`db = connect("<host><:port>/<dbname>")`	Open a new database connection.
`conn = new Mongo()` `db = conn.getDB("dbname")`	Open a connection to a new server using `new Mongo()`. Use `getDB()` method of the connection to select a database.

See also *Opening New Connections* (page 84) for more information on the opening new connections from the `mongo` shell.

Miscellaneous

The following table displays some miscellaneous methods:

Method	Description
`Object.bsonsize(<document>)`	Prints the *BSON* size of a <document> in bytes

See the MongoDB JavaScript API Documentation[4] for a full list of JavaScript methods .

Additional Resources

Consider the following reference material that addresses the `mongo` shell and its interface:

- `mongo`
- *js-administrative-methods*
- *database-commands*
- *Aggregation Reference* (page 225)
- Getting Started Guide[5]

Additionally, the MongoDB source code repository includes a jstests directory[6] which contains numerous `mongo` shell scripts.

[4] http://api.mongodb.org/js/index.html
[5] https://docs.mongodb.org/getting-started/shell
[6] https://github.com/mongodb/mongo/tree/master/jstests/

CHAPTER 4

MongoDB CRUD Operations

MongoDB provides rich semantics for reading and manipulating data. CRUD stands for *create*, *read*, *update*, and *delete*. These terms are the foundation for all interactions with the database.

MongoDB CRUD Introduction **(page 97)** An introduction to the MongoDB data model as well as queries and data manipulations.

MongoDB CRUD Concepts **(page 99)** The core documentation of query and data manipulation.

MongoDB CRUD Tutorials **(page 136)** Examples of basic query and data modification operations.

MongoDB CRUD Reference **(page 178)** Reference material for the query and data manipulation interfaces.

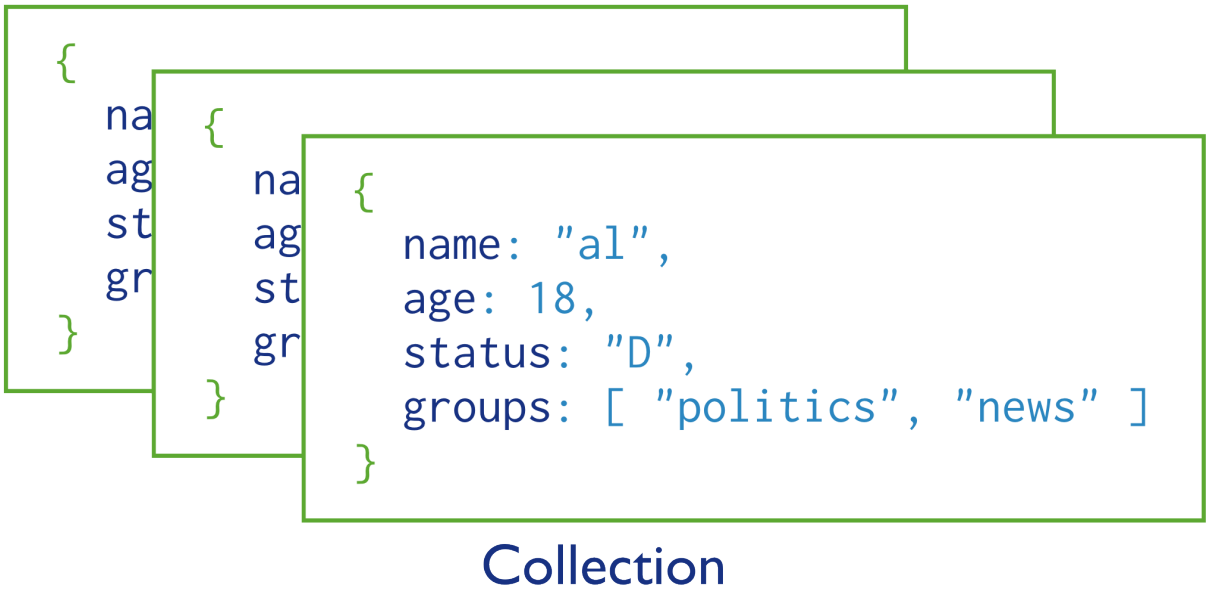
4.1 MongoDB CRUD Introduction

On this page

- Database Operations (page 98)

MongoDB stores data in the form of *documents*, which are JSON-like field and value pairs. Documents are analogous to structures in programming languages that associate keys with values (e.g. dictionaries, hashes, maps, and associative arrays). Formally, MongoDB documents are *BSON* documents. BSON is a binary representation of *JSON* with additional type information. In the documents, the value of a field can be any of the BSON data types, including other documents, arrays, and arrays of documents. For more information, see *Documents* (page 8).

MongoDB stores all documents in *collections*. A collection is a group of related documents that have a set of shared common indexes. Collections are analogous to a table in relational databases.

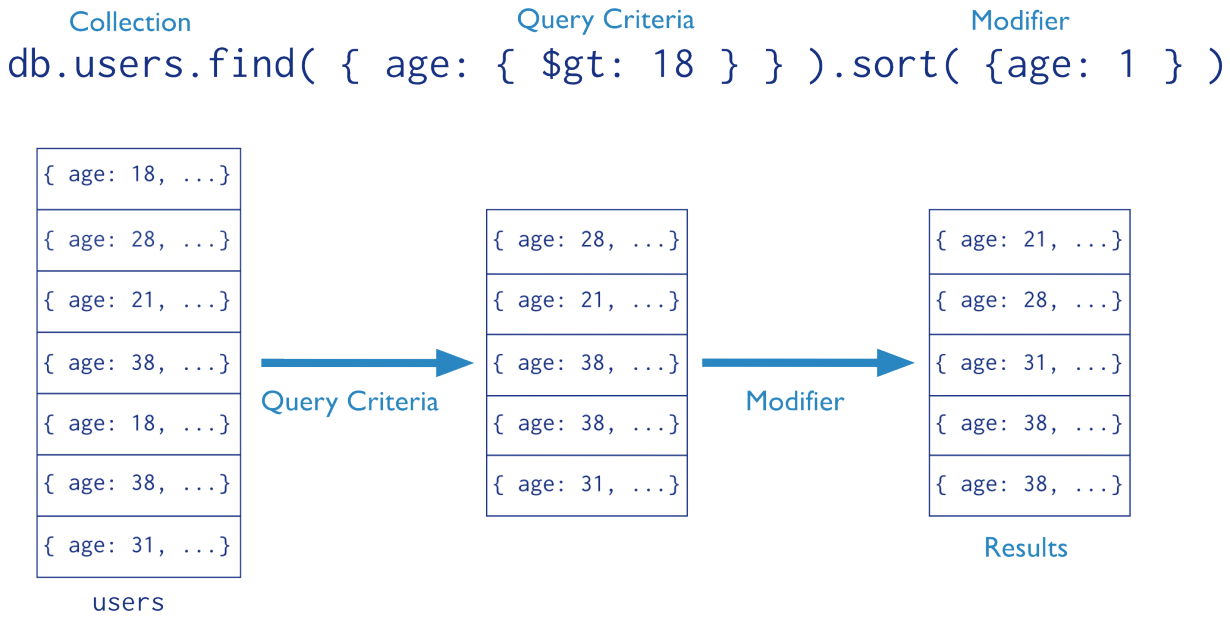


4.1.1 Database Operations

Query

In MongoDB a query targets a specific collection of documents. Queries specify criteria, or conditions, that identify the documents that MongoDB returns to the clients. A query may include a *projection* that specifies the fields from the matching documents to return. You can optionally modify queries to impose limits, skips, and sort orders.

In the following diagram, the query process specifies a query criteria and a sort modifier:

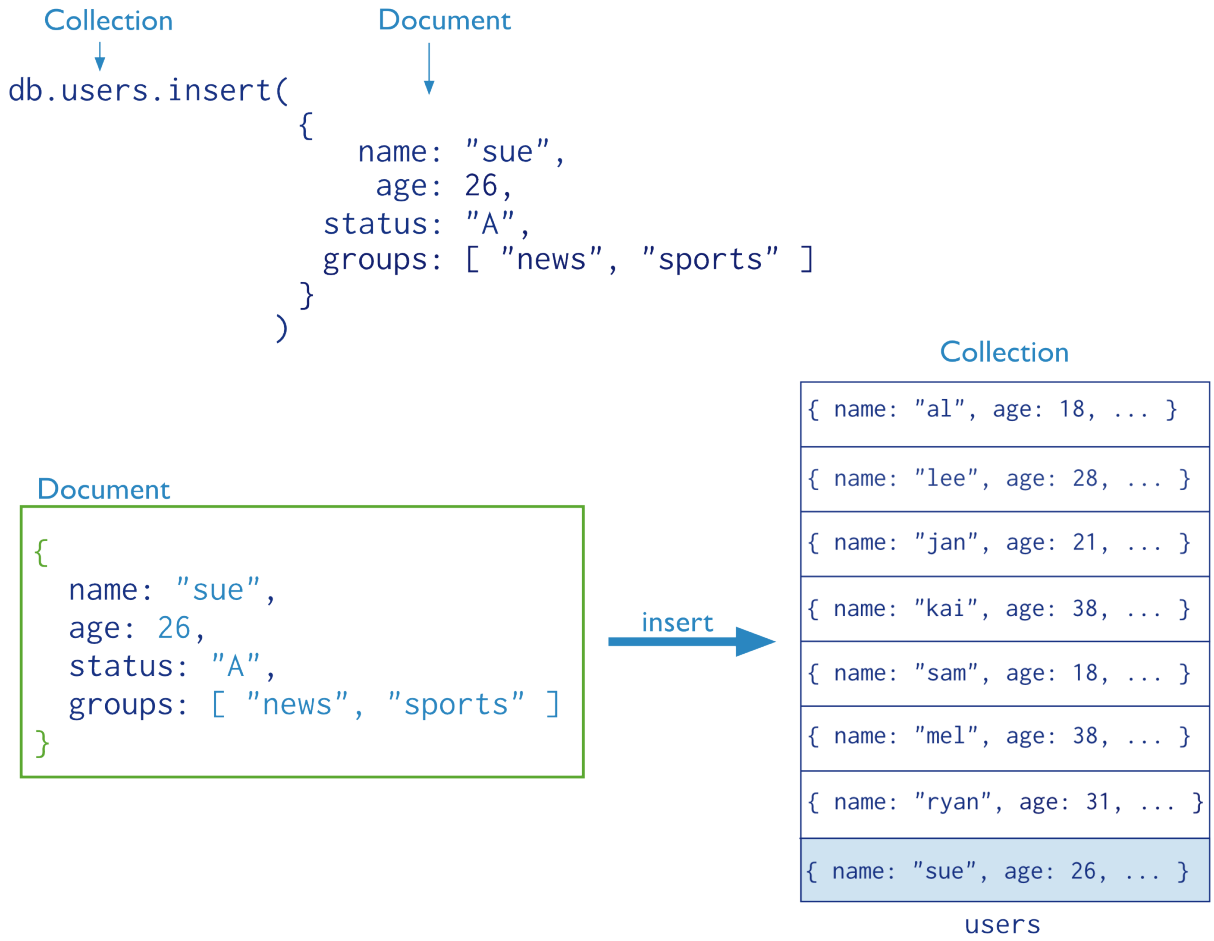


See *Read Operations Overview* (page 100) for more information.

Data Modification

Data modification refers to operations that create, update, or delete data. In MongoDB, these operations modify the data of a single *collection*. For the update and delete operations, you can specify the criteria to select the documents to update or remove.

In the following diagram, the insert operation adds a new document to the `users` collection.



See *Write Operations Overview* (page 114) for more information.

4.2 MongoDB CRUD Concepts

The *Read Operations* (page 100) and *Write Operations* (page 114) documents introduce the behavior and operations of read and write operations for MongoDB deployments.

Read Operations **(page 100)** Queries are the core operations that return data in MongoDB. Introduces queries, their behavior, and performances.

> ***Cursors*** **(page 103)** Queries return iterable objects, called cursors, that hold the full result set.
>
> ***Query Optimization*** **(page 105)** Analyze and improve query performance.
>
> ***Distributed Queries*** **(page 110)** Describes how *sharded clusters* and *replica sets* affect the performance of read operations.

Write Operations **(page 114)** Write operations insert, update, or remove documents in MongoDB. Introduces data create and modify operations, their behavior, and performances.

Atomicity and Transactions **(page 125)** Describes write operation atomicity in MongoDB.

Distributed Write Operations **(page 126)** Describes how MongoDB directs write operations on *sharded clusters* and *replica sets* and the performance characteristics of these operations.

Continue reading from *Write Operations* (page 114) for additional background on the behavior of data modification operations in MongoDB.

4.2.1 Read Operations

The following documents describe read operations:

Read Operations Overview **(page 100)** A high level overview of queries and projections in MongoDB, including a discussion of syntax and behavior.

Cursors **(page 103)** Queries return iterable objects, called cursors, that hold the full result set.

Query Optimization **(page 105)** Analyze and improve query performance.

Query Plans **(page 108)** MongoDB executes queries using optimal *plans*.

Distributed Queries **(page 110)** Describes how *sharded clusters* and *replica sets* affect the performance of read operations.

Read Operations Overview

On this page

- Query Interface (page 100)
- Query Behavior (page 101)
- Query Statements (page 101)
- Projections (page 102)

Read operations, or *queries*, retrieve data stored in the database. In MongoDB, queries select *documents* from a single *collection*.

Queries specify criteria, or conditions, that identify the documents that MongoDB returns to the clients. A query may include a *projection* that specifies the fields from the matching documents to return. The projection limits the amount of data that MongoDB returns to the client over the network.

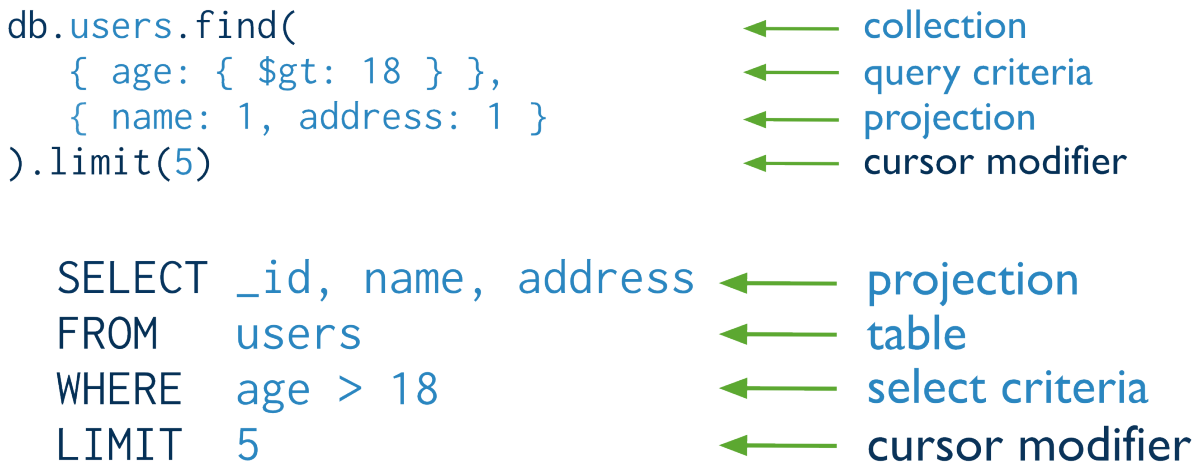
Query Interface

For query operations, MongoDB provides a `db.collection.find()` method. The method accepts both the query criteria and projections and returns a *cursor* (page 103) to the matching documents. You can optionally modify the query to impose limits, skips, and sort orders.

The following diagram highlights the components of a MongoDB query operation:

The next diagram shows the same query in SQL:

Example




```
db.users.find( { age: { $gt: 18 } }, { name: 1, address: 1 } ).limit(5)
```

This query selects the documents in the `users` collection that match the condition `age` is greater than `18`. To specify the greater than condition, query criteria uses the greater than (i.e. `$gt`) *query selection operator*. The query returns at most `5` matching documents (or more precisely, a cursor to those documents). The matching documents will return with only the `_id`, `name` and `address` fields. See *Projections* (page 102) for details.

See

SQL to MongoDB Mapping Chart (page 183) for additional examples of MongoDB queries and the corresponding SQL statements.

Query Behavior

MongoDB queries exhibit the following behavior:

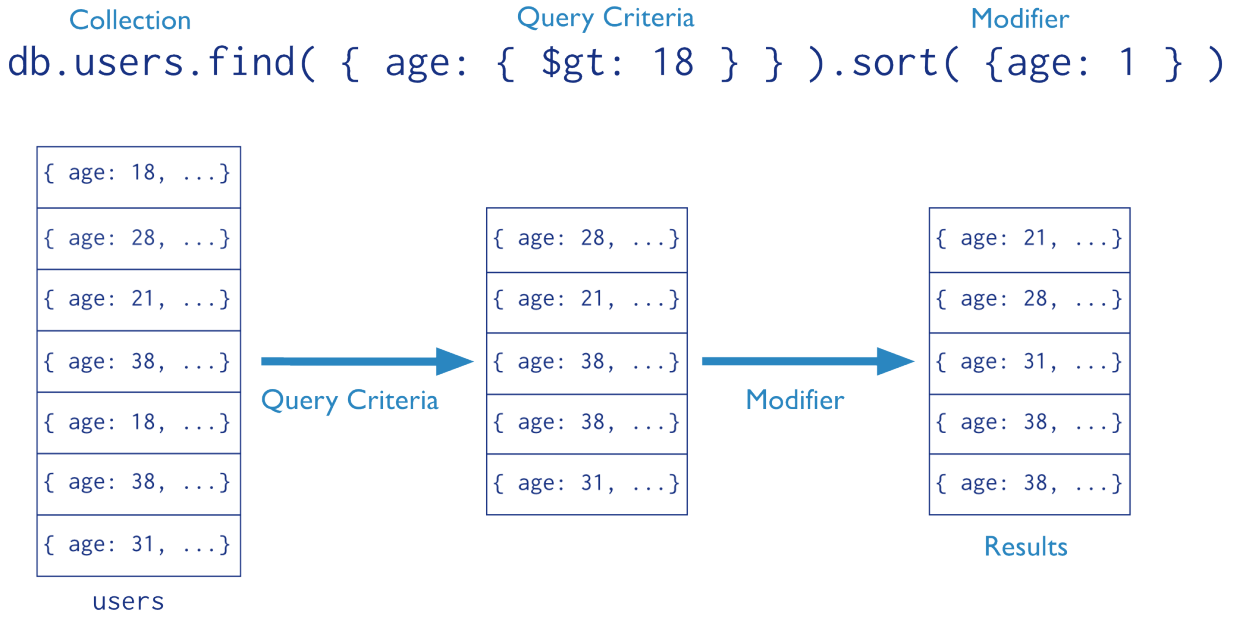
- All queries in MongoDB address a *single* collection.
- You can modify the query to impose `limits`, `skips`, and `sort orders`.
- The order of documents returned by a query is not defined unless you specify a `sort()`.
- Operations that *modify existing documents* (page 148) (i.e. *updates*) use the same query syntax as queries to select documents to update.
- In *aggregation pipeline* (page 199), the `$match` pipeline stage provides access to MongoDB queries.

MongoDB provides a `db.collection.findOne()` method as a special case of `find()` that returns a single document.

Query Statements

Consider the following diagram of the query process that specifies a query criteria and a sort modifier:

In the diagram, the query selects documents from the `users` collection. Using a `query selection operator` to define the conditions for matching documents, the query selects documents that have `age` greater than (i.e. `$gt`) `18`. Then the `sort()` modifier sorts the results by `age` in ascending order.



For additional examples of queries, see *Query Documents* (page 140).

Projections

Queries in MongoDB return all fields in all matching documents by default. To limit the amount of data that MongoDB sends to applications, include a *projection* in the queries. By projecting results with a subset of fields, applications reduce their network overhead and processing requirements.

Projections, which are the *second* argument to the `find()` method, may either specify a list of fields to return *or* list fields to exclude in the result documents.

Important: Except for excluding the `_id` field in inclusive projections, you cannot mix exclusive and inclusive projections.

Consider the following diagram of the query process that specifies a query criteria and a projection:

In the diagram, the query selects from the `users` collection. The criteria matches the documents that have `age` equal to `18`. Then the projection specifies that only the `name` field should return in the matching documents.

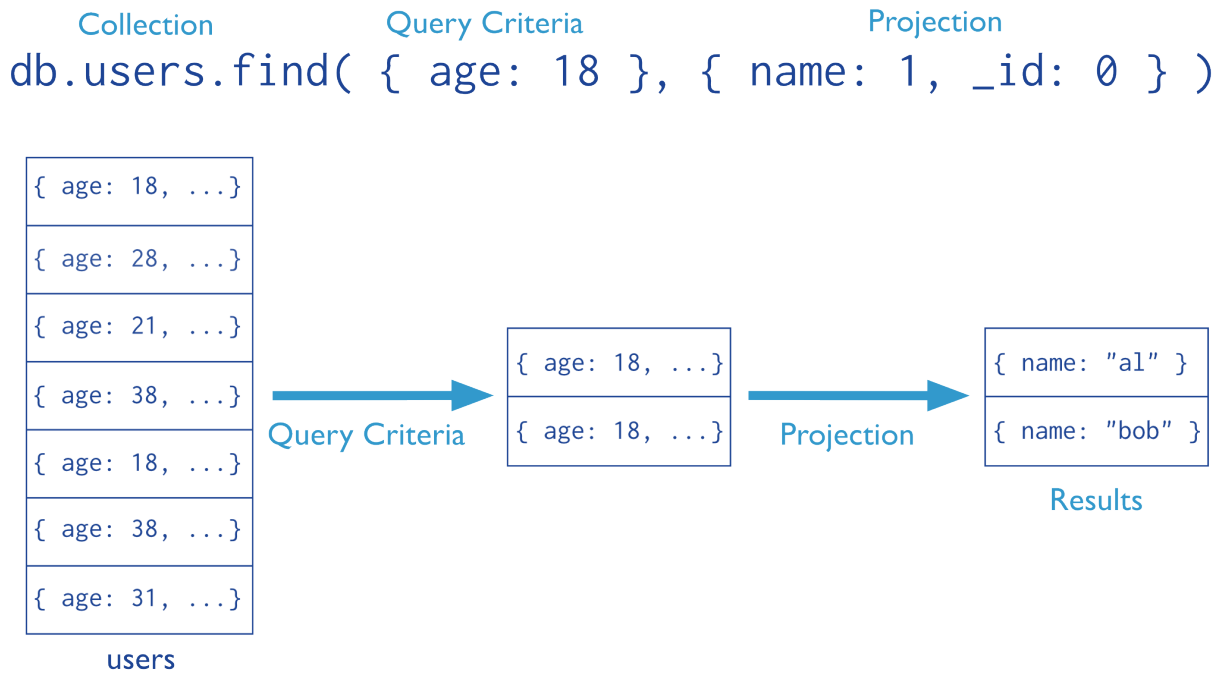
Projection Examples

Exclude One Field From a Result Set

```
db.records.find( { "user_id": { $lt: 42 } }, { "history": 0 } )
```

This query selects documents in the `records` collection that match the condition `{ "user_id": { $lt: 42 } }`, and uses the projection `{ "history": 0 }` to exclude the `history` field from the documents in the result set.

Return Two fields *and* the `_id` Field




```
db.records.find( { "user_id": { $lt: 42 } }, { "name": 1, "email": 1 } )
```

This query selects documents in the `records` collection that match the query `{ "user_id": { $lt: 42 } }` and uses the projection `{ "name": 1, "email": 1 }` to return just the `_id` field (implicitly included), `name` field, and the `email` field in the documents in the result set.

Return Two Fields *and* Exclude `_id`

```
db.records.find( { "user_id": { $lt: 42} }, { "_id": 0, "name": 1 , "email": 1 } )
```

This query selects documents in the `records` collection that match the query `{ "user_id": { $lt: 42} }`, and only returns the `name` and `email` fields in the documents in the result set.

See

Limit Fields to Return from a Query (page 153) for more examples of queries with projection statements.

Projection Behavior MongoDB projections have the following properties:

- By default, the `_id` field is included in the results. To suppress the `_id` field from the result set, specify `_id: 0` in the projection document.
- For fields that contain arrays, MongoDB provides the following projection operators: `$elemMatch`, `$slice`, and `$`.
- For related projection functionality in the *aggregation pipeline* (page 199) , use the `$project` pipeline stage.

Cursors

On this page

- Cursor Behaviors (page 104)
- Cursor Information (page 105)

In the `mongo` shell, the primary method for the read operation is the `db.collection.find()` method. This method queries a collection and returns a *cursor* to the returning documents.

To access the documents, you need to iterate the cursor. However, in the `mongo` shell, if the returned cursor is not assigned to a variable using the `var` keyword, then the cursor is automatically iterated up to 20 times [1] to print up to the first 20 documents in the results.

For example, in the `mongo` shell, the following read operation queries the `inventory` collection for documents that have `type` equal to `'food'` and automatically print up to the first 20 matching documents:

```
db.inventory.find( { type: 'food' } );
```

To manually iterate the cursor to access the documents, see *Iterate a Cursor in the mongo Shell* (page 158).

Cursor Behaviors

Closure of Inactive Cursors By default, the server will automatically close the cursor after 10 minutes of inactivity, or if client has exhausted the cursor. To override this behavior in the `mongo` shell, you can use the `cursor.noCursorTimeout()` method:

```
var myCursor = db.inventory.find().noCursorTimeout();
```

After setting the `noCursorTimeout` option, you must either close the cursor manually with `cursor.close()` or by exhausting the cursor's results.

See your `driver` documentation for information on setting the `noCursorTimeout` option.

Cursor Isolation As a cursor returns documents, other operations may interleave with the query. For the *MMAPv1* (page 603) storage engine, intervening write operations on a document may result in a cursor that returns a document more than once if that document has changed. To handle this situation, see the information on *snapshot mode* (page 135).

Cursor Batches The MongoDB server returns the query results in batches. Batch size will not exceed the *maximum BSON document size*. For most queries, the *first* batch returns 101 documents or just enough documents to exceed 1 megabyte. Subsequent batch size is 4 megabytes. To override the default size of the batch, see `batchSize()` and `limit()`.

For queries that include a sort operation *without* an index, the server must load all the documents in memory to perform the sort before returning any results.

As you iterate through the cursor and reach the end of the returned batch, if there are more results, `cursor.next()` will perform a `getmore operation` to retrieve the next batch. To see how many documents remain in the batch as you iterate the cursor, you can use the `objsLeftInBatch()` method, as in the following example:

```
var myCursor = db.inventory.find();

var myFirstDocument = myCursor.hasNext() ? myCursor.next() : null;
```

[1] You can use the `DBQuery.shellBatchSize` to change the number of iteration from the default value `20`. See *Working with the mongo Shell* (page 78) for more information.

```
myCursor.objsLeftInBatch();
```

Cursor Information

The `db.serverStatus()` method returns a document that includes a `metrics` field. The `metrics` field contains a `metrics.cursor` field with the following information:

- number of timed out cursors since the last server restart
- number of open cursors with the option `DBQuery.Option.noTimeout` set to prevent timeout after a period of inactivity
- number of "pinned" open cursors
- total number of open cursors

Consider the following example which calls the `db.serverStatus()` method and accesses the `metrics` field from the results and then the `cursor` field from the `metrics` field:

```
db.serverStatus().metrics.cursor
```

The result is the following document:

```
{
   "timedOut" : <number>
   "open" : {
      "noTimeout" : <number>,
      "pinned" : <number>,
      "total" : <number>
   }
}
```

See also:

`db.serverStatus()`

Query Optimization

On this page

- Create an Index to Support Read Operations (page 105)
- Query Selectivity (page 106)
- Covered Query (page 106)

Indexes improve the efficiency of read operations by reducing the amount of data that query operations need to process. This simplifies the work associated with fulfilling queries within MongoDB.

Create an Index to Support Read Operations

If your application queries a collection on a particular field or set of fields, then an index on the queried field or a *compound index* (page 522) on the set of fields can prevent the query from scanning the whole collection to find and return the query results. For more information about indexes, see the *complete documentation of indexes in MongoDB* (page 515).

Example

An application queries the `inventory` collection on the `type` field. The value of the `type` field is user-driven.

```
var typeValue = <someUserInput>;
db.inventory.find( { type: typeValue } );
```

To improve the performance of this query, add an ascending or a descending index to the `inventory` collection on the `type` field. [2] In the `mongo` shell, you can create indexes using the `db.collection.createIndex()` method:

```
db.inventory.createIndex( { type: 1 } )
```

This index can prevent the above query on `type` from scanning the whole collection to return the results.

To analyze the performance of the query with an index, see *Analyze Query Performance* (page 159).

In addition to optimizing read operations, indexes can support sort operations and allow for a more efficient storage utilization. See `db.collection.createIndex()` and *Indexes* (page 515) for more information about index creation.

Query Selectivity

Query selectivity refers to how well the query predicate excludes or filters out documents in a collection. Query selectivity can determine whether or not queries can use indexes effectively or even use indexes at all.

More selective queries match a smaller percentage of documents. For instance, an equality match on the unique `_id` field is highly selective as it can match at most one document.

Less selective queries match a larger percentage of documents. Less selective queries cannot use indexes effectively or even at all.

For instance, the inequality operators `$nin` and `$ne` are *not* very selective since they often match a large portion of the index. As a result, in many cases, a `$nin` or `$ne` query with an index may perform no better than a `$nin` or `$ne` query that must scan all documents in a collection.

The selectivity of `regular expressions` depends on the expressions themselves. For details, see *regular expression and index use*.

Covered Query

An index *covers* (page 106) a query when both of the following apply:

- all the fields in the *query* (page 140) are part of an index, **and**
- all the fields returned in the results are in the same index.

For example, a collection `inventory` has the following index on the `type` and `item` fields:

```
db.inventory.createIndex( { type: 1, item: 1 } )
```

This index will cover the following operation which queries on the `type` and `item` fields and returns only the `item` field:

[2] For single-field indexes, the selection between ascending and descending order is immaterial. For compound indexes, the selection is important. See *indexing order* (page 524) for more details.

```
db.inventory.find(
   { type: "food", item:/^c/ },
   { item: 1, _id: 0 }
)
```

For the specified index to cover the query, the projection document must explicitly specify `_id: 0` to exclude the `_id` field from the result since the index does not include the `_id` field.

Performance Because the index contains all fields required by the query, MongoDB can both match the *query conditions* (page 140) and return the results using only the index.

Querying *only* the index can be much faster than querying documents outside of the index. Index keys are typically smaller than the documents they catalog, and indexes are typically available in RAM or located sequentially on disk.

Limitations

Restrictions on Indexed Fields An index **cannot** cover a query if:

- any of the indexed fields in any of the documents in the collection includes an array. If an indexed field is an array, the index becomes a *multi-key index* (page 525) and cannot support a covered query.
- any of the indexed fields in the query predicate or returned in the projection are fields in embedded documents. [3] For example, consider a collection `users` with documents of the following form:

  ```
  { _id: 1, user: { login: "tester" } }
  ```

  The collection has the following index:

  ```
  { "user.login": 1 }
  ```

  The `{ "user.login": 1 }` index does **not** cover the following query:

  ```
  db.users.find( { "user.login": "tester" }, { "user.login": 1, _id: 0 } )
  ```

  However, the query can use the `{ "user.login": 1 }` index to find matching documents.

Restrictions on Sharded Collection An index cannot cover a query on a *sharded* collection when run against a `mongos` if the index does not contain the shard key, with the following exception for the `_id` index: If a query on a sharded collection only specifies a condition on the `_id` field and returns only the `_id` field, the `_id` index can cover the query when run against a `mongos` even if the `_id` field is not the shard key.

Changed in version 3.0: In previous versions, an index cannot *cover* (page 106) a query on a *sharded* collection when run against a `mongos`.

`explain` To determine whether a query is a covered query, use the `db.collection.explain()` or the `explain()` method and review the *results*.

`db.collection.explain()` provides information on the execution of other operations, such as `db.collection.update()`. See `db.collection.explain()` for details.

For more information see *Measure Index Use* (page 584).

[3] To index fields in embedded documents, use *dot notation*.

Query Plans

On this page

- Plan Cache Flushes (page 108)
- Index Filters (page 108)

The MongoDB query optimizer processes queries and chooses the most efficient query plan for a query given the available indexes. The query system then uses this query plan each time the query runs.

The query optimizer only caches the plans for those query shapes that can have more than one viable plan.

For each query, the query planner searches the query plan cache for an entry that fits the *query shape*. If there are no matching entries, the query planner generates candidate plans for evaluation over a trial period. The query planner chooses a winning plan, creates a cache entry containing the winning plan, and uses it to generate the result documents.

If a matching entry exists, the query planner generates a plan based on that entry and evaluates its performance through a `replanning` mechanism. This mechanism makes a `pass/fail` decision based on the plan performance and either keeps or evicts the cache entry. On eviction, the query planner selects a new plan using the normal planning process and caches it. The query planner executes the plan and returns the result documents for the query.

The following diagram illustrates the query planner logic:

See *Plan Cache Flushes* (page 108) for additional scenarios that trigger changes to the plan cache.

You can use the `db.collection.explain()` or the `cursor.explain()` method to view statistics about the query plan for a given query. This information can help as you develop *indexing strategies* (page 586).

`db.collection.explain()` provides information on the execution of other operations, such as `db.collection.update()`. See `db.collection.explain()` for details.

Changed in version 2.6: `explain()` operations no longer read from or write to the query planner cache.

Plan Cache Flushes

Catalog operations like index or collection drops flush the plan cache.

The plan cache does not persist if a `mongod` restarts or shuts down.

New in version 2.6: MongoDB provides `https://docs.mongodb.org/manual/reference/method/js-plan-cache` to view and modify the cached query plans. The `PlanCache.clear()` method flushes the entire plan cache. Users can also clear particular plan cache entries using `PlanCache.clearPlansByQuery()`.

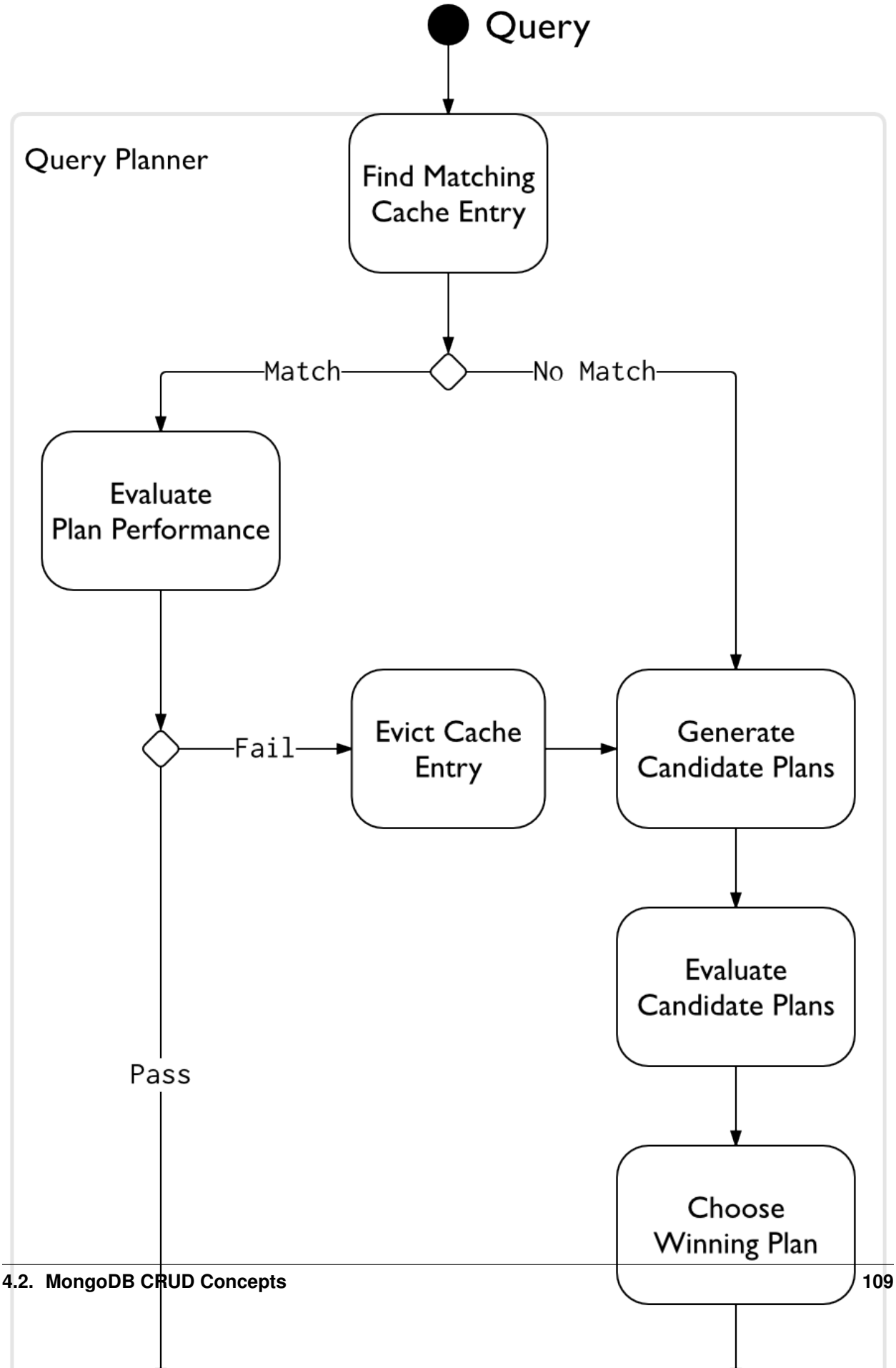
Index Filters

New in version 2.6.

Index filters determine which indexes the optimizer evaluates for a *query shape*. A query shape consists of a combination of query, sort, and projection specifications. If an index filter exists for a given query shape, the optimizer only considers those indexes specified in the filter.

When an index filter exists for the query shape, MongoDB ignores the `hint()`. To see whether MongoDB applied an index filter for a query shape, check the `indexFilterSet` field of either the `db.collection.explain()` or the `cursor.explain()` method.

Index filters only affects which indexes the optimizer evaluates; the optimizer may still select the collection scan as the winning plan for a given query shape.


Query
Query Planner
Find Matching Cache Entry
Match
No Match
Evaluate Plan Performance
Fail
Evict Cache Entry
Generate Candidate Plans
Evaluate Candidate Plans
Pass
Choose Winning Plan

Index filters exist for the duration of the server process and do not persist after shutdown. MongoDB also provides a command to manually remove filters.

Because index filters overrides the expected behavior of the optimizer as well as the `hint()` method, use index filters sparingly.

See `planCacheListFilters`, `planCacheClearFilters`, and `planCacheSetFilter`.

Distributed Queries

On this page

- Read Operations to Sharded Clusters (page 110)
- Read Operations to Replica Sets (page 110)

Read Operations to Sharded Clusters

Sharded clusters allow you to partition a data set among a cluster of `mongod` instances in a way that is nearly transparent to the application. For an overview of sharded clusters, see the *Sharding* (page 733) section of this manual.

For a sharded cluster, applications issue operations to one of the `mongos` instances associated with the cluster.

Read operations on sharded clusters are most efficient when directed to a specific shard. Queries to sharded collections should include the collection's *shard key* (page 747). When a query includes a shard key, the `mongos` can use cluster metadata from the *config database* (page 742) to route the queries to shards.

If a query does not include the shard key, the `mongos` must direct the query to *all* shards in the cluster. These *scatter gather* queries can be inefficient. On larger clusters, scatter gather queries are unfeasible for routine operations.

For replica set shards, read operations from secondary members of replica sets may not reflect the current state of the primary. Read preferences that direct read operations to different servers may result in non-monotonic reads.

For more information on read operations in sharded clusters, see the *Sharded Cluster Query Routing* (page 752) and *Shard Keys* (page 747) sections.

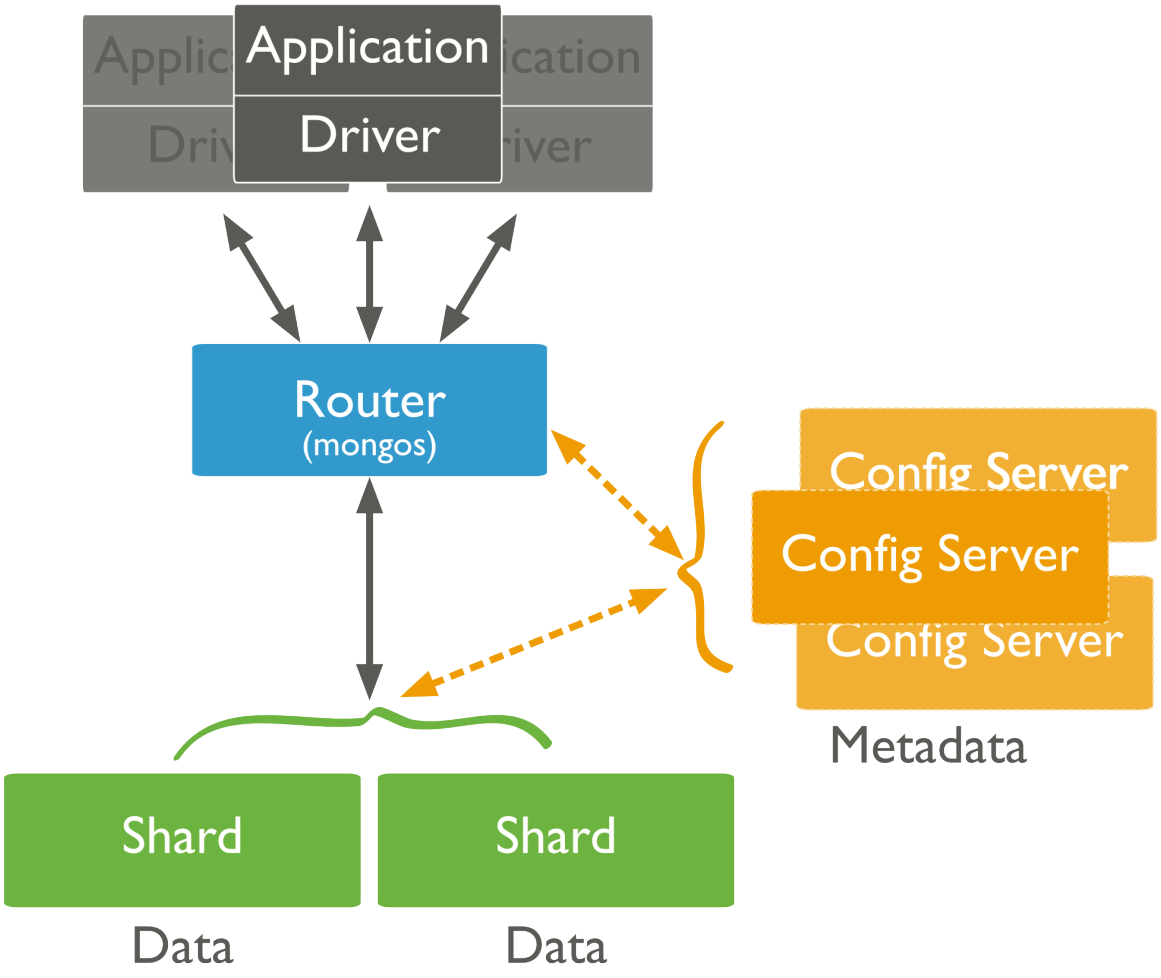
Read Operations to Replica Sets

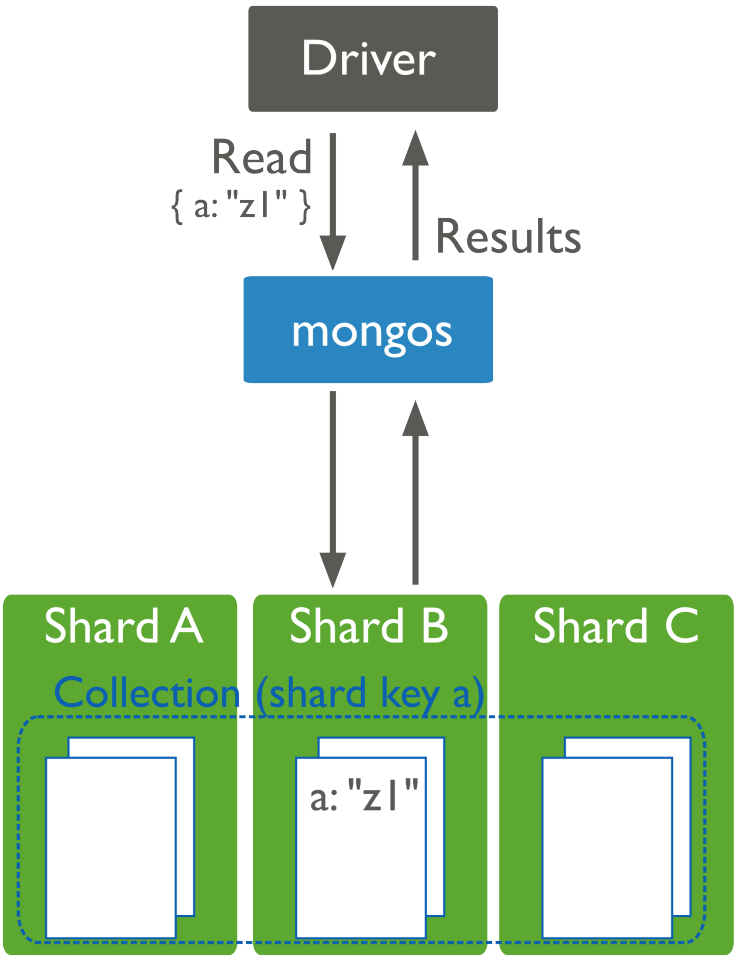
By default, clients reads from a replica set's *primary*; however, clients can specify a *read preference* (page 651) to direct read operations to other members. For example, clients can configure read preferences to read from secondaries or from nearest member to:

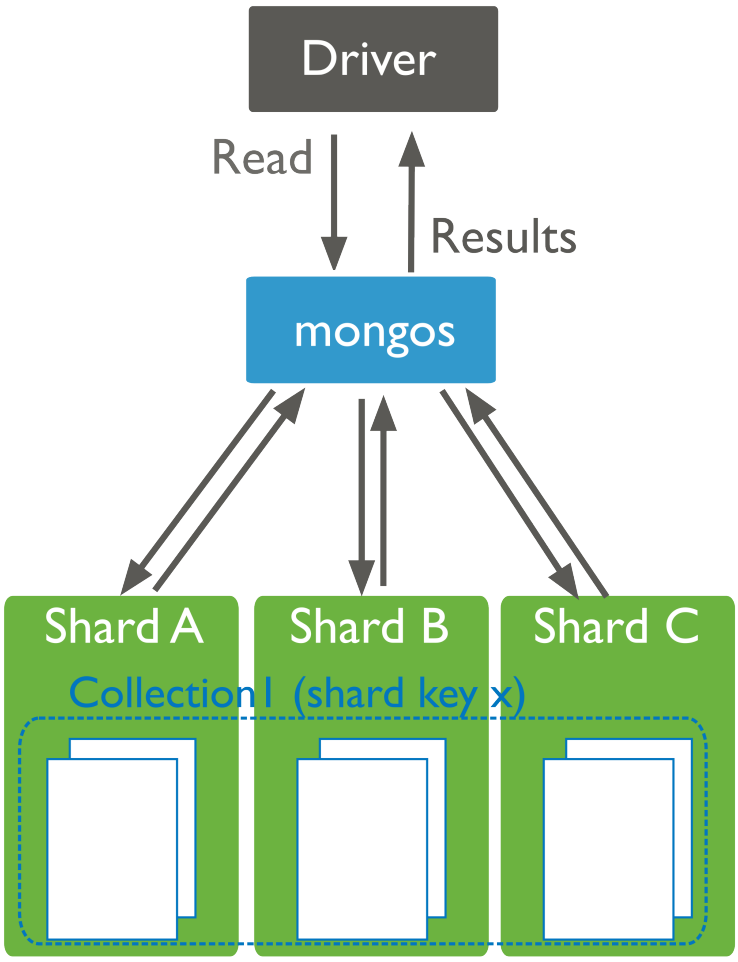
- reduce latency in multi-data-center deployments,
- improve read throughput by distributing high read-volumes (relative to write volume),
- perform backup operations, and/or
- allow reads until a *new primary is elected* (page 644).

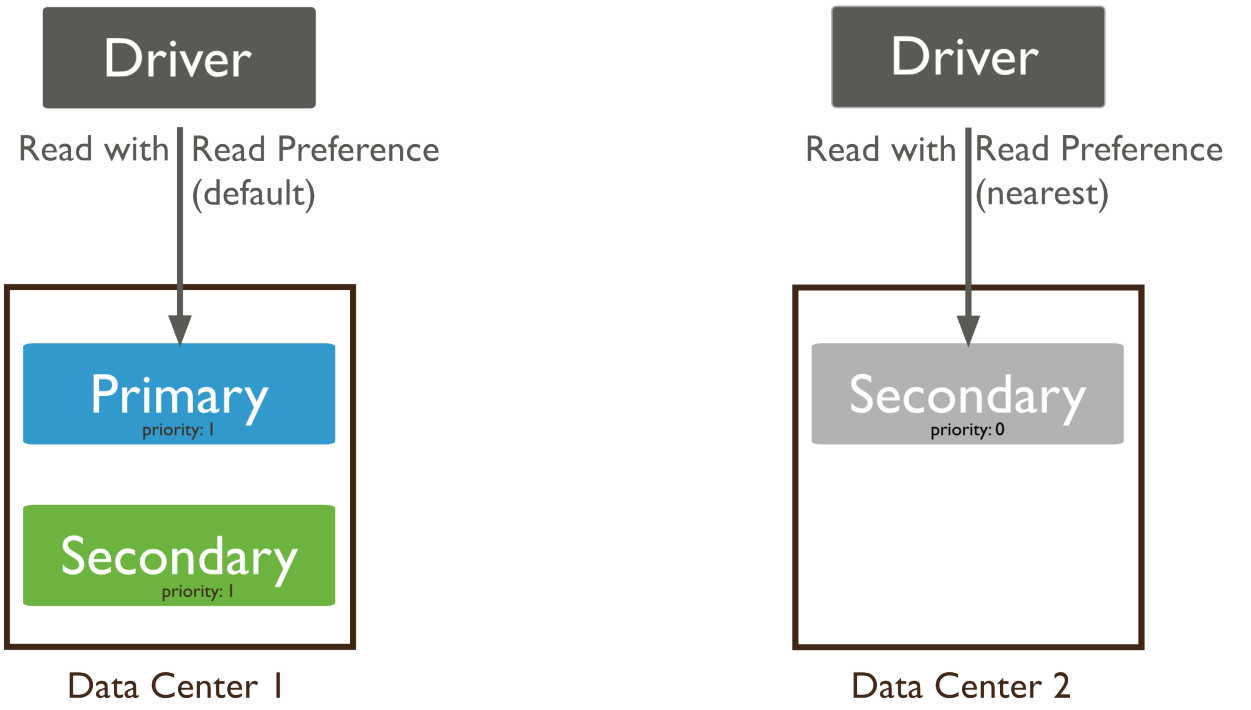
Read operations from secondary members of replica sets may not reflect the current state of the primary. Read preferences that direct read operations to different servers may result in non-monotonic reads.

You can configure the read preferece on a per-connection or per-operation basis. For more information on read preference or on the read preference modes, see *Read Preference* (page 651) and *Read Preference Modes* (page 728).


Application
Driver
Router
(mongos)
Config Server
Config Server
Config Server
Metadata
Shard
Shard
Data
Data


Driver
Read
{ a: "z | " }
Results
mongos
Shard A
Shard B
Shard C
Collection (shard key a)
a: "z | "


Driver
Read
Results
mongos
Shard A
Shard B
Shard C
Collection 1 (shard key x)



4.2.2 Write Operations

The following documents describe write operations:

Write Operations Overview **(page 114)** Provides an overview of MongoDB's data insertion and modification operations, including aspects of the syntax, and behavior.

Atomicity and Transactions **(page 125)** Describes write operation atomicity in MongoDB.

Distributed Write Operations **(page 126)** Describes how MongoDB directs write operations on *sharded clusters* and *replica sets* and the performance characteristics of these operations.

Write Operation Performance **(page 129)** Introduces the performance constraints and factors for writing data to MongoDB deployments.

Bulk Write Operations **(page 130)** Provides an overview of MongoDB's bulk write operations.

Write Operations Overview

On this page

- Insert (page 115)
- Update (page 118)
- Delete (page 122)
- Additional Methods (page 124)

A write operation is any operation that creates or modifies data in the MongoDB instance. In MongoDB, write operations target a single *collection*. All write operations in MongoDB are atomic on the level of a single *document*.

There are three classes of write operations in MongoDB: *insert* (page 115), *update* (page 118), and *delete* (page 122). Insert operations add new documents to a collection. Update operations modify existing documents, and delete operations delete documents from a collection. No insert, update, or delete can affect more than one document atomically.

For the update and remove operations, you can specify criteria, or filters, that identify the documents to update or remove. These operations use the same query syntax to specify the criteria as *read operations* (page 100).

MongoDB allows applications to determine the acceptable level of acknowledgement required of write operations. See *Write Concern* (page 179) for more information.

Insert

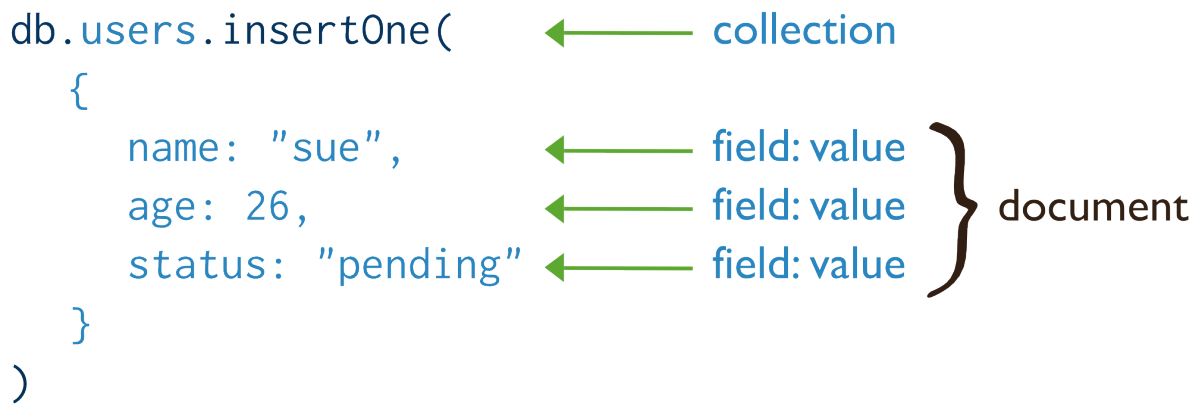
MongoDB provides the following methods for inserting documents into a collection:

- `db.collection.insertOne()`
- `db.collection.insertMany()`
- `db.collection.insert()`

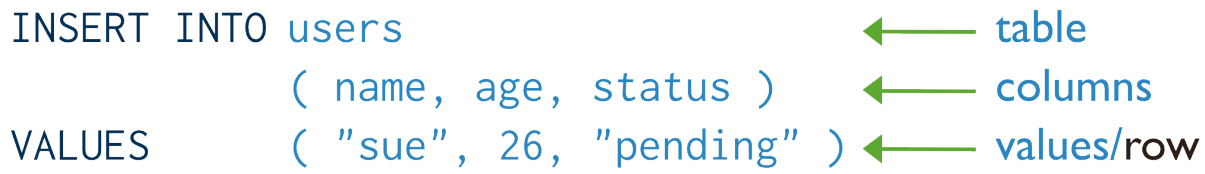
insertOne New in version 3.2.

`db.collection.insertOne()` inserts a *single* document

The following diagram highlights the components of the MongoDB `insertOne()` operation:



The following diagram shows the same query in SQL:



Example

The following operation inserts a new document into the `users` collection. The new document has three fields `name`, `age`, and `status`. Since the document does not specify an `_id` field, MongoDB adds the `_id` field and a generated value to the new document. See *Insert Behavior* (page 118).

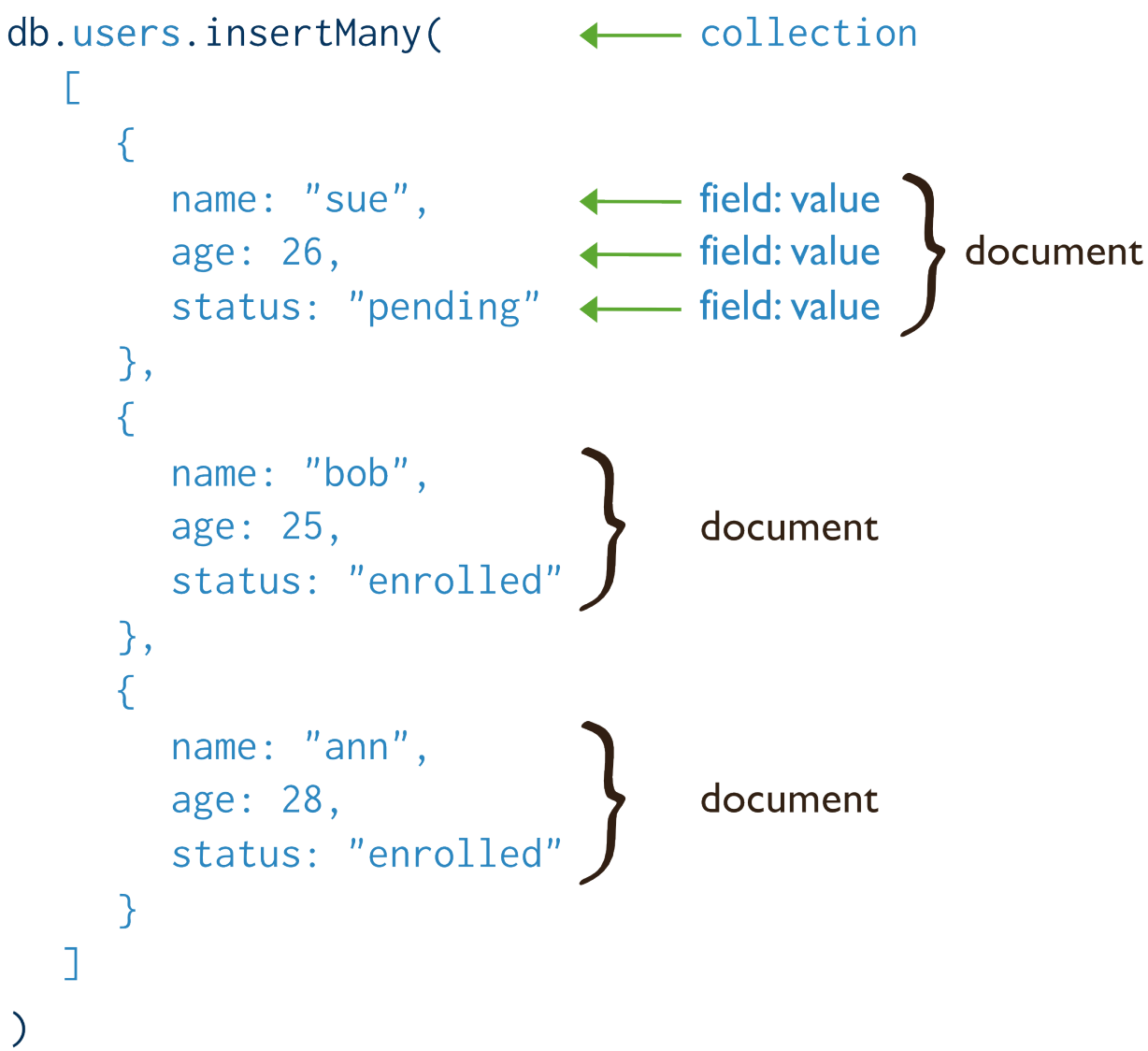
```
db.users.insertOne(
   {
      name: "sue",
      age: 26,
      status: "pending"
   }
)
```

For more information and examples, see `db.collection.insertOne()`.

insertMany New in version 3.2.

`db.collection.insertMany()` inserts *multiple* documents

The following diagram highlights the components of the MongoDB `insertMany()` operation:



The following diagram shows the same query in SQL:

Example

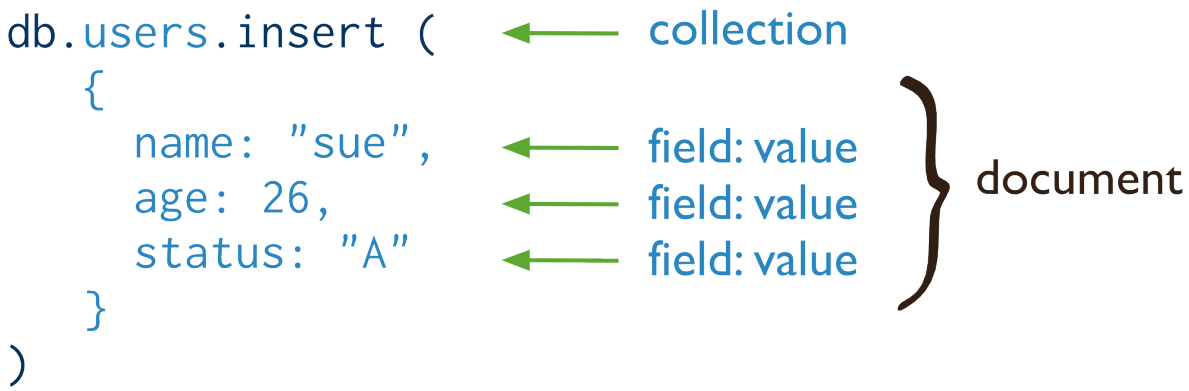
The following operation inserts three new documents into the `users` collection. Each document has three fields `name`, `age`, and `status`. Since the documents do not specify an `_id` field, MongoDB adds the `_id` field and a generated value to each document. See *Insert Behavior* (page 118).

```
db.users.insertMany(
   [
      { name: "sue", age: 26, status: "pending" },
      { name: "bob", age: 25, status: "enrolled" },
      { name: "ann", age: 28, status: "enrolled" }
   ]
)
```

For more information and examples, see `db.collection.insertMany()`.

insert In MongoDB, the `db.collection.insert()` method adds new *documents* to a collection. It can take either a single document or an array of documents to insert.

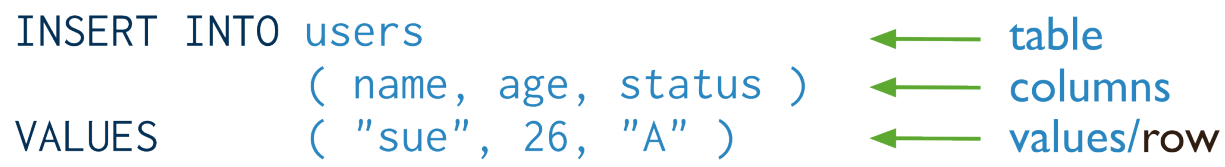
The following diagram highlights the components of a MongoDB insert operation:



The following diagram shows the same query in SQL:

Example

The following operation inserts a new document into the `users` collection. The new document has three fields `name`, `age`, and `status`. Since the document does not specify an `_id` field, MongoDB adds the `_id` field and a generated value to the new document. See *Insert Behavior* (page 118).




```
db.users.insert(
   {
      name: "sue",
      age: 26,
      status: "A"
   }
)
```

For more information and examples, see `db.collection.insert()`.

Insert Behavior The *_id* field is required in every MongoDB *document*. The `_id` field is like the document's *primary key*.

If you add a new document *without* the *_id* field, the client library or the `mongod` instance adds an `_id` field and populates the field with a unique *ObjectId*. If you pass in an `_id` value that already exists, an exception is thrown.

The `_id` field is *uniquely indexed* by default in every collection.

Other Methods to Add Documents The `updateOne()`, `updateMany()`, and `replaceOne()` operations accept the `upsert` parameter. When `upsert : true`, if no document in the collection matches the filter, a new document is created based on the information passed to the operation. See *Update Behavior with the upsert Option* (page 122).

Update

MongoDB provides the following methods for updating documents in a collection:

- `db.collection.updateOne()`
- `db.collection.updateMany()`
- `db.collection.replaceOne()`
- `db.collection.update()`

updateOne New in version 3.2.

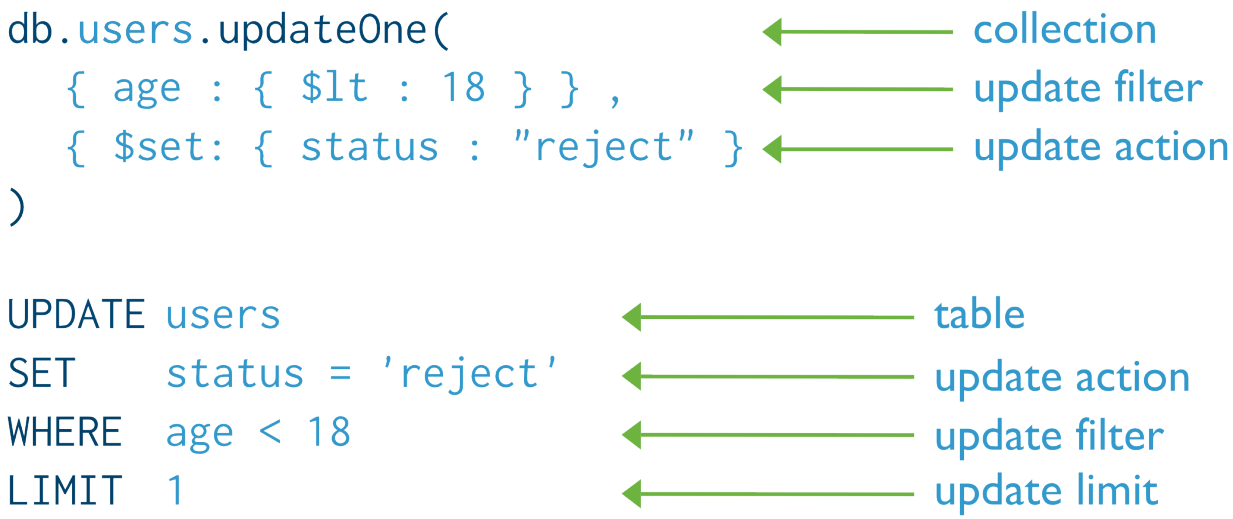
`db.collection.updateOne()` updates a *single* document.

The following diagram highlights the components of the MongoDB `updateOne()` operation:

The following diagram shows the same query in SQL:

Example

This update operation on the `users` collection sets the `status` field to `reject` for the *first* document that matches the filter of `age` less than `18`. See *Update Behavior* (page 121).




```
db.users.updateOne(
   { age: { $lt: 18 } },
   { $set: { status: "reject" } }
)
```

For more information and examples, see `db.collection.updateOne()`.

updateMany New in version 3.2.

`db.collection.updateMany()` updates *multiple* documents.

The following diagram highlights the components of the MongoDB `updateMany()` operation:

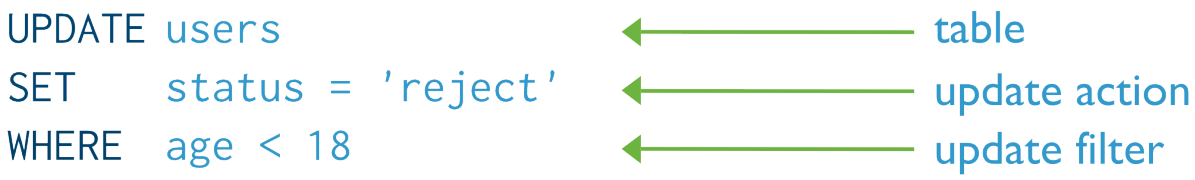
The following diagram shows the same query in SQL:

Example

This update operation on the `users` collection sets the `status` field to `reject` for *all* documents that match the filter of `age` less than `18`. See *Update Behavior* (page 121).

```
db.users.updateMany(
   { age: { $lt: 18 } },
   { $set: { status: "reject" } }
)
```

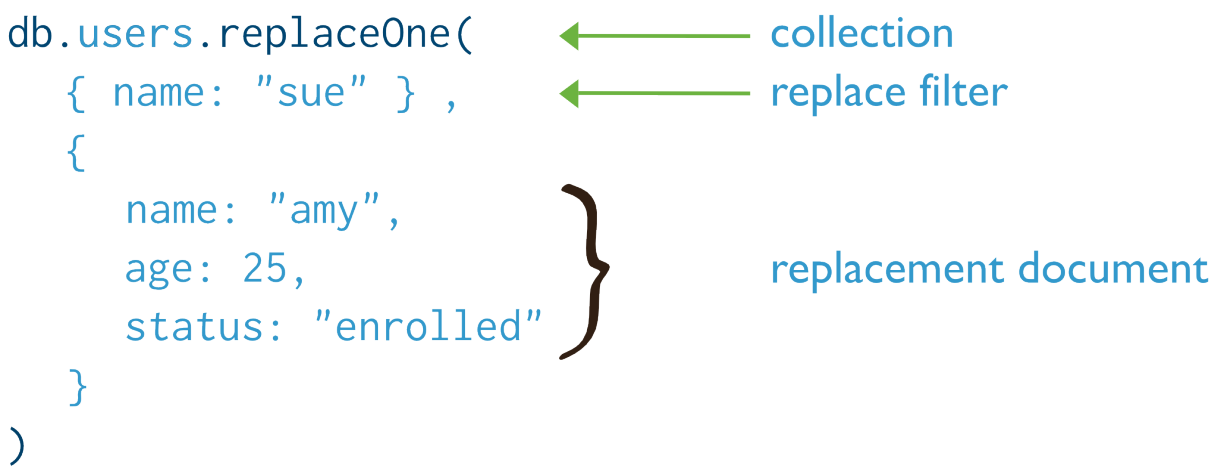
For more information and examples, see `db.collection.updateMany()`.



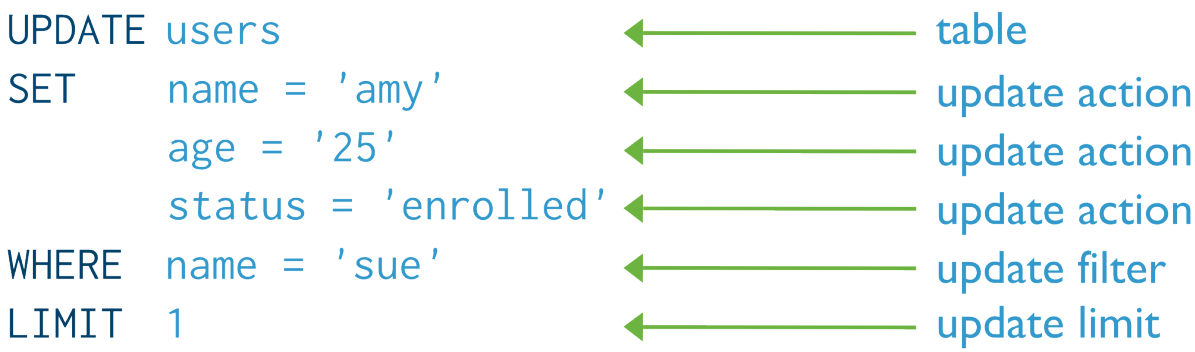
replaceOne New in version 3.2.

`db.collection.replaceOne()` replaces a *single* document.

The following diagram highlights the components of the MongoDB `replaceOne()` operation:



The following diagram shows the same query in SQL:



Example

This replace operation on the `users` collection replaces the *first* document that matches the filter of `name` is `sue` with a new document. See *Replace Behavior* (page 122).

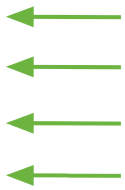
```
db.users.replaceOne(
   { name: "sue" },
   { name: "amy", age : 25, score: "enrolled" }
)
```

For more information and examples, see `db.collection.replaceOne()`.

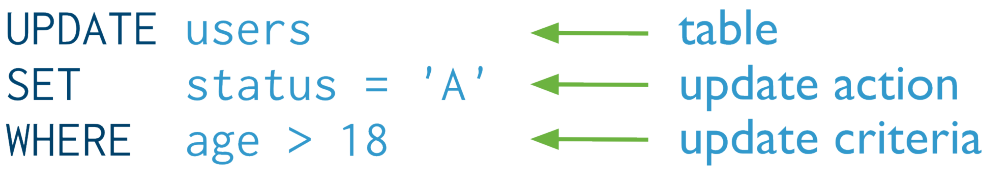
update In MongoDB, the `db.collection.update()` method modifies existing *documents* in a *collection*. The `db.collection.update()` method can accept query criteria to determine which documents to update as well as an options document that affects its behavior, such as the `multi` option to update multiple documents.

Operations performed by an update are atomic within a single document. For example, you can safely use the `$inc` and `$mul` operators to modify frequently-changed fields in concurrent applications.

The following diagram highlights the components of a MongoDB update operation:



The following diagram shows the same query in SQL:



Example

```
db.users.update(
   { age: { $gt: 18 } },
   { $set: { status: "A" } },
   { multi: true }
)
```

This update operation on the `users` collection sets the `status` field to `A` for the documents that match the criteria of `age` greater than `18`.

For more information, see `db.collection.update()` and *update() Examples*.

Update Behavior `updateOne()` and `updateMany()` use `https://docs.mongodb.org/manual/reference/operat` such as `$set`, `$unset`, or `$rename` to modify existing documents.

`updateOne()` will update the *first* document that is returned by the filter. `db.collection.findOneAndUpdate()` offers `sorting` of the filter results, allowing a degree of control over which document is updated.

By default, the `db.collection.update()` method updates a **single** document. However, with the `multi` option, `update()` can update all documents in a collection that match a query.

The db.collection.update() method either updates specific fields in the existing document or replaces the document. See db.collection.update() for details as well as examples.

When performing update operations that increase the document size beyond the allocated space for that document, the update operation relocates the document on disk.

MongoDB preserves the order of the document fields following write operations *except* for the following cases:

- The _id field is always the first field in the document.
- Updates that include renaming of field names may result in the reordering of fields in the document.

Changed in version 2.6: Starting in version 2.6, MongoDB actively attempts to preserve the field order in a document. Before version 2.6, MongoDB did not actively preserve the order of the fields in a document.

Replace Behavior replaceOne() cannot use https://docs.mongodb.org/manual/reference/operator/update in the replacement document. The replacement document must consist of only <field> : <value> assignments.

replaceOne() will replace the *first* document that matches the filter. db.collection.findOneAndReplace() offers sorting of the filter results, allowing a degree of control over which document is replaced.

You cannot replace the _id field.

Update Behavior with the upsert Option If update(), updateOne(), updateMany(), or replaceOne() include upsert : true **and** no documents match the filter portion of the operation, then the operation creates a new document and inserts it. If there are matching documents, then the operation modifies the matching document or documents.

Delete

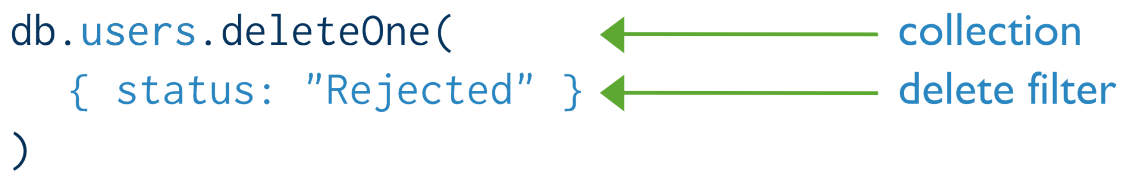
MongoDB provides the following methods for deleting documents from a collection:

- db.collection.deleteOne()
- db.collection.deleteMany()
- db.collection.remove()

deleteOne New in version 3.2.

db.collection.deleteOne() deletes a *single* document.

The following diagram highlights the components of the MongoDB deleteOne() operation:



The following diagram shows the same query in SQL:

Example

This delete operation on the `users` collection deletes the *first* document where `name` is `sue`. See *Delete Behavior* (page 124).

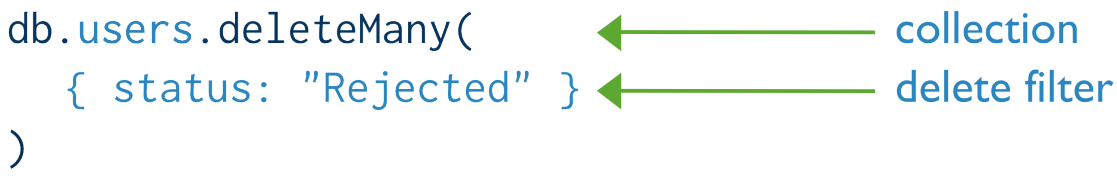
```
db.users.deleteOne(
   { status: "reject" }
)
```

For more information and examples, see `db.collection.deleteOne()`.

deleteMany New in version 3.2.

`db.collection.deleteMany()` deletes *multiple* documents.

The following diagram highlights the components of the MongoDB `deleteMany()` operation:



The following diagram shows the same query in SQL:

Example

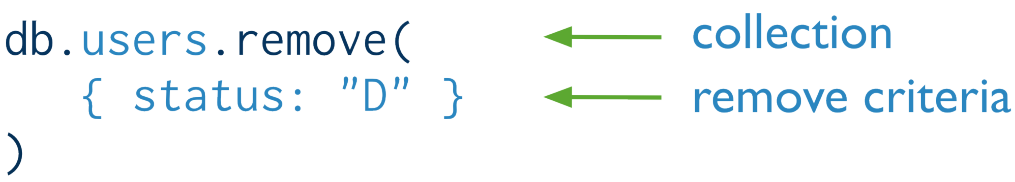
This delete operation on the `users` collection deletes *all* documents where `status` is `reject`. See *Delete Behavior* (page 124).

```
db.users.deleteMany(
   { status: "reject" }
)
```

For more information and examples, see `db.collection.deleteMany()`.

remove In MongoDB, the `db.collection.remove()` method deletes documents from a collection. The `db.collection.remove()` method accepts query criteria to determine which documents to remove as well as an options document that affects its behavior, such as the `justOne` option to remove only a single document.

The following diagram highlights the components of a MongoDB remove operation:



The following diagram shows the same query in SQL:

Example

```
db.users.remove(
   { status: "D" }
)
```

This delete operation on the `users` collection removes *all* documents that match the criteria of `status` equal to `D`.

For more information, see `db.collection.remove()` method and *Remove Documents* (page 152).

Delete Behavior `deleteOne()` will delete the *first* document that matches the filter. `db.collection.findOneAndDelete()` offers `sorting` of the filter results, allowing a degree of control over which document is deleted.

Remove Behavior By default, `db.collection.remove()` method removes all documents that match its query. If the optional `justOne` parameter is set to `true`, `remove()` will limit the delete operation to a single document.

Additional Methods

The `db.collection.save()` method can either update an existing document or insert a document if the document cannot be found by the `_id` field. See `db.collection.save()` for more information and examples.

Bulk Write MongoDB provides the `db.collection.bulkWrite()` method for executing multiple write operations in a group. Each write operation is still atomic on the level of a single *document*.

Example

The following `bulkWrite()` inserts several documents, performs an update, and then deletes several documents.

```
db.collection.bulkWrite(
   [
      { insertOne : { "document" : { name : "sue", age : 26 } } },
      { insertOne : { "document" : { name : "joe", age : 24 } } },
      { insertOne : { "document" : { name : "ann", age : 25 } } },
      { insertOne : { "document" : { name : "bob", age : 27 } } },
```

```
      { updateMany: {
         "filter" : { age : { $gt : 25} },
         "update" : { $set : { "status" : "enrolled" } }
         }
      },
      { deleteMany : { "filter" : { "status" : { $exists : true } } } }
   ]
)
```

Atomicity and Transactions

On this page

- `$isolated` Operator (page 125)
- Transaction-Like Semantics (page 125)
- Concurrency Control (page 126)

In MongoDB, a write operation is atomic on the level of a single document, even if the operation modifies multiple embedded documents *within* a single document.

When a single write operation modifies multiple documents, the modification of each document is atomic, but the operation as a whole is not atomic and other operations may interleave. However, you can *isolate* a single write operation that affects multiple documents using the `$isolated` operator.

`$isolated` Operator

Using the `$isolated` operator, a write operation that affects multiple documents can prevent other processes from interleaving once the write operation modifies the first document. This ensures that no client sees the changes until the write operation completes or errors out.

`$isolated` does **not** work with *sharded clusters*.

An isolated write operation does not provide "all-or-nothing" atomicity. That is, an error during the write operation does not roll back all its changes that preceded the error.

Note: `$isolated` operator causes write operations to acquire an exclusive lock on the collection, *even for document-level locking storage engines* such as WiredTiger. That is, `$isolated` operator will make WiredTiger single-threaded for the duration of the operation.

The `$isolated` operator does **not** work on sharded clusters.

For an example of an update operation that uses the `$isolated` operator, see `$isolated`. For an example of a remove operation that uses the `$isolated` operator, see *isolate-remove-operations*.

Transaction-Like Semantics

Since a single document can contain multiple embedded documents, single-document atomicity is sufficient for many practical use cases. For cases where a sequence of write operations must operate as if in a single transaction, you can implement a *two-phase commit* (page 164) in your application.

However, two-phase commits can only offer transaction-*like* semantics. Using two-phase commit ensures data consistency, but it is possible for applications to return intermediate data during the two-phase commit or rollback.

For more information on two-phase commit and rollback, see *Perform Two Phase Commits* (page 164).

Concurrency Control

Concurrency control allows multiple applications to run concurrently without causing data inconsistency or conflicts.

One approach is to create a *unique index* (page 568) on a field that can only have unique values. This prevents insertions or updates from creating duplicate data. Create a unique index on multiple fields to force uniqueness on that combination of field values. For examples of use cases, see *update() and Unique Index* and *findAndModify() and Unique Index*.

Another approach is to specify the expected current value of a field in the query predicate for the write operations. For an example, see *Update if Current* (page 170).

The two-phase commit pattern provides a variation where the query predicate includes the *application identifier* (page 168) as well as the expected state of the data in the write operation.

See also:

Read Isolation, Consistency, and Recency (page 133)

Distributed Write Operations

On this page

- Write Operations on Sharded Clusters (page 126)
- Write Operations on Replica Sets (page 126)

Write Operations on Sharded Clusters

For sharded collections in a *sharded cluster*, the `mongos` directs write operations from applications to the shards that are responsible for the specific *portion* of the data set. The `mongos` uses the cluster metadata from the *config database* (page 742) to route the write operation to the appropriate shards.

MongoDB partitions data in a sharded collection into *ranges* based on the values of the *shard key*. Then, MongoDB distributes these chunks to shards. The shard key determines the distribution of chunks to shards. This can affect the performance of write operations in the cluster.

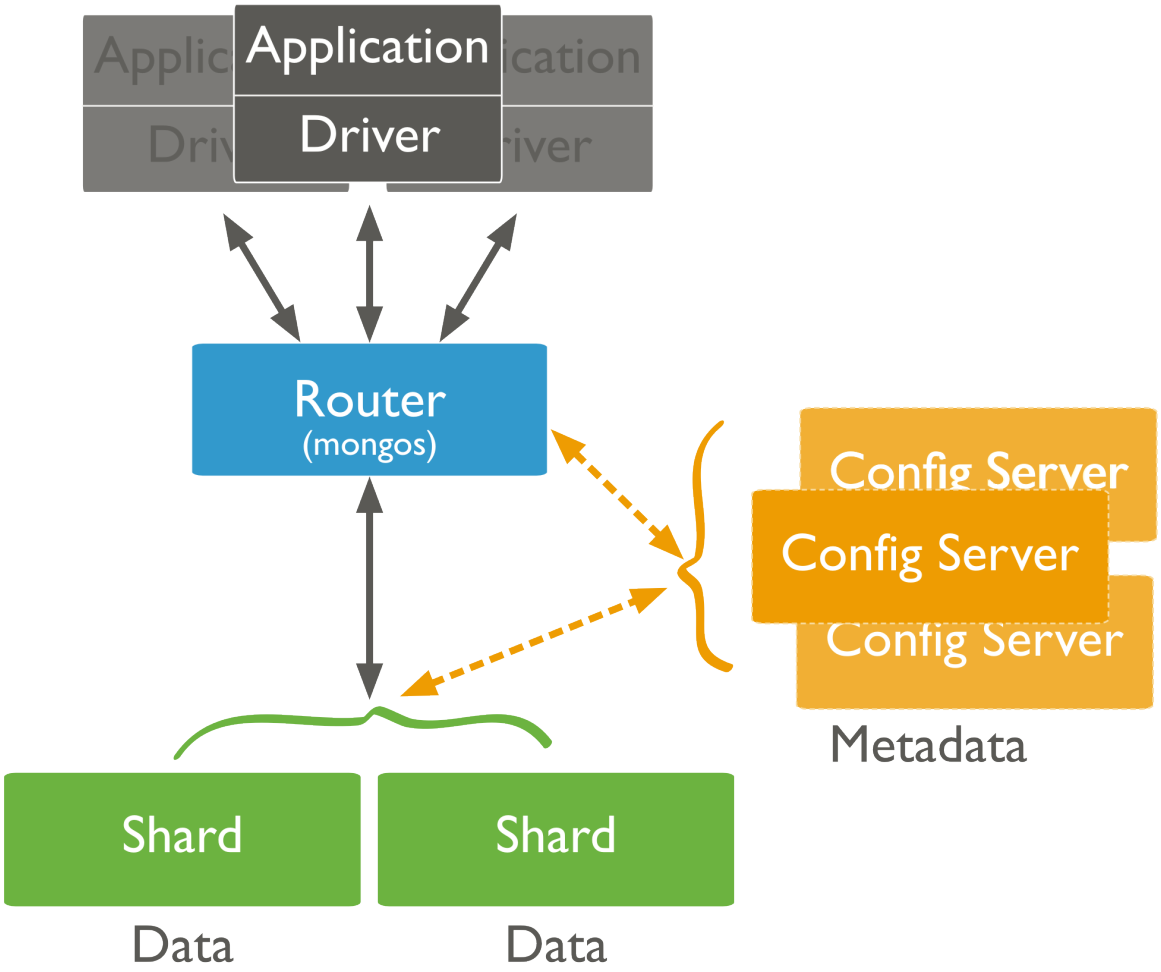
Important: Update operations that affect a *single* document **must** include the *shard key* or the `_id` field. Updates that affect multiple documents are more efficient in some situations if they have the *shard key*, but can be broadcast to all shards.

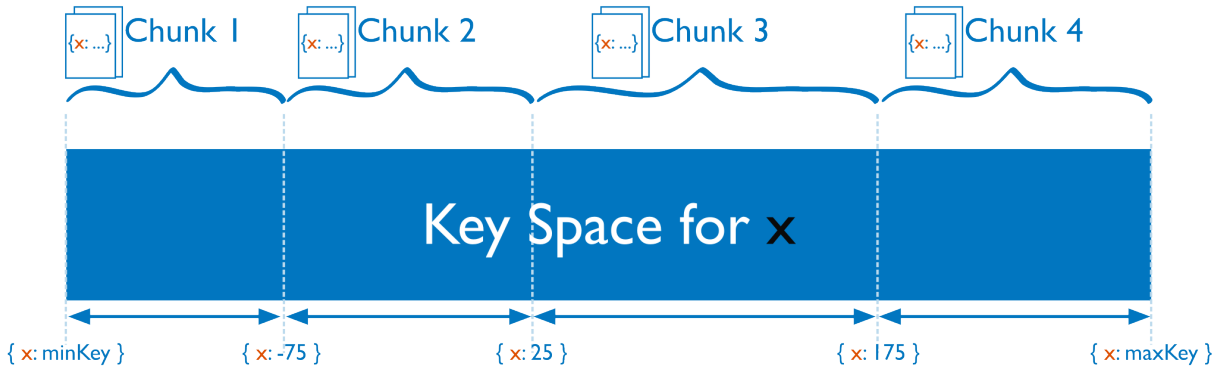
If the value of the shard key increases or decreases with every insert, all insert operations target a single shard. As a result, the capacity of a single shard becomes the limit for the insert capacity of the sharded cluster.

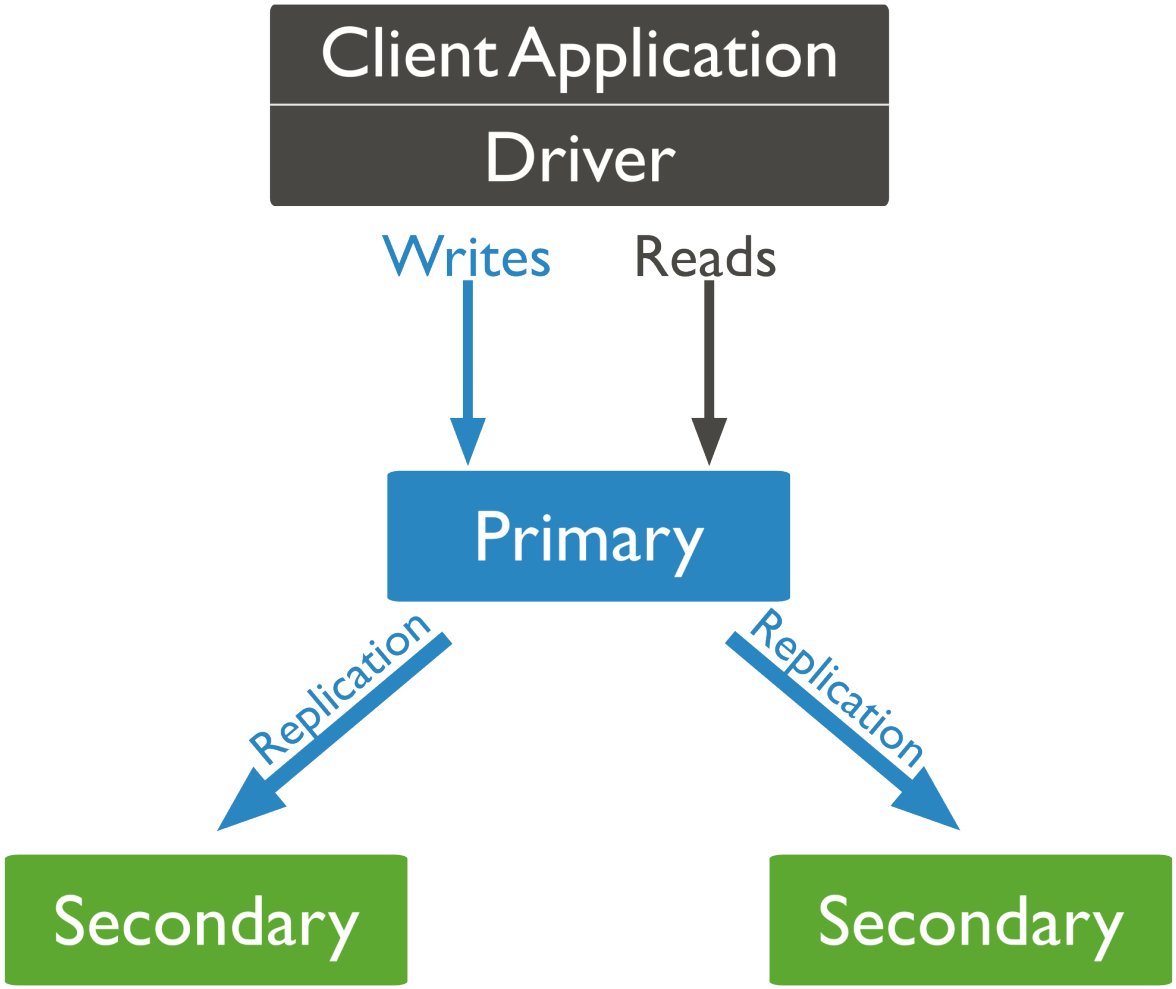
For more information, see *Sharded Cluster Tutorials* (page 764) and *Bulk Write Operations* (page 130).

Write Operations on Replica Sets

In *replica sets*, all write operations go to the set's *primary*. The primary applies the write operation and records the operations on the primary's operation log or *oplog*. The oplog is a reproducible sequence of operations to the data


Application
Driver
Router
(mongos)
Config Server
Config Server
Config Server
Metadata
Shard
Shard
Data
Data


Chunk 1
Chunk 2
Chunk 3
Chunk 4
Key Space for x
{ x: minKey }
{ x: -75 }
{ x: 25 }
{ x: 175 }
{ x: maxKey }


Client Application
Driver
Writes
Reads
Primary
Replication
Replication
Secondary
Secondary

set. *Secondary* members of the set continuously replicate the oplog and apply the operations to themselves in an asynchronous process.

For more information on replica sets and write operations, see *Replication Introduction* (page 623) and *Write Concern* (page 179).

Write Operation Performance

On this page

- Indexes (page 129)
- Document Growth and the MMAPv1 Storage Engine (page 129)
- Storage Performance (page 130)
- Additional Resources (page 130)

Indexes

After every insert, update, or delete operation, MongoDB must update *every* index associated with the collection in addition to the data itself. Therefore, every index on a collection adds some amount of overhead for the performance of write operations. [4]

In general, the performance gains that indexes provide for *read operations* are worth the insertion penalty. However, in order to optimize write performance when possible, be careful when creating new indexes and evaluate the existing indexes to ensure that your queries actually use these indexes.

For indexes and queries, see *Query Optimization* (page 105). For more information on indexes, see *Indexes* (page 515) and *Indexing Strategies* (page 586).

Document Growth and the MMAPv1 Storage Engine

Some update operations can increase the size of the document; for instance, if an update adds a new field to the document.

For the MMAPv1 storage engine, if an update operation causes a document to exceed the currently allocated *record size*, MongoDB relocates the document on disk with enough contiguous space to hold the document. Updates that require relocations take longer than updates that do not, particularly if the collection has indexes. If a collection has indexes, MongoDB must update all index entries. Thus, for a collection with many indexes, the move will impact the write throughput.

Changed in version 3.0.0: By default, MongoDB uses *Power of 2 Sized Allocations* (page 604) to add *padding automatically* (page 604) for the MMAPv1 storage engine. The *Power of 2 Sized Allocations* (page 604) ensures that MongoDB allocates document space in sizes that are powers of 2, which helps ensure that MongoDB can efficiently reuse free space created by document deletion or relocation as well as reduce the occurrences of reallocations in many cases.

Although *Power of 2 Sized Allocations* (page 604) minimizes the occurrence of re-allocation, it does not eliminate document re-allocation.

See *MMAPv1 Storage Engine* (page 603) for more information.

[4] For inserts and updates to un-indexed fields, the overhead for *sparse indexes* (page 574) is less than for non-sparse indexes. Also for non-sparse indexes, updates that do not change the record size have less indexing overhead.

Storage Performance

Hardware The capability of the storage system creates some important physical limits for the performance of MongoDB's write operations. Many unique factors related to the storage system of the drive affect write performance, including random access patterns, disk caches, disk readahead and RAID configurations.

Solid state drives (SSDs) can outperform spinning hard disks (HDDs) by 100 times or more for random workloads.

See

Production Notes (page 296) for recommendations regarding additional hardware and configuration options.

Journaling To provide durability in the event of a crash, MongoDB uses *write ahead logging* to an on-disk *journal*. MongoDB writes the in-memory changes first to the on-disk journal files. If MongoDB should terminate or encounter an error before committing the changes to the data files, MongoDB can use the journal files to apply the write operation to the data files.

While the durability assurance provided by the journal typically outweigh the performance costs of the additional write operations, consider the following interactions between the journal and performance:

- If the journal and the data file reside on the same block device, the data files and the journal may have to contend for a finite number of available I/O resources. Moving the journal to a separate device may increase the capacity for write operations.
- If applications specify *write concerns* (page 179) that include the `j option` (page 181), `mongod` will decrease the duration between journal writes, which can increase the overall write load.
- The duration between journal writes is configurable using the `commitIntervalMs` run-time option. Decreasing the period between journal commits will increase the number of write operations, which can limit MongoDB's capacity for write operations. Increasing the amount of time between journal commits may decrease the total number of write operation, but also increases the chance that the journal will not record a write operation in the event of a failure.

For additional information on journaling, see *Journaling* (page 606).

Additional Resources

- MongoDB Performance Evaluation and Tuning Consulting Package[5]

Bulk Write Operations

On this page

- Overview (page 131)
- Ordered vs Unordered Operations (page 131)
- bulkWrite() Methods (page 131)
- Strategies for Bulk Inserts to a Sharded Collection (page 133)

[5] https://www.mongodb.com/products/consulting?jmp=docs#performance_evaluation

Overview

MongoDB provides clients the ability to perform write operations in bulk. Bulk write operations affect a *single* collection. MongoDB allows applications to determine the acceptable level of acknowledgement required for bulk write operations.

New in version 3.2.

The `db.collection.bulkWrite()` method provides the ability to perform bulk insert, update, and remove operations. MongoDB also supports bulk insert through the `db.collection.insertMany()`.

Ordered vs Unordered Operations

Bulk write operations can be either *ordered* or *unordered*.

With an ordered list of operations, MongoDB executes the operations serially. If an error occurs during the processing of one of the write operations, MongoDB will return without processing any remaining write operations in the list. See *ordered Bulk Write*

With an unordered list of operations, MongoDB can execute the operations in parallel, but this behavior is not guaranteed. If an error occurs during the processing of one of the write operations, MongoDB will continue to process remaining write operations in the list. See *bulkwrite-example-unordered-bulk-write*.

Executing an ordered list of operations on a sharded collection will generally be slower than executing an unordered list since with an ordered list, each operation must wait for the previous operation to finish.

By default, `bulkWrite()` performs `ordered` operations. To specify `unordered` write operations, set `ordered : false` in the options document.

See *bulkwrite-write-operations-executionofoperations*

bulkWrite() Methods

`bulkWrite()` supports the following write operations:

- *bulkwrite-write-operations-insertOne*
- *updateOne*
- *updateMany*
- *bulkwrite-write-operations-replaceOne*
- *deleteOne*
- *deleteMany*

Each write operation is passed to `bulkWrite()` as a document in an array.

For example, the following performs multiple write operations:

The `characters` collection contains the following documents:

```
{ "_id" : 1, "char" : "Brisbane", "class" : "monk", "lvl" : 4 },
{ "_id" : 2, "char" : "Eldon", "class" : "alchemist", "lvl" : 3 },
{ "_id" : 3, "char" : "Meldane", "class" : "ranger", "lvl" : 3 }
```

The following `bulkWrite()` performs multiple operations on the collection:

```
try {
   db.characters.bulkWrite(
      [
         { insertOne :
            {
               "document" :
               {
                  "_id" : 4, "char" : "Dithras", "class" : "barbarian", "lvl" : 4
               }
            }
         },
         { insertOne :
            {
               "document" :
               {
                  "_id" : 5, "char" : "Taeln", "class" : "fighter", "lvl" : 3
               }
            }
         },
         { updateOne :
            {
               "filter" : { "char" : "Eldon" },
               "update" : { $set : { "status" : "Critical Injury" } }
            }
         },
         { deleteOne :
            { "filter" : { "char" : "Brisbane"} }
         },
         { replaceOne :
            {
               "filter" : { "char" : "Meldane" },
               "replacement" : { "char" : "Tanys", "class" : "oracle", "lvl" : 4 }
            }
         }
      ]
   );
}
catch (e) {
   print(e);
}
```

The operation returns the following:

```
{
   "acknowledged" : true,
   "deletedCount" : 1,
   "insertedCount" : 2,
   "matchedCount" : 2,
   "upsertedCount" : 0,
   "insertedIds" : {
      "0" : 4,
      "1" : 5
   },
   "upsertedIds" : {

   }
}
```

For more examples, see *bulkWrite() Examples*

Strategies for Bulk Inserts to a Sharded Collection

Large bulk insert operations, including initial data inserts or routine data import, can affect *sharded cluster* performance. For bulk inserts, consider the following strategies:

Pre-Split the Collection If the sharded collection is empty, then the collection has only one initial *chunk*, which resides on a single shard. MongoDB must then take time to receive data, create splits, and distribute the split chunks to the available shards. To avoid this performance cost, you can pre-split the collection, as described in *Split Chunks in a Sharded Cluster* (page 808).

Unordered Writes to `mongos` To improve write performance to sharded clusters, use `bulkWrite()` with the optional parameter `ordered` set to `false`. `mongos` can attempt to send the writes to multiple shards simultaneously. For *empty* collections, first pre-split the collection as described in *Split Chunks in a Sharded Cluster* (page 808).

Avoid Monotonic Throttling If your shard key increases monotonically during an insert, then all inserted data goes to the last chunk in the collection, which will always end up on a single shard. Therefore, the insert capacity of the cluster will never exceed the insert capacity of that single shard.

If your insert volume is larger than what a single shard can process, and if you cannot avoid a monotonically increasing shard key, then consider the following modifications to your application:

- Reverse the binary bits of the shard key. This preserves the information and avoids correlating insertion order with increasing sequence of values.
- Swap the first and last 16-bit words to "shuffle" the inserts.

Example

The following example, in C++, swaps the leading and trailing 16-bit word of *BSON ObjectIds* generated so they are no longer monotonically increasing.

```
using namespace mongo;
OID make_an_id() {
  OID x = OID::gen();
  const unsigned char *p = x.getData();
  swap( (unsigned short&) p[0], (unsigned short&) p[10] );
  return x;
}

void foo() {
  // create an object
  BSONObj o = BSON( "_id" << make_an_id() << "x" << 3 << "name" << "jane" );
  // now we may insert o into a sharded collection
}
```

See also:

Shard Keys (page 747) for information on choosing a sharded key. Also see *Shard Key Internals* (page 747) (in particular, *Choosing a Shard Key* (page 771)).

4.2.3 Read Isolation, Consistency, and Recency

On this page

- Isolation Guarantees (page 134)
- Consistency Guarantees (page 135)
- Recency (page 136)

Isolation Guarantees

Read Uncommitted

In MongoDB, clients can see the results of writes before the writes are *durable*:

- Regardless of *write concern* (page 179), other clients using `"local"` (page 182) (i.e. the default) readConcern can see the result of a write operation before the write operation is acknowledged to the issuing client.
- Clients using `"local"` (page 182) (i.e. the default) readConcern can read data which may be subsequently *rolled back* (page 647).

Read uncommitted is the default isolation level and applies to `mongod` standalone instances as well as to replica sets and sharded clusters.

Read Uncommitted And Single Document Atomicity

Write operations are atomic with respect to a single document; i.e. if a write is updating multiple fields in the document, a reader will never see the document with only some of the fields updated.

With a standalone `mongod` instance, a set of read and write operations to a single document is serializable. With a replica set, a set of read and write operations to a single document is serializable *only* in the absence of a rollback.

However, although the readers may not see a *partially* updated document, read uncommitted means that concurrent readers may still see the updated document before the changes are *durable*.

Read Uncommitted And Multiple Document Write

When a single write operation modifies multiple documents, the modification of each document is atomic, but the operation as a whole is not atomic and other operations may interleave. However, you can *isolate* a single write operation that affects multiple documents using the `$isolated` operator.

Without isolating the multi-document write operations, MongoDB exhibits the following behavior:

1. Non-point-in-time read operations. Suppose a read operation begins at time t_1 and starts reading documents. A write operation then commits an update to one of the documents at some later time t_2. The reader may see the updated version of the document, and therefore does not see a point-in-time snapshot of the data.
2. Non-serializable operations. Suppose a read operation reads a document d_1 at time t_1 and a write operation updates d_1 at some later time t_3. This introduces a read-write dependency such that, if the operations were to be serialized, the read operation must precede the write operation. But also suppose that the write operation updates document d_2 at time t_2 and the read operation subsequently reads d_2 at some later time t_4. This introduces a write-read dependency which would instead require the read operation to come *after* the write operation in a serializable schedule. There is a dependency cycle which makes serializability impossible.
3. Reads may miss matching documents that are updated during the course of the read operation.

Using the `$isolated` operator, a write operation that affects multiple documents can prevent other processes from interleaving once the write operation modifies the first document. This ensures that no client sees the changes until the write operation completes or errors out.

`$isolated` does **not** work with *sharded clusters*.

An isolated write operation does not provide "all-or-nothing" atomicity. That is, an error during the write operation does not roll back all its changes that preceded the error.

Note: `$isolated` operator causes write operations to acquire an exclusive lock on the collection, *even for document-level locking storage engines* such as WiredTiger. That is, `$isolated` operator will make WiredTiger single-threaded for the duration of the operation.

See also:

Atomicity and Transactions (page 125)

Cursor Snapshot

MongoDB cursors can return the same document more than once in some situations. As a cursor returns documents other operations may interleave with the query. If some of these operations are *updates* (page 114) that cause the document to move (in the case of MMAPv1, caused by document growth) or that change the indexed field on the index used by the query; then the cursor will return the same document more than once.

In very specific cases, you can isolate the cursor from returning the same document more than once by using the `cursor.snapshot()` method. `snapshot()` guarantees that the query will return each document no more than once.

Warning:

- The `snapshot()` does not guarantee that the data returned by the query will reflect a single moment in time *nor* does it provide isolation from insert or delete operations.
- You **cannot** use `snapshot()` with *sharded collections*.
- You **cannot** use `snapshot()` with the `sort()` or `hint()` cursor methods.

As an alternative, if your collection has a field or fields that are never modified, you can use a *unique* index on this field or these fields to achieve a similar result as the `snapshot()`. Query with `hint()` to explicitly force the query to use that index.

Consistency Guarantees

Monotonic Reads

MongoDB provides monotonic reads from a standalone `mongod` instance. Suppose an application performs a sequence of operations that consists of a read operation R_1 followed later in the sequence by another read operation R_2. If the application performs the sequence on a standalone `mongod` instance, the later read R_2 never returns results that reflect an earlier state than that returned from R_1; i.e. R_2 returns data that is monotonically increasing in recency from R_1.

Changed in version 3.2: For replica sets and sharded clusters, MongoDB provides monotonic reads if read operations specify *Read Concern* (page 181) `"majority"` and read preference `primary` (page 728).

In previous versions, MongoDB cannot make monotonic read guarantees from replica sets and sharded clusters.

Monotonic Writes

MongoDB provides monotonic write guarantees for standalone `mongod` instances, replica sets, and sharded clusters.

Suppose an application performs a sequence of operations that consists of a write operation W_1 followed later in the sequence by a write operation W_2. MongoDB guarantees that W_1 operation precedes W_2.

Recency

In MongoDB, in a replica set with one primary member [6],

- With `"local"` (page 182) `readConcern`, reads from the primary reflect the latest writes in absence of a failover;
- With `"majority"` (page 182) `readConcern`, read operations from the primary or the secondaries have *eventual consistency*.

4.3 MongoDB CRUD Tutorials

The following tutorials provide instructions for querying and modifying data. For a higher-level overview of these operations, see *MongoDB CRUD Operations* (page 97).

Insert Documents **(page 137)** Insert new documents into a collection.

Query Documents **(page 140)** Find documents in a collection using search criteria.

Modify Documents **(page 148)** Modify documents in a collection

Remove Documents **(page 152)** Remove documents from a collection.

Limit Fields to Return from a Query **(page 153)** Limit which fields are returned by a query.

Limit Number of Elements in an Array after an Update **(page 156)** Use `$push` with modifiers to sort and maintain an array of fixed size.

Iterate a Cursor in the mongo Shell **(page 158)** Access documents returned by a `find` query by iterating the cursor, either manually or using the iterator index.

Analyze Query Performance **(page 159)** Use query introspection (i.e. `explain`) to analyze the efficiency of queries and determine how a query uses available indexes.

Perform Two Phase Commits **(page 164)** Use two-phase commits when writing data to multiple documents.

Update Document if Current **(page 170)** Update a document only if it has not changed since it was last read.

Create Tailable Cursor **(page 172)** Create tailable cursors for use in capped collections with high numbers of write operations for which an index would be too expensive.

Create an Auto-Incrementing Sequence Field **(page 173)** Describes how to create an incrementing sequence number for the `_id` field using a Counters Collection or an Optimistic Loop.

Perform Quorum Reads on Replica Sets **(page 176)** Perform quorum reads using `findAndModify`.

[6] In *some circumstances* (page 729), two nodes in a replica set may *transiently* believe that they are the primary, but at most, one of them will be able to complete writes with `{ w: "majority" }` (page 180) write concern. The node that can complete `{ w: "majority" }` (page 180) writes is the current primary, and the other node is a former primary that has not yet recognized its demotion, typically due to a *network partition*. When this occurs, clients that connect to the former primary may observe stale data despite having requested read preference `primary` (page 728), and new writes to the former primary will eventually roll back.

4.3.1 Insert Documents

On this page

- Insert a Document (page 137)
- Insert an Array of Documents (page 138)
- Insert Multiple Documents with `Bulk` (page 139)
- Additional Examples and Methods (page 140)

In MongoDB, the `db.collection.insert()` method adds new documents into a collection.

Insert a Document

Step 1: Insert a document into a collection.

Insert a document into a collection named `inventory`. The operation will create the collection if the collection does not currently exist.

```
db.inventory.insert(
   {
     item: "ABC1",
     details: {
        model: "14Q3",
        manufacturer: "XYZ Company"
     },
     stock: [ { size: "S", qty: 25 }, { size: "M", qty: 50 } ],
     category: "clothing"
   }
)
```

The operation returns a `WriteResult` object with the status of the operation. A successful insert of the document returns the following object:

```
WriteResult({ "nInserted" : 1 })
```

The `nInserted` field specifies the number of documents inserted. If the operation encounters an error, the `WriteResult` object will contain the error information.

Step 2: Review the inserted document.

If the insert operation is successful, verify the insertion by querying the collection.

```
db.inventory.find()
```

The document you inserted should return.

```
{ "_id" : ObjectId("53d98f133bb604791249ca99"), "item" : "ABC1", "details" : { "model" : "14Q3", "mar
```

The returned document shows that MongoDB added an `_id` field to the document. If a client inserts a document that does not contain the `_id` field, MongoDB adds the field with the value set to a generated ObjectId[7]. The ObjectId[8] values in your documents will differ from the ones shown.

[7] https://docs.mongodb.org/manual/reference/method/ObjectId
[8] https://docs.mongodb.org/manual/reference/method/ObjectId

Insert an Array of Documents

You can pass an array of documents to the `db.collection.insert()` method to insert multiple documents.

Step 1: Create an array of documents.

Define a variable `mydocuments` that holds an array of documents to insert.

```
var mydocuments =
    [
      {
        item: "ABC2",
        details: { model: "14Q3", manufacturer: "M1 Corporation" },
        stock: [ { size: "M", qty: 50 } ],
        category: "clothing"
      },
      {
        item: "MNO2",
        details: { model: "14Q3", manufacturer: "ABC Company" },
        stock: [ { size: "S", qty: 5 }, { size: "M", qty: 5 }, { size: "L", qty: 1 } ],
        category: "clothing"
      },
      {
        item: "IJK2",
        details: { model: "14Q2", manufacturer: "M5 Corporation" },
        stock: [ { size: "S", qty: 5 }, { size: "L", qty: 1 } ],
        category: "houseware"
      }
    ];
```

Step 2: Insert the documents.

Pass the `mydocuments` array to the `db.collection.insert()` to perform a bulk insert.

```
db.inventory.insert( mydocuments );
```

The method returns a `BulkWriteResult` object with the status of the operation. A successful insert of the documents returns the following object:

```
BulkWriteResult({
   "writeErrors" : [ ],
   "writeConcernErrors" : [ ],
   "nInserted" : 3,
   "nUpserted" : 0,
   "nMatched" : 0,
   "nModified" : 0,
   "nRemoved" : 0,
   "upserted" : [ ]
})
```

The `nInserted` field specifies the number of documents inserted. If the operation encounters an error, the `BulkWriteResult` object will contain information regarding the error.

The inserted documents will each have an `_id` field added by MongoDB.

Insert Multiple Documents with `Bulk`

New in version 2.6.

MongoDB provides a `Bulk()` API that you can use to perform multiple write operations in bulk. The following sequence of operations describes how you would use the `Bulk()` API to insert a group of documents into a MongoDB collection.

Step 1: Initialize a `Bulk` operations builder.

Initialize a `Bulk` operations builder for the collection `inventory`.

```
var bulk = db.inventory.initializeUnorderedBulkOp();
```

The operation returns an unordered operations builder which maintains a list of operations to perform. Unordered operations means that MongoDB can execute in parallel as well as in nondeterministic order. If an error occurs during the processing of one of the write operations, MongoDB will continue to process remaining write operations in the list.

You can also initialize an ordered operations builder; see `db.collection.initializeOrderedBulkOp()` for details.

Step 2: Add insert operations to the `bulk` object.

Add two insert operations to the `bulk` object using the `Bulk.insert()` method.

```
bulk.insert(
   {
     item: "BE10",
     details: { model: "14Q2", manufacturer: "XYZ Company" },
     stock: [ { size: "L", qty: 5 } ],
     category: "clothing"
   }
);
bulk.insert(
   {
     item: "ZYT1",
     details: { model: "14Q1", manufacturer: "ABC Company"  },
     stock: [ { size: "S", qty: 5 }, { size: "M", qty: 5 } ],
     category: "houseware"
   }
);
```

Step 3: Execute the bulk operation.

Call the `execute()` method on the `bulk` object to execute the operations in its list.

```
bulk.execute();
```

The method returns a `BulkWriteResult` object with the status of the operation. A successful insert of the documents returns the following object:

```
BulkWriteResult({
   "writeErrors" : [ ],
   "writeConcernErrors" : [ ],
```

```
    "nInserted" : 2,
    "nUpserted" : 0,
    "nMatched" : 0,
    "nModified" : 0,
    "nRemoved" : 0,
    "upserted" : [ ]
})
```

The `nInserted` field specifies the number of documents inserted. If the operation encounters an error, the `BulkWriteResult` object will contain information regarding the error.

Additional Examples and Methods

For more examples, see `db.collection.insert()`.

The `db.collection.update()` method, the `db.collection.findAndModify()`, and the `db.collection.save()` method can also add new documents. See the individual reference pages for the methods for more information and examples.

4.3.2 Query Documents

On this page

- Select All Documents in a Collection (page 140)
- Specify Equality Condition (page 141)
- Specify Conditions Using Query Operators (page 141)
- Specify `AND` Conditions (page 141)
- Specify `OR` Conditions (page 141)
- Specify `AND` as well as `OR` Conditions (page 142)
- Embedded Documents (page 142)
- Arrays (page 143)
- Null or Missing Fields (page 147)

In MongoDB, the `db.collection.find()` method retrieves documents from a collection. [9] The `db.collection.find()` method returns a *cursor* (page 103) to the retrieved documents.

This tutorial provides examples of read operations using the `db.collection.find()` method in the `mongo` shell. In these examples, the retrieved documents contain all their fields. To restrict the fields to return in the retrieved documents, see *Limit Fields to Return from a Query* (page 153).

Select All Documents in a Collection

An empty query document (`{}`) selects all documents in the collection:

```
db.inventory.find( {} )
```

Not specifying a query document to the `find()` is equivalent to specifying an empty query document. Therefore the following operation is equivalent to the previous operation:

[9] The `db.collection.findOne()` method also performs a read operation to return a single document. Internally, the `db.collection.findOne()` method is the `db.collection.find()` method with a limit of 1.

```
db.inventory.find()
```

Specify Equality Condition

To specify equality condition, use the query document `{ <field>: <value> }` to select all documents that contain the `<field>` with the specified `<value>`.

The following example retrieves from the `inventory` collection all documents where the `type` field has the value `snacks`:

```
db.inventory.find( { type: "snacks" } )
```

Specify Conditions Using Query Operators

A query document can use the *query operators* to specify conditions in a MongoDB query.

The following example selects all documents in the `inventory` collection where the value of the `type` field is either `'food'` or `'snacks'`:

```
db.inventory.find( { type: { $in: [ 'food', 'snacks' ] } } )
```

Although you can express this query using the `$or` operator, use the `$in` operator rather than the `$or` operator when performing equality checks on the same field.

Refer to the `https://docs.mongodb.org/manual/reference/operator/query` document for the complete list of query operators.

Specify AND Conditions

A compound query can specify conditions for more than one field in the collection's documents. Implicitly, a logical `AND` conjunction connects the clauses of a compound query so that the query selects the documents in the collection that match all the conditions.

In the following example, the query document specifies an equality match on the field `type` **and** a less than (`$lt`) comparison match on the field `price`:

```
db.inventory.find( { type: 'food', price: { $lt: 9.95 } } )
```

This query selects all documents where the `type` field has the value `'food'` **and** the value of the `price` field is less than `9.95`. See *comparison operators* for other comparison operators.

Specify OR Conditions

Using the `$or` operator, you can specify a compound query that joins each clause with a logical `OR` conjunction so that the query selects the documents in the collection that match at least one condition.

In the following example, the query document selects all documents in the collection where the field `qty` has a value greater than (`$gt`) `100` **or** the value of the `price` field is less than (`$lt`) `9.95`:

```
db.inventory.find(
   {
     $or: [ { qty: { $gt: 100 } }, { price: { $lt: 9.95 } } ]
   }
)
```

Note: Queries which use *comparison operators* are subject to *type-bracketing*.

Specify AND as well as OR Conditions

With additional clauses, you can specify precise conditions for matching documents.

In the following example, the compound query document selects all documents in the collection where the value of the `type` field is `'food'` **and** *either* the `qty` has a value greater than (`$gt`) `100` *or* the value of the `price` field is less than (`$lt`) `9.95`:

```
db.inventory.find(
   {
     type: 'food',
     $or: [ { qty: { $gt: 100 } }, { price: { $lt: 9.95 } } ]
   }
)
```

Embedded Documents

When the field holds an embedded document, a query can either specify an exact match on the embedded document or specify a match by individual fields in the embedded document using the *dot notation*.

Exact Match on the Embedded Document

To specify an equality match on the whole embedded document, use the query document `{ <field>: <value> }` where `<value>` is the document to match. Equality matches on an embedded document require an *exact* match of the specified `<value>`, including the field order.

In the following example, the query matches all documents where the value of the field `producer` is an embedded document that contains *only* the field `company` with the value `'ABC123'` and the field `address` with the value `'123 Street'`, in the exact order:

```
db.inventory.find(
    {
      producer:
        {
          company: 'ABC123',
          address: '123 Street'
        }
    }
)
```

Equality Match on Fields within an Embedded Document

Use the *dot notation* to match by specific fields in an embedded document. Equality matches for specific fields in an embedded document will select documents in the collection where the embedded document contains the specified fields with the specified values. The embedded document can contain additional fields.

In the following example, the query uses the *dot notation* to match all documents where the value of the field `producer` is an embedded document that contains a field `company` with the value `'ABC123'` and may contain other fields:

```
db.inventory.find( { 'producer.company': 'ABC123' } )
```

Arrays

When the field holds an array, you can query for an exact array match or for specific values in the array. If the array holds embedded documents, you can query for specific fields in the embedded documents using *dot notation*.

If you specify multiple conditions using the `$elemMatch` operator, the array must contain at least one element that satisfies all the conditions. See *Single Element Satisfies the Criteria* (page 144).

If you specify multiple conditions without using the `$elemMatch` operator, then some combination of the array elements, not necessarily a single element, must satisfy all the conditions; i.e. different elements in the array can satisfy different parts of the conditions. See *Combination of Elements Satisfies the Criteria* (page 144).

Consider an `inventory` collection that contains the following documents:

```
{ _id: 5, type: "food", item: "aaa", ratings: [ 5, 8, 9 ] }
{ _id: 6, type: "food", item: "bbb", ratings: [ 5, 9 ] }
{ _id: 7, type: "food", item: "ccc", ratings: [ 9, 5, 8 ] }
```

Exact Match on an Array

To specify equality match on an array, use the query document `{ <field>: <value> }` where `<value>` is the array to match. Equality matches on the array require that the array field match *exactly* the specified `<value>`, including the element order.

The following example queries for all documents where the field `ratings` is an array that holds exactly three elements, `5`, `8`, and `9`, in this order:

```
db.inventory.find( { ratings: [ 5, 8, 9 ] } )
```

The operation returns the following document:

```
{ "_id" : 5, "type" : "food", "item" : "aaa", "ratings" : [ 5, 8, 9 ] }
```

Match an Array Element

Equality matches can specify a single element in the array to match. These specifications match if the array contains at least *one* element with the specified value.

The following example queries for all documents where `ratings` is an array that contains `5` as one of its elements:

```
db.inventory.find( { ratings: 5 } )
```

The operation returns the following documents:

```
{ "_id" : 5, "type" : "food", "item" : "aaa", "ratings" : [ 5, 8, 9 ] }
{ "_id" : 6, "type" : "food", "item" : "bbb", "ratings" : [ 5, 9 ] }
{ "_id" : 7, "type" : "food", "item" : "ccc", "ratings" : [ 9, 5, 8 ] }
```

Match a Specific Element of an Array

Equality matches can specify equality matches for an element at a particular index or position of the array using the *dot notation*.

In the following example, the query uses the *dot notation* to match all documents where the `ratings` array contains `5` as the first element:

```
db.inventory.find( { 'ratings.0': 5 } )
```

The operation returns the following documents:

```
{ "_id" : 5, "type" : "food", "item" : "aaa", "ratings" : [ 5, 8, 9 ] }
{ "_id" : 6, "type" : "food", "item" : "bbb", "ratings" : [ 5, 9 ] }
```

Specify Multiple Criteria for Array Elements

Single Element Satisfies the Criteria Use `$elemMatch` operator to specify multiple criteria on the elements of an array such that at least one array element satisfies all the specified criteria.

The following example queries for documents where the `ratings` array contains at least one element that is greater than (`$gt`) `5` and less than (`$lt`) `9`:

```
db.inventory.find( { ratings: { $elemMatch: { $gt: 5, $lt: 9 } } } )
```

The operation returns the following documents, whose `ratings` array contains the element `8` which meets the criteria:

```
{ "_id" : 5, "type" : "food", "item" : "aaa", "ratings" : [ 5, 8, 9 ] }
{ "_id" : 7, "type" : "food", "item" : "ccc", "ratings" : [ 9, 5, 8 ] }
```

Combination of Elements Satisfies the Criteria The following example queries for documents where the `ratings` array contains elements that in some combination satisfy the query conditions; e.g., one element can satisfy the greater than `5` condition and another element can satisfy the less than `9` condition, or a single element can satisfy both:

```
db.inventory.find( { ratings: { $gt: 5, $lt: 9 } } )
```

The operation returns the following documents:

```
{ "_id" : 5, "type" : "food", "item" : "aaa", "ratings" : [ 5, 8, 9 ] }
{ "_id" : 6, "type" : "food", "item" : "bbb", "ratings" : [ 5, 9 ] }
{ "_id" : 7, "type" : "food", "item" : "ccc", "ratings" : [ 9, 5, 8 ] }
```

The document with the `"ratings" : [ 5, 9 ]` matches the query since the element `9` is greater than `5` (the first condition) and the element `5` is less than `9` (the second condition).

Array of Embedded Documents

Consider that the `inventory` collection includes the following documents:

```
{
  _id: 100,
  type: "food",
  item: "xyz",
  qty: 25,
  price: 2.5,
  ratings: [ 5, 8, 9 ],
  memos: [ { memo: "on time", by: "shipping" }, { memo: "approved", by: "billing" } ]
}
```

```
{
  _id: 101,
  type: "fruit",
  item: "jkl",
  qty: 10,
  price: 4.25,
  ratings: [ 5, 9 ],
  memos: [ { memo: "on time", by: "payment" }, { memo: "delayed", by: "shipping" } ]
}
```

Match a Field in the Embedded Document Using the Array Index If you know the array index of the embedded document, you can specify the document using the embedded document's position using the *dot notation*.

The following example selects all documents where the `memos` contains an array whose first element (i.e. index is `0`) is a document that contains the field `by` whose value is `'shipping'`:

```
db.inventory.find( { 'memos.0.by': 'shipping' } )
```

The operation returns the following document:

```
{
   _id: 100,
   type: "food",
   item: "xyz",
   qty: 25,
   price: 2.5,
   ratings: [ 5, 8, 9 ],
   memos: [ { memo: "on time", by: "shipping" }, { memo: "approved", by: "billing" } ]
}
```

Match a Field Without Specifying Array Index If you do not know the index position of the document in the array, concatenate the name of the field that contains the array, with a dot (`.`) and the name of the field in the embedded document.

The following example selects all documents where the `memos` field contains an array that contains at least one embedded document that contains the field `by` with the value `'shipping'`:

```
db.inventory.find( { 'memos.by': 'shipping' } )
```

The operation returns the following documents:

```
{
  _id: 100,
  type: "food",
  item: "xyz",
  qty: 25,
  price: 2.5,
  ratings: [ 5, 8, 9 ],
  memos: [ { memo: "on time", by: "shipping" }, { memo: "approved", by: "billing" } ]
}
{
  _id: 101,
  type: "fruit",
  item: "jkl",
  qty: 10,
  price: 4.25,
  ratings: [ 5, 9 ],
```

```
  memos: [ { memo: "on time", by: "payment" }, { memo: "delayed", by: "shipping" } ]
}
```

Specify Multiple Criteria for Array of Documents

Single Element Satisfies the Criteria Use `$elemMatch` operator to specify multiple criteria on an array of embedded documents such that at least one embedded document satisfies all the specified criteria.

The following example queries for documents where the `memos` array has at least one embedded document that contains both the field `memo` equal to `'on time'` and the field `by` equal to `'shipping'`:

```
db.inventory.find(
   {
     memos:
       {
          $elemMatch:
            {
               memo: 'on time',
               by: 'shipping'
            }
       }
    }
)
```

The operation returns the following document:

```
{
   _id: 100,
   type: "food",
   item: "xyz",
   qty: 25,
   price: 2.5,
   ratings: [ 5, 8, 9 ],
   memos: [ { memo: "on time", by: "shipping" }, { memo: "approved", by: "billing" } ]
}
```

Combination of Elements Satisfies the Criteria The following example queries for documents where the `memos` array contains elements that in some combination satisfy the query conditions; e.g. one element satisfies the field `memo` equal to `'on time'` condition and another element satisfies the field `by` equal to `'shipping'` condition, or a single element can satisfy both criteria:

```
db.inventory.find(
  {
    'memos.memo': 'on time',
    'memos.by': 'shipping'
  }
)
```

The query returns the following documents:

```
{
  _id: 100,
  type: "food",
  item: "xyz",
  qty: 25,
  price: 2.5,
```

```
  ratings: [ 5, 8, 9 ],
  memos: [ { memo: "on time", by: "shipping" }, { memo: "approved", by: "billing" } ]
}
{
  _id: 101,
  type: "fruit",
  item: "jkl",
  qty: 10,
  price: 4.25,
  ratings: [ 5, 9 ],
  memos: [ { memo: "on time", by: "payment" }, { memo: "delayed", by: "shipping" } ]

}
```

See also:

Limit Fields to Return from a Query (page 153)

Null or Missing Fields

Different query operators in MongoDB treat `null` values differently.

Given the collection `inventory` with the following documents:

```
{ "_id" : 900, "item" : null }
{ "_id" : 901 }
```

Equality Filter

The `{ item : null }` query matches documents that either contain the `item` field whose value is `null` *or* that do not contain the `item` field.

Given the following query:

```
db.inventory.find( { item: null } )
```

The query returns both documents:

```
{ "_id" : 900, "item" : null }
{ "_id" : 901 }
```

If the query uses an index that is *sparse* (page 574), however, then the query will only match `null` values, not missing fields.

Changed in version 2.6: If using the sparse index results in an incomplete result, MongoDB will not use the index unless a `hint()` explicitly specifies the index. See *Sparse Indexes* (page 574) for more information.

Type Check

The `{ item : { $type: 10 } }` query matches documents that contains the `item` field whose value is `null` *only*; i.e. the value of the `item` field is of BSON Type `Null` (i.e. `10`) :

```
db.inventory.find( { item : { $type: 10 } } )
```

The query returns only the document where the `item` field has a `null` value:

```
{ "_id" : 900, "item" : null }
```

Existence Check

The `{ item : { $exists: false } }` query matches documents that do not contain the `item` field:

```
db.inventory.find( { item : { $exists: false } } )
```

The query returns only the document that does *not* contain the `item` field:

```
{ "_id" : 901 }
```

See also:

The reference documentation for the `$type` and `$exists` operators.

4.3.3 Modify Documents

On this page

- Update Specific Fields in a Document (page 148)
- Replace the Document (page 150)
- `upsert` Option (page 150)
- Additional Examples and Methods (page 152)

MongoDB provides the `update()` method to update the documents of a collection. The method accepts as its parameters:

- an query filter document to determine which documents to update,
- an update document to specify the modification to perform or a replacement document that wholly replaces the matching documents except for the `_id` field, and
- an options document.

By default, `update()` updates a single document. To update multiple documents, use the *multi* option.

Update Specific Fields in a Document

To change a field value, MongoDB provides update operators[10], such as `$set` to modify values.

To specify the modification to perform using update operators, use an update document of the form:

```
{
   <update operator>: { <field1>: <value1>, ... },
   <update operator>: { <field2>: <value2>, ... },
   ...
}
```

Some update operators, such as `$set`, will create the field if the field does not exist. See the individual update operator[11] reference.

[10] https://docs.mongodb.org/manual/reference/operator/update
[11] https://docs.mongodb.org/manual/reference/operator/update

Step 1: Use update operators to change field values.

For the document with `item` equal to `"MNO2"`, use the `$set` operator to update the `category` field and the `details` field to the specified values and the `$currentDate` operator to update the field `lastModified` with the current date.

```
db.inventory.update(
    { item: "MNO2" },
    {
      $set: {
        category: "apparel",
        details: { model: "14Q3", manufacturer: "XYZ Company" }
      },
      $currentDate: { lastModified: true }
    }
)
```

The update operation returns a `WriteResult` object which contains the status of the operation. A successful update of the document returns the following object:

```
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
```

The `nMatched` field specifies the number of existing documents matched for the update, and `nModified` specifies the number of existing documents modified.

Step 2: Update an embedded field.

To update a field within an embedded document, use the *dot notation*. When using the dot notation, enclose the whole dotted field name in quotes.

The following updates the `model` field within the embedded `details` document.

```
db.inventory.update(
  { item: "ABC1" },
  { $set: { "details.model": "14Q2" } }
)
```

The update operation returns a `WriteResult` object which contains the status of the operation. A successful update of the document returns the following object:

```
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
```

Step 3: Update multiple documents.

By default, the `update()` method updates a single document. To update multiple documents, use the `multi` option in the `update()` method.

Update the `category` field to `"apparel"` and update the `lastModified` field to the current date for *all* documents that have `category` field equal to `"clothing"`.

```
db.inventory.update(
   { category: "clothing" },
   {
     $set: { category: "apparel" },
     $currentDate: { lastModified: true }
   },
```

```
   { multi: true }
)
```

The update operation returns a `WriteResult` object which contains the status of the operation. A successful update of the document returns the following object:

```
WriteResult({ "nMatched" : 3, "nUpserted" : 0, "nModified" : 3 })
```

Replace the Document

To replace the entire content of a document except for the `_id` field, pass an entirely new document as the second argument to `update()`.

The replacement document can have different fields from the original document. In the replacement document, you can omit the `_id` field since the `_id` field is immutable. If you do include the `_id` field, it must be the same value as the existing value.

Step 1: Replace a document.

The following operation replaces the document with `item` equal to `"BE10"`. The newly replaced document will only contain the `_id` field and the fields in the replacement document.

```
db.inventory.update(
   { item: "BE10" },
   {
     item: "BE05",
     stock: [ { size: "S", qty: 20 }, { size: "M", qty: 5 } ],
     category: "apparel"
   }
)
```

The update operation returns a `WriteResult` object which contains the status of the operation. A successful update of the document returns the following object:

```
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
```

`upsert` Option

By default, if no document matches the update query, the `update()` method does nothing.

However, by specifying *upsert: true*, the `update()` method either updates matching document or documents, or inserts a new document using the update specification if no matching document exists.

Step 1: Specify `upsert: true` for the update replacement operation.

When you specify `upsert: true` for an update operation to replace a document and no matching documents are found, MongoDB creates a new document using the equality conditions in the update conditions document, and replaces this document, except for the `_id` field if specified, with the update document.

The following operation either updates a matching document by replacing it with a new document or adds a new document if no matching document exists.

```
db.inventory.update(
   { item: "TBD1" },
   {
     item: "TBD1",
     details: { "model" : "14Q4", "manufacturer" : "ABC Company" },
     stock: [ { "size" : "S", "qty" : 25 } ],
     category: "houseware"
   },
   { upsert: true }
)
```

The update operation returns a `WriteResult` object which contains the status of the operation, including whether the `db.collection.update()` method modified an existing document or added a new document.

```
WriteResult({
    "nMatched" : 0,
    "nUpserted" : 1,
    "nModified" : 0,
    "_id" : ObjectId("53dbd684babeaec6342ed6c7")
})
```

The `nMatched` field shows that the operation matched `0` documents.

The `nUpserted` of `1` shows that the update added a document.

The `nModified` of `0` specifies that no existing documents were updated.

The `_id` field shows the generated `_id` field for the added document.

Step 2: Specify `upsert: true` for the update specific fields operation.

When you specify `upsert: true` for an update operation that modifies specific fields and no matching documents are found, MongoDB creates a new document using the equality conditions in the update conditions document, and applies the modification as specified in the update document.

The following update operation either updates specific fields of a matching document or adds a new document if no matching document exists.

```
db.inventory.update(
   { item: "TBD2" },
   {
     $set: {
        details: { "model" : "14Q3", "manufacturer" : "IJK Co." },
        category: "houseware"
     }
   },
   { upsert: true }
)
```

The update operation returns a `WriteResult` object which contains the status of the operation, including whether the `db.collection.update()` method modified an existing document or added a new document.

```
WriteResult({
    "nMatched" : 0,
    "nUpserted" : 1,
    "nModified" : 0,
    "_id" : ObjectId("53dbd7c8babeaec6342ed6c8")
})
```

The `nMatched` field shows that the operation matched `0` documents.

The `nUpserted` of `1` shows that the update added a document.

The `nModified` of `0` specifies that no existing documents were updated.

The information above indicates that the operation has created one new document. The `_id` field shows the generated `_id` field for the added document; you can perform a query to confirm the result:

```
db.inventory.findOne( { _id: ObjectId("53dbd7c8babeaec6342ed6c8") } )
```

The result matches the document specified in the `update()`:

```
{
    "_id" : ObjectId("56a12ec8242ae5d73c07b15e"),
    "item" : "TBD2",
    "details" : {
        "model" : "14Q3",
        "manufacturer" : "IJK Co."
    },
    "category" : "houseware"
}
```

Additional Examples and Methods

For more examples, see *Update examples* in the `db.collection.update()` reference page.

The `db.collection.findAndModify()` and the `db.collection.save()` method can also modify existing documents or insert a new one. See the individual reference pages for the methods for more information and examples.

4.3.4 Remove Documents

On this page

- Remove All Documents (page 152)
- Remove Documents that Match a Condition (page 153)
- Remove a Single Document that Matches a Condition (page 153)

In MongoDB, the `db.collection.remove()` method removes documents from a collection. You can remove all documents from a collection, remove all documents that match a condition, or limit the operation to remove just a single document.

This tutorial provides examples of remove operations using the `db.collection.remove()` method in the `mongo` shell.

Remove All Documents

To remove all documents from a collection, pass an empty query document `{}` to the `remove()` method. The `remove()` method does not remove the indexes.

The following example removes all documents from the `inventory` collection:

```
db.inventory.remove({})
```

To remove all documents from a collection, it may be more efficient to use the `drop()` method to drop the entire collection, including the indexes, and then recreate the collection and rebuild the indexes.

Remove Documents that Match a Condition

To remove the documents that match a deletion criteria, call the `remove()` method with the `<query>` parameter.

The following example removes all documents from the `inventory` collection where the `type` field equals `food`:

```
db.inventory.remove( { type : "food" } )
```

For large deletion operations, it may be more efficient to copy the documents that you want to keep to a new collection and then use `drop()` on the original collection.

Remove a Single Document that Matches a Condition

To remove a single document, call the `remove()` method with the `justOne` parameter set to `true` or `1`.

The following example removes one document from the `inventory` collection where the `type` field equals `food`:

```
db.inventory.remove( { type : "food" }, 1 )
```

To delete a single document sorted by some specified order, use the *findAndModify()* method.

4.3.5 Limit Fields to Return from a Query

On this page

- Return All Fields in Matching Documents (page 154)
- Return the Specified Fields and the `_id` Field Only (page 154)
- Return Specified Fields Only (page 154)
- Return All But the Excluded Field (page 154)
- Return Specific Fields in Embedded Documents (page 154)
- Suppress Specific Fields in Embedded Documents (page 155)
- Projection for Array Fields (page 156)

The *projection* document limits the fields to return for all matching documents. The projection document can specify the inclusion of fields or the exclusion of fields.

The specifications have the following forms:

Syntax	Description
`<field>: <1 or true>`	Specify the inclusion of a field.
`<field>: <0 or false>`	Specify the suppression of the field.

Important: The `_id` field is, by default, included in the result set. To suppress the `_id` field from the result set, specify `_id: 0` in the projection document.

You cannot combine inclusion and exclusion semantics in a single projection with the *exception* of the `_id` field.

This tutorial offers various query examples that limit the fields to return for all matching documents. The examples in this tutorial use a collection `inventory` and use the `db.collection.find()` method in the `mongo` shell. The `db.collection.find()` method returns a *cursor* (page 103) to the retrieved documents. For examples on query selection criteria, see *Query Documents* (page 140).

Return All Fields in Matching Documents

If you specify no projection, the `find()` method returns all fields of all documents that match the query.

```
db.inventory.find( { type: 'food' } )
```

This operation will return all documents in the `inventory` collection where the value of the `type` field is `'food'`. The returned documents contain all fields.

Return the Specified Fields and the `_id` Field Only

A projection can explicitly include several fields. In the following operation, the `find()` method returns all documents that match the query. In the result set, only the `item` and `qty` fields and, by default, the `_id` field return in the matching documents.

```
db.inventory.find( { type: 'food' }, { item: 1, qty: 1 } )
```

Return Specified Fields Only

You can remove the `_id` field from the results by specifying its exclusion in the projection, as in the following example:

```
db.inventory.find( { type: 'food' }, { item: 1, qty: 1, _id:0 } )
```

This operation returns all documents that match the query. In the result set, *only* the `item` and `qty` fields return in the matching documents.

Return All But the Excluded Field

To exclude a single field or group of fields you can use a projection in the following form:

```
db.inventory.find( { type: 'food' }, { type:0 } )
```

This operation returns all documents where the value of the `type` field is `food`. In the result set, the `type` field does not return in the matching documents.

With the exception of the `_id` field you cannot combine inclusion and exclusion statements in projection documents.

Return Specific Fields in Embedded Documents

Use the *dot notation* (page 9) to return specific fields inside an embedded document. For example, the `inventory` collection contains the following document:

```
{
  "_id" : 3,
  "type" : "food",
  "item" : "aaa",
  "classification": { dept: "grocery", category: "chocolate"  }
}
```

The following operation returns all documents that match the query. The specified projection returns only the `category` field in the `classification` document. The returned `category` field remains inside the `classification` document.

```
db.inventory.find(
   { type: 'food', _id: 3 },
   { "classification.category": 1, _id: 0 }
)
```

The operation returns the following document:

```
{ "classification" : { "category" : "chocolate" } }
```

Suppress Specific Fields in Embedded Documents

Use *dot notation* (page 9) to suppress specific fields inside an embedded document using a `0` instead of `1`. For example, the `inventory` collection contains the following document:

```
{
   "_id" : 3,
   "type" : "food",
   "item" : "Super Dark Chocolate",
   "classification" : { "dept" : "grocery", "category" : "chocolate"},
   "vendor" : {
      "primary" : {
         "name" : "Marsupial Vending Co",
         "address" : "Wallaby Rd",
         "delivery" : ["M","W","F"]
      },
      "secondary":{
         "name" : "Intl. Chocolatiers",
         "address" : "Cocoa Plaza",
         "delivery" : ["Sa"]
      }
   }
}
```

The following operation returns all documents where the value of the `type` field is `food` and the `_id` field is `3`. The projection suppresses only the `category` field in the `classification` document. The `dept` field remains inside the `classification` document.

```
db.inventory.find(
   { type: 'food', _id: 3 },
   { "classification.category": 0}
)
```

The operation returns the following document:

```
{
   "_id" : 3,
   "type" : "food",
   "item" : "Super Dark Chocolate",
   "classification" : { "dept" : "grocery"},
   "vendor" : {
      "primary" : {
         "name" : "Bobs Vending",
         "address" : "Wallaby Rd",
         "delivery" : ["M","W","F"]
      },
      "secondary":{
         "name" : "Intl. Chocolatiers",
         "address" : "Cocoa Plaza",
```

```
            "delivery" : ["Sa"]
        }
    }
}
```

You can suppress nested subdocuments at any depth using *dot notation* (page 9). The following specifies a projection to suppress the `delivery` array only for the `secondary` document.

```
db.inventory.find(
    { "type" : "food" },
    { "vendor.secondary.delivery" : 0 }
)
```

This returns all documents except the `delivery` array in the `secondary` document

```
{
    "_id" : 3,
    "type" : "food",
    "item" : "Super Dark Chocolate",
    "classification" : { "dept" : "grocery", "category" : "chocolate"},
    "vendor" : {
        "primary" : {
            "name" : "Bobs Vending",
            "address" : "Wallaby Rd",
            "delivery" : ["M","W","F"]
        },
        "secondary":{
            "name" : "Intl. Chocolatiers",
            "address" : "Cocoa Plaza"
        }
    }
}
```

Projection for Array Fields

For fields that contain arrays, MongoDB provides the following projection operators: `$elemMatch`, `$slice`, and `$`.

For example, the `inventory` collection contains the following document:

```
{ "_id" : 5, "type" : "food", "item" : "aaa", "ratings" : [ 5, 8, 9 ] }
```

Then the following operation uses the `$slice` projection operator to return just the first two elements in the `ratings` array.

```
db.inventory.find( { _id: 5 }, { ratings: { $slice: 2 } } )
```

`$elemMatch`, `$slice`, and `$` are the *only* way to project *portions* of an array. For instance, you *cannot* project a portion of an array using the array index; e.g. `{ "ratings.0": 1 }` projection will *not* project the array with the first element.

See also:

Query Documents (page 140)

4.3.6 Limit Number of Elements in an Array after an Update

On this page

- Synopsis (page 157)
- Pattern (page 157)

New in version 2.4.

Synopsis

Consider an application where users may submit many scores (e.g. for a test), but the application only needs to track the top three test scores.

This pattern uses the `$push` operator with the `$each`, `$sort`, and `$slice` modifiers to sort and maintain an array of fixed size.

Pattern

Consider the following document in the collection `students`:

```
{
  _id: 1,
  scores: [
    { attempt: 1, score: 10 },
    { attempt: 2 , score:8 }
  ]
}
```

The following update uses the `$push` operator with:

- the `$each` modifier to append to the array 2 new elements,
- the `$sort` modifier to order the elements by ascending (`1`) score, and
- the `$slice` modifier to keep the last `3` elements of the ordered array.

```
db.students.update(
   { _id: 1 },
   {
     $push: {
        scores: {
           $each: [ { attempt: 3, score: 7 }, { attempt: 4, score: 4 } ],
           $sort: { score: 1 },
           $slice: -3
        }
     }
   }
)
```

Note: When using the `$sort` modifier on the array element, access the field in the embedded document element directly instead of using the *dot notation* on the array field.

After the operation, the document contains only the top 3 scores in the `scores` array:

```
{
   "_id" : 1,
   "scores" : [
```

```
      { "attempt" : 3, "score" : 7 },
      { "attempt" : 2, "score" : 8 },
      { "attempt" : 1, "score" : 10 }
   ]
}
```

See also:

- `$push` operator,
- `$each` modifier,
- `$sort` modifier, and
- `$slice` modifier.

4.3.7 Iterate a Cursor in the `mongo` Shell

On this page

- Manually Iterate the Cursor (page 158)
- Iterator Index (page 159)

The `db.collection.find()` method returns a cursor. To access the documents, you need to iterate the cursor. However, in the `mongo` shell, if the returned cursor is not assigned to a variable using the `var` keyword, then the cursor is automatically iterated up to 20 times to print up to the first 20 documents in the results. The following describes ways to manually iterate the cursor to access the documents or to use the iterator index.

Manually Iterate the Cursor

In the `mongo` shell, when you assign the cursor returned from the `find()` method to a variable using the `var` keyword, the cursor does not automatically iterate.

You can call the cursor variable in the shell to iterate up to 20 times [12] and print the matching documents, as in the following example:

```
var myCursor = db.inventory.find( { type: 'food' } );

myCursor
```

You can also use the cursor method `next()` to access the documents, as in the following example:

```
var myCursor = db.inventory.find( { type: 'food' } );

while (myCursor.hasNext()) {
   print(tojson(myCursor.next()));
}
```

As an alternative print operation, consider the `printjson()` helper method to replace `print(tojson())`:

```
var myCursor = db.inventory.find( { type: 'food' } );

while (myCursor.hasNext()) {
```

[12] You can use the `DBQuery.shellBatchSize` to change the number of iteration from the default value `20`. See *Working with the mongo Shell* (page 78) for more information.

```
    printjson(myCursor.next());
}
```

You can use the cursor method `forEach()` to iterate the cursor and access the documents, as in the following example:

```
var myCursor =  db.inventory.find( { type: 'food' } );

myCursor.forEach(printjson);
```

See *JavaScript cursor methods* and your `driver` documentation for more information on cursor methods.

Iterator Index

In the `mongo` shell, you can use the `toArray()` method to iterate the cursor and return the documents in an array, as in the following:

```
var myCursor = db.inventory.find( { type: 'food' } );
var documentArray = myCursor.toArray();
var myDocument = documentArray[3];
```

The `toArray()` method loads into RAM all documents returned by the cursor; the `toArray()` method exhausts the cursor.

Additionally, some `drivers` provide access to the documents by using an index on the cursor (i.e. `cursor[index]`). This is a shortcut for first calling the `toArray()` method and then using an index on the resulting array.

Consider the following example:

```
var myCursor = db.inventory.find( { type: 'food' } );
var myDocument = myCursor[3];
```

The `myCursor[3]` is equivalent to the following example:

```
myCursor.toArray() [3];
```

4.3.8 Analyze Query Performance

On this page

- Evaluate the Performance of a Query (page 160)
- Compare Performance of Indexes (page 162)
- Additional Resources (page 164)

The `cursor.explain("executionStats")` and the `db.collection.explain("executionStats")` methods provide statistics about the performance of a query. This data output can be useful in measuring if and how a query uses an index.

`db.collection.explain()` provides information on the execution of other operations, such as `db.collection.update()`. See `db.collection.explain()` for details.

Evaluate the Performance of a Query

Consider a collection `inventory` with the following documents:

```
{ "_id" : 1, "item" : "f1", type: "food", quantity: 500 }
{ "_id" : 2, "item" : "f2", type: "food", quantity: 100 }
{ "_id" : 3, "item" : "p1", type: "paper", quantity: 200 }
{ "_id" : 4, "item" : "p2", type: "paper", quantity: 150 }
{ "_id" : 5, "item" : "f3", type: "food", quantity: 300 }
{ "_id" : 6, "item" : "t1", type: "toys", quantity: 500 }
{ "_id" : 7, "item" : "a1", type: "apparel", quantity: 250 }
{ "_id" : 8, "item" : "a2", type: "apparel", quantity: 400 }
{ "_id" : 9, "item" : "t2", type: "toys", quantity: 50 }
{ "_id" : 10, "item" : "f4", type: "food", quantity: 75 }
```

Query with No Index

The following query retrieves documents where the `quantity` field has a value between `100` and `200`, inclusive:

```
db.inventory.find( { quantity: { $gte: 100, $lte: 200 } } )
```

The query returns the following documents:

```
{ "_id" : 2, "item" : "f2", "type" : "food", "quantity" : 100 }
{ "_id" : 3, "item" : "p1", "type" : "paper", "quantity" : 200 }
{ "_id" : 4, "item" : "p2", "type" : "paper", "quantity" : 150 }
```

To view the query plan selected, use the `explain("executionStats")` method:

```
db.inventory.find(
   { quantity: { $gte: 100, $lte: 200 } }
).explain("executionStats")
```

`explain()` returns the following results:

```
{
   "queryPlanner" : {
         "plannerVersion" : 1,
         ...
         "winningPlan" : {
            "stage" : "COLLSCAN",
            ...
         }
   },
   "executionStats" : {
      "executionSuccess" : true,
      "nReturned" : 3,
      "executionTimeMillis" : 0,
      "totalKeysExamined" : 0,
      "totalDocsExamined" : 10,
      "executionStages" : {
         "stage" : "COLLSCAN",
         ...
      },
      ...
   },
   ...
}
```

- `queryPlanner.winningPlan.stage` displays `COLLSCAN` to indicate a collection scan.
- `executionStats.nReturned` displays `3` to indicate that the query matches and returns three documents.
- `executionStats.totalDocsExamined` display `10` to indicate that MongoDB had to scan ten documents (i.e. all documents in the collection) to find the three matching documents.

The difference between the number of matching documents and the number of examined documents may suggest that, to improve efficiency, the query might benefit from the use of an index.

Query with Index

To support the query on the `quantity` field, add an index on the `quantity` field:

```
db.inventory.createIndex( { quantity: 1 } )
```

To view the query plan statistics, use the `explain("executionStats")` method:

```
db.inventory.find(
   { quantity: { $gte: 100, $lte: 200 } }
).explain("executionStats")
```

The `explain()` method returns the following results:

```
{
   "queryPlanner" : {
         "plannerVersion" : 1,
         ...
         "winningPlan" : {
               "stage" : "FETCH",
               "inputStage" : {
                  "stage" : "IXSCAN",
                  "keyPattern" : {
                     "quantity" : 1
                  },
                  ...
               }
         },
         "rejectedPlans" : [ ]
   },
   "executionStats" : {
         "executionSuccess" : true,
         "nReturned" : 3,
         "executionTimeMillis" : 0,
         "totalKeysExamined" : 3,
         "totalDocsExamined" : 3,
         "executionStages" : {
            ...
         },
         ...
   },
   ...
}
```

- `queryPlanner.winningPlan.inputStage.stage` displays `IXSCAN` to indicate index use.
- `executionStats.nReturned` displays `3` to indicate that the query matches and returns three documents.
- `executionStats.totalKeysExamined` display `3` to indicate that MongoDB scanned three index entries.

- `executionStats.totalDocsExamined` display `3` to indicate that MongoDB scanned three documents.

When run with an index, the query scanned `3` index entries and `3` documents to return `3` matching documents. Without the index, to return the `3` matching documents, the query had to scan the whole collection, scanning `10` documents.

Compare Performance of Indexes

To manually compare the performance of a query using more than one index, you can use the `hint()` method in conjunction with the `explain()` method.

Consider the following query:

```
db.inventory.find( { quantity: { $gte: 100, $lte: 300 }, type: "food" } )
```

The query returns the following documents:

```
{ "_id" : 2, "item" : "f2", "type" : "food", "quantity" : 100 }
{ "_id" : 5, "item" : "f3", "type" : "food", "quantity" : 300 }
```

To support the query, add a *compound index* (page 522). With *compound indexes* (page 522), the order of the fields matter.

For example, add the following two compound indexes. The first index orders by `quantity` field first, and then the `type` field. The second index orders by `type` first, and then the `quantity` field.

```
db.inventory.createIndex( { quantity: 1, type: 1 } )
db.inventory.createIndex( { type: 1, quantity: 1 } )
```

Evaluate the effect of the first index on the query:

```
db.inventory.find(
   { quantity: { $gte: 100, $lte: 300 }, type: "food" }
).hint({ quantity: 1, type: 1 }).explain("executionStats")
```

The `explain()` method returns the following output:

```
{
   "queryPlanner" : {
      ...
      "winningPlan" : {
         "stage" : "FETCH",
         "inputStage" : {
            "stage" : "IXSCAN",
            "keyPattern" : {
               "quantity" : 1,
               "type" : 1
            },
            ...
            }
         }
      },
      "rejectedPlans" : [ ]
   },
   "executionStats" : {
      "executionSuccess" : true,
      "nReturned" : 2,
      "executionTimeMillis" : 0,
      "totalKeysExamined" : 5,
      "totalDocsExamined" : 2,
      "executionStages" : {
```

```
      ...
      }
   },
   ...
}
```

MongoDB scanned `5` index keys (`executionStats.totalKeysExamined`) to return `2` matching documents (`executionStats.nReturned`).

Evaluate the effect of the second index on the query:

```
db.inventory.find(
   { quantity: { $gte: 100, $lte: 300 }, type: "food" }
).hint({ type: 1, quantity: 1 }).explain("executionStats")
```

The `explain()` method returns the following output:

```
{
   "queryPlanner" : {
      ...
      "winningPlan" : {
         "stage" : "FETCH",
         "inputStage" : {
            "stage" : "IXSCAN",
            "keyPattern" : {
               "type" : 1,
               "quantity" : 1
            },
            ...
         }
      },
      "rejectedPlans" : [ ]
   },
   "executionStats" : {
      "executionSuccess" : true,
      "nReturned" : 2,
      "executionTimeMillis" : 0,
      "totalKeysExamined" : 2,
      "totalDocsExamined" : 2,
      "executionStages" : {
         ...
      }
   },
   ...
}
```

MongoDB scanned `2` index keys (`executionStats.totalKeysExamined`) to return `2` matching documents (`executionStats.nReturned`).

For this example query, the compound index `{ type: 1, quantity: 1 }` is more efficient than the compound index `{ quantity: 1, type: 1 }`.

See also:

Query Optimization (page 105), *Query Plans* (page 108), *Optimize Query Performance* (page 314), *Indexing Strategies* (page 586)

Additional Resources

- MongoDB Performance Evaluation and Tuning Consulting Package[13]

4.3.9 Perform Two Phase Commits

On this page

- Synopsis (page 164)
- Background (page 164)
- Pattern (page 165)
- Recovering from Failure Scenarios (page 167)
- Multiple Applications (page 169)
- Using Two-Phase Commits in Production Applications (page 170)

Synopsis

This document provides a pattern for doing multi-document updates or "multi-document transactions" using a two-phase commit approach for writing data to multiple documents. Additionally, you can extend this process to provide a *rollback-like* (page 168) functionality.

Background

Operations on a single *document* are always atomic with MongoDB databases; however, operations that involve multiple documents, which are often referred to as "multi-document transactions", are not atomic. Since documents can be fairly complex and contain multiple "nested" documents, single-document atomicity provides the necessary support for many practical use cases.

Despite the power of single-document atomic operations, there are cases that require multi-document transactions. When executing a transaction composed of sequential operations, certain issues arise, such as:

- Atomicity: if one operation fails, the previous operation within the transaction must "rollback" to the previous state (i.e. the "nothing," in "all or nothing").
- Consistency: if a major failure (i.e. network, hardware) interrupts the transaction, the database must be able to recover a consistent state.

For situations that require multi-document transactions, you can implement two-phase commit in your application to provide support for these kinds of multi-document updates. Using two-phase commit ensures that data is consistent and, in case of an error, the state that preceded the transaction is *recoverable* (page 168). During the procedure, however, documents can represent pending data and states.

Note: Because only single-document operations are atomic with MongoDB, two-phase commits can only offer transaction-*like* semantics. It is possible for applications to return intermediate data at intermediate points during the two-phase commit or rollback.

[13] https://www.mongodb.com/products/consulting?jmp=docs#performance_evaluation

Pattern

Overview

Consider a scenario where you want to transfer funds from account `A` to account `B`. In a relational database system, you can subtract the funds from `A` and add the funds to `B` in a single multi-statement transaction. In MongoDB, you can emulate a two-phase commit to achieve a comparable result.

The examples in this tutorial use the following two collections:

1. A collection named `accounts` to store account information.
2. A collection named `transactions` to store information on the fund transfer transactions.

Initialize Source and Destination Accounts

Insert into the `accounts` collection a document for account `A` and a document for account `B`.

```
db.accounts.insert(
   [
     { _id: "A", balance: 1000, pendingTransactions: [] },
     { _id: "B", balance: 1000, pendingTransactions: [] }
   ]
)
```

The operation returns a `BulkWriteResult()` object with the status of the operation. Upon successful insert, the `BulkWriteResult()` has `nInserted` set to `2` .

Initialize Transfer Record

For each fund transfer to perform, insert into the `transactions` collection a document with the transfer information. The document contains the following fields:

- `source` and `destination` fields, which refer to the `_id` fields from the `accounts` collection,
- `value` field, which specifies the amount of transfer affecting the `balance` of the `source` and `destination` accounts,
- `state` field, which reflects the current state of the transfer. The `state` field can have the value of `initial`, `pending`, `applied`, `done`, `canceling`, and `canceled`.
- `lastModified` field, which reflects last modification date.

To initialize the transfer of `100` from account `A` to account `B`, insert into the `transactions` collection a document with the transfer information, the transaction `state` of `"initial"`, and the `lastModified` field set to the current date:

```
db.transactions.insert(
    { _id: 1, source: "A", destination: "B", value: 100, state: "initial", lastModified: new Date() }
)
```

The operation returns a `WriteResult()` object with the status of the operation. Upon successful insert, the `WriteResult()` object has `nInserted` set to `1`.

Transfer Funds Between Accounts Using Two-Phase Commit

Step 1: Retrieve the transaction to start. From the `transactions` collection, find a transaction in the `initial` state. Currently the `transactions` collection has only one document, namely the one added in the *Initialize Transfer Record* (page 165) step. If the collection contains additional documents, the query will return any transaction with an `initial` state unless you specify additional query conditions.

```
var t = db.transactions.findOne( { state: "initial" } )
```

Type the variable `t` in the `mongo` shell to print the contents of the variable. The operation should print a document similar to the following except the `lastModified` field should reflect date of your insert operation:

```
{ "_id" : 1, "source" : "A", "destination" : "B", "value" : 100, "state" : "initial", "lastModified"
```

Step 2: Update transaction state to pending. Set the transaction `state` from `initial` to `pending` and use the `$currentDate` operator to set the `lastModified` field to the current date.

```
db.transactions.update(
    { _id: t._id, state: "initial" },
    {
      $set: { state: "pending" },
      $currentDate: { lastModified: true }
    }
)
```

The operation returns a `WriteResult()` object with the status of the operation. Upon successful update, the `nMatched` and `nModified` displays `1`.

In the update statement, the `state: "initial"` condition ensures that no other process has already updated this record. If `nMatched` and `nModified` is `0`, go back to the first step to get a different transaction and restart the procedure.

Step 3: Apply the transaction to both accounts. Apply the transaction `t` to both accounts using the `update()` method *if* the transaction has not been applied to the accounts. In the update condition, include the condition `pendingTransactions: { $ne: t._id }` in order to avoid re-applying the transaction if the step is run more than once.

To apply the transaction to the account, update both the `balance` field and the `pendingTransactions` field.

Update the source account, subtracting from its `balance` the transaction `value` and adding to its `pendingTransactions` array the transaction `_id`.

```
db.accounts.update(
   { _id: t.source, pendingTransactions: { $ne: t._id } },
   { $inc: { balance: -t.value }, $push: { pendingTransactions: t._id } }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Update the destination account, adding to its `balance` the transaction `value` and adding to its `pendingTransactions` array the transaction `_id`.

```
db.accounts.update(
   { _id: t.destination, pendingTransactions: { $ne: t._id } },
   { $inc: { balance: t.value }, $push: { pendingTransactions: t._id } }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Step 4: Update transaction state to applied. Use the following `update()` operation to set the transaction's `state` to `applied` and update the `lastModified` field:

```
db.transactions.update(
   { _id: t._id, state: "pending" },
   {
     $set: { state: "applied" },
     $currentDate: { lastModified: true }
   }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Step 5: Update both accounts' list of pending transactions. Remove the applied transaction `_id` from the `pendingTransactions` array for both accounts.

Update the source account.

```
db.accounts.update(
   { _id: t.source, pendingTransactions: t._id },
   { $pull: { pendingTransactions: t._id } }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Update the destination account.

```
db.accounts.update(
   { _id: t.destination, pendingTransactions: t._id },
   { $pull: { pendingTransactions: t._id } }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Step 6: Update transaction state to done. Complete the transaction by setting the `state` of the transaction to `done` and updating the `lastModified` field:

```
db.transactions.update(
   { _id: t._id, state: "applied" },
   {
     $set: { state: "done" },
     $currentDate: { lastModified: true }
   }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Recovering from Failure Scenarios

The most important part of the transaction procedure is not the prototypical example above, but rather the possibility for recovering from the various failure scenarios when transactions do not complete successfully. This section presents an overview of possible failures and provides steps to recover from these kinds of events.

Recovery Operations

The two-phase commit pattern allows applications running the sequence to resume the transaction and arrive at a consistent state. Run the recovery operations at application startup, and possibly at regular intervals, to catch any unfinished transactions.

The time required to reach a consistent state depends on how long the application needs to recover each transaction.

The following recovery procedures uses the `lastModified` date as an indicator of whether the pending transaction requires recovery; specifically, if the pending or applied transaction has not been updated in the last 30 minutes, the procedures determine that these transactions require recovery. You can use different conditions to make this determination.

Transactions in Pending State To recover from failures that occur after step "Update transaction state to pending. (page **??**)" but before "Update transaction state to applied. (page **??**)" step, retrieve from the `transactions` collection a pending transaction for recovery:

```
var dateThreshold = new Date();
dateThreshold.setMinutes(dateThreshold.getMinutes() - 30);

var t = db.transactions.findOne( { state: "pending", lastModified: { $lt: dateThreshold } } );
```

And resume from step "Apply the transaction to both accounts. (page **??**)"

Transactions in Applied State To recover from failures that occur after step "Update transaction state to applied. (page **??**)" but before "Update transaction state to done. (page **??**)" step, retrieve from the `transactions` collection an applied transaction for recovery:

```
var dateThreshold = new Date();
dateThreshold.setMinutes(dateThreshold.getMinutes() - 30);

var t = db.transactions.findOne( { state: "applied", lastModified: { $lt: dateThreshold } } );
```

And resume from "Update both accounts' list of pending transactions. (page **??**)"

Rollback Operations

In some cases, you may need to "roll back" or undo a transaction; e.g., if the application needs to "cancel" the transaction or if one of the accounts does not exist or stops existing during the transaction.

Transactions in Applied State After the "Update transaction state to applied. (page **??**)" step, you should **not** roll back the transaction. Instead, complete that transaction and *create a new transaction* (page 165) to reverse the transaction by switching the values in the source and the destination fields.

Transactions in Pending State After the "Update transaction state to pending. (page **??**)" step, but before the "Update transaction state to applied. (page **??**)" step, you can rollback the transaction using the following procedure:

Step 1: Update transaction state to canceling. Update the transaction `state` from `pending` to `canceling`.

```
db.transactions.update(
   { _id: t._id, state: "pending" },
   {
     $set: { state: "canceling" },
```

```
      $currentDate: { lastModified: true }
   }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Step 2: Undo the transaction on both accounts. To undo the transaction on both accounts, reverse the transaction `t` if the transaction has been applied. In the update condition, include the condition `pendingTransactions: t._id` in order to update the account only if the pending transaction has been applied.

Update the destination account, subtracting from its `balance` the transaction `value` and removing the transaction `_id` from the `pendingTransactions` array.

```
db.accounts.update(
   { _id: t.destination, pendingTransactions: t._id },
   {
     $inc: { balance: -t.value },
     $pull: { pendingTransactions: t._id }
   }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`. If the pending transaction has not been previously applied to this account, no document will match the update condition and `nMatched` and `nModified` will be `0`.

Update the source account, adding to its `balance` the transaction `value` and removing the transaction `_id` from the `pendingTransactions` array.

```
db.accounts.update(
   { _id: t.source, pendingTransactions: t._id },
   {
     $inc: { balance: t.value},
     $pull: { pendingTransactions: t._id }
   }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`. If the pending transaction has not been previously applied to this account, no document will match the update condition and `nMatched` and `nModified` will be `0`.

Step 3: Update transaction state to canceled. To finish the rollback, update the transaction `state` from `canceling` to `cancelled`.

```
db.transactions.update(
   { _id: t._id, state: "canceling" },
   {
     $set: { state: "cancelled" },
     $currentDate: { lastModified: true }
   }
)
```

Upon successful update, the method returns a `WriteResult()` object with `nMatched` and `nModified` set to `1`.

Multiple Applications

Transactions exist, in part, so that multiple applications can create and run operations concurrently without causing data inconsistency or conflicts. In our procedure, to update or retrieve the transaction document, the update conditions

include a condition on the `state` field to prevent reapplication of the transaction by multiple applications.

For example, applications `App1` and `App2` both grab the same transaction, which is in the `initial` state. `App1` applies the whole transaction before `App2` starts. When `App2` attempts to perform the "Update transaction state to pending. (page **??**)" step, the update condition, which includes the `state: "initial"` criterion, will not match any document, and the `nMatched` and `nModified` will be `0`. This should signal to `App2` to go back to the first step to restart the procedure with a different transaction.

When multiple applications are running, it is crucial that only one application can handle a given transaction at any point in time. As such, in addition including the expected state of the transaction in the update condition, you can also create a marker in the transaction document itself to identify the application that is handling the transaction. Use `findAndModify()` method to modify the transaction and get it back in one step:

```
t = db.transactions.findAndModify(
       {
         query: { state: "initial", application: { $exists: false } },
         update:
           {
             $set: { state: "pending", application: "App1" },
             $currentDate: { lastModified: true }
           },
         new: true
       }
    )
```

Amend the transaction operations to ensure that only applications that match the identifier in the `application` field apply the transaction.

If the application `App1` fails during transaction execution, you can use the *recovery procedures* (page 167), but applications should ensure that they "own" the transaction before applying the transaction. For example to find and resume the pending job, use a query that resembles the following:

```
var dateThreshold = new Date();
dateThreshold.setMinutes(dateThreshold.getMinutes() - 30);

db.transactions.find(
   {
     application: "App1",
     state: "pending",
     lastModified: { $lt: dateThreshold }
   }
)
```

Using Two-Phase Commits in Production Applications

The example transaction above is intentionally simple. For example, it assumes that it is always possible to roll back operations to an account and that account balances can hold negative values.

Production implementations would likely be more complex. Typically, accounts need information about current balance, pending credits, and pending debits.

For all transactions, ensure that you use the appropriate level of *write concern* (page 179) for your deployment.

4.3.10 Update Document if Current

On this page

- Overview (page 171)
- Pattern (page 171)
- Example (page 171)
- Modifications to the Pattern (page 172)

Overview

The *Update if Current* pattern is an approach to *concurrency control* (page 126) when multiple applications have access to the data.

Pattern

The pattern queries for the document to update. Then, for each field to modify, the pattern includes the field and its value in the returned document in the query predicate for the update operation. This way, the update only modifies the document fields *if* the fields have not changed since the query.

Example

Consider the following example in the `mongo` shell. The example updates the `quantity` and the `reordered` fields of a document *only* if the fields have not changed since the query.

Changed in version 2.6: The `db.collection.update()` method now returns a `WriteResult()` object that contains the status of the operation. Previous versions required an extra `db.getLastErrorObj()` method call.

```
var myDocument = db.products.findOne( { sku: "abc123" } );

if ( myDocument ) {
   var oldQuantity = myDocument.quantity;
   var oldReordered = myDocument.reordered;

   var results = db.products.update(
      {
        _id: myDocument._id,
        quantity: oldQuantity,
        reordered: oldReordered
      },
      {
        $inc: { quantity: 50 },
        $set: { reordered: true }
      }
   )

   if ( results.hasWriteError() ) {
      print( "unexpected error updating document: " + tojson(results) );
   }
   else if ( results.nMatched === 0 ) {
      print( "No matching document for " +
             "{ _id: "+ myDocument._id.toString() +
             ", quantity: " + oldQuantity +
             ", reordered: " + oldReordered
             + " } "
```

```
        );
    }
}
```

Modifications to the Pattern

Another approach is to add a `version` field to the documents. Applications increment this field upon each update operation to the documents. You must be able to ensure that *all* clients that connect to your database include the `version` field in the query predicate. To associate increasing numbers with documents in a collection, you can use one of the methods described in *Create an Auto-Incrementing Sequence Field* (page 173).

For more approaches, see *Concurrency Control* (page 126).

4.3.11 Create Tailable Cursor

On this page

- Overview (page 172)

Overview

By default, MongoDB will automatically close a cursor when the client has exhausted all results in the cursor. However, for *capped collections* (page 6) you may use a *Tailable Cursor* that remains open after the client exhausts the results in the initial cursor. Tailable cursors are conceptually equivalent to the `tail` Unix command with the `-f` option (i.e. with "follow" mode). After clients insert new additional documents into a capped collection, the tailable cursor will continue to retrieve documents.

Use tailable cursors on capped collections that have high write volumes where indexes aren't practical. For instance, MongoDB *replication* (page 623) uses tailable cursors to tail the primary's *oplog*.

Note: If your query is on an indexed field, do not use tailable cursors, but instead, use a regular cursor. Keep track of the last value of the indexed field returned by the query. To retrieve the newly added documents, query the collection again using the last value of the indexed field in the query criteria, as in the following example:

```
db.<collection>.find( { indexedField: { $gt: <lastvalue> } } )
```

Consider the following behaviors related to tailable cursors:

- Tailable cursors do not use indexes and return documents in *natural order*.
- Because tailable cursors do not use indexes, the initial scan for the query may be expensive; but, after initially exhausting the cursor, subsequent retrievals of the newly added documents are inexpensive.
- Tailable cursors may become *dead*, or invalid, if either:
  - the query returns no match.
  - the cursor returns the document at the "end" of the collection and then the application deletes that document.

  A *dead* cursor has an id of `0`.

See your `driver documentation` for the driver-specific method to specify the tailable cursor.

4.3.12 Create an Auto-Incrementing Sequence Field

On this page

- Synopsis (page 173)
- Considerations (page 173)
- Procedures (page 173)

Synopsis

MongoDB reserves the `_id` field in the top level of all documents as a primary key. `_id` must be unique, and always has an index with a *unique constraint* (page 568). However, except for the unique constraint you can use any value for the `_id` field in your collections. This tutorial describes two methods for creating an incrementing sequence number for the `_id` field using the following:

- *Use Counters Collection* (page 173)
- *Optimistic Loop* (page 175)

Considerations

Generally in MongoDB, you would not use an auto-increment pattern for the `_id` field, or any field, because it does not scale for databases with large numbers of documents. Typically the default value *ObjectId* is more ideal for the `_id`.

Procedures

Use Counters Collection

Counter Collection Implementation Use a separate `counters` collection to track the *last* number sequence used. The `_id` field contains the sequence name and the `seq` field contains the last value of the sequence.

1. Insert into the `counters` collection, the initial value for the `userid`:

```
db.counters.insert(
   {
      _id: "userid",
      seq: 0
   }
)
```

2. Create a `getNextSequence` function that accepts a `name` of the sequence. The function uses the `findAndModify()` method to atomically increment the `seq` value and return this new value:

```
function getNextSequence(name) {
   var ret = db.counters.findAndModify(
          {
            query: { _id: name },
            update: { $inc: { seq: 1 } },
            new: true
          }
   );
```

```
    return ret.seq;
}
```

3. Use this `getNextSequence()` function during `insert()`.

```
db.users.insert(
   {
     _id: getNextSequence("userid"),
     name: "Sarah C."
   }
)

db.users.insert(
   {
     _id: getNextSequence("userid"),
     name: "Bob D."
   }
)
```

You can verify the results with `find()`:

```
db.users.find()
```

The `_id` fields contain incrementing sequence values:

```
{
  _id : 1,
  name : "Sarah C."
}
{
  _id : 2,
  name : "Bob D."
}
```

`findAndModify` Behavior When `findAndModify()` includes the `upsert: true` option **and** the query field(s) is not uniquely indexed, the method could insert a document multiple times in certain circumstances. For instance, if multiple clients each invoke the method with the same query condition and these methods complete the find phase before any of methods perform the modify phase, these methods could insert the same document.

In the `counters` collection example, the query field is the `_id` field, which always has a unique index. Consider that the `findAndModify()` includes the `upsert: true` option, as in the following modified example:

```
function getNextSequence(name) {
   var ret = db.counters.findAndModify(
          {
            query: { _id: name },
            update: { $inc: { seq: 1 } },
            new: true,
            upsert: true
          }
   );

   return ret.seq;
}
```

If multiple clients were to invoke the `getNextSequence()` method with the same `name` parameter, then the methods would observe one of the following behaviors:

- Exactly one `findAndModify()` would successfully insert a new document.

- Zero or more `findAndModify()` methods would update the newly inserted document.
- Zero or more `findAndModify()` methods would fail when they attempted to insert a duplicate.

If the method fails due to a unique index constraint violation, retry the method. Absent a delete of the document, the retry should not fail.

Optimistic Loop

In this pattern, an *Optimistic Loop* calculates the incremented `_id` value and attempts to insert a document with the calculated `_id` value. If the insert is successful, the loop ends. Otherwise, the loop will iterate through possible `_id` values until the insert is successful.

1. Create a function named `insertDocument` that performs the "insert if not present" loop. The function wraps the `insert()` method and takes a `doc` and a `targetCollection` arguments.

   Changed in version 2.6: The `db.collection.insert()` method now returns a *writeresults-insert* object that contains the status of the operation. Previous versions required an extra `db.getLastErrorObj()` method call.

   ```
   function insertDocument(doc, targetCollection) {

       while (1) {

           var cursor = targetCollection.find( {}, { _id: 1 } ).sort( { _id: -1 } ).limit(1);

           var seq = cursor.hasNext() ? cursor.next()._id + 1 : 1;

           doc._id = seq;

           var results = targetCollection.insert(doc);

           if( results.hasWriteError() ) {
               if( results.writeError.code == 11000 /* dup key */ )
                   continue;
               else
                   print( "unexpected error inserting data: " + tojson( results ) );
           }

           break;
       }
   }
   ```

   The `while (1)` loop performs the following actions:

   - Queries the `targetCollection` for the document with the maximum `_id` value.
   - Determines the next sequence value for `_id` by:
     - adding `1` to the returned `_id` value if the returned cursor points to a document.
     - otherwise: it sets the next sequence value to `1` if the returned cursor points to no document.
   - For the `doc` to insert, set its `_id` field to the calculated sequence value `seq`.
   - Insert the `doc` into the `targetCollection`.
   - If the insert operation errors with duplicate key, repeat the loop. Otherwise, if the insert operation encounters some other error or if the operation succeeds, break out of the loop.

2. Use the `insertDocument()` function to perform an insert:

```
var myCollection = db.users2;

insertDocument(
   {
     name: "Grace H."
   },
   myCollection
);

insertDocument(
   {
     name: "Ted R."
   },
   myCollection
)
```

You can verify the results with `find()`:

```
db.users2.find()
```

The `_id` fields contain incrementing sequence values:

```
{
  _id: 1,
  name: "Grace H."
}
{
  _id : 2,
  "name" : "Ted R."
}
```

The `while` loop may iterate many times in collections with larger insert volumes.

4.3.13 Perform Quorum Reads on Replica Sets

New in version 3.2.

Overview

When reading from the primary of a replica set, it is possible to read data that is stale or not durable, depending on the read concern used [14]. With a read concern level of `"local"` (page 182), a client can read data before it is *durable*; that is, before they have propagated to enough replica set members to avoid a rollback. A read concern level of `"majority"` (page 182) guarantees durable reads but may return stale data that has been overwritten by another write operation.

This tutorial outlines a procedure that uses `db.collection.findAndModify()` to read data that is not stale and cannot be rolled back. To do so, the procedure uses the `findAndModify()` method with a *write concern* (page 179) to modify a dummy field in a document. Specifically, the procedure requires that:

- `db.collection.findAndModify()` use an **exact** match query, and a *unique index* (page 568) **must exist** to satisfy the query.

[14] In *some circumstances* (page 729), two nodes in a replica set may *transiently* believe that they are the primary, but at most, one of them will be able to complete writes with `{ w: "majority" }` (page 180) write concern. The node that can complete `{ w: "majority" }` (page 180) writes is the current primary, and the other node is a former primary that has not yet recognized its demotion, typically due to a *network partition*. When this occurs, clients that connect to the former primary may observe stale data despite having requested read preference `primary` (page 728), and new writes to the former primary will eventually roll back.

- `findAndModify()` must actually modify a document; i.e. result in a change to the document.
- `findAndModify()` must use the write concern `{ w: "majority" }` (page 180).

Important: The "quorum read" procedure has a substantial cost over simply using a read concern of `"majority"` (page 182) because it incurs write latency rather than read latency. This technique should only be used if staleness is absolutely intolerable.

Prerequisites

This tutorial reads from a collection named `products`. Initialize the collection using the following operation.

```
db.products.insert( [
   {
     _id: 1,
     sku: "xyz123",
     description: "hats",
     available: [ { quantity: 25, size: "S" }, { quantity: 50, size: "M" } ],
     _dummy_field: 0
   },
   {
     _id: 2,
     sku: "abc123",
     description: "socks",
     available: [ { quantity: 10, size: "L" } ],
     _dummy_field: 0
   },
   {
     _id: 3,
     sku: "ijk123",
     description: "t-shirts",
     available: [ { quantity: 30, size: "M" }, { quantity: 5, size: "L" } ],
     _dummy_field: 0
   }
] )
```

The documents in this collection contain a dummy field named `_dummy_field` that will be incremented by the `db.collection.findAndModify()` in the tutorial. If the field does not exist, the `db.collection.findAndModify()` operation will add the field to the document. The purpose of the field is to ensure that the `db.collection.findAndModify()` results in a modification to the document.

Procedure

Step 1: Create a unique index.

Create a unique index on the fields that will be used to specify an exact match in the `db.collection.findAndModify()` operation.

This tutorial will use an exact match on the `sku` field. As such, create a unique index on the `sku` field.

```
db.products.createIndex( { sku: 1 }, { unique: true } )
```

Step 2: Use `findAndModify` to read committed data.

Use the `db.collection.findAndModify()` method to make a trivial update to the document you want to read and return the modified document. A write concern of `{ w: "majority" }` (page 180) is required. To specify the document to read, you must use an exact match query that is supported by a unique index.

The following `findAndModify()` operation specifies an exact match on the uniquely indexed field `sku` and increments the field named `_dummy_field` in the matching document. While not necessary, the write concern for this command also includes a *wtimeout* (page 181) value of `5000` milliseconds to prevent the operation from blocking forever if the write cannot propagate to a majority of voting members.

```
var updatedDocument = db.products.findAndModify(
   {
     query: { sku: "abc123" },
     update: { $inc: { _dummy_field: 1 } },
     new: true,
     writeConcern: { w: "majority", wtimeout: 5000 }
   }
);
```

Even in situations where two nodes in the replica set believe that they are the primary, only one will be able to complete the write with `w: "majority"` (page 180). As such, the `findAndModify()` method with `"majority"` (page 180) write concern will be successful only when the client has connected to the true primary to perform the operation.

Since the quorum read procedure only increments a dummy field in the document, you can safely repeat invocations of `findAndModify()`, adjusting the *wtimeout* (page 181) as necessary.

4.4 MongoDB CRUD Reference

On this page

- Query Cursor Methods (page 178)
- Query and Data Manipulation Collection Methods (page 179)
- MongoDB CRUD Reference Documentation (page 179)

4.4.1 Query Cursor Methods

Name	Description
`cursor.count()`	Modifies the cursor to return the number of documents in the result set rather than the documents themselves.
`cursor.explain()`	Reports on the query execution plan for a cursor.
`cursor.hint()`	Forces MongoDB to use a specific index for a query.
`cursor.limit()`	Constrains the size of a cursor's result set.
`cursor.next()`	Returns the next document in a cursor.
`cursor.skip()`	Returns a cursor that begins returning results only after passing or skipping a number of documents.
`cursor.sort()`	Returns results ordered according to a sort specification.
`cursor.toArray()`	Returns an array that contains all documents returned by the cursor.

4.4.2 Query and Data Manipulation Collection Methods

Name	Description
`db.collection.count()`	Wraps `count` to return a count of the number of documents in a collection or matching a query.
`db.collection.distinct()`	Returns an array of documents that have distinct values for the specified field.
`db.collection.find()`	Performs a query on a collection and returns a cursor object.
`db.collection.findOne()`	Performs a query and returns a single document.
`db.collection.insert()`	Creates a new document in a collection.
`db.collection.remove()`	Deletes documents from a collection.
`db.collection.save()`	Provides a wrapper around an `insert()` and `update()` to insert new documents.
`db.collection.update()`	Modifies a document in a collection.

4.4.3 MongoDB CRUD Reference Documentation

Write Concern **(page 179)** Description of the write operation acknowledgements returned by MongoDB.

Read Concern **(page 181)** Description of the `readConcern` option.

SQL to MongoDB Mapping Chart **(page 183)** An overview of common database operations showing both the MongoDB operations and SQL statements.

The bios Example Collection **(page 189)** Sample data for experimenting with MongoDB. `insert()`, `update()` and `find()` pages use the data for some of their examples.

Write Concern

On this page

- Write Concern Specification (page 179)

Write concern describes the level of acknowledgement requested from MongoDB for write operations to a standalone `mongod` or to *replica sets* (page 623) or to *sharded clusters* (page 733). In sharded clusters, `mongos` instances will pass the write concern on to the shards.

Changed in version 3.2: For replica sets using `protocolVersion: 1` (page 718) **and** running with the *journal* enabled:

- `w: "majority"` (page 180) implies *j: true* (page 181).
- *Secondary members* acknowledge replicated write operations after the secondary members have written to their respective on-disk journals, regardless of the *j* (page 181) option used for the write on the *primary*.

Changed in version 2.6: A new protocol for *write operations* (page 995) integrates write concerns with the write operations and eliminates the need to call the `getLastError` command. Previous versions required a `getLastError` command immediately after a write operation to specify the write concern.

Write Concern Specification

Write concern can include the following fields:

```
{ w: <value>, j: <boolean>, wtimeout: <number> }
```

- the *w* (page 180) option to request acknowledgment that the write operation has propagated to a specified number of `mongod` instances or to `mongod` instances with specified tags.
- the *j* (page 181) option to request acknowledgement that the write operation has been written to the journal, and
- *wtimeout* (page 181) option to specify a time limit to prevent write operations from blocking indefinitely.

w Option The `w` option requests acknowledgement that the write operation has propagated to a specified number of `mongod` instances or to `mongod` instances with specified tags.

Using the `w` option, the following `w: <value>` write concerns are available:

Note: Standalone `mongod` instances and primaries of replica sets acknowledge write operations after applying the write in memory, unless *j:true* (page 181).

Changed in version 3.2: For replica sets using `protocolVersion: 1` (page 718), secondaries acknowledge write operations after the secondary members have written to their respective on-disk *journals* (page 606), regardless of the *j* (page 181) option.

Value	Description
<number>	Requests acknowledgement that the write operation has propagated to the specified number of `mongod` instances. For example: **w: 1** Requests acknowledgement that the write operation has propagated to the standalone `mongod` or the primary in a replica set. `w: 1` is the default write concern for MongoDB. **w: 0** Requests no acknowledgment of the write operation. However, `w: 0` may return information about socket exceptions and networking errors to the application. If you specify `w: 0` but include *j: true* (page 181), the *j: true* (page 181) prevails to request acknowledgement from the standalone `mongod` or the primary of a replica set. Numbers greater than 1 are valid only for replica sets to request acknowledgement from specified number of members, including the primary.
"majority"	*Changed in version 3.2* Requests acknowledgment that write operations have propagated to the majority of voting nodes [15], including the primary, and have been written to the on-disk *journal* (page 606) for these nodes. For replica sets using `protocolVersion: 1` (page 718), `w: "majority"` (page 180) implies *j: true* (page 181). So, unlike `w: <number>`, with `w: "majority"` (page 180), the primary also writes to the on-disk journal before acknowledging the write. After the write operation returns with a `w: "majority"` (page 180) acknowledgement to the client, the client can read the result of that write with a `"majority"` (page 182) readConcern.
<tag set>	Requests acknowledgement that the write operations have propagated to a replica set member with the specified *tag* (page 700).

`j` Option The *j* (page 181) option requests acknowledgement from MongoDB that the write operation has been written to the *journal* (page 606).

`j`	Requests acknowledgement that the `mongod` instances, as specified in the *w: <value>* (page 180), have written to the on-disk journal. `j: true` does not by itself guarantee that the write will not be rolled back due to replica set primary failover. Changed in version 3.2: With `j: true` (page 181), MongoDB returns only after the requested number of members, including the primary, have written to the journal. Previously `j: true` (page 181) write concern in a replica set only requires the *primary* to write to the journal, regardless of the *w: <value>* (page 180) write concern. For replica sets using `protocolVersion: 1` (page 718), `w: "majority"` (page 180) implies *j: true* (page 181), if journaling is enabled. Journaling is enabled by default.

Changed in version 2.6: Specifying a write concern that includes `j: true` to a `mongod` or `mongos` running with `--nojournal` option produces an error. Previous versions would ignore the `j: true`.

`wtimeout` This option specifies a time limit, in milliseconds, for the write concern. `wtimeout` is only applicable for `w` values greater than `1`.

`wtimeout` causes write operations to return with an error after the specified limit, even if the required write concern will eventually succeed. When these write operations return, MongoDB **does not** undo successful data modifications performed before the write concern exceeded the `wtimeout` time limit.

If you do not specify the `wtimeout` option and the level of write concern is unachievable, the write operation will block indefinitely. Specifying a `wtimeout` value of `0` is equivalent to a write concern without the `wtimeout` option.

Read Concern

On this page

- Storage Engine and Drivers Support (page 181)
- Read Concern Levels (page 182)
- `readConcern` Option (page 182)

New in version 3.2.

MongoDB 3.2 introduces the `readConcern` query option for replica sets and replica set shards. By default, MongoDB uses a read concern of `"local"` to return the most recent data available to the MongoDB instance at the time of the query, even if the data has not been persisted to a majority of replica set members and may be rolled back.

Storage Engine and Drivers Support

For the *WiredTiger storage engine* (page 595), the `readConcern` option allows clients to choose a level of isolation for their reads. You can specify a read concern of `"majority"` to read data that has been written to a majority of replica set members and thus cannot be rolled back.

With the *MMAPv1 storage engine* (page 603), you can only specify a `readConcern` option of `"local"`.

Tip

The `serverStatus` command returns the `storageEngine.supportsCommittedReads` field which indicates whether the storage engine supports `"majority"` read concern.

`readConcern` requires MongoDB drivers updated for 3.2.

Read Concern Levels

By default, MongoDB uses a `readConcern` of `"local"` which does not guarantee that the read data would not be rolled back.

You can specify a `readConcern` of `"majority"` to read data that has been written to a majority of replica set members and thus cannot be rolled back.

level	Description
`"local"`	Default. The query returns the instance's most recent copy of data. Provides no guarantee that the data has been written to a majority of the replica set members.
`"majority"`	The query returns the instance's most recent copy of data confirmed as written to a majority of members in the replica set. To use a *read concern* level of `"majority"` (page 182), you must use the WiredTiger storage engine and start the `mongod` instances with the *`--enableMajorityReadConcern`* command line option (or the `replication.enableMajorityReadConcern` setting if using a configuration file). Only replica sets using `protocol version 1` (page 718) support `"majority"` (page 182) read concern. Replica sets running protocol version 0 do not support `"majority"` (page 182) read concern. To ensure that a single thread can read its own writes, use `"majority"` (page 182) read concern and `"majority"` (page 180) write concern against the primary of the replica set.

Regardless of the *read concern* level, the most recent data on a node may not reflect the most recent version of the data in the system.

`readConcern` Option

Use the `readConcern` option to specify the read concern level.

```
readConcern: { level: <"majority"|"local"> }
```

For the `level` field, specify either the string `"majority"` or `"local"`.

The `readConcern` option is available for the following operations:

- `find` command
- `aggregate` command and the `db.collection.aggregate()` method
- `distinct` command

- `count` command
- `parallelCollectionScan` command
- `geoNear` command
- `geoSearch` command

To specify the read concern for the `mongo` shell method `db.collection.find()`, use the `cursor.readConcern()` method.

SQL to MongoDB Mapping Chart

On this page

- Terminology and Concepts (page 183)
- Executables (page 183)
- Examples (page 184)
- Additional Resources (page 188)

In addition to the charts that follow, you might want to consider the *Frequently Asked Questions* (page 831) section for a selection of common questions about MongoDB.

Terminology and Concepts

The following table presents the various SQL terminology and concepts and the corresponding MongoDB terminology and concepts.

SQL Terms/Concepts	MongoDB Terms/Concepts
database	*database*
table	*collection*
row	*document* or *BSON* document
column	*field*
index	*index*
table joins	embedded documents and linking
primary key	*primary key*
Specify any unique column or column combination as primary key.	In MongoDB, the primary key is automatically set to the *_id* field.
aggregation (e.g. group by)	aggregation pipeline See the *SQL to Aggregation Mapping Chart* (page 234).

Executables

The following table presents some database executables and the corresponding MongoDB executables. This table is *not* meant to be exhaustive.

	MongoDB	MySQL	Oracle	Informix	DB2
Database Server	`mongod`	`mysqld`	`oracle`	`IDS`	`DB2 Server`
Database Client	`mongo`	`mysql`	`sqlplus`	`DB-Access`	`DB2 Client`

Examples

The following table presents the various SQL statements and the corresponding MongoDB statements. The examples in the table assume the following conditions:

- The SQL examples assume a table named `users`.
- The MongoDB examples assume a collection named `users` that contain documents of the following prototype:

```
{
  _id: ObjectId("509a8fb2f3f4948bd2f983a0"),
  user_id: "abc123",
  age: 55,
  status: 'A'
}
```

Create and Alter The following table presents the various SQL statements related to table-level actions and the corresponding MongoDB statements.

SQL Schema Statements	MongoDB Schema Statements

`CREATE TABLE users (`
`  id MEDIUMINT NOT NULL`
`  AUTO_INCREMENT,`
`  user_id Varchar(30),`
`  age Number,`
`  status char(1),`
`  PRIMARY KEY (id)`
`)`
	Implicitly created on first `insert()` operation. The primary key `_id` is automatically added if `_id` field is not specified.
	`db.users.insert( {`
	`  user_id: "abc123",`
	`  age: 55,`
	`  status: "A"`
	`} )`
	However, you can also explicitly create a collection:
	`db.createCollection("users")`

`ALTER TABLE users`
`ADD join_date DATETIME`
	Collections do not describe or enforce the structure of its documents; i.e. there is no structural alteration at the collection level.
	However, at the document level, `update()` operations can add fields to existing documents using the `$set` operator.
	`db.users.update(`
	`  { },`
	`  { $set: { join_date: new Date() } },`
	`  { multi: true }`
	`)`

`ALTER TABLE users`
`DROP COLUMN join_date`
	Collections do not describe or enforce the structure of its documents; i.e. there is no structural alteration at the collection level.
	However, at the document level, `update()` operations can remove fields from documents using the `$unset` operator.
	`db.users.update(`
	`  { },`
	`  { $unset: { join_date: "" } },`
	`  { multi: true }`
	`)`

`CREATE INDEX idx_user_id_asc` `ON users(user_id)`	`db.users.createIndex( { user_id: 1 } )`

`CREATE INDEX`
`  idx_user_id_asc_age_desc`
`ON users(user_id, age DESC)`
	`db.users.createIndex( { user_id: 1, age: -1 } )`

`DROP TABLE users`	`db.users.drop()`

For more information, see `db.collection.insert()`, `db.createCollection()`, `db.collection.update()`, `$set`, `$unset`, `db.collection.createIndex()`, *indexes* (page 515), `db.collection.drop()`, and *Data Modeling Concepts* (page 252).

Insert The following table presents the various SQL statements related to inserting records into tables and the corresponding MongoDB statements.

SQL INSERT Statements	MongoDB insert() Statements

`INSERT INTO users(user_id,`
`  age,`
`  status)`
`VALUES ("bcd001",`
`  45,`
`  "A")`
	`db.users.insert(`
	`  { user_id: "bcd001", age: 45, status: "A" }`
	`)`

For more information, see `db.collection.insert()`.

Select The following table presents the various SQL statements related to reading records from tables and the corresponding MongoDB statements.

Note: The `find()` method always includes the `_id` field in the returned documents unless specifically excluded through *projection* (page 153). Some of the SQL queries below may include an `_id` field to reflect this, even if the field is not included in the corresponding `find()` query.

SQL SELECT Statements	MongoDB find() Statements
`SELECT * FROM users`	`db.users.find()`
`SELECT id, user_id, status FROM users`	`db.users.find( { }, { user_id: 1, status: 1 } )`
`SELECT user_id, status FROM users`	`db.users.find( { }, { user_id: 1, status: 1, _id: 0 } )`
`SELECT * FROM users WHERE status = "A"`	`db.users.find( { status: "A" } )`
`SELECT user_id, status FROM users WHERE status = "A"`	`db.users.find( { status: "A" }, { user_id: 1, status: 1, _id: 0 } )`
`SELECT * FROM users WHERE status != "A"`	`db.users.find( { status: { $ne: "A" } } )`
`SELECT * FROM users WHERE status = "A" AND age = 50`	`db.users.find( { status: "A", age: 50 } )`
`SELECT * FROM users WHERE status = "A" OR age = 50`	`db.users.find( { $or: [ { status: "A" } , { age: 50 } ] } )`
`SELECT * FROM users WHERE age > 25`	`db.users.find( { age: { $gt: 25 } } )`
`SELECT * FROM users WHERE age < 25`	`db.users.find( { age: { $lt: 25 } } )`
`SELECT * FROM users WHERE age > 25 AND age <= 50`	`db.users.find( { age: { $gt: 25, $lte: 50 } } )`
`SELECT * FROM users WHERE user_id like "%bc%"`	`db.users.find( { user_id: /bc/ } )`

For more information, see `db.collection.find()`, `db.collection.distinct()`, `db.collection.findOne()`, `$ne` `$and`, `$or`, `$gt`, `$lt`, `$exists`, `$lte`, `$regex`, `limit()`, `skip()`, `explain()`, `sort()`, and `count()`.

Update Records The following table presents the various SQL statements related to updating existing records in tables and the corresponding MongoDB statements.

SQL Update Statements	MongoDB update() Statements

`UPDATE users`
`SET status = "C"`
`WHERE age > 25`
	`db.users.update(`
	`{ age: { $gt: 25 } },`
	`{ $set: { status: "C" } },`
	`{ multi: true }`
	`)`

`UPDATE users`
`SET age = age + 3`
`WHERE status = "A"`
	`db.users.update(`
	`{ status: "A" } ,`
	`{ $inc: { age: 3 } },`
	`{ multi: true }`
	`)`

For more information, see `db.collection.update()`, `$set`, `$inc`, and `$gt`.

Delete Records The following table presents the various SQL statements related to deleting records from tables and the corresponding MongoDB statements.

SQL Delete Statements	MongoDB remove() Statements
`DELETE FROM users` `WHERE status = "D"`	`db.users.remove( { status: "D" } )`
`DELETE FROM users`	`db.users.remove({})`

For more information, see `db.collection.remove()`.

Additional Resources

- Transitioning from SQL to MongoDB (Presentation)[16]
- Best Practices for Migrating from RDBMS to MongoDB (Webinar)[17]
- SQL vs. MongoDB Day 1-2[18]
- SQL vs. MongoDB Day 3-5[19]
- MongoDB vs. SQL Day 14[20]
- MongoDB and MySQL Compared[21]

[16] http://www.mongodb.com/presentations/webinar-transitioning-sql-mongodb?jmp=docs
[17] http://www.mongodb.com/webinar/best-practices-migration?jmp=docs
[18] http://www.mongodb.com/blog/post/mongodb-vs-sql-day-1-2?jmp=docs
[19] http://www.mongodb.com/blog/post/mongodb-vs-sql-day-3-5?jmp=docs
[20] http://www.mongodb.com/blog/post/mongodb-vs-sql-day-14?jmp=docs
[21] http://www.mongodb.com/mongodb-and-mysql-compared?jmp=docs

- Quick Reference Cards[22]
- MongoDB Database Modernization Consulting Package[23]

The `bios` Example Collection

The `bios` collection provides example data for experimenting with MongoDB. Many of this guide's examples on `insert`, `update` and `read` operations create or query data from the `bios` collection.

The following documents comprise the `bios` collection. In the examples, the data might be different, as the examples themselves make changes to the data.

```
{
    "_id" : 1,
    "name" : {
        "first" : "John",
        "last" : "Backus"
    },
    "birth" : ISODate("1924-12-03T05:00:00Z"),
    "death" : ISODate("2007-03-17T04:00:00Z"),
    "contribs" : [
        "Fortran",
        "ALGOL",
        "Backus-Naur Form",
        "FP"
    ],
    "awards" : [
        {
            "award" : "W.W. McDowell Award",
            "year" : 1967,
            "by" : "IEEE Computer Society"
        },
        {
            "award" : "National Medal of Science",
            "year" : 1975,
            "by" : "National Science Foundation"
        },
        {
            "award" : "Turing Award",
            "year" : 1977,
            "by" : "ACM"
        },
        {
            "award" : "Draper Prize",
            "year" : 1993,
            "by" : "National Academy of Engineering"
        }
    ]
}

{
    "_id" : ObjectId("51df07b094c6acd67e492f41"),
    "name" : {
        "first" : "John",
        "last" : "McCarthy"
    },
```

[22] https://www.mongodb.com/lp/misc/quick-reference-cards?jmp=docs
[23] https://www.mongodb.com/products/consulting?jmp=docs#database_modernization

```
    "birth" : ISODate("1927-09-04T04:00:00Z"),
    "death" : ISODate("2011-12-24T05:00:00Z"),
    "contribs" : [
        "Lisp",
        "Artificial Intelligence",
        "ALGOL"
    ],
    "awards" : [
        {
            "award" : "Turing Award",
            "year" : 1971,
            "by" : "ACM"
        },
        {
            "award" : "Kyoto Prize",
            "year" : 1988,
            "by" : "Inamori Foundation"
        },
        {
            "award" : "National Medal of Science",
            "year" : 1990,
            "by" : "National Science Foundation"
        }
    ]
}

{
    "_id" : 3,
    "name" : {
        "first" : "Grace",
        "last" : "Hopper"
    },
    "title" : "Rear Admiral",
    "birth" : ISODate("1906-12-09T05:00:00Z"),
    "death" : ISODate("1992-01-01T05:00:00Z"),
    "contribs" : [
        "UNIVAC",
        "compiler",
        "FLOW-MATIC",
        "COBOL"
    ],
    "awards" : [
        {
            "award" : "Computer Sciences Man of the Year",
            "year" : 1969,
            "by" : "Data Processing Management Association"
        },
        {
            "award" : "Distinguished Fellow",
            "year" : 1973,
            "by" : " British Computer Society"
        },
        {
            "award" : "W. W. McDowell Award",
            "year" : 1976,
            "by" : "IEEE Computer Society"
        },
        {
```

```
            "award" : "National Medal of Technology",
            "year" : 1991,
            "by" : "United States"
        }
    ]
}

{
    "_id" : 4,
    "name" : {
        "first" : "Kristen",
        "last" : "Nygaard"
    },
    "birth" : ISODate("1926-08-27T04:00:00Z"),
    "death" : ISODate("2002-08-10T04:00:00Z"),
    "contribs" : [
        "OOP",
        "Simula"
    ],
    "awards" : [
        {
            "award" : "Rosing Prize",
            "year" : 1999,
            "by" : "Norwegian Data Association"
        },
        {
            "award" : "Turing Award",
            "year" : 2001,
            "by" : "ACM"
        },
        {
            "award" : "IEEE John von Neumann Medal",
            "year" : 2001,
            "by" : "IEEE"
        }
    ]
}

{
    "_id" : 5,
    "name" : {
        "first" : "Ole-Johan",
        "last" : "Dahl"
    },
    "birth" : ISODate("1931-10-12T04:00:00Z"),
    "death" : ISODate("2002-06-29T04:00:00Z"),
    "contribs" : [
        "OOP",
        "Simula"
    ],
    "awards" : [
        {
            "award" : "Rosing Prize",
            "year" : 1999,
            "by" : "Norwegian Data Association"
        },
        {
            "award" : "Turing Award",
```

```
            "year" : 2001,
            "by" : "ACM"
        },
        {
            "award" : "IEEE John von Neumann Medal",
            "year" : 2001,
            "by" : "IEEE"
        }
    ]
}

{
    "_id" : 6,
    "name" : {
        "first" : "Guido",
        "last" : "van Rossum"
    },
    "birth" : ISODate("1956-01-31T05:00:00Z"),
    "contribs" : [
        "Python"
    ],
    "awards" : [
        {
            "award" : "Award for the Advancement of Free Software",
            "year" : 2001,
            "by" : "Free Software Foundation"
        },
        {
            "award" : "NLUUG Award",
            "year" : 2003,
            "by" : "NLUUG"
        }
    ]
}

{
    "_id" : ObjectId("51e062189c6ae665454e301d"),
    "name" : {
        "first" : "Dennis",
        "last" : "Ritchie"
    },
    "birth" : ISODate("1941-09-09T04:00:00Z"),
    "death" : ISODate("2011-10-12T04:00:00Z"),
    "contribs" : [
        "UNIX",
        "C"
    ],
    "awards" : [
        {
            "award" : "Turing Award",
            "year" : 1983,
            "by" : "ACM"
        },
        {
            "award" : "National Medal of Technology",
            "year" : 1998,
            "by" : "United States"
        },
```

```
        {
            "award" : "Japan Prize",
            "year" : 2011,
            "by" : "The Japan Prize Foundation"
        }
    ]
}

{
    "_id" : 8,
    "name" : {
        "first" : "Yukihiro",
        "aka" : "Matz",
        "last" : "Matsumoto"
    },
    "birth" : ISODate("1965-04-14T04:00:00Z"),
    "contribs" : [
        "Ruby"
    ],
    "awards" : [
        {
            "award" : "Award for the Advancement of Free Software",
            "year" : "2011",
            "by" : "Free Software Foundation"
        }
    ]
}

{
    "_id" : 9,
    "name" : {
        "first" : "James",
        "last" : "Gosling"
    },
    "birth" : ISODate("1955-05-19T04:00:00Z"),
    "contribs" : [
        "Java"
    ],
    "awards" : [
        {
            "award" : "The Economist Innovation Award",
            "year" : 2002,
            "by" : "The Economist"
        },
        {
            "award" : "Officer of the Order of Canada",
            "year" : 2007,
            "by" : "Canada"
        }
    ]
}

{
    "_id" : 10,
    "name" : {
        "first" : "Martin",
        "last" : "Odersky"
    },
```

```
    "contribs" : [
        "Scala"
    ]
}
```

CHAPTER 5

Aggregation

On this page

- Aggregation Pipeline (page 195)
- Map-Reduce (page 197)
- Single Purpose Aggregation Operations (page 198)
- Additional Features and Behaviors (page 198)
- Additional Resources (page 237)

Aggregations operations process data records and return computed results. Aggregation operations group values from multiple documents together, and can perform a variety of operations on the grouped data to return a single result. MongoDB provides three ways to perform aggregation: the *aggregation pipeline* (page 195), the *map-reduce function* (page 197), and *single purpose aggregation methods* (page 198).

5.1 Aggregation Pipeline

MongoDB's *aggregation framework* (page 199) is modeled on the concept of data processing pipelines. Documents enter a multi-stage pipeline that transforms the documents into an aggregated result.

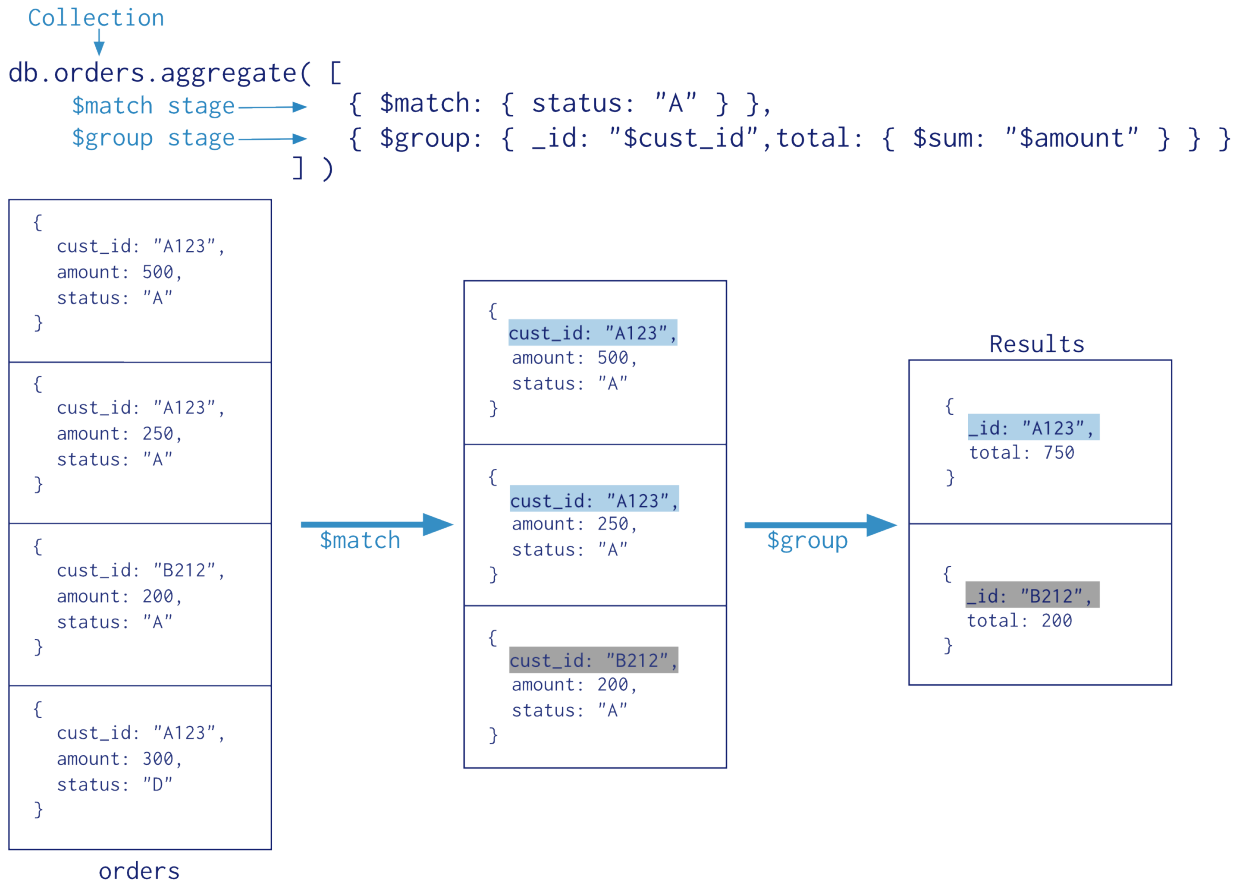
The most basic pipeline stages provide *filters* that operate like queries and *document transformations* that modify the form of the output document.

Other pipeline operations provide tools for grouping and sorting documents by specific field or fields as well as tools for aggregating the contents of arrays, including arrays of documents. In addition, pipeline stages can use *operators* for tasks such as calculating the average or concatenating a string.

The pipeline provides efficient data aggregation using native operations within MongoDB, and is the preferred method for data aggregation in MongoDB.

The aggregation pipeline can operate on a *sharded collection* (page 733).

The aggregation pipeline can use indexes to improve its performance during some of its stages. In addition, the aggregation pipeline has an internal optimization phase. See *Pipeline Operators and Indexes* (page 200) and *Aggregation Pipeline Optimization* (page 201) for details.


Collection
db.orders.aggregate([
$match stage
{ $match: { status: "A" } },
$group stage
{ $group: { _id: "$cust_id",total: { $sum: "$amount" } } }
])
{ cust_id: "A123", amount: 500, status: "A" }
{ cust_id: "A123", amount: 250, status: "A" }
{ cust_id: "B212", amount: 200, status: "A" }
{ cust_id: "A123", amount: 300, status: "D" }
orders
$match
{ cust_id: "A123", amount: 500, status: "A" }
{ cust_id: "A123", amount: 250, status: "A" }
{ cust_id: "B212", amount: 200, status: "A" }
$group
Results
{ _id: "A123", total: 750 }
{ _id: "B212", total: 200 }

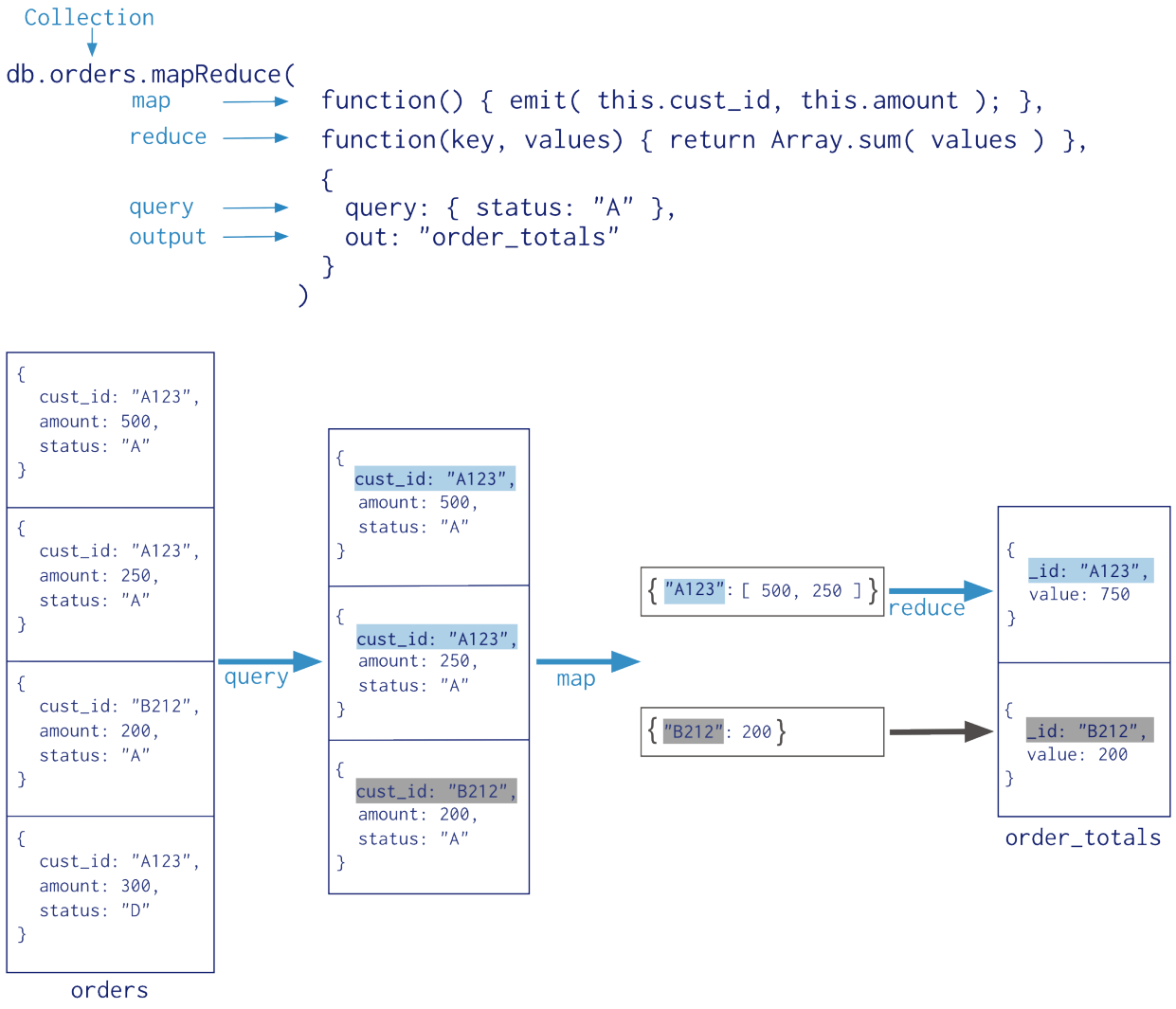
5.2 Map-Reduce

MongoDB also provides *map-reduce* (page 214) operations to perform aggregation. In general, map-reduce operations have two phases: a *map* stage that processes each document and *emits* one or more objects for each input document, and *reduce* phase that combines the output of the map operation. Optionally, map-reduce can have a *finalize* stage to make final modifications to the result. Like other aggregation operations, map-reduce can specify a query condition to select the input documents as well as sort and limit the results.

Map-reduce uses custom JavaScript functions to perform the map and reduce operations, as well as the optional *finalize* operation. While the custom JavaScript provide great flexibility compared to the aggregation pipeline, in general, map-reduce is less efficient and more complex than the aggregation pipeline.

Map-reduce can operate on a *sharded collection* (page 733). Map reduce operations can also output to a sharded collection. See *Aggregation Pipeline and Sharded Collections* (page 205) and *Map-Reduce and Sharded Collections* (page 215) for details.

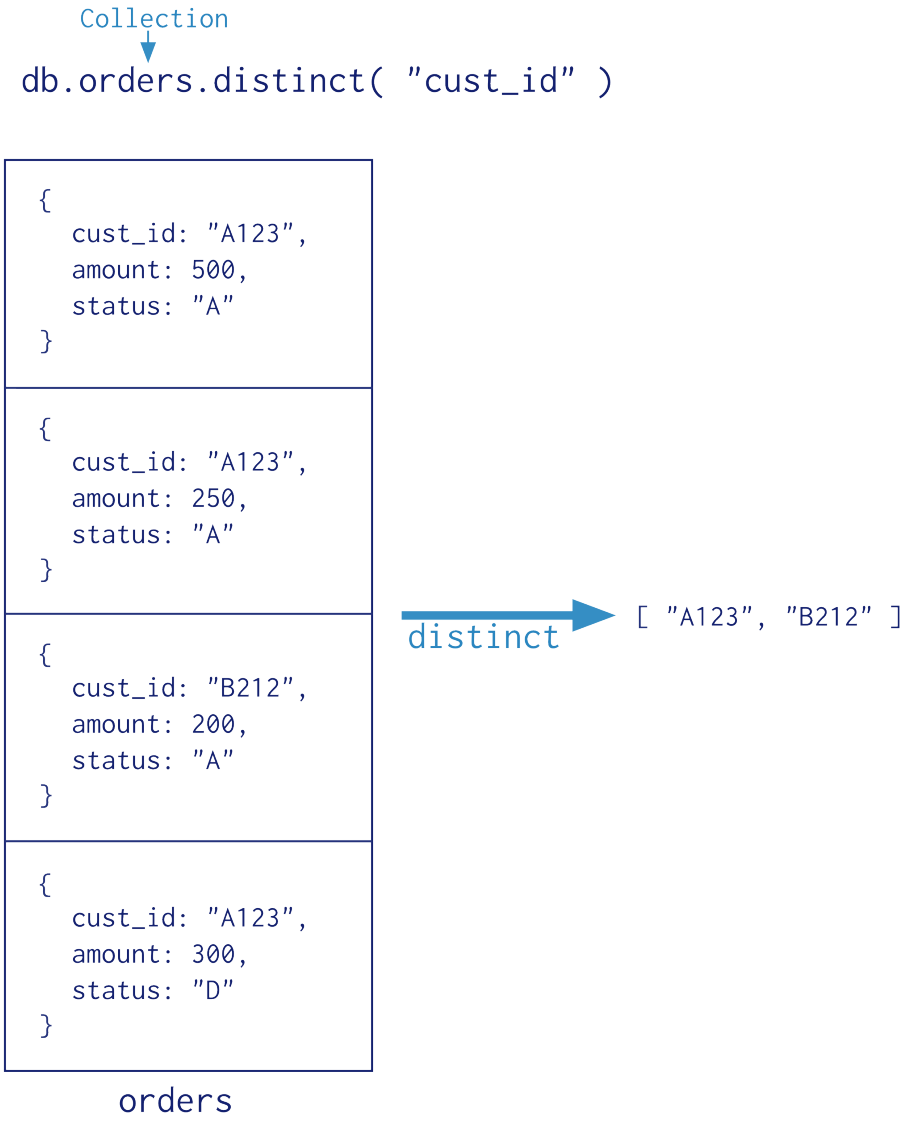
Note: Starting in MongoDB 2.4, certain `mongo` shell functions and properties are inaccessible in map-reduce operations. MongoDB 2.4 also provides support for multiple JavaScript operations to run at the same time. Before MongoDB 2.4, JavaScript code executed in a single thread, raising concurrency issues for map-reduce.



5.3 Single Purpose Aggregation Operations

MongoDB also provides `db.collection.count()`, `db.collection.group()`, `db.collection.distinct()`. special purpose database commands.

All of these operations aggregate documents from a single collection. While these operations provide simple access to common aggregation processes, they lack the flexibility and capabilities of the aggregation pipeline and map-reduce.



5.4 Additional Features and Behaviors

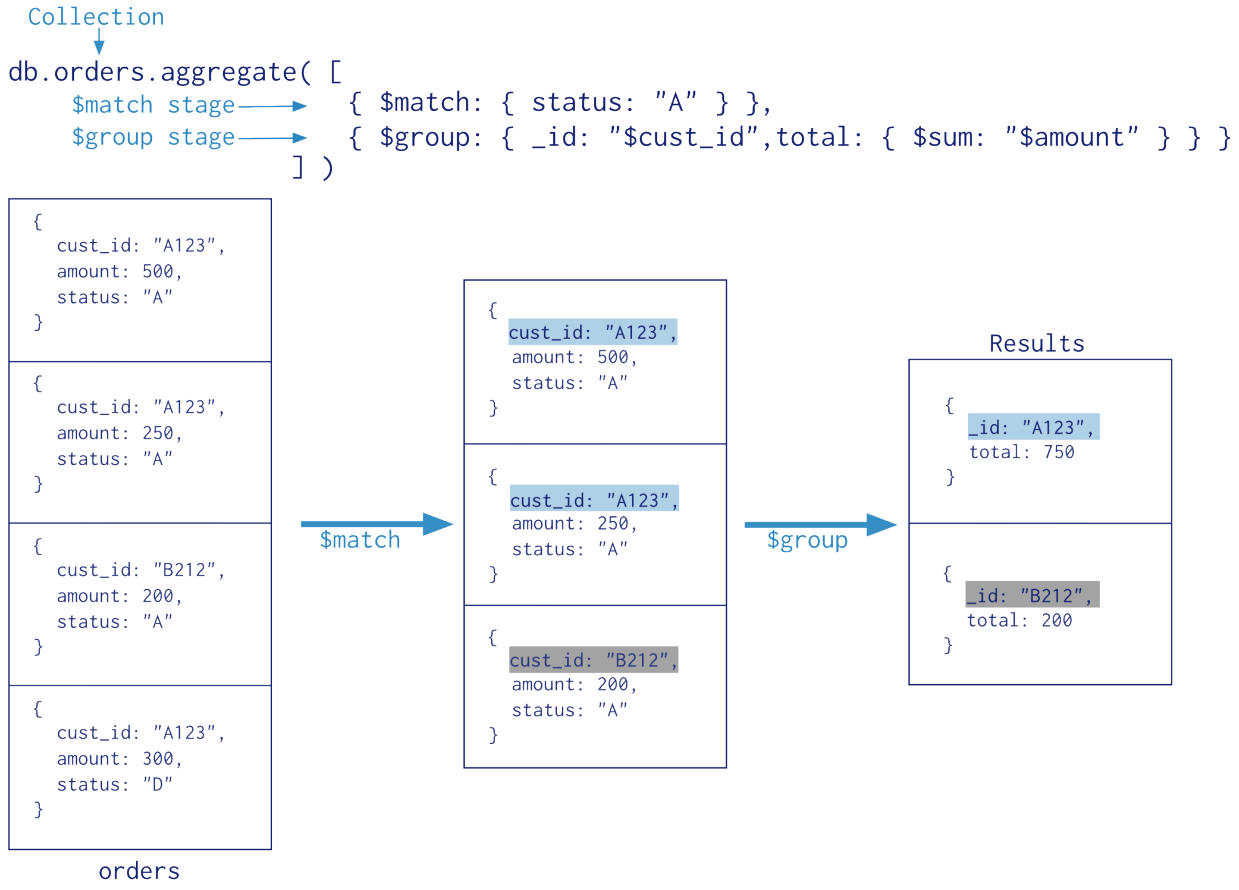
For a feature comparison of the aggregation pipeline, map-reduce, and the special group functionality, see *Aggregation Commands Comparison* (page 232).

5.4.1 Aggregation Pipeline

On this page

- Pipeline (page 199)
- Pipeline Expressions (page 200)
- Aggregation Pipeline Behavior (page 200)
- Additional Resources (page 213)

The aggregation pipeline is a framework for data aggregation modeled on the concept of data processing pipelines. Documents enter a multi-stage pipeline that transforms the documents into aggregated results.



The aggregation pipeline provides an alternative to *map-reduce* and may be the preferred solution for aggregation tasks where the complexity of map-reduce may be unwarranted.

Aggregation pipeline have some limitations on value types and result size. See *Aggregation Pipeline Limits* (page 205) for details on limits and restrictions on the aggregation pipeline.

Pipeline

The MongoDB aggregation pipeline consists of *stages*. Each stage transforms the documents as they pass through the pipeline. Pipeline stages do not need to produce one output document for every input document; e.g., some stages may generate new documents or filter out documents. Pipeline stages can appear multiple times in the pipeline.

MongoDB provides the `db.collection.aggregate()` method in the `mongo` shell and the `aggregate` command for aggregation pipeline. See *aggregation-pipeline-operator-reference* for the available stages.

For example usage of the aggregation pipeline, consider *Aggregation with User Preference Data* (page 209) and *Aggregation with the Zip Code Data Set* (page 206).

Pipeline Expressions

Some pipeline stages takes a pipeline expression as its operand. Pipeline expressions specify the transformation to apply to the input documents. Expressions have a *document* (page 8) structure and can contain other *expression* (page 226).

Pipeline expressions can only operate on the current document in the pipeline and cannot refer to data from other documents: expression operations provide in-memory transformation of documents.

Generally, expressions are stateless and are only evaluated when seen by the aggregation process with one exception: *accumulator* expressions.

The accumulators, used in the `$group` stage, maintain their state (e.g. totals, maximums, minimums, and related data) as documents progress through the pipeline.

Changed in version 3.2: Some accumulators are available in the `$project` stage; however, when used in the `$project` stage, the accumulators do not maintain their state across documents.

For more information on expressions, see *Expressions* (page 226).

Aggregation Pipeline Behavior

In MongoDB, the `aggregate` command operates on a single collection, logically passing the *entire* collection into the aggregation pipeline. To optimize the operation, wherever possible, use the following strategies to avoid scanning the entire collection.

Pipeline Operators and Indexes

The `$match` and `$sort` pipeline operators can take advantage of an index when they occur at the **beginning** of the pipeline.

New in version 2.4: The `$geoNear` pipeline operator takes advantage of a geospatial index. When using `$geoNear`, the `$geoNear` pipeline operation must appear as the first stage in an aggregation pipeline.

Changed in version 3.2: Starting in MongoDB 3.2, indexes can *cover* (page 106) an aggregation pipeline. In MongoDB 2.6 and 3.0, indexes could not cover an aggregation pipeline since even when the pipeline uses an index, aggregation still requires access to the actual documents.

Early Filtering

If your aggregation operation requires only a subset of the data in a collection, use the `$match`, `$limit`, and `$skip` stages to restrict the documents that enter at the beginning of the pipeline. When placed at the beginning of a pipeline, `$match` operations use suitable indexes to scan only the matching documents in a collection.

Placing a `$match` pipeline stage followed by a `$sort` stage at the start of the pipeline is logically equivalent to a single query with a sort and can use an index. When possible, place `$match` operators at the beginning of the pipeline.

Additional Features

The aggregation pipeline has an internal optimization phase that provides improved performance for certain sequences of operators. For details, see *Aggregation Pipeline Optimization* (page 201).

The aggregation pipeline supports operations on sharded collections. See *Aggregation Pipeline and Sharded Collections* (page 205).

Aggregation Pipeline Optimization

On this page

- Projection Optimization (page 201)
- Pipeline Sequence Optimization (page 201)
- Pipeline Coalescence Optimization (page 202)
- Examples (page 204)

Aggregation pipeline operations have an optimization phase which attempts to reshape the pipeline for improved performance.

To see how the optimizer transforms a particular aggregation pipeline, include the `explain` option in the `db.collection.aggregate()` method.

Optimizations are subject to change between releases.

Projection Optimization The aggregation pipeline can determine if it requires only a subset of the fields in the documents to obtain the results. If so, the pipeline will only use those required fields, reducing the amount of data passing through the pipeline.

Pipeline Sequence Optimization

`$sort` + `$match` Sequence Optimization When you have a sequence with `$sort` followed by a `$match`, the `$match` moves before the `$sort` to minimize the number of objects to sort. For example, if the pipeline consists of the following stages:

```
{ $sort: { age : -1 } },
{ $match: { status: 'A' } }
```

During the optimization phase, the optimizer transforms the sequence to the following:

```
{ $match: { status: 'A' } },
{ $sort: { age : -1 } }
```

`$skip` + `$limit` Sequence Optimization When you have a sequence with `$skip` followed by a `$limit`, the `$limit` moves before the `$skip`. With the reordering, the `$limit` value increases by the `$skip` amount.

For example, if the pipeline consists of the following stages:

```
{ $skip: 10 },
{ $limit: 5 }
```

During the optimization phase, the optimizer transforms the sequence to the following:

```
{ $limit: 15 },
{ $skip: 10 }
```

This optimization allows for more opportunities for *$sort + $limit Coalescence* (page 202), such as with `$sort` + `$skip` + `$limit` sequences. See *$sort + $limit Coalescence* (page 202) for details on the coalescence and *$sort + $skip + $limit Sequence* (page 204) for an example.

For aggregation operations on *sharded collections* (page 205), this optimization reduces the results returned from each shard.

`$redact` + `$match` Sequence Optimization When possible, when the pipeline has the `$redact` stage immediately followed by the `$match` stage, the aggregation can sometimes add a portion of the `$match` stage before the `$redact` stage. If the added `$match` stage is at the start of a pipeline, the aggregation can use an index as well as query the collection to limit the number of documents that enter the pipeline. See *Pipeline Operators and Indexes* (page 200) for more information.

For example, if the pipeline consists of the following stages:

```
{ $redact: { $cond: { if: { $eq: [ "$level", 5 ] }, then: "$$PRUNE", else: "$$DESCEND" } } },
{ $match: { year: 2014, category: { $ne: "Z" } } }
```

The optimizer can add the same `$match` stage before the `$redact` stage:

```
{ $match: { year: 2014 } },
{ $redact: { $cond: { if: { $eq: [ "$level", 5 ] }, then: "$$PRUNE", else: "$$DESCEND" } } },
{ $match: { year: 2014, category: { $ne: "Z" } } }
```

`$project` + `$skip` or `$limit` Sequence Optimization New in version 3.2.

When you have a sequence with `$project` followed by either `$skip` or `$limit`, the `$skip` or `$limit` moves before `$project`. For example, if the pipeline consists of the following stages:

```
{ $sort: { age : -1 } },
{ $project: { status: 1, name: 1 } },
{ $limit: 5 }
```

During the optimization phase, the optimizer transforms the sequence to the following:

```
{ $sort: { age : -1 } },
{ $limit: 5 }
{ $project: { status: 1, name: 1 } },
```

This optimization allows for more opportunities for *$sort + $limit Coalescence* (page 202), such as with `$sort` + `$limit` sequences. See *$sort + $limit Coalescence* (page 202) for details on the coalescence.

Pipeline Coalescence Optimization When possible, the optimization phase coalesces a pipeline stage into its predecessor. Generally, coalescence occurs *after* any sequence reordering optimization.

`$sort` + `$limit` Coalescence When a `$sort` immediately precedes a `$limit`, the optimizer can coalesce the `$limit` into the `$sort`. This allows the sort operation to only maintain the top `n` results as it progresses, where `n` is the specified limit, and MongoDB only needs to store `n` items in memory [1]. See *sort-and-memory* for more information.

[1] The optimization will still apply when `allowDiskUse` is `true` and the `n` items exceed the *aggregation memory limit* (page 205).

`$limit` + `$limit` Coalescence When a `$limit` immediately follows another `$limit`, the two stages can coalesce into a single `$limit` where the limit amount is the *smaller* of the two initial limit amounts. For example, a pipeline contains the following sequence:

```
{ $limit: 100 },
{ $limit: 10 }
```

Then the second `$limit` stage can coalesce into the first `$limit` stage and result in a single `$limit` stage where the limit amount `10` is the minimum of the two initial limits `100` and `10`.

```
{ $limit: 10 }
```

`$skip` + `$skip` Coalescence When a `$skip` immediately follows another `$skip`, the two stages can coalesce into a single `$skip` where the skip amount is the *sum* of the two initial skip amounts. For example, a pipeline contains the following sequence:

```
{ $skip: 5 },
{ $skip: 2 }
```

Then the second `$skip` stage can coalesce into the first `$skip` stage and result in a single `$skip` stage where the skip amount `7` is the sum of the two initial limits `5` and `2`.

```
{ $skip: 7 }
```

`$match` + `$match` Coalescence When a `$match` immediately follows another `$match`, the two stages can coalesce into a single `$match` combining the conditions with an `$and`. For example, a pipeline contains the following sequence:

```
{ $match: { year: 2014 } },
{ $match: { status: "A" } }
```

Then the second `$match` stage can coalesce into the first `$match` stage and result in a single `$match` stage

```
{ $match: { $and: [ { "year" : 2014 }, { "status" : "A" } ] } }
```

`$lookup` + `$unwind` Coalescence New in version 3.2.

When a `$unwind` immediately follows another `$lookup`, and the `$unwind` operates on the `as` field of the `$lookup`, the optimizer can coalesce the `$unwind` into the `$lookup` stage. This avoids creating large intermediate documents.

For example, a pipeline contains the following sequence:

```
{
  $lookup: {
    from: "otherCollection",
    as: "resultingArray",
    localField: "x",
    foreignField: "y"
  }
},
{ $unwind: "$resultingArray"}
```

The optimizer can coalesce the `$unwind` stage into the `$lookup` stage. If you run the aggregation with `explain` option, the `explain` output shows the coalesced stage:

```
{
  $lookup: {
    from: "otherCollection",
    as: "resultingArray",
    localField: "x",
    foreignField: "y",
    unwinding: { preserveNullAndEmptyArrays: false }
  }
}
```

Examples The following examples are some sequences that can take advantage of both sequence reordering and coalescence. Generally, coalescence occurs *after* any sequence reordering optimization.

`$sort` + `$skip` + `$limit` Sequence A pipeline contains a sequence of `$sort` followed by a `$skip` followed by a `$limit`:

```
{ $sort: { age : -1 } },
{ $skip: 10 },
{ $limit: 5 }
```

First, the optimizer performs the *$skip + $limit Sequence Optimization* (page 201) to transforms the sequence to the following:

```
{ $sort: { age : -1 } },
{ $limit: 15 }
{ $skip: 10 }
```

The *$skip + $limit Sequence Optimization* (page 201) increases the `$limit` amount with the reordering. See *$skip + $limit Sequence Optimization* (page 201) for details.

The reordered sequence now has `$sort` immediately preceding the `$limit`, and the pipeline can coalesce the two stages to decrease memory usage during the sort operation. See *$sort + $limit Coalescence* (page 202) for more information.

`$limit` + `$skip` + `$limit` + `$skip` Sequence A pipeline contains a sequence of alternating `$limit` and `$skip` stages:

```
{ $limit: 100 },
{ $skip: 5 },
{ $limit: 10 },
{ $skip: 2 }
```

The *$skip + $limit Sequence Optimization* (page 201) reverses the position of the `{ $skip: 5 }` and `{ $limit: 10 }` stages and increases the limit amount:

```
{ $limit: 100 },
{ $limit: 15},
{ $skip: 5 },
{ $skip: 2 }
```

The optimizer then coalesces the two `$limit` stages into a single `$limit` stage and the two `$skip` stages into a single `$skip` stage. The resulting sequence is the following:

```
{ $limit: 15 },
{ $skip: 7 }
```

See *$limit + $limit Coalescence* (page 203) and *$skip + $skip Coalescence* (page 203) for details.

See also:

`explain` option in the `db.collection.aggregate()`

Aggregation Pipeline Limits

On this page

- Result Size Restrictions (page 205)
- Memory Restrictions (page 205)

Aggregation operations with the `aggregate` command have the following limitations.

Result Size Restrictions Changed in version 2.6.

Starting in MongoDB 2.6, the `aggregate` command can return a cursor or store the results in a collection. When returning a cursor or storing the results in a collection, each document in the result set is subject to the `BSON Document Size` limit, currently 16 megabytes; if any single document that exceeds the `BSON Document Size` limit, the command will produce an error. The limit only applies to the returned documents; during the pipeline processing, the documents may exceed this size. The `db.collection.aggregate()` method returns a cursor by default starting in MongoDB 2.6

If you do not specify the cursor option or store the results in a collection, the `aggregate` command returns a single BSON document that contains a field with the result set. As such, the command will produce an error if the total size of the result set exceeds the `BSON Document Size` limit.

Earlier versions of the `aggregate` command can only return a single BSON document that contains the result set and will produce an error if the if the total size of the result set exceeds the `BSON Document Size` limit.

Memory Restrictions Changed in version 2.6.

Pipeline stages have a limit of 100 megabytes of RAM. If a stage exceeds this limit, MongoDB will produce an error. To allow for the handling of large datasets, use the `allowDiskUse` option to enable aggregation pipeline stages to write data to temporary files.

See also:

sort-memory-limit and *group-memory-limit*.

Aggregation Pipeline and Sharded Collections

On this page

- Behavior (page 205)
- Optimization (page 206)

The aggregation pipeline supports operations on *sharded* collections. This section describes behaviors specific to the *aggregation pipeline* (page 199) and sharded collections.

Behavior Changed in version 3.2.

If the pipeline starts with an exact `$match` on a shard key, the entire pipeline runs on the matching shard only. Previously, the pipeline would have been split, and the work of merging it would have to be done on the primary shard.

For aggregation operations that must run on multiple shards, if the operations do not require running on the database's primary shard, these operations will route the results to a random shard to merge the results to avoid overloading the

primary shard for that database. The `$out` stage and the `$lookup` stage require running on the database's primary shard.

Optimization When splitting the aggregation pipeline into two parts, the pipeline is split to ensure that the shards perform as many stages as possible with consideration for optimization.

To see how the pipeline was split, include the `explain` option in the `db.collection.aggregate()` method.

Optimizations are subject to change between releases.

Aggregation with the Zip Code Data Set

On this page

- Data Model (page 206)
- `aggregate()` Method (page 206)
- Return States with Populations above 10 Million (page 207)
- Return Average City Population by State (page 207)
- Return Largest and Smallest Cities by State (page 208)

The examples in this document use the `zipcodes` collection. This collection is available at: media.mongodb.org/zips.json[2]. Use `mongoimport` to load this data set into your `mongod` instance.

Data Model Each document in the `zipcodes` collection has the following form:

```
{
  "_id": "10280",
  "city": "NEW YORK",
  "state": "NY",
  "pop": 5574,
  "loc": [
    -74.016323,
    40.710537
  ]
}
```

- The `_id` field holds the zip code as a string.
- The `city` field holds the city name. A city can have more than one zip code associated with it as different sections of the city can each have a different zip code.
- The `state` field holds the two letter state abbreviation.
- The `pop` field holds the population.
- The `loc` field holds the location as a latitude longitude pair.

`aggregate()` Method All of the following examples use the `aggregate()` helper in the `mongo` shell.

The `aggregate()` method uses the *aggregation pipeline* (page 199) to processes documents into aggregated results. An *aggregation pipeline* (page 199) consists of *stages* with each stage processing the documents as they pass along the pipeline. Documents pass through the stages in sequence.

The `aggregate()` method in the `mongo` shell provides a wrapper around the `aggregate` database command. See the documentation for your `driver` for a more idiomatic interface for data aggregation operations.

[2] http://media.mongodb.org/zips.json

Return States with Populations above 10 Million The following aggregation operation returns all states with total population greater than 10 million:

```
db.zipcodes.aggregate( [
   { $group: { _id: "$state", totalPop: { $sum: "$pop" } } },
   { $match: { totalPop: { $gte: 10*1000*1000 } } }
] )
```

In this example, the *aggregation pipeline* (page 199) consists of the `$group` stage followed by the `$match` stage:

- The `$group` stage groups the documents of the `zipcode` collection by the `state` field, calculates the `totalPop` field for each state, and outputs a document for each unique state.

  The new per-state documents have two fields: the `_id` field and the `totalPop` field. The `_id` field contains the value of the `state`; i.e. the group by field. The `totalPop` field is a calculated field that contains the total population of each state. To calculate the value, `$group` uses the `$sum` operator to add the population field (`pop`) for each state.

  After the `$group` stage, the documents in the pipeline resemble the following:

  ```
  {
    "_id" : "AK",
    "totalPop" : 550043
  }
  ```

- The `$match` stage filters these grouped documents to output only those documents whose `totalPop` value is greater than or equal to 10 million. The `$match` stage does not alter the matching documents but outputs the matching documents unmodified.

The equivalent *SQL* for this aggregation operation is:

```
SELECT state, SUM(pop) AS totalPop
FROM zipcodes
GROUP BY state
HAVING totalPop >= (10*1000*1000)
```

See also:

`$group`, `$match`, `$sum`

Return Average City Population by State The following aggregation operation returns the average populations for cities in each state:

```
db.zipcodes.aggregate( [
   { $group: { _id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } } },
   { $group: { _id: "$_id.state", avgCityPop: { $avg: "$pop" } } }
] )
```

In this example, the *aggregation pipeline* (page 199) consists of the `$group` stage followed by another `$group` stage:

- The first `$group` stage groups the documents by the combination of `city` and `state`, uses the `$sum` expression to calculate the population for each combination, and outputs a document for each `city` and `state` combination. [3]

  After this stage in the pipeline, the documents resemble the following:

[3] A city can have more than one zip code associated with it as different sections of the city can each have a different zip code.

```
{
  "_id" : {
    "state" : "CO",
    "city" : "EDGEWATER"
  },
  "pop" : 13154
}
```

- A second `$group` stage groups the documents in the pipeline by the `_id.state` field (i.e. the `state` field inside the `_id` document), uses the `$avg` expression to calculate the average city population (`avgCityPop`) for each state, and outputs a document for each state.

The documents that result from this aggregation operation resembles the following:

```
{
  "_id" : "MN",
  "avgCityPop" : 5335
}
```

See also:

`$group`, `$sum`, `$avg`

Return Largest and Smallest Cities by State The following aggregation operation returns the smallest and largest cities by population for each state:

```
db.zipcodes.aggregate( [
   { $group:
      {
        _id: { state: "$state", city: "$city" },
        pop: { $sum: "$pop" }
      }
   },
   { $sort: { pop: 1 } },
   { $group:
      {
        _id : "$_id.state",
        biggestCity:  { $last: "$_id.city" },
        biggestPop:   { $last: "$pop" },
        smallestCity: { $first: "$_id.city" },
        smallestPop:  { $first: "$pop" }
      }
   },

  // the following $project is optional, and
  // modifies the output format.

  { $project:
    { _id: 0,
      state: "$_id",
      biggestCity:  { name: "$biggestCity",  pop: "$biggestPop" },
      smallestCity: { name: "$smallestCity", pop: "$smallestPop" }
    }
  }
] )
```

In this example, the *aggregation pipeline* (page 199) consists of a `$group` stage, a `$sort` stage, another `$group` stage, and a `$project` stage:

- The first `$group` stage groups the documents by the combination of the `city` and `state`, calculates the `sum` of the `pop` values for each combination, and outputs a document for each `city` and `state` combination.

  At this stage in the pipeline, the documents resemble the following:

  ```
  {
    "_id" : {
      "state" : "CO",
      "city" : "EDGEWATER"
    },
    "pop" : 13154
  }
  ```

- The `$sort` stage orders the documents in the pipeline by the `pop` field value, from smallest to largest; i.e. by increasing order. This operation does not alter the documents.

- The next `$group` stage groups the now-sorted documents by the `_id.state` field (i.e. the `state` field inside the `_id` document) and outputs a document for each state.

  The stage also calculates the following four fields for each state. Using the `$last` expression, the `$group` operator creates the `biggestCity` and `biggestPop` fields that store the city with the largest population and that population. Using the `$first` expression, the `$group` operator creates the `smallestCity` and `smallestPop` fields that store the city with the smallest population and that population.

  The documents, at this stage in the pipeline, resemble the following:

  ```
  {
    "_id" : "WA",
    "biggestCity" : "SEATTLE",
    "biggestPop" : 520096,
    "smallestCity" : "BENGE",
    "smallestPop" : 2
  }
  ```

- The final `$project` stage renames the `_id` field to `state` and moves the `biggestCity`, `biggestPop`, `smallestCity`, and `smallestPop` into `biggestCity` and `smallestCity` embedded documents.

The output documents of this aggregation operation resemble the following:

```
{
  "state" : "RI",
  "biggestCity" : {
    "name" : "CRANSTON",
    "pop" : 176404
  },
  "smallestCity" : {
    "name" : "CLAYVILLE",
    "pop" : 45
  }
}
```

Aggregation with User Preference Data

On this page

- Data Model (page 210)
- Normalize and Sort Documents (page 210)
- Return Usernames Ordered by Join Month (page 210)
- Return Total Number of Joins per Month (page 211)
- Return the Five Most Common "Likes" (page 212)

Data Model Consider a hypothetical sports club with a database that contains a `users` collection that tracks the user's join dates, sport preferences, and stores these data in documents that resemble the following:

```
{
  _id : "jane",
  joined : ISODate("2011-03-02"),
  likes : ["golf", "racquetball"]
}
{
  _id : "joe",
  joined : ISODate("2012-07-02"),
  likes : ["tennis", "golf", "swimming"]
}
```

Normalize and Sort Documents The following operation returns user names in upper case and in alphabetical order. The aggregation includes user names for all documents in the `users` collection. You might do this to normalize user names for processing.

```
db.users.aggregate(
  [
    { $project : { name:{$toUpper:"$_id"} , _id:0 } },
    { $sort : { name : 1 } }
  ]
)
```

All documents from the `users` collection pass through the pipeline, which consists of the following operations:

- The `$project` operator:
  - creates a new field called `name`.
  - converts the value of the `_id` to upper case, with the `$toUpper` operator. Then the `$project` creates a new field, named `name` to hold this value.
  - suppresses the `id` field. `$project` will pass the `_id` field by default, unless explicitly suppressed.
- The `$sort` operator orders the results by the `name` field.

The results of the aggregation would resemble the following:

```
{
  "name" : "JANE"
},
{
  "name" : "JILL"
},
{
  "name" : "JOE"
}
```

Return Usernames Ordered by Join Month The following aggregation operation returns user names sorted by the month they joined. This kind of aggregation could help generate membership renewal notices.

```
db.users.aggregate(
  [
    { $project :
       {
         month_joined : { $month : "$joined" },
         name : "$_id",
```

```
          _id : 0
        }
    },
    { $sort : { month_joined : 1 } }
  ]
)
```

The pipeline passes all documents in the `users` collection through the following operations:

- The `$project` operator:
  - Creates two new fields: `month_joined` and `name`.
  - Suppresses the `id` from the results. The `aggregate()` method includes the `_id`, unless explicitly suppressed.
- The `$month` operator converts the values of the `joined` field to integer representations of the month. Then the `$project` operator assigns those values to the `month_joined` field.
- The `$sort` operator sorts the results by the `month_joined` field.

The operation returns results that resemble the following:

```
{
  "month_joined" : 1,
  "name" : "ruth"
},
{
  "month_joined" : 1,
  "name" : "harold"
},
{
  "month_joined" : 1,
  "name" : "kate"
}
{
  "month_joined" : 2,
  "name" : "jill"
}
```

Return Total Number of Joins per Month The following operation shows how many people joined each month of the year. You might use this aggregated data for recruiting and marketing strategies.

```
db.users.aggregate(
  [
    { $project : { month_joined : { $month : "$joined" } } } ,
    { $group : { _id : {month_joined:"$month_joined"} , number : { $sum : 1 } } },
    { $sort : { "_id.month_joined" : 1 } }
  ]
)
```

The pipeline passes all documents in the `users` collection through the following operations:

- The `$project` operator creates a new field called `month_joined`.
- The `$month` operator converts the values of the `joined` field to integer representations of the month. Then the `$project` operator assigns the values to the `month_joined` field.
- The `$group` operator collects all documents with a given `month_joined` value and counts how many documents there are for that value. Specifically, for each unique value, `$group` creates a new "per-month" document with two fields:

  - _id, which contains a nested document with the month_joined field and its value.
  - number, which is a generated field. The $sum operator increments this field by 1 for every document containing the given month_joined value.

- The $sort operator sorts the documents created by $group according to the contents of the month_joined field.

The result of this aggregation operation would resemble the following:

```
{
  "_id" : {
    "month_joined" : 1
  },
  "number" : 3
},
{
  "_id" : {
    "month_joined" : 2
  },
  "number" : 9
},
{
  "_id" : {
    "month_joined" : 3
  },
  "number" : 5
}
```

Return the Five Most Common "Likes" The following aggregation collects top five most "liked" activities in the data set. This type of analysis could help inform planning and future development.

```
db.users.aggregate(
  [
    { $unwind : "$likes" },
    { $group : { _id : "$likes" , number : { $sum : 1 } } },
    { $sort : { number : -1 } },
    { $limit : 5 }
  ]
)
```

The pipeline begins with all documents in the users collection, and passes these documents through the following operations:

- The $unwind operator separates each value in the likes array, and creates a new version of the source document for every element in the array.

  Example

  Given the following document from the users collection:

  ```
  {
    _id : "jane",
    joined : ISODate("2011-03-02"),
    likes : ["golf", "racquetball"]
  }
  ```

  The $unwind operator would create the following documents:

```
{
  _id : "jane",
  joined : ISODate("2011-03-02"),
  likes : "golf"
}
{
  _id : "jane",
  joined : ISODate("2011-03-02"),
  likes : "racquetball"
}
```

- The `$group` operator collects all documents the same value for the `likes` field and counts each grouping. With this information, `$group` creates a new document with two fields:
  - `_id`, which contains the `likes` value.
  - `number`, which is a generated field. The `$sum` operator increments this field by 1 for every document containing the given `likes` value.
- The `$sort` operator sorts these documents by the `number` field in reverse order.
- The `$limit` operator only includes the first 5 result documents.

The results of aggregation would resemble the following:

```
{
  "_id" : "golf",
  "number" : 33
},
{
  "_id" : "racquetball",
  "number" : 31
},
{
  "_id" : "swimming",
  "number" : 24
},
{
  "_id" : "handball",
  "number" : 19
},
{
  "_id" : "tennis",
  "number" : 18
}
```

Additional Resources

- MongoDB Analytics: Learn Aggregation by Example: Exploratory Analytics and Visualization Using Flight Data[4]
- MongoDB for Time Series Data: Analyzing Time Series Data Using the Aggregation Framework and Hadoop[5]
- The Aggregation Framework[6]

[4] http://www.mongodb.com/presentations/mongodb-analytics-learn-aggregation-example-exploratory-analytics-and-visualization?jmp=docs
[5] http://www.mongodb.com/presentations/mongodb-time-series-data-part-2-analyzing-time-series-data-using-aggregation-framework?jmp=docs
[6] https://www.mongodb.com/presentations/aggregation-framework-0?jmp=docs

- Webinar: Exploring the Aggregation Framework[7]
- Quick Reference Cards[8]

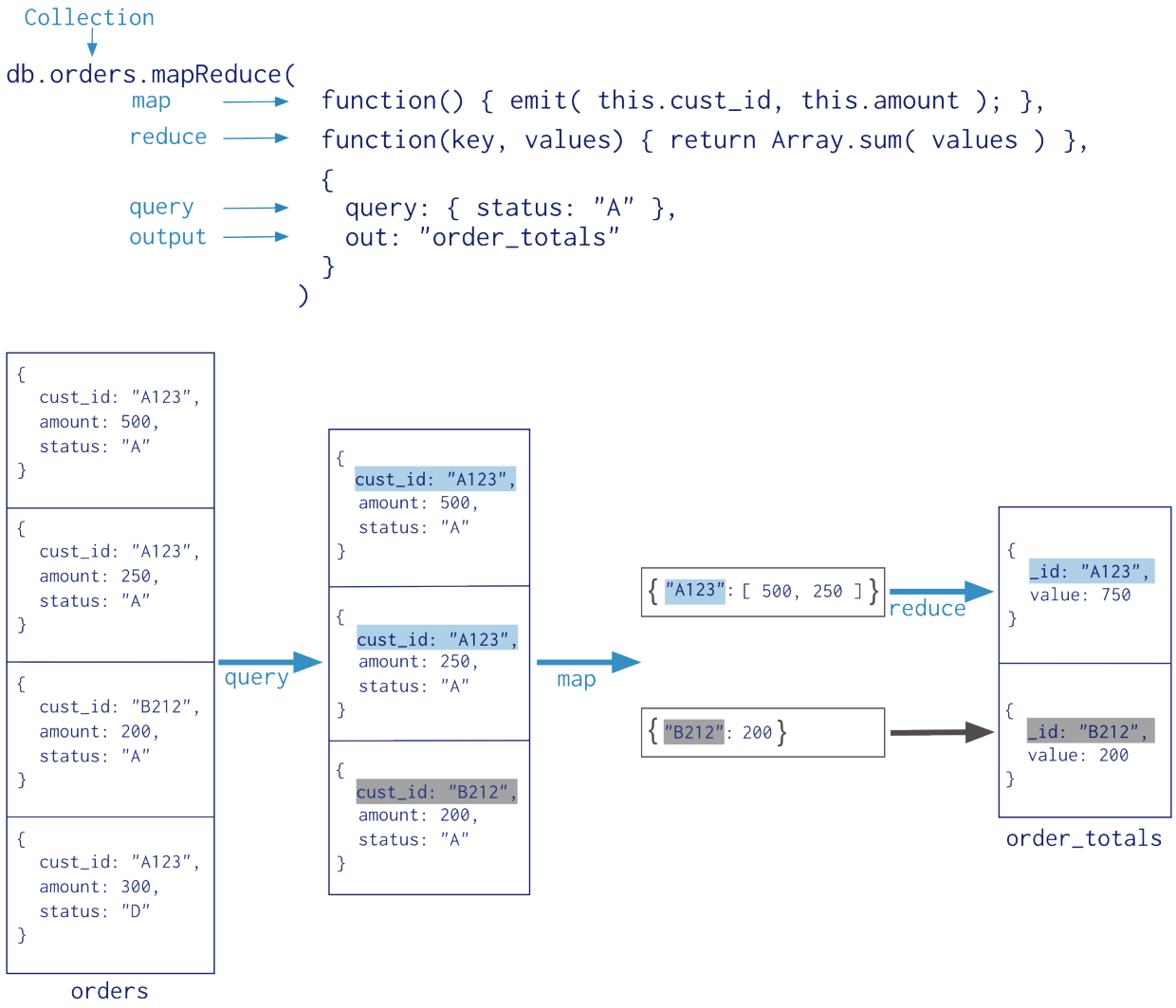
5.4.2 Map-Reduce

On this page

- Map-Reduce JavaScript Functions (page 215)
- Map-Reduce Behavior (page 215)

Map-reduce is a data processing paradigm for condensing large volumes of data into useful *aggregated* results. For map-reduce operations, MongoDB provides the `mapReduce` database command.

Consider the following map-reduce operation:



[7] https://www.mongodb.com/webinar/exploring-the-aggregation-framework?jmp=docs
[8] https://www.mongodb.com/lp/misc/quick-reference-cards?jmp=docs

In this map-reduce operation, MongoDB applies the *map* phase to each input document (i.e. the documents in the collection that match the query condition). The map function emits key-value pairs. For those keys that have multiple values, MongoDB applies the *reduce* phase, which collects and condenses the aggregated data. MongoDB then stores the results in a collection. Optionally, the output of the reduce function may pass through a *finalize* function to further condense or process the results of the aggregation.

All map-reduce functions in MongoDB are JavaScript and run within the `mongod` process. Map-reduce operations take the documents of a single *collection* as the *input* and can perform any arbitrary sorting and limiting before beginning the map stage. `mapReduce` can return the results of a map-reduce operation as a document, or may write the results to collections. The input and the output collections may be sharded.

Note: For most aggregation operations, the *Aggregation Pipeline* (page 199) provides better performance and more coherent interface. However, map-reduce operations provide some flexibility that is not presently available in the aggregation pipeline.

Map-Reduce JavaScript Functions

In MongoDB, map-reduce operations use custom JavaScript functions to *map*, or associate, values to a key. If a key has multiple values mapped to it, the operation *reduces* the values for the key to a single object.

The use of custom JavaScript functions provide flexibility to map-reduce operations. For instance, when processing a document, the map function can create more than one key and value mapping or no mapping. Map-reduce operations can also use a custom JavaScript function to make final modifications to the results at the end of the map and reduce operation, such as perform additional calculations.

Map-Reduce Behavior

In MongoDB, the map-reduce operation can write results to a collection or return the results inline. If you write map-reduce output to a collection, you can perform subsequent map-reduce operations on the same input collection that merge replace, merge, or reduce new results with previous results. See `mapReduce` and *Perform Incremental Map-Reduce* (page 219) for details and examples.

When returning the results of a map reduce operation *inline*, the result documents must be within the `BSON Document Size` limit, which is currently 16 megabytes. For additional information on limits and restrictions on map-reduce operations, see the `https://docs.mongodb.org/manual/reference/command/mapReduce` reference page.

MongoDB supports map-reduce operations on *sharded collections* (page 733). Map-reduce operations can also output the results to a sharded collection. See *Map-Reduce and Sharded Collections* (page 215).

Map-Reduce and Sharded Collections

On this page

- Sharded Collection as Input (page 216)
- Sharded Collection as Output (page 216)

Map-reduce supports operations on sharded collections, both as an input and as an output. This section describes the behaviors of `mapReduce` specific to sharded collections.

Sharded Collection as Input When using sharded collection as the input for a map-reduce operation, `mongos` will automatically dispatch the map-reduce job to each shard in parallel. There is no special option required. `mongos` will wait for jobs on all shards to finish.

Sharded Collection as Output If the `out` field for `mapReduce` has the `sharded` value, MongoDB shards the output collection using the `_id` field as the shard key.

To output to a sharded collection:

- If the output collection does not exist, MongoDB creates and shards the collection on the `_id` field.
- For a new or an empty sharded collection, MongoDB uses the results of the first stage of the map-reduce operation to create the initial *chunks* distributed among the shards.
- `mongos` dispatches, in parallel, a map-reduce post-processing job to every shard that owns a chunk. During the post-processing, each shard will pull the results for its own chunks from the other shards, run the final reduce/finalize, and write locally to the output collection.

Note:

- During later map-reduce jobs, MongoDB splits chunks as needed.
- Balancing of chunks for the output collection is automatically prevented during post-processing to avoid concurrency issues.

In MongoDB 2.0:

- `mongos` retrieves the results from each shard, performs a merge sort to order the results, and proceeds to the reduce/finalize phase as needed. `mongos` then writes the result to the output collection in sharded mode.
- This model requires only a small amount of memory, even for large data sets.
- Shard chunks are not automatically split during insertion. This requires manual intervention until the chunks are granular and balanced.

Important: For best results, only use the sharded output options for `mapReduce` in version 2.2 or later.

Map Reduce Concurrency

The map-reduce operation is composed of many tasks, including reads from the input collection, executions of the `map` function, executions of the `reduce` function, writes to a temporary collection during processing, and writes to the output collection.

During the operation, map-reduce takes the following locks:

- The read phase takes a read lock. It yields every 100 documents.
- The insert into the temporary collection takes a write lock for a single write.
- If the output collection does not exist, the creation of the output collection takes a write lock.
- If the output collection exists, then the output actions (i.e. `merge`, `replace`, `reduce`) take a write lock. This write lock is *global*, and blocks all operations on the `mongod` instance.

Note: The final write lock during post-processing makes the results appear atomically. However, output actions `merge` and `reduce` may take minutes to process. For the `merge` and `reduce`, the `nonAtomic` flag is available, which releases the lock between writing each output document. See the `db.collection.mapReduce()` reference for more information.

Map-Reduce Examples

On this page

- Return the Total Price Per Customer (page 217)
- Calculate Order and Total Quantity with Average Quantity Per Item (page 218)

In the `mongo` shell, the `db.collection.mapReduce()` method is a wrapper around the `mapReduce` command. The following examples use the `db.collection.mapReduce()` method:

Consider the following map-reduce operations on a collection `orders` that contains documents of the following prototype:

```
{
     _id: ObjectId("50a8240b927d5d8b5891743c"),
     cust_id: "abc123",
     ord_date: new Date("Oct 04, 2012"),
     status: 'A',
     price: 25,
     items: [ { sku: "mmm", qty: 5, price: 2.5 },
              { sku: "nnn", qty: 5, price: 2.5 } ]
}
```

Return the Total Price Per Customer Perform the map-reduce operation on the `orders` collection to group by the `cust_id`, and calculate the sum of the `price` for each `cust_id`:

1. Define the map function to process each input document:
   - In the function, `this` refers to the document that the map-reduce operation is processing.
   - The function maps the `price` to the `cust_id` for each document and emits the `cust_id` and `price` pair.

   ```
   var mapFunction1 = function() {
                          emit(this.cust_id, this.price);
                      };
   ```

2. Define the corresponding reduce function with two arguments `keyCustId` and `valuesPrices`:
   - The `valuesPrices` is an array whose elements are the `price` values emitted by the map function and grouped by `keyCustId`.
   - The function reduces the `valuesPrice` array to the sum of its elements.

   ```
   var reduceFunction1 = function(keyCustId, valuesPrices) {
                             return Array.sum(valuesPrices);
                         };
   ```

3. Perform the map-reduce on all documents in the `orders` collection using the `mapFunction1` map function and the `reduceFunction1` reduce function.

   ```
   db.orders.mapReduce(
                          mapFunction1,
                          reduceFunction1,
                          { out: "map_reduce_example" }
                        )
   ```

This operation outputs the results to a collection named `map_reduce_example`. If the `map_reduce_example` collection already exists, the operation will replace the contents with the results of this map-reduce operation:

Calculate Order and Total Quantity with Average Quantity Per Item In this example, you will perform a map-reduce operation on the `orders` collection for all documents that have an `ord_date` value greater than `01/01/2012`. The operation groups by the `item.sku` field, and calculates the number of orders and the total quantity ordered for each `sku`. The operation concludes by calculating the average quantity per order for each `sku` value:

1. Define the map function to process each input document:
   - In the function, `this` refers to the document that the map-reduce operation is processing.
   - For each item, the function associates the `sku` with a new object `value` that contains the `count` of `1` and the item `qty` for the order and emits the `sku` and `value` pair.

```
var mapFunction2 = function() {
                       for (var idx = 0; idx < this.items.length; idx++) {
                           var key = this.items[idx].sku;
                           var value = {
                                         count: 1,
                                         qty: this.items[idx].qty
                                       };
                           emit(key, value);
                       }
                   };
```

2. Define the corresponding reduce function with two arguments `keySKU` and `countObjVals`:
   - `countObjVals` is an array whose elements are the objects mapped to the grouped `keySKU` values passed by map function to the reducer function.
   - The function reduces the `countObjVals` array to a single object `reducedValue` that contains the `count` and the `qty` fields.
   - In `reducedVal`, the `count` field contains the sum of the `count` fields from the individual array elements, and the `qty` field contains the sum of the `qty` fields from the individual array elements.

```
var reduceFunction2 = function(keySKU, countObjVals) {
                        reducedVal = { count: 0, qty: 0 };

                        for (var idx = 0; idx < countObjVals.length; idx++) {
                            reducedVal.count += countObjVals[idx].count;
                            reducedVal.qty += countObjVals[idx].qty;
                        }

                        return reducedVal;
                    };
```

3. Define a finalize function with two arguments `key` and `reducedVal`. The function modifies the `reducedVal` object to add a computed field named `avg` and returns the modified object:

```
var finalizeFunction2 = function (key, reducedVal) {

                          reducedVal.avg = reducedVal.qty/reducedVal.count;

                          return reducedVal;

                        };
```

4. Perform the map-reduce operation on the `orders` collection using the `mapFunction2`, `reduceFunction2`, and `finalizeFunction2` functions.

```
db.orders.mapReduce( mapFunction2,
                     reduceFunction2,
                     {
                       out: { merge: "map_reduce_example" },
                       query: { ord_date:
                                   { $gt: new Date('01/01/2012') }
                              },
                       finalize: finalizeFunction2
                     }
                   )
```

This operation uses the `query` field to select only those documents with `ord_date` greater than `new Date(01/01/2012)`. Then it output the results to a collection `map_reduce_example`. If the `map_reduce_example` collection already exists, the operation will merge the existing contents with the results of this map-reduce operation.

Perform Incremental Map-Reduce

On this page

- Data Setup (page 219)
- Initial Map-Reduce of Current Collection (page 220)
- Subsequent Incremental Map-Reduce (page 221)

Map-reduce operations can handle complex aggregation tasks. To perform map-reduce operations, MongoDB provides the `mapReduce` command and, in the `mongo` shell, the `db.collection.mapReduce()` wrapper method.

If the map-reduce data set is constantly growing, you may want to perform an incremental map-reduce rather than performing the map-reduce operation over the entire data set each time.

To perform incremental map-reduce:

1. Run a map-reduce job over the current collection and output the result to a separate collection.
2. When you have more data to process, run subsequent map-reduce job with:
   - the `query` parameter that specifies conditions that match *only* the new documents.
   - the `out` parameter that specifies the `reduce` action to merge the new results into the existing output collection.

Consider the following example where you schedule a map-reduce operation on a `sessions` collection to run at the end of each day.

Data Setup The `sessions` collection contains documents that log users' sessions each day, for example:

```
db.sessions.save( { userid: "a", ts: ISODate('2011-11-03 14:17:00'), length: 95 } );
db.sessions.save( { userid: "b", ts: ISODate('2011-11-03 14:23:00'), length: 110 } );
db.sessions.save( { userid: "c", ts: ISODate('2011-11-03 15:02:00'), length: 120 } );
db.sessions.save( { userid: "d", ts: ISODate('2011-11-03 16:45:00'), length: 45 } );

db.sessions.save( { userid: "a", ts: ISODate('2011-11-04 11:05:00'), length: 105 } );
db.sessions.save( { userid: "b", ts: ISODate('2011-11-04 13:14:00'), length: 120 } );
```

```
db.sessions.save( { userid: "c", ts: ISODate('2011-11-04 17:00:00'), length: 130 } );
db.sessions.save( { userid: "d", ts: ISODate('2011-11-04 15:37:00'), length: 65 } );
```

Initial Map-Reduce of Current Collection Run the first map-reduce operation as follows:

1. Define the map function that maps the `userid` to an object that contains the fields `userid`, `total_time`, `count`, and `avg_time`:

```
var mapFunction = function() {
                      var key = this.userid;
                      var value = {
                                    userid: this.userid,
                                    total_time: this.length,
                                    count: 1,
                                    avg_time: 0
                                   };

                      emit( key, value );
                  };
```

2. Define the corresponding reduce function with two arguments `key` and `values` to calculate the total time and the count. The `key` corresponds to the `userid`, and the `values` is an array whose elements corresponds to the individual objects mapped to the `userid` in the `mapFunction`.

```
var reduceFunction = function(key, values) {

                       var reducedObject = {
                                            userid: key,
                                            total_time: 0,
                                            count:0,
                                            avg_time:0
                                           };

                       values.forEach( function(value) {
                                            reducedObject.total_time += value.total_time;
                                            reducedObject.count += value.count;
                                       }
                                     );
                       return reducedObject;
                     };
```

3. Define the finalize function with two arguments `key` and `reducedValue`. The function modifies the `reducedValue` document to add another field `average` and returns the modified document.

```
var finalizeFunction = function (key, reducedValue) {

                         if (reducedValue.count > 0)
                             reducedValue.avg_time = reducedValue.total_time / reducedValue.cou

                         return reducedValue;
                       };
```

4. Perform map-reduce on the `session` collection using the `mapFunction`, the `reduceFunction`, and the `finalizeFunction` functions. Output the results to a collection `session_stat`. If the `session_stat` collection already exists, the operation will replace the contents:

```
db.sessions.mapReduce( mapFunction,
                       reduceFunction,
```

```
                    {
                      out: "session_stat",
                      finalize: finalizeFunction
                    }
                  )
```

Subsequent Incremental Map-Reduce Later, as the `sessions` collection grows, you can run additional map-reduce operations. For example, add new documents to the `sessions` collection:

```
db.sessions.save( { userid: "a", ts: ISODate('2011-11-05 14:17:00'), length: 100 } );
db.sessions.save( { userid: "b", ts: ISODate('2011-11-05 14:23:00'), length: 115 } );
db.sessions.save( { userid: "c", ts: ISODate('2011-11-05 15:02:00'), length: 125 } );
db.sessions.save( { userid: "d", ts: ISODate('2011-11-05 16:45:00'), length: 55 } );
```

At the end of the day, perform incremental map-reduce on the `sessions` collection, but use the `query` field to select only the new documents. Output the results to the collection `session_stat`, but `reduce` the contents with the results of the incremental map-reduce:

```
db.sessions.mapReduce( mapFunction,
                       reduceFunction,
                       {
                         query: { ts: { $gt: ISODate('2011-11-05 00:00:00') } },
                         out: { reduce: "session_stat" },
                         finalize: finalizeFunction
                       }
                     );
```

Troubleshoot the Map Function

The `map` function is a JavaScript function that associates or "maps" a value with a key and emits the key and value pair during a *map-reduce* (page 214) operation.

To verify the `key` and `value` pairs emitted by the `map` function, write your own `emit` function.

Consider a collection `orders` that contains documents of the following prototype:

```
{
     _id: ObjectId("50a8240b927d5d8b5891743c"),
     cust_id: "abc123",
     ord_date: new Date("Oct 04, 2012"),
     status: 'A',
     price: 250,
     items: [ { sku: "mmm", qty: 5, price: 2.5 },
              { sku: "nnn", qty: 5, price: 2.5 } ]
}
```

1. Define the `map` function that maps the `price` to the `cust_id` for each document and emits the `cust_id` and `price` pair:

   ```
   var map = function() {
       emit(this.cust_id, this.price);
   };
   ```

2. Define the `emit` function to print the key and value:

   ```
   var emit = function(key, value) {
       print("emit");
   ```

```
    print("key: " + key + "  value: " + tojson(value));
}
```

3. Invoke the `map` function with a single document from the `orders` collection:

```
var myDoc = db.orders.findOne( { _id: ObjectId("50a8240b927d5d8b5891743c") } );
map.apply(myDoc);
```

4. Verify the key and value pair is as you expected.

```
emit
key: abc123 value:250
```

5. Invoke the `map` function with multiple documents from the `orders` collection:

```
var myCursor = db.orders.find( { cust_id: "abc123" } );

while (myCursor.hasNext()) {
    var doc = myCursor.next();
    print ("document _id= " + tojson(doc._id));
    map.apply(doc);
    print();
}
```

6. Verify the key and value pairs are as you expected.

See also:

The `map` function must meet various requirements. For a list of all the requirements for the `map` function, see `mapReduce`, or the `mongo` shell helper method `db.collection.mapReduce()`.

Troubleshoot the Reduce Function

On this page

- Confirm Output Type (page 222)
- Ensure Insensitivity to the Order of Mapped Values (page 223)
- Ensure Reduce Function Idempotence (page 224)

The `reduce` function is a JavaScript function that "reduces" to a single object all the values associated with a particular key during a *map-reduce* (page 214) operation. The `reduce` function must meet various requirements. This tutorial helps verify that the `reduce` function meets the following criteria:

- The `reduce` function must return an object whose *type* must be **identical** to the type of the `value` emitted by the `map` function.
- The order of the elements in the `valuesArray` should not affect the output of the `reduce` function.
- The `reduce` function must be *idempotent*.

For a list of all the requirements for the `reduce` function, see `mapReduce`, or the `mongo` shell helper method `db.collection.mapReduce()`.

Confirm Output Type You can test that the `reduce` function returns a value that is the same type as the value emitted from the `map` function.

1. Define a `reduceFunction1` function that takes the arguments `keyCustId` and `valuesPrices`. `valuesPrices` is an array of integers:

```
var reduceFunction1 = function(keyCustId, valuesPrices) {
                          return Array.sum(valuesPrices);
                      };
```

2. Define a sample array of integers:

```
var myTestValues = [ 5, 5, 10 ];
```

3. Invoke the `reduceFunction1` with `myTestValues`:

```
reduceFunction1('myKey', myTestValues);
```

4. Verify the `reduceFunction1` returned an integer:

```
20
```

5. Define a `reduceFunction2` function that takes the arguments `keySKU` and `valuesCountObjects`. `valuesCountObjects` is an array of documents that contain two fields `count` and `qty`:

```
var reduceFunction2 = function(keySKU, valuesCountObjects) {
                          reducedValue = { count: 0, qty: 0 };

                          for (var idx = 0; idx < valuesCountObjects.length; idx++) {
                              reducedValue.count += valuesCountObjects[idx].count;
                              reducedValue.qty += valuesCountObjects[idx].qty;
                          }

                          return reducedValue;
                      };
```

6. Define a sample array of documents:

```
var myTestObjects = [
                        { count: 1, qty: 5 },
                        { count: 2, qty: 10 },
                        { count: 3, qty: 15 }
                    ];
```

7. Invoke the `reduceFunction2` with `myTestObjects`:

```
reduceFunction2('myKey', myTestObjects);
```

8. Verify the `reduceFunction2` returned a document with exactly the `count` and the `qty` field:

```
{ "count" : 6, "qty" : 30 }
```

Ensure Insensitivity to the Order of Mapped Values The `reduce` function takes a `key` and a `values` array as its argument. You can test that the result of the `reduce` function does not depend on the order of the elements in the `values` array.

1. Define a sample `values1` array and a sample `values2` array that only differ in the order of the array elements:

```
var values1 = [
                  { count: 1, qty: 5 },
                  { count: 2, qty: 10 },
                  { count: 3, qty: 15 }
              ];
```

```
var values2 = [
                { count: 3, qty: 15 },
                { count: 1, qty: 5 },
                { count: 2, qty: 10 }
              ];
```

2. Define a `reduceFunction2` function that takes the arguments `keySKU` and `valuesCountObjects`. `valuesCountObjects` is an array of documents that contain two fields `count` and `qty`:

```
var reduceFunction2 = function(keySKU, valuesCountObjects) {
                          reducedValue = { count: 0, qty: 0 };

                          for (var idx = 0; idx < valuesCountObjects.length; idx++) {
                              reducedValue.count += valuesCountObjects[idx].count;
                              reducedValue.qty += valuesCountObjects[idx].qty;
                          }

                          return reducedValue;
                      };
```

3. Invoke the `reduceFunction2` first with `values1` and then with `values2`:

```
reduceFunction2('myKey', values1);
reduceFunction2('myKey', values2);
```

4. Verify the `reduceFunction2` returned the same result:

```
{ "count" : 6, "qty" : 30 }
```

Ensure Reduce Function Idempotence Because the map-reduce operation may call a `reduce` multiple times for the same key, and won't call a `reduce` for single instances of a key in the working set, the `reduce` function must return a value of the same type as the value emitted from the `map` function. You can test that the `reduce` function process "reduced" values without affecting the *final* value.

1. Define a `reduceFunction2` function that takes the arguments `keySKU` and `valuesCountObjects`. `valuesCountObjects` is an array of documents that contain two fields `count` and `qty`:

```
var reduceFunction2 = function(keySKU, valuesCountObjects) {
                          reducedValue = { count: 0, qty: 0 };

                          for (var idx = 0; idx < valuesCountObjects.length; idx++) {
                              reducedValue.count += valuesCountObjects[idx].count;
                              reducedValue.qty += valuesCountObjects[idx].qty;
                          }

                          return reducedValue;
                      };
```

2. Define a sample key:

```
var myKey = 'myKey';
```

3. Define a sample `valuesIdempotent` array that contains an element that is a call to the `reduceFunction2` function:

```
var valuesIdempotent = [
                           { count: 1, qty: 5 },
                           { count: 2, qty: 10 },
```

```
                reduceFunction2(myKey, [ { count:3, qty: 15 } ] )
              ];
```

4. Define a sample `values1` array that combines the values passed to `reduceFunction2`:

```
var values1 = [
                { count: 1, qty: 5 },
                { count: 2, qty: 10 },
                { count: 3, qty: 15 }
              ];
```

5. Invoke the `reduceFunction2` first with `myKey` and `valuesIdempotent` and then with `myKey` and `values1`:

```
reduceFunction2(myKey, valuesIdempotent);
reduceFunction2(myKey, values1);
```

6. Verify the `reduceFunction2` returned the same result:

```
{ "count" : 6, "qty" : 30 }
```

5.4.3 Aggregation Reference

Aggregation Pipeline Quick Reference **(page 225)** Quick reference card for aggregation pipeline.

Aggregation Commands **(page 231)** The reference for the data aggregation commands, which provide the interfaces to MongoDB's aggregation capability.

Aggregation Commands Comparison **(page 232)** A comparison of `group`, `mapReduce` and `aggregate` that explores the strengths and limitations of each aggregation modality.

`https://docs.mongodb.org/manual/reference/operator/aggregation` Aggregation pipeline operations have a collection of operators available to define and manipulate documents in pipeline stages.

Variables in Aggregation Expressions **(page 234)** Use of variables in aggregation pipeline expressions.

SQL to Aggregation Mapping Chart **(page 234)** An overview common aggregation operations in SQL and MongoDB using the aggregation pipeline and operators in MongoDB and common SQL statements.

Aggregation Pipeline Quick Reference

On this page

- Stages (page 225)
- Expressions (page 226)
- Accumulators (page 230)

Stages

In the `db.collection.aggregate` method, pipeline stages appear in an array. Documents pass through the stages in sequence. All except the `$out` and `$geoNear` stages can appear multiple times in a pipeline.

```
db.collection.aggregate( [ { <stage> }, ... ] )
```

Name	Description
`$project`	Reshapes each document in the stream, such as by adding new fields or removing existing fields. For each input document, outputs one document.
`$match`	Filters the document stream to allow only matching documents to pass unmodified into the next pipeline stage. `$match` uses standard MongoDB queries. For each input document, outputs either one document (a match) or zero documents (no match).
`$redact`	Reshapes each document in the stream by restricting the content for each document based on information stored in the documents themselves. Incorporates the functionality of `$project` and `$match`. Can be used to implement field level redaction. For each input document, outputs either one or zero documents.
`$limit`	Passes the first *n* documents unmodified to the pipeline where *n* is the specified limit. For each input document, outputs either one document (for the first *n* documents) or zero documents (after the first *n* documents).
`$skip`	Skips the first *n* documents where *n* is the specified skip number and passes the remaining documents unmodified to the pipeline. For each input document, outputs either zero documents (for the first *n* documents) or one document (if after the first *n* documents).
`$unwind`	Deconstructs an array field from the input documents to output a document for *each* element. Each output document replaces the array with an element value. For each input document, outputs *n* documents where *n* is the number of array elements and can be zero for an empty array.
`$group`	Groups input documents by a specified identifier expression and applies the accumulator expression(s), if specified, to each group. Consumes all input documents and outputs one document per each distinct group. The output documents only contain the identifier field and, if specified, accumulated fields.
`$sample`	Randomly selects the specified number of documents from its input.
`$sort`	Reorders the document stream by a specified sort key. Only the order changes; the documents remain unmodified. For each input document, outputs one document.
`$geoNear`	Returns an ordered stream of documents based on the proximity to a geospatial point. Incorporates the functionality of `$match`, `$sort`, and `$limit` for geospatial data. The output documents include an additional distance field and can include a location identifier field.
`$lookup`	Performs a left outer join to another collection in the *same* database to filter in documents from the "joined" collection for processing.
`$out`	Writes the resulting documents of the aggregation pipeline to a collection. To use the `$out` stage, it must be the last stage in the pipeline.
`$indexStats`	Returns statistics regarding the use of each index for the collection.

Expressions

Expressions can include *field paths and system variables* (page 226), *literals* (page 227), *expression objects* (page 227), and *expression operators* (page 227). Expressions can be nested.

Field Path and System Variables Aggregation expressions use *field path* to access fields in the input documents. To specify a field path, use a string that prefixes with a dollar sign `$` the field name or the dotted field name, if the field is in embedded document. For example, `"$user"` to specify the field path for the `user` field or `"$user.name"` to specify the field path to `"user.name"` field.

`"$<field>"` is equivalent to `"$$CURRENT.<field>"` where the `CURRENT` (page 234) is a system variable that defaults to the root of the current object in the most stages, unless stated otherwise in specific stages. `CURRENT` (page 234) can be rebound.

Along with the `CURRENT` (page 234) system variable, other *system variables* (page 234) are also available for use in expressions. To use user-defined variables, use `$let` and `$map` expressions. To access variables in expressions, use a string that prefixes the variable name with `$$`.

Literals Literals can be of any type. However, MongoDB parses string literals that start with a dollar sign `$` as a path to a field and numeric/boolean literals in *expression objects* (page 227) as projection flags. To avoid parsing literals, use the `$literal` expression.

Expression Objects Expression objects have the following form:

```
{ <field1>: <expression1>, ... }
```

If the expressions are numeric or boolean literals, MongoDB treats the literals as projection flags (e.g. `1` or `true` to include the field), valid only in the `$project` stage. To avoid treating numeric or boolean literals as projection flags, use the `$literal` expression to wrap the numeric or boolean literals.

Operator Expressions Operator expressions are similar to functions that take arguments. In general, these expressions take an array of arguments and have the following form:

```
{ <operator>: [ <argument1>, <argument2> ... ] }
```

If operator accepts a single argument, you can omit the outer array designating the argument list:

```
{ <operator>: <argument> }
```

To avoid parsing ambiguity if the argument is a literal array, you must wrap the literal array in a `$literal` expression or keep the outer array that designates the argument list.

Boolean Expressions Boolean expressions evaluate their argument expressions as booleans and return a boolean as the result.

In addition to the `false` boolean value, Boolean expression evaluates as `false` the following: `null`, `0`, and `undefined` values. The Boolean expression evaluates all other values as `true`, including non-zero numeric values and arrays.

Name	Description
`$and`	Returns `true` only when *all* its expressions evaluate to `true`. Accepts any number of argument expressions.
`$or`	Returns `true` when *any* of its expressions evaluates to `true`. Accepts any number of argument expressions.
`$not`	Returns the boolean value that is the opposite of its argument expression. Accepts a single argument expression.

Set Expressions Set expressions performs set operation on arrays, treating arrays as sets. Set expressions ignores the duplicate entries in each input array and the order of the elements.

If the set operation returns a set, the operation filters out duplicates in the result to output an array that contains only unique entries. The order of the elements in the output array is unspecified.

If a set contains a nested array element, the set expression does *not* descend into the nested array but evaluates the array at top-level.

Name	Description
`$setEquals`	Returns `true` if the input sets have the same distinct elements. Accepts two or more argument expressions.
`$setIntersection`	Returns a set with elements that appear in *all* of the input sets. Accepts any number of argument expressions.
`$setUnion`	Returns a set with elements that appear in *any* of the input sets. Accepts any number of argument expressions.
`$setDifference`	Returns a set with elements that appear in the first set but not in the second set; i.e. performs a relative complement[9] of the second set relative to the first. Accepts exactly two argument expressions.
`$setIsSubset`	Returns `true` if all elements of the first set appear in the second set, including when the first set equals the second set; i.e. not a strict subset[10]. Accepts exactly two argument expressions.
`$anyElementTrue`	Returns `true` if *any* elements of a set evaluate to `true`; otherwise, returns `false`. Accepts a single argument expression.
`$allElementsTrue`	Returns `true` if *no* element of a set evaluates to `false`, otherwise, returns `false`. Accepts a single argument expression.

Comparison Expressions Comparison expressions return a boolean except for `$cmp` which returns a number.

The comparison expressions take two argument expressions and compare both value and type, using the *specified BSON comparison order* (page 13) for values of different types.

Name	Description
`$cmp`	Returns: `0` if the two values are equivalent, `1` if the first value is greater than the second, and `-1` if the first value is less than the second.
`$eq`	Returns `true` if the values are equivalent.
`$gt`	Returns `true` if the first value is greater than the second.
`$gte`	Returns `true` if the first value is greater than or equal to the second.
`$lt`	Returns `true` if the first value is less than the second.
`$lte`	Returns `true` if the first value is less than or equal to the second.
`$ne`	Returns `true` if the values are *not* equivalent.

Arithmetic Expressions Arithmetic expressions perform mathematic operations on numbers. Some arithmetic expressions can also support date arithmetic.

[9] http://en.wikipedia.org/wiki/Complement_(set_theory)
[10] http://en.wikipedia.org/wiki/Subset

Name	Description
$abs	Returns the absolute value of a number.
$add	Adds numbers to return the sum, or adds numbers and a date to return a new date. If adding numbers and a date, treats the numbers as milliseconds. Accepts any number of argument expressions, but at most, one expression can resolve to a date.
$ceil	Returns the smallest integer greater than or equal to the specified number.
$divide	Returns the result of dividing the first number by the second. Accepts two argument expressions.
$exp	Raises *e* to the specified exponent.
$floor	Returns the largest integer less than or equal to the specified number.
$ln	Calculates the natural log of a number.
$log	Calculates the log of a number in the specified base.
$log10	Calculates the log base 10 of a number.
$mod	Returns the remainder of the first number divided by the second. Accepts two argument expressions.
$multiply	Multiplies numbers to return the product. Accepts any number of argument expressions.
$pow	Raises a number to the specified exponent.
$sqrt	Calculates the square root.
$subtract	Returns the result of subtracting the second value from the first. If the two values are numbers, return the difference. If the two values are dates, return the difference in milliseconds. If the two values are a date and a number in milliseconds, return the resulting date. Accepts two argument expressions. If the two values are a date and a number, specify the date argument first as it is not meaningful to subtract a date from a number.
$trunc	Truncates a number to its integer.

String Expressions String expressions, with the exception of `$concat`, only have a well-defined behavior for strings of ASCII characters.

`$concat` behavior is well-defined regardless of the characters used.

Name	Description
$concat	Concatenates any number of strings.
$substr	Returns a substring of a string, starting at a specified index position up to a specified length. Accepts three expressions as arguments: the first argument must resolve to a string, and the second and third arguments must resolve to integers.
$toLower	Converts a string to lowercase. Accepts a single argument expression.
$toUpper	Converts a string to uppercase. Accepts a single argument expression.
$strcasecmp	Performs case-insensitive string comparison and returns: `0` if two strings are equivalent, `1` if the first string is greater than the second, and `-1` if the first string is less than the second.

Text Search Expressions

Name	Description
$meta	Access text search metadata.

Array Expressions

Name	Description
$arrayElemAt	Returns the element at the specified array index.
$concatArrays	Concatenates arrays to return the concatenated array.
$filter	Selects a subset of the array to return an array with only the elements that match the filter condition.
$isArray	Determines if the operand is an array. Returns a boolean.
$size	Returns the number of elements in the array. Accepts a single expression as argument.
$slice	Returns a subset of an array.

Variable Expressions

Name	Description
`$map`	Applies a subexpression to each element of an array and returns the array of resulting values in orde Accepts named parameters.
`$let`	Defines variables for use within the scope of a subexpression and returns the result of the subexpress Accepts named parameters.

Literal Expressions

Name	Description
`$literal`	Return a value without parsing. Use for values that the aggregation pipeline may interpret as an expression. For example, use a `$literal` expression to a string that starts with a `$` to avoid parsing a field path.

Date Expressions

Name	Description
`$dayOfYear`	Returns the day of the year for a date as a number between 1 and 366 (leap year).
`$dayOfMonth`	Returns the day of the month for a date as a number between 1 and 31.
`$dayOfWeek`	Returns the day of the week for a date as a number between 1 (Sunday) and 7 (Saturday).
`$year`	Returns the year for a date as a number (e.g. 2014).
`$month`	Returns the month for a date as a number between 1 (January) and 12 (December).
`$week`	Returns the week number for a date as a number between 0 (the partial week that precedes the first Sunday of the year) and 53 (leap year).
`$hour`	Returns the hour for a date as a number between 0 and 23.
`$minute`	Returns the minute for a date as a number between 0 and 59.
`$second`	Returns the seconds for a date as a number between 0 and 60 (leap seconds).
`$millisecond`	Returns the milliseconds of a date as a number between 0 and 999.
`$dateToString`	Returns the date as a formatted string.

Conditional Expressions

Name	Description
`$cond`	A ternary operator that evaluates one expression, and depending on the result, returns the value o the other two expressions. Accepts either three expressions in an ordered list or three named par
`$ifNull`	Returns either the non-null result of the first expression or the result of the second expression if expression results in a null result. Null result encompasses instances of undefined values or miss fields. Accepts two expressions as arguments. The result of the second expression can be null.

Accumulators

Changed in version 3.2: Some accumulators are now available in the `$project` stage. In previous versions of MongoDB , accumulators are available only for the `$group` stage.

Accumulators, when used in the `$group` stage, maintain their state (e.g. totals, maximums, minimums, and related data) as documents progress through the pipeline.

When used in the `$group` stage, accumulators take as input a single expression, evaluating the expression once for each input document, and maintain their stage for the group of documents that share the same group key.

When used in the `$project` stage, the accumulators do not maintain their state. When used in the `$project` stage, accumulators take as input either a single argument or multiple arguments.

Name	Description
`$sum`	Returns a sum of numerical values. Ignores non-numeric values. Changed in version 3.2: Available in both `$group` and `$project` stages.
`$avg`	Returns an average of numerical values. Ignores non-numeric values. Changed in version 3.2: Available in both `$group` and `$project` stages.
`$first`	Returns a value from the first document for each group. Order is only defined if the documents are in a defined order. Available in `$group` stage only.
`$last`	Returns a value from the last document for each group. Order is only defined if the documents are in a defined order. Available in `$group` stage only.
`$max`	Returns the highest expression value for each group. Changed in version 3.2: Available in both `$group` and `$project` stages.
`$min`	Returns the lowest expression value for each group. Changed in version 3.2: Available in both `$group` and `$project` stages.
`$push`	Returns an array of expression values for each group. Available in `$group` stage only.
`$addToSet`	Returns an array of *unique* expression values for each group. Order of the array elements is undefined. Available in `$group` stage only.
`$stdDevPop`	Returns the population standard deviation of the input values. Changed in version 3.2: Available in both `$group` and `$project` stages.
`$stdDevSamp`	Returns the sample standard deviation of the input values. Changed in version 3.2: Available in both `$group` and `$project` stages.

Aggregation Commands

On this page

- Aggregation Commands (page 231)
- Aggregation Methods (page 231)

Aggregation Commands

Name	Description
`aggregate`	Performs *aggregation tasks* (page 199) such as group using the aggregation framework.
`count`	Counts the number of documents in a collection.
`distinct`	Displays the distinct values found for a specified key in a collection.
`group`	Groups documents in a collection by the specified key and performs simple aggregation.
`mapReduce`	Performs *map-reduce* (page 214) aggregation for large data sets.

Aggregation Methods

Name	Description
`db.collection.aggregate()`	Provides access to the *aggregation pipeline* (page 199).
`db.collection.group()`	Groups documents in a collection by the specified key and performs simple aggregation.
`db.collection.mapReduce()`	Performs *map-reduce* (page 214) aggregation for large data sets.

Aggregation Commands Comparison

The following table provides a brief overview of the features of the MongoDB aggregation commands.

	aggregate	mapReduce	group
Description	Designed with specific goals of improving performance and usability for aggregation tasks. Uses a "pipeline" approach where objects are transformed as they pass through a series of pipeline operators such as `$group`, `$match`, and `$sort`. See `https://docs.mongodb.org/manual/reference/operator/aggregation` for more information on the pipeline operators.	Implements the Map-Reduce aggregation for processing large data sets.	Provides grouping functionality. Is slower than the `aggregate` command and has less functionality than the `mapReduce` command.
Key Features	Pipeline operators can be repeated as needed. Pipeline operators need not produce one output document for every input document. Can also generate new documents or filter out documents.	In addition to grouping operations, can perform complex aggregation tasks as well as perform incremental aggregation on continuously growing datasets. See *Map-Reduce Examples* (page 217) and *Perform Incremental Map-Reduce* (page 219).	Can either group by existing fields or with a custom `keyf` JavaScript function, can group by calculated fields. See `group` for information and example using the `keyf` function.
Flexibility	Limited to the operators and expressions supported by the aggregation pipeline. However, can add computed fields, create new virtual sub-objects, and extract sub-fields into the top-level of results by using the `$project` pipeline operator. See `$project` for more information as well as `https://docs.mongodb.org/manual/reference/operator/aggregation` for more information on all the available pipeline operators.	Custom `map`, `reduce` and `finalize` JavaScript functions offer flexibility to aggregation logic. See `mapReduce` for details and restrictions on the functions.	Custom `reduce` and `finalize` JavaScript functions offer flexibility to grouping logic. See `group` for details and restrictions on these functions.
Output Results	Returns results in various options (inline as a document that contains the result set, a cursor to the result set) or stores the results in a collection. The result is subject to the *BSON Document size* limit if returned inline as a document that contains the result set. Changed in version 2.6: Can return results as a cursor or store the results to a collection.	Returns results in various options (inline, new collection, merge, replace, reduce). See `mapReduce` for details on the output options.	Returns results inline as an array of grouped items. The result set must fit within the *maximum BSON document size limit*. The returned array can contain at most 20,000 elements; i.e. at most 20,000 unique groupings.
Sharding Notes	Supports non-sharded and sharded input collections.	Supports non-sharded and sharded input collections. Prior to 2.4, JavaScript code executed in a single thread.	Does **not** support sharded collection. Prior to 2.4, JavaScript code executed in a single thread.
More Information	See *Aggregation Pipeline* (page 199) and `aggregate`.	See *Map-Reduce* (page 214) and `mapReduce`.	See `group`.

Variables in Aggregation Expressions

On this page

- User Variables (page 234)
- System Variables (page 234)

Aggregation expressions (page 226) can use both user-defined and system variables.

Variables can hold any *BSON type data* (page 12). To access the value of the variable, use a string with the variable name prefixed with double dollar signs (`$$`).

If the variable references an object, to access a specific field in the object, use the dot notation; i.e. `"$$<variable>.<field>"`.

User Variables

User variable names can contain the ascii characters `[_a-zA-Z0-9]` and any non-ascii character.

User variable names must begin with a lowercase ascii letter `[a-z]` or a non-ascii character.

System Variables

MongoDB offers the following system variables:

Variable	Description
ROOT	References the root document, i.e. the top-level document, currently being processed in the aggregation pipeline stage.
CURRENT	References the start of the field path being processed in the aggregation pipeline stage. Unless documented otherwise, all stages start with `CURRENT` (page 234) the same as `ROOT` (page 234). `CURRENT` (page 234) is modifiable. However, since `$<field>` is equivalent to `$$CURRENT.<field>`, rebinding `CURRENT` (page 234) changes the meaning of `$` accesses.
DESCEND	One of the allowed results of a `$redact` expression.
PRUNE	One of the allowed results of a `$redact` expression.
KEEP	One of the allowed results of a `$redact` expression.

See also:

`$let`, `$redact`, `$map`

SQL to Aggregation Mapping Chart

On this page

- Examples (page 235)
- Additional Resources (page 237)

The *aggregation pipeline* (page 199) allows MongoDB to provide native aggregation capabilities that corresponds to many common data aggregation operations in SQL.

The following table provides an overview of common SQL aggregation terms, functions, and concepts and the corresponding MongoDB *aggregation operators*:

SQL Terms, Functions, and Concepts	MongoDB Aggregation Operators
WHERE	`$match`
GROUP BY	`$group`
HAVING	`$match`
SELECT	`$project`
ORDER BY	`$sort`
LIMIT	`$limit`
SUM()	`$sum`
COUNT()	`$sum`
join	`$lookup` New in version 3.2.

Examples

The following table presents a quick reference of SQL aggregation statements and the corresponding MongoDB statements. The examples in the table assume the following conditions:

- The SQL examples assume *two* tables, `orders` and `order_lineitem` that join by the `order_lineitem.order_id` and the `orders.id` columns.
- The MongoDB examples assume *one* collection `orders` that contain documents of the following prototype:

```
{
  cust_id: "abc123",
  ord_date: ISODate("2012-11-02T17:04:11.102Z"),
  status: 'A',
  price: 50,
  items: [ { sku: "xxx", qty: 25, price: 1 },
           { sku: "yyy", qty: 25, price: 1 } ]
}
```

SQL Example	MongoDB Example	Description
`SELECT COUNT(*) AS count FROM orders`	`db.orders.aggregate( [ { $group: { _id: null, count: { $sum: 1 } } } ] )`	Count all records from `orders`
`SELECT SUM(price) AS total FROM orders`	`db.orders.aggregate( [ { $group: { _id: null, total: { $sum: "$price" } } } ] )`	Sum the `price` field from `orders`
`SELECT cust_id, SUM(price) AS total FROM orders GROUP BY cust_id`	`db.orders.aggregate( [ { $group: { _id: "$cust_id", total: { $sum: "$price" } } } ] )`	For each unique `cust_id`, sum the `price` field.
`SELECT cust_id, SUM(price) AS total FROM orders GROUP BY cust_id ORDER BY total`	`db.orders.aggregate( [ { $group: { _id: "$cust_id", total: { $sum: "$price" } } }, { $sort: { total: 1 } } ] )`	For each unique `cust_id`, sum the `price` field, results sorted by sum.
`SELECT cust_id, ord_date, SUM(price) AS total FROM orders GROUP BY cust_id, ord_date`	`db.orders.aggregate( [ { $group: { _id: { cust_id: "$cust_id", ord_date: { month: { $month: "$ord_date" }, day: { $dayOfMonth: "$ord_date" }, year: { $year: "$ord_date"} } }, total: { $sum: "$price" } } } ] )`	For each unique `cust_id`, `ord_date` grouping, sum the `price` field. Excludes the time portion of the date.
`SELECT cust_id, count(*) FROM orders`	`db.orders.aggregate( [ { $group: {`	For `cust_id` with multiple records, return the `cust_id` and the corresponding record count.

Additional Resources

- MongoDB and MySQL Compared[11]
- Quick Reference Cards[12]
- MongoDB Database Modernization Consulting Package[13]

5.5 Additional Resources

- MongoDB Analytics: Learn Aggregation by Example: Exploratory Analytics and Visualization Using Flight Data[14]
- MongoDB for Time Series Data: Analyzing Time Series Data Using the Aggregation Framework and Hadoop[15]
- The Aggregation Framework[16]
- Webinar: Exploring the Aggregation Framework[17]
- Quick Reference Cards[18]

[11] http://www.mongodb.com/mongodb-and-mysql-compared?jmp=docs
[12] https://www.mongodb.com/lp/misc/quick-reference-cards?jmp=docs
[13] https://www.mongodb.com/products/consulting?jmp=docs#database_modernization
[14] http://www.mongodb.com/presentations/mongodb-analytics-learn-aggregation-example-exploratory-analytics-and-visualization?jmp=docs
[15] http://www.mongodb.com/presentations/mongodb-time-series-data-part-2-analyzing-time-series-data-using-aggregation-framework?jmp=docs
[16] https://www.mongodb.com/presentations/aggregation-framework-0?jmp=docs
[17] https://www.mongodb.com/webinar/exploring-the-aggregation-framework?jmp=docs
[18] https://www.mongodb.com/lp/misc/quick-reference-cards?jmp=docs

CHAPTER 6

Text Search

On this page

- Overview (page 239)
- Example (page 239)
- Language Support (page 240)

6.1 Overview

MongoDB supports query operations that perform a text search of string content. To perform text search, MongoDB uses a *text index* (page 533) and the `$text` operator.

6.2 Example

This example demonstrates how to build a text index and use it to find coffee shops, given only text fields.

Create a collection `stores` with the following documents:

```
db.stores.insert(
   [
     { _id: 1, name: "Java Hut", description: "Coffee and cakes" },
     { _id: 2, name: "Burger Buns", description: "Gourmet hamburgers" },
     { _id: 3, name: "Coffee Shop", description: "Just coffee" },
     { _id: 4, name: "Clothes Clothes Clothes", description: "Discount clothing" },
     { _id: 5, name: "Java Shopping", description: "Indonesian goods" }
   ]
)
```

6.2.1 Text Index

MongoDB provides *text indexes* (page 533) to support text search queries on string content. `text` indexes can include any field whose value is a string or an array of string elements.

To perform text search queries, you must have a `text` index on your collection. A collection can only have **one** text search index, but that index can cover multiple fields.

For example you can run the following in a `mongo` shell to allow text search over the `name` and `description` fields:

```
db.stores.createIndex( { name: "text", description: "text" } )
```

6.2.2 `$text` Operator

Use the `$text` query operator to perform text searches on a collection with a *text index* (page 533).

`$text` will tokenize the search string using whitespace and most punctuation as delimiters, and perform a logical `OR` of all such tokens in the search string.

For example, you could use the following query to find all stores containing any terms from the list "coffee", "shop", and "java":

```
db.stores.find( { $text: { $search: "java coffee shop" } } )
```

Exact Phrase

You can also search for exact phrases by wrapping them in double-quotes. For example, the following will find all documents containing "java" or "coffee shop":

```
db.stores.find( { $text: { $search: "java \"coffee shop\"" } } )
```

Term Exclusion

To exclude a word, you can prepend a "–" character. For example, to find all stores containing "java" or "shop" but not "coffee", use the following:

```
db.stores.find( { $text: { $search: "java shop -coffee" } } )
```

Sorting

MongoDB will return its results in unsorted order by default. However, text search queries will compute a relevance score for each document that specifies how well a document matches the query.

To sort the results in order of relevance score, you must explicitly project the `$meta textScore` field and sort on it:

```
db.stores.find(
   { $text: { $search: "java coffee shop" } },
   { score: { $meta: "textScore" } }
).sort( { score: { $meta: "textScore" } } )
```

Text search is also available in the aggregation pipeline.

6.3 Language Support

MongoDB supports text search for various languages. See *Text Search Languages* (page 245) for a list of supported languages.

6.3.1 Text Indexes

MongoDB provides *text indexes* (page 533) to support text search queries on string content. `text` indexes can include any field whose value is a string or an array of string elements.

To perform text search queries, you must have a `text` index on your collection. A collection can only have **one** text search index, but that index can cover multiple fields.

For example you can run the following in a `mongo` shell to allow text search over the `name` and `description` fields:

```
db.stores.createIndex( { name: "text", description: "text" } )
```

See the *Text Indexes* (page 533) section for a full reference on text indexes, including behavior, tokenization, and properties.

6.3.2 Text Search Operators

On this page

- Query Framework (page 241)
- Aggregation Framework (page 241)

Query Framework

Use the `$text` query operator to perform text searches on a collection with a *text index* (page 533).

`$text` will tokenize the search string using whitespace and most punctuation as delimiters, and perform a logical `OR` of all such tokens in the search string.

For example, you could use the following query to find all stores containing any terms from the list "coffee", "shop", and "java":

```
db.stores.find( { $text: { $search: "java coffee shop" } } )
```

Use the `$meta` query operator to obtain and sort by the relevance score of each matching document. For example, to order a list of coffee shops in order of relevance, run the following:

```
db.stores.find(
   { $text: { $search: "coffee shop cake" } },
   { score: { $meta: "textScore" } }
).sort( { score: { $meta: "textScore" } } )
```

For more information on the `$text` and `$meta` operators, including restrictions and behavior, see:

- `$text Reference Page`
- *$text Query Examples*
- `$meta` projection operator

Aggregation Framework

When working with the *Aggregation* (page 195) framework, use `$match` with a `$text` expression to execute a text search query. To sort the results in order of relevance score,use the `$meta` *aggregation operator* in the `$sort` stage

[1].

For more information and examples of text search in the *Aggregation* (page 195) framework, see *Text Search in the Aggregation Pipeline* (page 242).

6.3.3 Text Search in the Aggregation Pipeline

On this page

- Restrictions (page 242)
- Text Score (page 242)
- Calculate the Total Views for Articles that Contains a Word (page 243)
- Return Results Sorted by Text Search Score (page 243)
- Match on Text Score (page 243)
- Specify a Language for Text Search (page 243)

New in version 2.6. In the aggregation pipeline, text search is available via the use of the `$text` query operator in the `$match` stage.

Restrictions

Text search in the aggregation pipeline has the following restrictions:

- The `$match` stage that includes a `$text` must be the **first** stage in the pipeline.
- A `text` operator can only occur once in the stage.
- The `text` operator expression cannot appear in `$or` or `$not` expressions.
- The text search, by default, does not return the matching documents in order of matching scores. Use the `$meta` aggregation expression in the `$sort` stage.

Text Score

The `$text` operator assigns a score to each document that contains the search term in the indexed fields. The score represents the relevance of a document to a given text search query. The score can be part of a `$sort` pipeline specification as well as part of the projection expression. The `{ $meta: "textScore" }` expression provides information on the processing of the `$text` operation. See `$meta` aggregation for details on accessing the score for projection or sort.

The metadata is only available after the `$match` stage that includes the `$text` operation.

Examples

The following examples assume a collection `articles` that has a text index on the field `subject`:

```
db.articles.createIndex( { subject: "text" } )
```

[1] The behavior and requirements of the `$meta` projection operator differ from that of the `$meta` aggregation operator. For details on the `$meta` aggregation operator, see the `$meta` aggregation operator reference page.

Calculate the Total Views for Articles that Contains a Word

The following aggregation searches for the term `cake` in the `$match` stage and calculates the total `views` for the matching documents in the `$group` stage.

```
db.articles.aggregate(
   [
     { $match: { $text: { $search: "cake" } } },
     { $group: { _id: null, views: { $sum: "$views" } } }
   ]
)
```

Return Results Sorted by Text Search Score

To sort by the text search score, include a `$meta` expression in the `$sort` stage. The following example matches on *either* the term `cake` or `tea`, sorts by the `textScore` in descending order, and returns only the `title` field in the results set.

```
db.articles.aggregate(
   [
     { $match: { $text: { $search: "cake tea" } } },
     { $sort: { score: { $meta: "textScore" } } },
     { $project: { title: 1, _id: 0 } }
   ]
)
```

The specified metadata determines the sort order. For example, the `"textScore"` metadata sorts in descending order. See `$meta` for more information on metadata as well as an example of overriding the default sort order of the metadata.

Match on Text Score

The `"textScore"` metadata is available for projections, sorts, and conditions subsequent the `$match` stage that includes the `$text` operation.

The following example matches on *either* the term `cake` or `tea`, projects the `title` and the `score` fields, and then returns only those documents with a `score` greater than `1.0`.

```
db.articles.aggregate(
   [
     { $match: { $text: { $search: "cake tea" } } },
     { $project: { title: 1, _id: 0, score: { $meta: "textScore" } } },
     { $match: { score: { $gt: 1.0 } } }
   ]
)
```

Specify a Language for Text Search

The following aggregation searches in spanish for documents that contain the term `saber` but not the term `claro` in the `$match` stage and calculates the total `views` for the matching documents in the `$group` stage.

```
db.articles.aggregate(
   [
     { $match: { $text: { $search: "saber -claro", $language: "es" } } },
     { $group: { _id: null, views: { $sum: "$views" } } }
```

```
    ]
)
```

6.3.4 Text Search with Basis Technology Rosette Linguistics Platform

On this page

- Overview (page 244)
- Prerequisites (page 244)
- Procedure (page 245)
- Additional Information (page 245)

Enterprise Feature

Available in MongoDB Enterprise only.

Overview

New in version 3.2.

In addition to the languages supported by text search in MongoDB, MongoDB Enterprise provides support for the following additional languages: Arabic, Farsi (specifically Dari and Iranian Persian dialects), Urdu, Simplified Chinese, and Traditional Chinese.

To provide support for these six additional languages, MongoDB Enterprise integrates Basis Technology Rosette Linguistics Platform (RLP) to perform normalization, word breaking, sentence breaking, and stemming or tokenization depending on the language.

MongoDB Enterprise supports RLP SDK 7.11.1 on Red Hat Enterprise Linux 6.x. For information on providing support on other platforms, contact your sales representative.

See also:

Text Search Languages (page 245), *Specify a Language for Text Index* (page 538)

Prerequisites

To use MongoDB with RLP, MongoDB requires a license for the Base Linguistics component of RLP and one or more languages specified above. MongoDB does not require a license for all six languages listed above.

Support for any of the specified languages is conditional on having a valid RLP license for the language. For instance, if there is only an RLP license provided for Arabic, then MongoDB will only enable support for Arabic and will not enable support for any other RLP based languages. For any language which lacks a valid license, the MongoDB log will contain a warning message. Additionally, you can set the MongoDB log verbosity level to `2` to log debug messages that identify each supported language.

You do not need the Language Extension Pack as MongoDB does not support these RLP languages at this time.

Contact Basis Technology at info@basistech.com[2] to get a copy of RLP and a license for one or more languages. For more information on how to contact Basis Technology, see http://www.basistech.com/contact/.

[2] info@basistech.com

Procedure

Step 1: Download Rosette Linguistics Platform from Basis Technology.

From Basis Technology, obtain the links to download the RLP C++ SDK package file, the documentation package file, and the license file (`rlp-license.xml`) for Linux x64. Basis Technology provides the download links in an email.

Using the links, download the RLP C++ SDK package file, the documentation package file, and the license file (`rlp-license.xml`) for Linux x64.

Note: These links automatically expire after 30 days.

Step 2: Install the RLP binaries.

Untar the RLP binaries and place them in a directory; this directory is referred to as the installation directory or `BT_ROOT`. For this example, we will use `/opt/basis` as the `BT_ROOT`.

```
tar zxvC /opt/basis rlp-7.11.1-sdk-amd64-glibc25-gcc41.tar.gz
```

Step 3: Move the RLP license into the RLP licenses directory.

Move the RLP license file `rlp-license.xml` to the `<BT_ROOT>/rlp/rlp/licenses` directory; in our example, move the file to the `/opt/basis/rlp/rlp/licenses/` directory.

```
mv rlp-license.xml /opt/basis/rlp/rlp/licenses/
```

Step 4: Run `mongod` with RLP support.

To enable support for RLP, use the `--basisTechRootDirectory` option to specify the `BT_ROOT` directory.

Include any additional settings as appropriate for your deployment.

```
mongod --basisTechRootDirectory=/opt/basis
```

Additional Information

For installation help, see the RLP Quick Start manual or Chapter 2 of the Rosette Linguistics Platform Application Developer's Guide.

For debugging any RLP specific issues, you can set the `rlpVerbose` parameter to `true` (i.e. `--setParameter rlpVerbose=true`) to view `INFO` messages from RLP.

Warning: Enabling `rlpVerbose` has a performance overhead and should only be enabled for troubleshooting installation issues.

6.3.5 Text Search Languages

The *text index* (page 533) and the `$text` operator supports the following languages:

Changed in version 2.6: MongoDB introduces version 2 of the text search feature. With version 2, text search feature supports using the two-letter language codes defined in ISO 639-1. Version 1 of text search only supported the long form of each language name.

Changed in version 3.2: MongoDB Enterprise includes support for Arabic, Farsi (specifically Dari and Iranian Persian dialects), Urdu, Simplified Chinese, and Traditional Chinese. To support the new languages, the text search feature uses the three-letter language codes defined in ISO 636-3. To enable support for these languages, see *Text Search with Basis Technology Rosette Linguistics Platform* (page 244).

Language Name	ISO 639-1 (Two letter codes)	ISO 636-3 (Three letter codes)	RLP names (Three letter codes)
`danish`	`da`		
`dutch`	`nl`		
`english`	`en`		
`finnish`	`fi`		
`french`	`fr`		
`german`	`de`		
`hungarian`	`hu`		
`italian`	`it`		
`norwegian`	`nb`		
`portuguese`	`pt`		
`romanian`	`ro`		
`russian`	`ru`		
`spanish`	`es`		
`swedish`	`sv`		
`turkish`	`tr`		
`arabic`		`ara`	
`dari`		`prs`	
`iranian persian`		`pes`	
`urdu`		`urd`	
`simplified chinese` or `hans`			`zhs`
`traditional chinese` or `hant`			`zht`

Note: If you specify a language value of `"none"`, then the text search uses simple tokenization with no list of stop words and no stemming.

See also:

Specify a Language for Text Index (page 538)

CHAPTER 7

Data Models

Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. This flexibility gives you data-modeling choices to match your application and its performance requirements.

Data Modeling Introduction **(page 247)** An introduction to data modeling in MongoDB.

Document Validation **(page 250)** MongoDB provides the capability to validate documents during updates and insertions.

Data Modeling Concepts **(page 252)** The core documentation detailing the decisions you must make when determining a data model, and discussing considerations that should be taken into account.

Data Model Examples and Patterns **(page 258)** Examples of possible data models that you can use to structure your MongoDB documents.

Data Model Reference **(page 276)** Reference material for data modeling for developers of MongoDB applications.

7.1 Data Modeling Introduction

On this page

- Document Structure (page 247)
- Atomicity of Write Operations (page 248)
- Document Growth (page 249)
- Data Use and Performance (page 249)
- Additional Resources (page 249)

Data in MongoDB has a *flexible schema*. Unlike SQL databases, where you must determine and declare a table's schema before inserting data, MongoDB's *collections* do not enforce *document* structure. This flexibility facilitates the mapping of documents to an entity or an object. Each document can match the data fields of the represented entity, even if the data has substantial variation. In practice, however, the documents in a collection share a similar structure.

The key challenge in data modeling is balancing the needs of the application, the performance characteristics of the database engine, and the data retrieval patterns. When designing data models, always consider the application usage of the data (i.e. queries, updates, and processing of the data) as well as the inherent structure of the data itself.

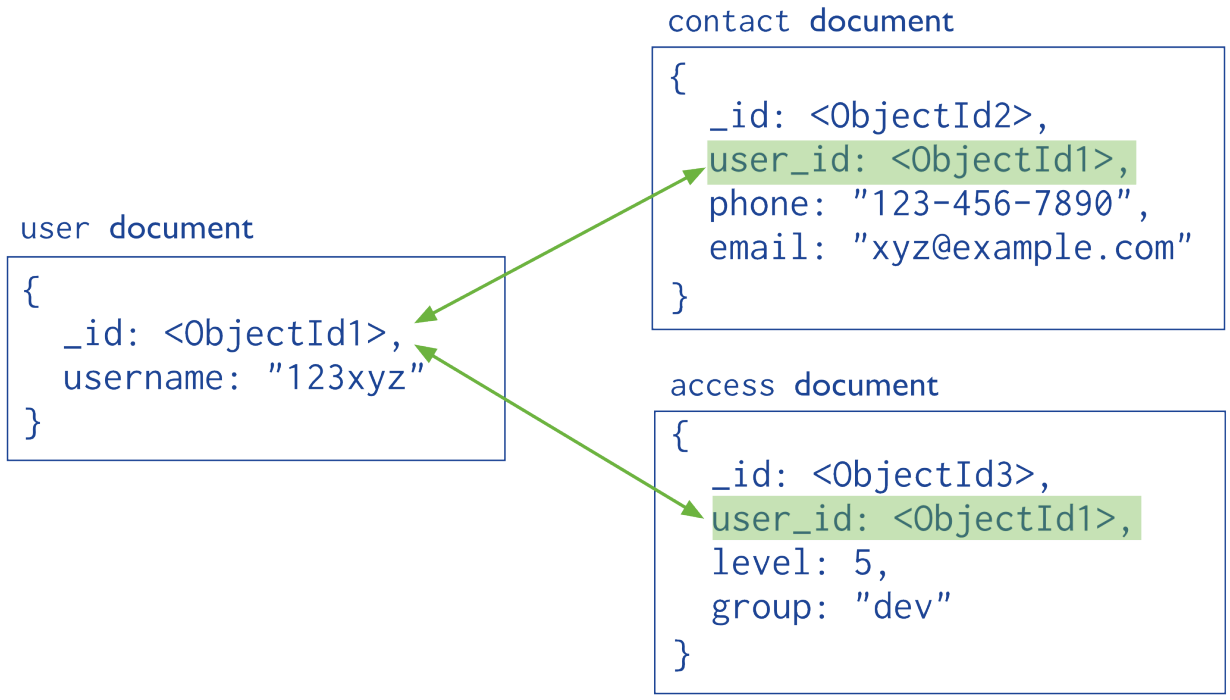
7.1.1 Document Structure

The key decision in designing data models for MongoDB applications revolves around the structure of documents and how the application represents relationships between data. There are two tools that allow applications to represent

these relationships: *references* and *embedded documents*.

References

References store the relationships between data by including links or *references* from one document to another. Applications can resolve these *references* (page 277) to access the related data. Broadly, these are *normalized* data models.



See *Normalized Data Models* (page 254) for the strengths and weaknesses of using references.

Embedded Data

Embedded documents capture relationships between data by storing related data in a single document structure. MongoDB documents make it possible to embed document structures in a field or array within a document. These *denormalized* data models allow applications to retrieve and manipulate related data in a single database operation.

See *Embedded Data Models* (page 253) for the strengths and weaknesses of embedding documents.

7.1.2 Atomicity of Write Operations

In MongoDB, write operations are atomic at the *document* level, and no single write operation can atomically affect more than one document or more than one collection. A denormalized data model with embedded data combines all related data for a represented entity in a single document. This facilitates atomic write operations since a single write operation can insert or update the data for an entity. Normalizing the data would split the data across multiple collections and would require multiple write operations that are not atomic collectively.

However, schemas that facilitate atomic writes may limit ways that applications can use the data or may limit ways to modify applications. The *Atomicity Considerations* (page 256) documentation describes the challenge of designing a schema that balances flexibility and atomicity.

7.1.3 Document Growth

Some updates, such as pushing elements to an array or adding new fields, increase a *document's* size.

For the MMAPv1 storage engine, if the document size exceeds the allocated space for that document, MongoDB relocates the document on disk. When using the MMAPv1 storage engine, growth consideration can affect the decision to normalize or denormalize data. See *Document Growth Considerations* (page 255) for more about planning for and managing document growth for MMAPv1.

7.1.4 Data Use and Performance

When designing a data model, consider how applications will use your database. For instance, if your application only uses recently inserted documents, consider using *Capped Collections* (page 6). Or if your application needs are mainly read operations to a collection, adding indexes to support common queries can improve performance.

See *Operational Factors and Data Models* (page 255) for more information on these and other operational considerations that affect data model designs.

7.1.5 Additional Resources

- Thinking in Documents Part 1 (Blog Post)[1]

[1] https://www.mongodb.com/blog/post/thinking-documents-part-1?jmp=docs

7.2 Document Validation

On this page

- Behavior (page 250)
- Restrictions (page 252)
- Bypass Document Validation (page 252)
- Additional Information (page 252)

New in version 3.2.

MongoDB provides the capability to validate documents during updates and insertions. Validation rules are specified on a per-collection basis using the `validator` option, which takes a document that specifies the validation rules or expressions. Specify the expressions using any *query operators*, with the exception of `$near`, `$nearSphere`, `$text`, and `$where`.

Add document validation to an existing collection using the `collMod` command with the `validator` option. You can also specify document validation rules when creating a new collection using `db.createCollection()` with the `validator` option, as in the following:

```
db.createCollection( "contacts",
   { validator: { $or:
      [
         { phone: { $type: "string" } },
         { email: { $regex: /@mongodb\.com$/ } },
         { status: { $in: [ "Unknown", "Incomplete" ] } }
      ]
   }
} )
```

MongoDB also provides the `validationLevel` option, which determines how strictly MongoDB applies validation rules to existing documents during an update, and the `validationAction` option, which determines whether MongoDB should `error` and reject documents that violate the validation rules or `warn` about the violations in the log but allow invalid documents.

7.2.1 Behavior

Validation occurs during updates and inserts. When you add validation to a collection, existing documents do not undergo validation checks until modification.

Existing Documents

You can control how MongoDB handles existing documents using the `validationLevel` option.

By default, `validationLevel` is `strict` and MongoDB applies validation rules to all inserts and updates. Setting `validationLevel` to `moderate` applies validation rules to inserts and to updates to existing documents that fulfill the validation criteria. With the `moderate` level, updates to existing documents that do not fulfill the validation criteria are not checked for validity.

Example

Consider the following documents in a `contacts` collection:

```
{
   "_id": "125876",
   "name": "Anne",
   "phone": "+1 555 123 456",
   "city": "London",
   "status": "Complete"
},
{
   "_id": "860000",
   "name": "Ivan",
   "city": "Vancouver"
}
```

Issue the following command to add a validator to the `contacts` collection:

```
db.runCommand( {
   collMod: "contacts",
   validator: { $or: [ { phone: { $exists: true } }, { email: { $exists: true } } ] },
   validationLevel: "moderate"
} )
```

The `contacts` collection now has a validator with the `moderate` validationLevel. If you attempted to update the document with `_id` of `125876`, MongoDB would apply validation rules since the existing document matches the criteria. In contrast, MongoDB will not apply validation rules to updates to the document with `_id` of `860000` as it does not meet the validation rules.

To disable validation entirely, you can set `validationLevel` to `off`.

Accept or Reject Invalid Documents

The `validationAction` option determines how MongoDB handles documents that violate the validation rules.

By default, `validationAction` is `error` and MongoDB rejects any insertion or update that violates the validation criteria. When `validationAction` is set to `warn`, MongoDB logs any violations but allows the insertion or update to proceed.

Example

The following example creates a `contacts` collection with a validator that specifies that inserted or updated documents should match at least one of three following conditions:

- the `phone` field is a string
- the `email` field matches the regular expression
- the `status` field is either `Unknown` or `Incomplete`.

```
db.createCollection( "contacts",
   {
      validator: { $or:
         [
            { phone: { $type: "string" } },
            { email: { $regex: /@mongodb\.com$/ } },
            { status: { $in: [ "Unknown", "Incomplete" ] } }
         ]
      },
      validationAction: "warn"
   }
)
```

With the validator in place, the following insert operation fails the validation rules, but since the `validationAction` is `warn`, the write operation logs the failure and succeeds.

```
db.contacts.insert( { name: "Amanda", status: "Updated" } )
```

The log includes the full namespace of the collection and the document that failed the validation rules, as well as the time of the operation:

```
2015-10-15T11:20:44.260-0400 W STORAGE  [conn3] Document would fail validation collection: example.co
```

7.2.2 Restrictions

You cannot specify a validator for collections in the `admin`, `local`, and `config` databases.

You cannot specify a validator for `system.*` collections.

7.2.3 Bypass Document Validation

User can bypass document validation using the `bypassDocumentValidation` option. For a list of commands that support the `bypassDocumentValidation` option, see *Document Validation* (page 890).

For deployments that have enabled access control, to bypass document validation, the authenticated user must have `bypassDocumentValidation` (page 501) action. The built-in roles `dbAdmin` (page 487) and `restore` (page 492) provide this action.

7.2.4 Additional Information

See also:

`collMod`, `db.createCollection()`, `db.getCollectionInfos()`.

7.3 Data Modeling Concepts

Consider the following aspects of data modeling in MongoDB:

Data Model Design **(page 252)** Presents the different strategies that you can choose from when determining your data model, their strengths and their weaknesses.

Operational Factors and Data Models **(page 255)** Details features you should keep in mind when designing your data model, such as lifecycle management, indexing, horizontal scalability, and document growth.

For a general introduction to data modeling in MongoDB, see the *Data Modeling Introduction* (page 247). For example data models, see *Data Modeling Examples and Patterns* (page 258).

7.3.1 Data Model Design

On this page

- Embedded Data Models (page 253)
- Normalized Data Models (page 254)
- Additional Resources (page 254)

Effective data models support your application needs. The key consideration for the structure of your documents is the decision to *embed* (page 253) or to *use references* (page 254).

Embedded Data Models

With MongoDB, you may embed related data in a single structure or document. These schema are generally known as "denormalized" models, and take advantage of MongoDB's rich documents. Consider the following diagram:

Embedded data models allow applications to store related pieces of information in the same database record. As a result, applications may need to issue fewer queries and updates to complete common operations.

In general, use embedded data models when:

- you have "contains" relationships between entities. See *Model One-to-One Relationships with Embedded Documents* (page 259).
- you have one-to-many relationships between entities. In these relationships the "many" or child documents always appear with or are viewed in the context of the "one" or parent documents. See *Model One-to-Many Relationships with Embedded Documents* (page 260).

In general, embedding provides better performance for read operations, as well as the ability to request and retrieve related data in a single database operation. Embedded data models make it possible to update related data in a single atomic write operation.

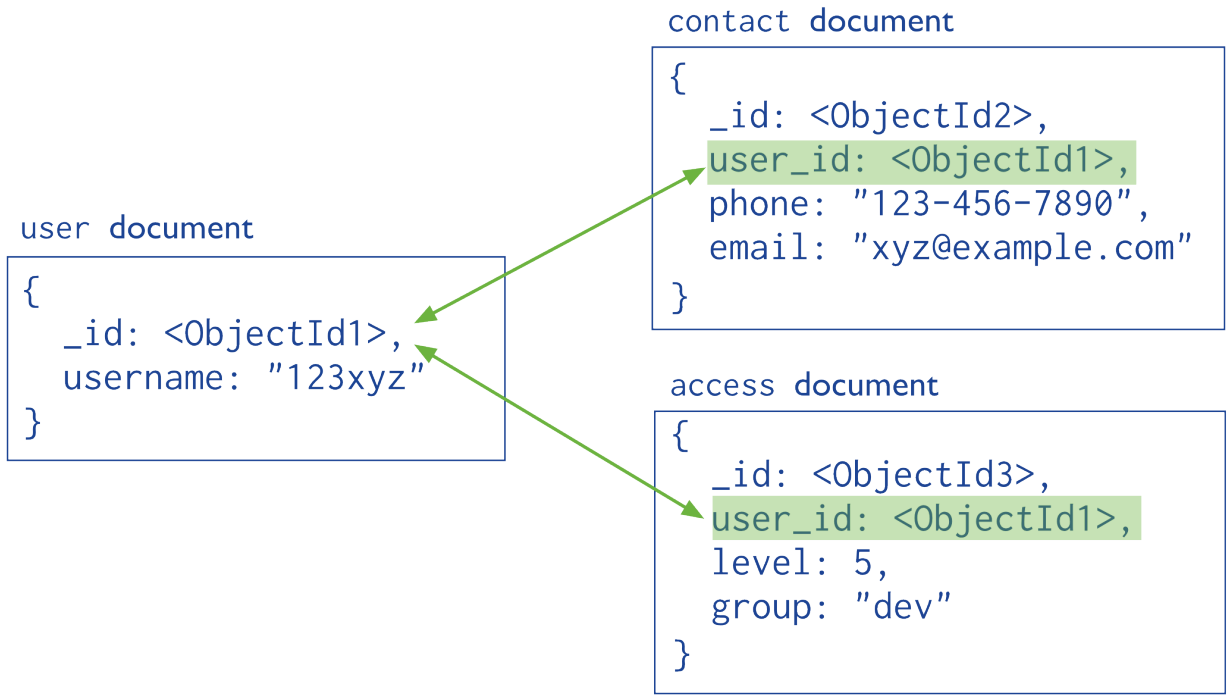
However, embedding related data in documents may lead to situations where documents grow after creation. With the MMAPv1 storage engine, document growth can impact write performance and lead to data fragmentation.

In version 3.0.0, MongoDB uses *Power of 2 Sized Allocations* (page 604) as the default allocation strategy for MMAPv1 in order to account for document growth, minimizing the likelihood of data fragmentation. See *Power of 2 Sized Allocations* (page 604) for details. Furthermore, documents in MongoDB must be smaller than the `maximum BSON document size`. For bulk binary data, consider *GridFS* (page 611).

To interact with embedded documents, use *dot notation* to "reach into" embedded documents. See *query for data in arrays* (page 143) and *query data in embedded documents* (page 142) for more examples on accessing data in arrays and embedded documents.

Normalized Data Models

Normalized data models describe relationships using *references* (page 277) between documents.



In general, use normalized data models:

- when embedding would result in duplication of data but would not provide sufficient read performance advantages to outweigh the implications of the duplication.
- to represent more complex many-to-many relationships.
- to model large hierarchical data sets.

References provides more flexibility than embedding. However, client-side applications must issue follow-up queries to resolve the references. In other words, normalized data models can require more round trips to the server.

See *Model One-to-Many Relationships with Document References* (page 261) for an example of referencing. For examples of various tree models using references, see *Model Tree Structures* (page 263).

Additional Resources

- Thinking in Documents Part 1 (Blog Post)[2]
- Thinking in Documents (Presentation)[3]

[2] https://www.mongodb.com/blog/post/thinking-documents-part-1?jmp=docs
[3] http://www.mongodb.com/presentations/webinar-back-basics-1-thinking-documents?jmp=docs

- Schema Design for Time Series Data (Presentation)[4]
- Socialite, the Open Source Status Feed - Storing a Social Graph (Presentation)[5]
- MongoDB Rapid Start Consultation Services[6]

7.3.2 Operational Factors and Data Models

On this page

- Document Growth (page 255)
- Atomicity (page 256)
- Sharding (page 256)
- Indexes (page 256)
- Large Number of Collections (page 256)
- Collection Contains Large Number of Small Documents (page 257)
- Storage Optimization for Small Documents (page 257)
- Data Lifecycle Management (page 258)

Modeling application data for MongoDB depends on both the data itself, as well as the characteristics of MongoDB itself. For example, different data models may allow applications to use more efficient queries, increase the throughput of insert and update operations, or distribute activity to a sharded cluster more effectively.

These factors are *operational* or address requirements that arise outside of the application but impact the performance of MongoDB based applications. When developing a data model, analyze all of your application's *read operations* (page 100) and *write operations* (page 114) in conjunction with the following considerations.

Document Growth

Changed in version 3.0.0.

Some updates to documents can increase the size of documents. These updates include pushing elements to an array (i.e. `$push`) and adding new fields to a document.

When using the MMAPv1 storage engine, document growth can be a consideration for your data model. For MMAPv1, if the document size exceeds the allocated space for that document, MongoDB will relocate the document on disk. With MongoDB 3.0.0, however, the default use of the *Power of 2 Sized Allocations* (page 604) minimizes the occurrences of such re-allocations as well as allows for the effective reuse of the freed record space.

When using MMAPv1, if your applications require updates that will frequently cause document growth to exceeds the current power of 2 allocation, you may want to refactor your data model to use references between data in distinct documents rather than a denormalized data model.

You may also use a *pre-allocation* strategy to explicitly avoid document growth. Refer to the Pre-Aggregated Reports Use Case[7] for an example of the *pre-allocation* approach to handling document growth.

See *MMAPv1 Storage Engine* (page 603) for more information on MMAPv1.

[4] http://www.mongodb.com/presentations/webinar-time-series-data-mongodb?jmp=docs
[5] http://www.mongodb.com/presentations/socialite-open-source-status-feed-part-2-managing-social-graph?jmp=docs
[6] https://www.mongodb.com/products/consulting?jmp=docs#rapid_start
[7] https://docs.mongodb.org/ecosystem/use-cases/pre-aggregated-reports

Atomicity

In MongoDB, operations are atomic at the *document* level. No **single** write operation can change more than one document. Operations that modify more than a single document in a collection still operate on one document at a time. [8] Ensure that your application stores all fields with atomic dependency requirements in the same document. If the application can tolerate non-atomic updates for two pieces of data, you can store these data in separate documents.

A data model that embeds related data in a single document facilitates these kinds of atomic operations. For data models that store references between related pieces of data, the application must issue separate read and write operations to retrieve and modify these related pieces of data.

See *Model Data for Atomic Operations* (page 272) for an example data model that provides atomic updates for a single document.

Sharding

MongoDB uses *sharding* to provide horizontal scaling. These clusters support deployments with large data sets and high-throughput operations. Sharding allows users to *partition* a *collection* within a database to distribute the collection's documents across a number of `mongod` instances or *shards*.

To distribute data and application traffic in a sharded collection, MongoDB uses the *shard key* (page 747). Selecting the proper *shard key* (page 747) has significant implications for performance, and can enable or prevent query isolation and increased write capacity. It is important to consider carefully the field or fields to use as the shard key.

See *Sharding Introduction* (page 733) and *Shard Keys* (page 747) for more information.

Indexes

Use indexes to improve performance for common queries. Build indexes on fields that appear often in queries and for all operations that return sorted results. MongoDB automatically creates a unique index on the `_id` field.

As you create indexes, consider the following behaviors of indexes:

- Each index requires at least 8 kB of data space.
- Adding an index has some negative performance impact for write operations. For collections with high write-to-read ratio, indexes are expensive since each insert must also update any indexes.
- Collections with high read-to-write ratio often benefit from additional indexes. Indexes do not affect un-indexed read operations.
- When active, each index consumes disk space and memory. This usage can be significant and should be tracked for capacity planning, especially for concerns over working set size.

See *Indexing Strategies* (page 586) for more information on indexes as well as *Analyze Query Performance* (page 159). Additionally, the MongoDB *database profiler* (page 326) may help identify inefficient queries.

Large Number of Collections

In certain situations, you might choose to store related information in several collections rather than in a single collection.

Consider a sample collection `logs` that stores log documents for various environment and applications. The `logs` collection contains documents of the following form:

[8] Document-level atomic operations include all operations within a single MongoDB document record: operations that affect multiple embedded documents within that single record are still atomic.

```
{ log: "dev", ts: ..., info: ... }
{ log: "debug", ts: ..., info: ...}
```

If the total number of documents is low, you may group documents into collection by type. For logs, consider maintaining distinct log collections, such as `logs_dev` and `logs_debug`. The `logs_dev` collection would contain only the documents related to the dev environment.

Generally, having a large number of collections has no significant performance penalty and results in very good performance. Distinct collections are very important for high-throughput batch processing.

When using models that have a large number of collections, consider the following behaviors:

- Each collection has a certain minimum overhead of a few kilobytes.
- Each index, including the index on `_id`, requires at least 8 kB of data space.
- For each *database*, a single namespace file (i.e. `<database>.ns`) stores all meta-data for that database, and each index and collection has its own entry in the namespace file. MongoDB places `limits on the size of namespace files`.
- MongoDB using the `mmapv1` storage engine has `limits on the number of namespaces`. You may wish to know the current number of namespaces in order to determine how many additional namespaces the database can support. To get the current number of namespaces, run the following in the `mongo` shell:

  ```
  db.system.namespaces.count()
  ```

  The limit on the number of namespaces depend on the `<database>.ns` size. The namespace file defaults to 16 MB.

  To change the size of the *new* namespace file, start the server with the option `--nssize <new size MB>`. For existing databases, after starting up the server with `--nssize`, run the `db.repairDatabase()` command from the `mongo` shell. For impacts and considerations on running `db.repairDatabase()`, see `repairDatabase`.

Collection Contains Large Number of Small Documents

You should consider embedding for performance reasons if you have a collection with a large number of small documents. If you can group these small documents by some logical relationship *and* you frequently retrieve the documents by this grouping, you might consider "rolling-up" the small documents into larger documents that contain an array of embedded documents.

"Rolling up" these small documents into logical groupings means that queries to retrieve a group of documents involve sequential reads and fewer random disk accesses. Additionally, "rolling up" documents and moving common fields to the larger document benefit the index on these fields. There would be fewer copies of the common fields *and* there would be fewer associated key entries in the corresponding index. See *Indexes* (page 515) for more information on indexes.

However, if you often only need to retrieve a subset of the documents within the group, then "rolling-up" the documents may not provide better performance. Furthermore, if small, separate documents represent the natural model for the data, you should maintain that model.

Storage Optimization for Small Documents

Each MongoDB document contains a certain amount of overhead. This overhead is normally insignificant but becomes significant if all documents are just a few bytes, as might be the case if the documents in your collection only have one or two fields.

Consider the following suggestions and strategies for optimizing storage utilization for these collections:

- Use the `_id` field explicitly.

  MongoDB clients automatically add an `_id` field to each document and generate a unique 12-byte *ObjectId* for the `_id` field. Furthermore, MongoDB always indexes the `_id` field. For smaller documents this may account for a significant amount of space.

  To optimize storage use, users can specify a value for the `_id` field explicitly when inserting documents into the collection. This strategy allows applications to store a value in the `_id` field that would have occupied space in another portion of the document.

  You can store any value in the `_id` field, but because this value serves as a primary key for documents in the collection, it must uniquely identify them. If the field's value is not unique, then it cannot serve as a primary key as there would be collisions in the collection.

- Use shorter field names.

  Note: Shortening field names reduces expressiveness and does not provide considerable benefit for larger documents and where document overhead is not of significant concern. Shorter field names do not reduce the size of indexes, because indexes have a predefined structure.

  In general, it is not necessary to use short field names.

  MongoDB stores all field names in every document. For most documents, this represents a small fraction of the space used by a document; however, for small documents the field names may represent a proportionally large amount of space. Consider a collection of small documents that resemble the following:

  ```
  { last_name : "Smith", best_score: 3.9 }
  ```

  If you shorten the field named `last_name` to `lname` and the field named `best_score` to `score`, as follows, you could save 9 bytes per document.

  ```
  { lname : "Smith", score : 3.9 }
  ```

- Embed documents.

  In some cases you may want to embed documents in other documents and save on the per-document overhead. See *Collection Contains Large Number of Small Documents* (page 257).

Data Lifecycle Management

Data modeling decisions should take data lifecycle management into consideration.

The *Time to Live or TTL feature* (page 567) of collections expires documents after a period of time. Consider using the TTL feature if your application requires some data to persist in the database for a limited period of time.

Additionally, if your application only uses recently inserted documents, consider *Capped Collections* (page 6). Capped collections provide *first-in-first-out* (FIFO) management of inserted documents and efficiently support operations that insert and read documents based on insertion order.

7.4 Data Model Examples and Patterns

The following documents provide overviews of various data modeling patterns and common schema design considerations:

***Model Relationships Between Documents* (page 259)** Examples for modeling relationships between documents.

> ***Model One-to-One Relationships with Embedded Documents* (page 259)** Presents a data model that uses *embedded documents* (page 253) to describe one-to-one relationships between connected data.

Model One-to-Many Relationships with Embedded Documents **(page 260)** Presents a data model that uses *embedded documents* (page 253) to describe one-to-many relationships between connected data.

Model One-to-Many Relationships with Document References **(page 261)** Presents a data model that uses *references* (page 254) to describe one-to-many relationships between documents.

Model Tree Structures **(page 263)** Examples for modeling tree structures.

Model Tree Structures with Parent References **(page 264)** Presents a data model that organizes documents in a tree-like structure by storing *references* (page 254) to "parent" nodes in "child" nodes.

Model Tree Structures with Child References **(page 266)** Presents a data model that organizes documents in a tree-like structure by storing *references* (page 254) to "child" nodes in "parent" nodes.

See *Model Tree Structures* (page 263) for additional examples of data models for tree structures.

Model Specific Application Contexts **(page 272)** Examples for models for specific application contexts.

Model Data for Atomic Operations **(page 272)** Illustrates how embedding fields related to an atomic update within the same document ensures that the fields are in sync.

Model Data to Support Keyword Search **(page 273)** Describes one method for supporting keyword search by storing keywords in an array in the same document as the text field. Combined with a multi-key index, this pattern can support application's keyword search operations.

7.4.1 Model Relationships Between Documents

Model One-to-One Relationships with Embedded Documents **(page 259)** Presents a data model that uses *embedded documents* (page 253) to describe one-to-one relationships between connected data.

Model One-to-Many Relationships with Embedded Documents **(page 260)** Presents a data model that uses *embedded documents* (page 253) to describe one-to-many relationships between connected data.

Model One-to-Many Relationships with Document References **(page 261)** Presents a data model that uses *references* (page 254) to describe one-to-many relationships between documents.

Model One-to-One Relationships with Embedded Documents

On this page

- Overview (page 259)
- Pattern (page 260)

Overview

Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that uses *embedded* (page 253) documents to describe relationships between connected data.

Pattern

Consider the following example that maps patron and address relationships. The example illustrates the advantage of embedding over referencing if you need to view one data entity in context of the other. In this one-to-one relationship between `patron` and `address` data, the `address` belongs to the `patron`.

In the normalized data model, the `address` document contains a reference to the `patron` document.

```
{
   _id: "joe",
   name: "Joe Bookreader"
}

{
   patron_id: "joe",
   street: "123 Fake Street",
   city: "Faketon",
   state: "MA",
   zip: "12345"
}
```

If the `address` data is frequently retrieved with the `name` information, then with referencing, your application needs to issue multiple queries to resolve the reference. The better data model would be to embed the `address` data in the `patron` data, as in the following document:

```
{
   _id: "joe",
   name: "Joe Bookreader",
   address: {
              street: "123 Fake Street",
              city: "Faketon",
              state: "MA",
              zip: "12345"
            }
}
```

With the embedded data model, your application can retrieve the complete patron information with one query.

Model One-to-Many Relationships with Embedded Documents

On this page

- Overview (page 260)
- Pattern (page 261)

Overview

Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that uses *embedded* (page 253) documents to describe relationships between connected data.

Pattern

Consider the following example that maps patron and multiple address relationships. The example illustrates the advantage of embedding over referencing if you need to view many data entities in context of another. In this one-to-many relationship between `patron` and `address` data, the `patron` has multiple `address` entities.

In the normalized data model, the `address` documents contain a reference to the `patron` document.

```
{
   _id: "joe",
   name: "Joe Bookreader"
}

{
   patron_id: "joe",
   street: "123 Fake Street",
   city: "Faketon",
   state: "MA",
   zip: "12345"
}

{
   patron_id: "joe",
   street: "1 Some Other Street",
   city: "Boston",
   state: "MA",
   zip: "12345"
}
```

If your application frequently retrieves the `address` data with the `name` information, then your application needs to issue multiple queries to resolve the references. A more optimal schema would be to embed the `address` data entities in the `patron` data, as in the following document:

```
{
   _id: "joe",
   name: "Joe Bookreader",
   addresses: [
                {
                  street: "123 Fake Street",
                  city: "Faketon",
                  state: "MA",
                  zip: "12345"
                },
                {
                  street: "1 Some Other Street",
                  city: "Boston",
                  state: "MA",
                  zip: "12345"
                }
              ]
 }
```

With the embedded data model, your application can retrieve the complete patron information with one query.

Model One-to-Many Relationships with Document References

On this page

- Overview (page 262)
- Pattern (page 262)

Overview

Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that uses *references* (page 254) between documents to describe relationships between connected data.

Pattern

Consider the following example that maps publisher and book relationships. The example illustrates the advantage of referencing over embedding to avoid repetition of the publisher information.

Embedding the publisher document inside the book document would lead to **repetition** of the publisher data, as the following documents show:

```
{
   title: "MongoDB: The Definitive Guide",
   author: [ "Kristina Chodorow", "Mike Dirolf" ],
   published_date: ISODate("2010-09-24"),
   pages: 216,
   language: "English",
   publisher: {
              name: "O'Reilly Media",
              founded: 1980,
              location: "CA"
            }
}

{
   title: "50 Tips and Tricks for MongoDB Developer",
   author: "Kristina Chodorow",
   published_date: ISODate("2011-05-06"),
   pages: 68,
   language: "English",
   publisher: {
              name: "O'Reilly Media",
              founded: 1980,
              location: "CA"
            }
}
```

To avoid repetition of the publisher data, use *references* and keep the publisher information in a separate collection from the book collection.

When using references, the growth of the relationships determine where to store the reference. If the number of books per publisher is small with limited growth, storing the book reference inside the publisher document may sometimes be useful. Otherwise, if the number of books per publisher is unbounded, this data model would lead to mutable, growing arrays, as in the following example:

```
{
   name: "O'Reilly Media",
   founded: 1980,
   location: "CA",
   books: [12346789, 234567890, ...]
}

{
    _id: 123456789,
    title: "MongoDB: The Definitive Guide",
    author: [ "Kristina Chodorow", "Mike Dirolf" ],
    published_date: ISODate("2010-09-24"),
    pages: 216,
    language: "English"
}

{
   _id: 234567890,
   title: "50 Tips and Tricks for MongoDB Developer",
   author: "Kristina Chodorow",
   published_date: ISODate("2011-05-06"),
   pages: 68,
   language: "English"
}
```

To avoid mutable, growing arrays, store the publisher reference inside the book document:

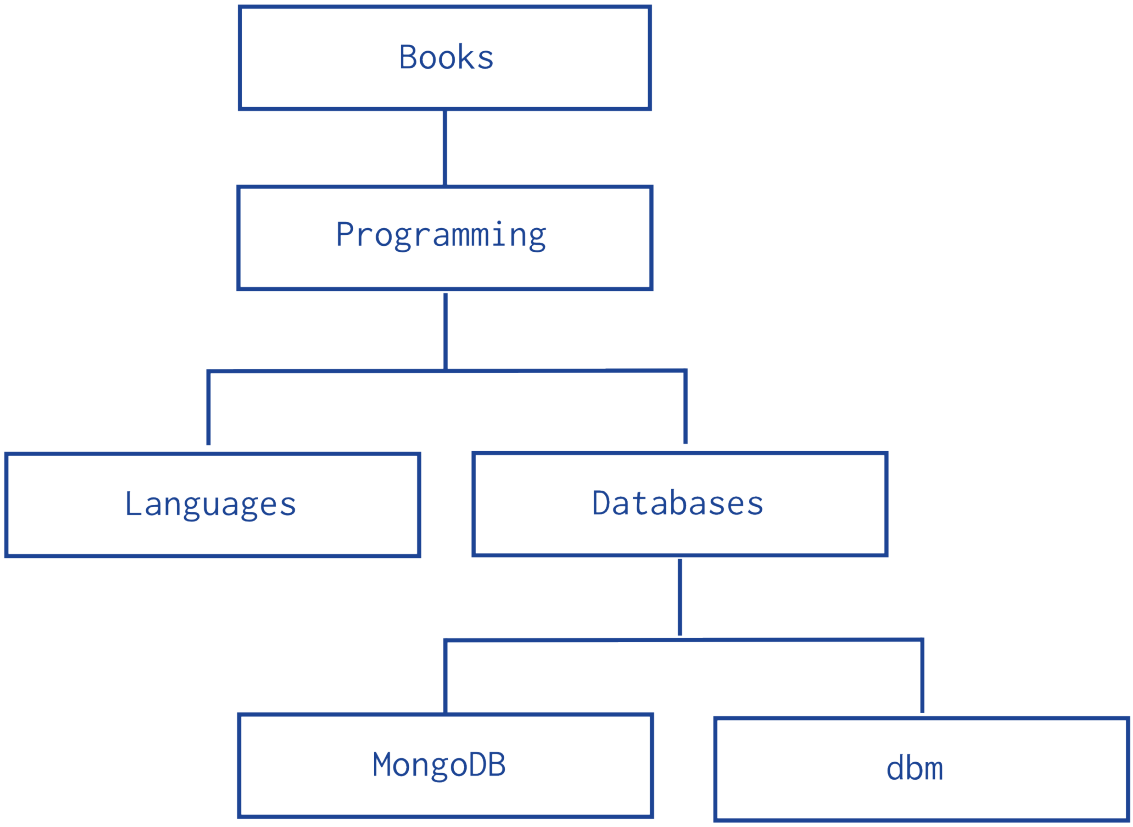
```
{
   _id: "oreilly",
   name: "O'Reilly Media",
   founded: 1980,
   location: "CA"
}

{
   _id: 123456789,
   title: "MongoDB: The Definitive Guide",
   author: [ "Kristina Chodorow", "Mike Dirolf" ],
   published_date: ISODate("2010-09-24"),
   pages: 216,
   language: "English",
   publisher_id: "oreilly"
}

{
   _id: 234567890,
   title: "50 Tips and Tricks for MongoDB Developer",
   author: "Kristina Chodorow",
   published_date: ISODate("2011-05-06"),
   pages: 68,
   language: "English",
   publisher_id: "oreilly"
}
```

7.4.2 Model Tree Structures

MongoDB allows various ways to use tree data structures to model large hierarchical or nested data relationships.



Model Tree Structures with Parent References **(page 264)** Presents a data model that organizes documents in a tree-like structure by storing *references* (page 254) to "parent" nodes in "child" nodes.

Model Tree Structures with Child References **(page 266)** Presents a data model that organizes documents in a tree-like structure by storing *references* (page 254) to "child" nodes in "parent" nodes.

Model Tree Structures with an Array of Ancestors **(page 267)** Presents a data model that organizes documents in a tree-like structure by storing *references* (page 254) to "parent" nodes and an array that stores all ancestors.

Model Tree Structures with Materialized Paths **(page 269)** Presents a data model that organizes documents in a tree-like structure by storing full relationship paths between documents. In addition to the tree node, each document stores the `_id` of the nodes ancestors or path as a string.

Model Tree Structures with Nested Sets **(page 270)** Presents a data model that organizes documents in a tree-like structure using the *Nested Sets* pattern. This optimizes discovering subtrees at the expense of tree mutability.

Model Tree Structures with Parent References

On this page

- Overview (page 265)
- Pattern (page 265)

Overview

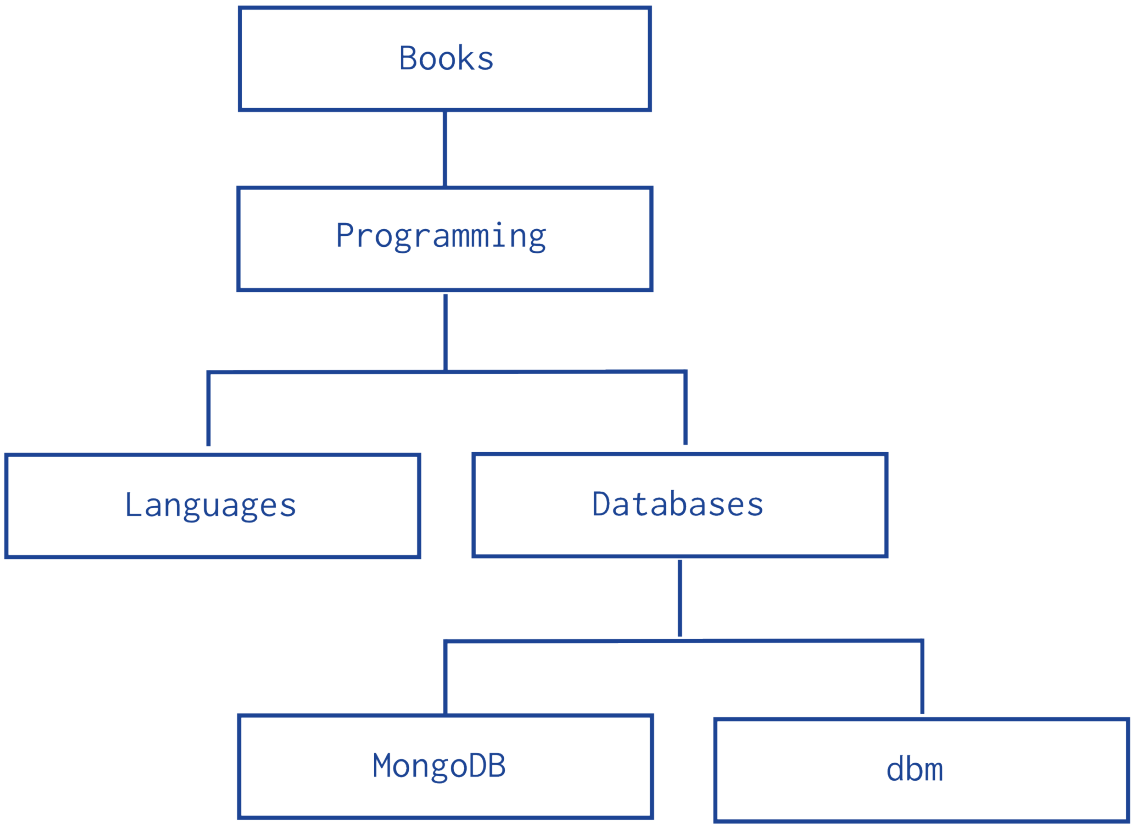
Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that describes a tree-like structure in MongoDB documents by storing *references* (page 254) to "parent" nodes in children nodes.

Pattern

The *Parent References* pattern stores each tree node in a document; in addition to the tree node, the document stores the id of the node's parent.

Consider the following hierarchy of categories:



The following example models the tree using *Parent References*, storing the reference to the parent category in the field `parent`:

```
db.categories.insert( { _id: "MongoDB", parent: "Databases" } )
db.categories.insert( { _id: "dbm", parent: "Databases" } )
db.categories.insert( { _id: "Databases", parent: "Programming" } )
db.categories.insert( { _id: "Languages", parent: "Programming" } )
db.categories.insert( { _id: "Programming", parent: "Books" } )
db.categories.insert( { _id: "Books", parent: null } )
```

- The query to retrieve the parent of a node is fast and straightforward:

```
db.categories.findOne( { _id: "MongoDB" } ).parent
```

- You can create an index on the field `parent` to enable fast search by the parent node:

```
db.categories.createIndex( { parent: 1 } )
```

- You can query by the `parent` field to find its immediate children nodes:

```
db.categories.find( { parent: "Databases" } )
```

The *Parent Links* pattern provides a simple solution to tree storage but requires multiple queries to retrieve subtrees.

Model Tree Structures with Child References

On this page

- Overview (page 266)
- Pattern (page 266)

Overview

Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that describes a tree-like structure in MongoDB documents by storing *references* (page 254) in the parent-nodes to children nodes.

Pattern

The *Child References* pattern stores each tree node in a document; in addition to the tree node, document stores in an array the id(s) of the node's children.

Consider the following hierarchy of categories:

The following example models the tree using *Child References*, storing the reference to the node's children in the field `children`:

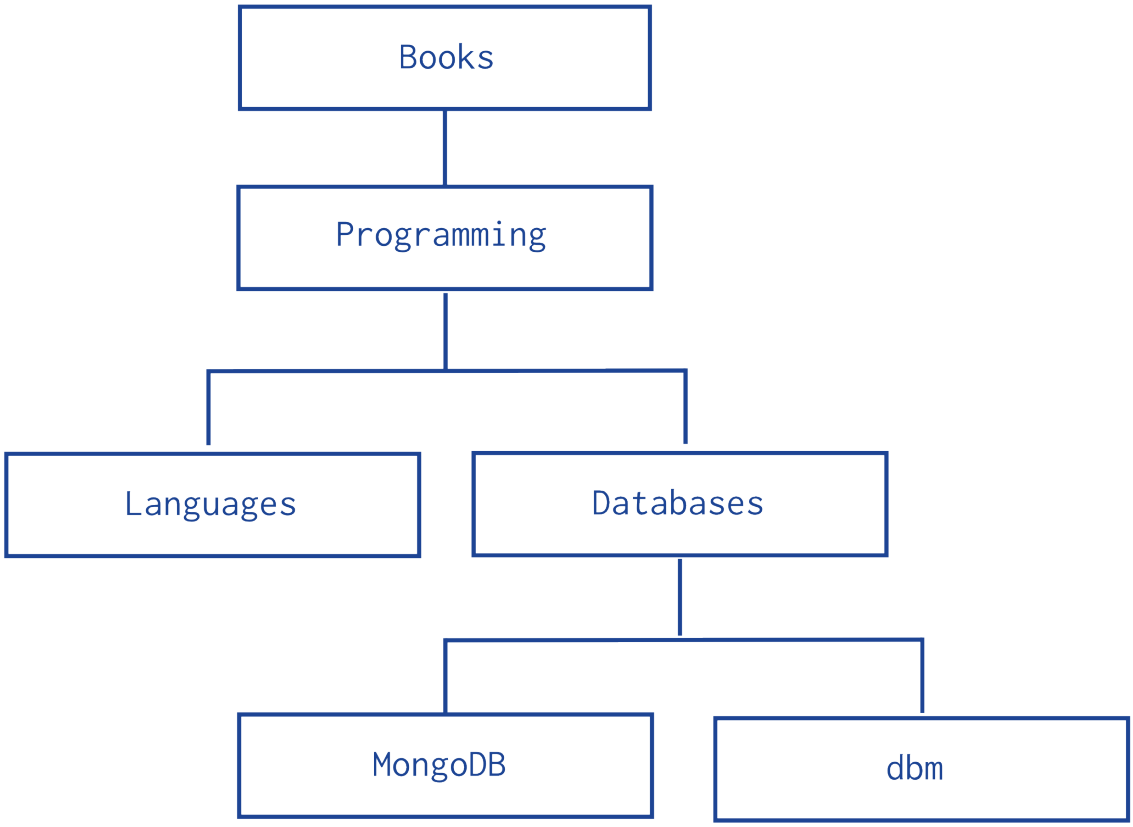
```
db.categories.insert( { _id: "MongoDB", children: [] } )
db.categories.insert( { _id: "dbm", children: [] } )
db.categories.insert( { _id: "Databases", children: [ "MongoDB", "dbm" ] } )
db.categories.insert( { _id: "Languages", children: [] } )
db.categories.insert( { _id: "Programming", children: [ "Databases", "Languages" ] } )
db.categories.insert( { _id: "Books", children: [ "Programming" ] } )
```

- The query to retrieve the immediate children of a node is fast and straightforward:

```
db.categories.findOne( { _id: "Databases" } ).children
```

- You can create an index on the field `children` to enable fast search by the child nodes:

```
db.categories.createIndex( { children: 1 } )
```



- You can query for a node in the `children` field to find its parent node as well as its siblings:

```
db.categories.find( { children: "MongoDB" } )
```

The *Child References* pattern provides a suitable solution to tree storage as long as no operations on subtrees are necessary. This pattern may also provide a suitable solution for storing graphs where a node may have multiple parents.

Model Tree Structures with an Array of Ancestors

On this page

- Overview (page 267)
- Pattern (page 268)

Overview

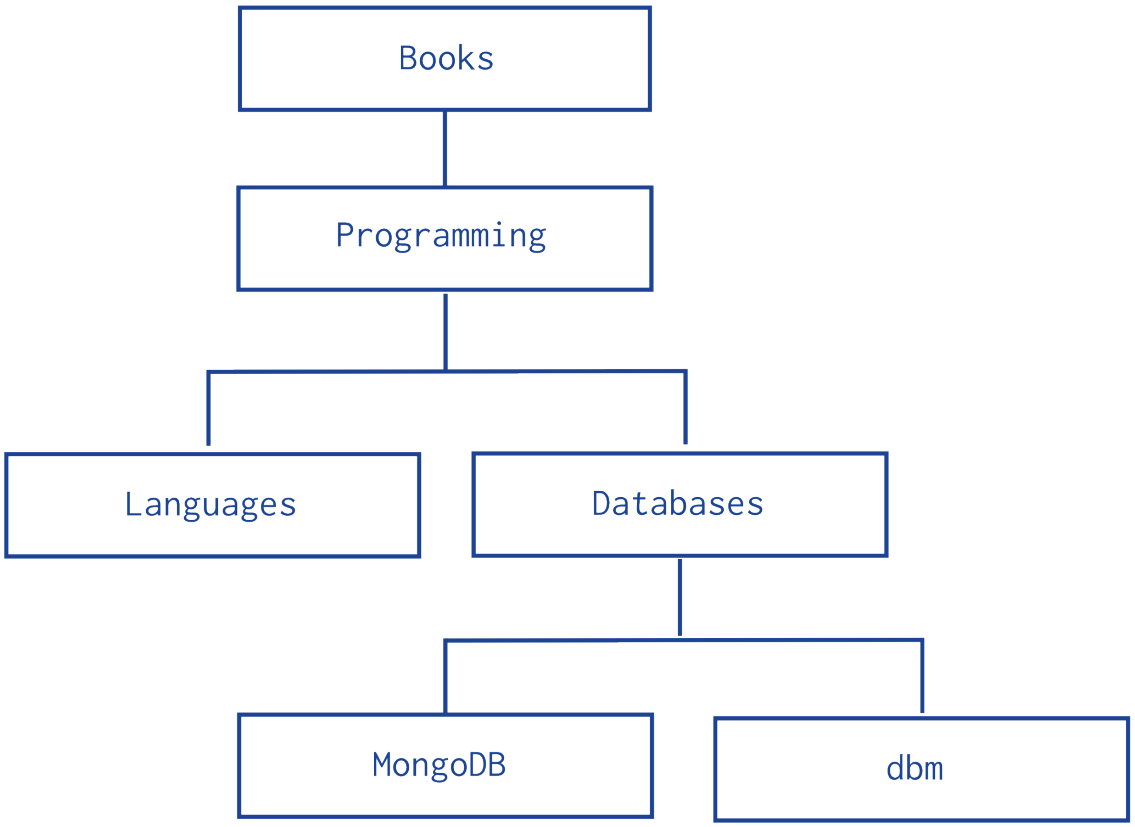
Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that describes a tree-like structure in MongoDB documents using *references* (page 254) to parent nodes and an array that stores all ancestors.

Pattern

The *Array of Ancestors* pattern stores each tree node in a document; in addition to the tree node, document stores in an array the id(s) of the node's ancestors or path.

Consider the following hierarchy of categories:



The following example models the tree using *Array of Ancestors*. In addition to the `ancestors` field, these documents also store the reference to the immediate parent category in the `parent` field:

```
db.categories.insert( { _id: "MongoDB", ancestors: [ "Books", "Programming", "Databases" ], parent: "
db.categories.insert( { _id: "dbm", ancestors: [ "Books", "Programming", "Databases" ], parent: "Data
db.categories.insert( { _id: "Databases", ancestors: [ "Books", "Programming" ], parent: "Programming
db.categories.insert( { _id: "Languages", ancestors: [ "Books", "Programming" ], parent: "Programming
db.categories.insert( { _id: "Programming", ancestors: [ "Books" ], parent: "Books" } )
db.categories.insert( { _id: "Books", ancestors: [ ], parent: null } )
```

- The query to retrieve the ancestors or path of a node is fast and straightforward:

  ```
  db.categories.findOne( { _id: "MongoDB" } ).ancestors
  ```

- You can create an index on the field `ancestors` to enable fast search by the ancestors nodes:

  ```
  db.categories.createIndex( { ancestors: 1 } )
  ```

- You can query by the field `ancestors` to find all its descendants:

```
db.categories.find( { ancestors: "Programming" } )
```

The *Array of Ancestors* pattern provides a fast and efficient solution to find the descendants and the ancestors of a node by creating an index on the elements of the ancestors field. This makes *Array of Ancestors* a good choice for working with subtrees.

The *Array of Ancestors* pattern is slightly slower than the *Materialized Paths* (page 269) pattern but is more straightforward to use.

Model Tree Structures with Materialized Paths

On this page

- Overview (page 269)
- Pattern (page 269)

Overview

Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that describes a tree-like structure in MongoDB documents by storing full relationship paths between documents.

Pattern

The *Materialized Paths* pattern stores each tree node in a document; in addition to the tree node, document stores as a string the id(s) of the node's ancestors or path. Although the *Materialized Paths* pattern requires additional steps of working with strings and regular expressions, the pattern also provides more flexibility in working with the path, such as finding nodes by partial paths.

Consider the following hierarchy of categories:

The following example models the tree using *Materialized Paths*, storing the path in the field `path`; the path string uses the comma `,` as a delimiter:

```
db.categories.insert( { _id: "Books", path: null } )
db.categories.insert( { _id: "Programming", path: ",Books," } )
db.categories.insert( { _id: "Databases", path: ",Books,Programming," } )
db.categories.insert( { _id: "Languages", path: ",Books,Programming," } )
db.categories.insert( { _id: "MongoDB", path: ",Books,Programming,Databases," } )
db.categories.insert( { _id: "dbm", path: ",Books,Programming,Databases," } )
```

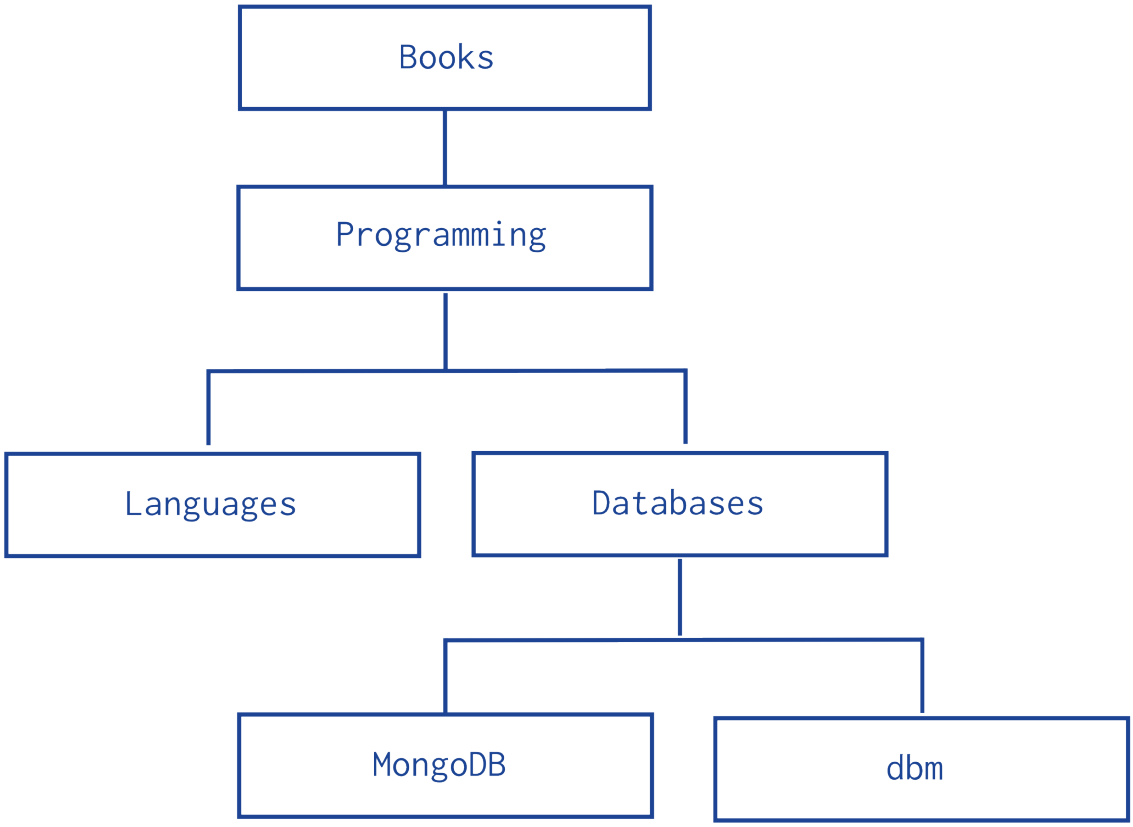
- You can query to retrieve the whole tree, sorting by the field `path`:

  ```
  db.categories.find().sort( { path: 1 } )
  ```

- You can use regular expressions on the `path` field to find the descendants of `Programming`:

  ```
  db.categories.find( { path: /,Programming,/ } )
  ```

- You can also retrieve the descendants of `Books` where the `Books` is also at the topmost level of the hierarchy:




```
db.categories.find( { path: /^,Books,/ } )
```

- To create an index on the field `path` use the following invocation:

  ```
  db.categories.createIndex( { path: 1 } )
  ```

  This index may improve performance depending on the query:

  – For queries from the root `Books` sub-tree (e.g. `https://docs.mongodb.org/manual/^,Books,/` or `https://docs.mongodb.org/manual/^,Books,Programming,/`), an index on the `path` field improves the query performance significantly.

  – For queries of sub-trees where the path from the root is not provided in the query (e.g. `https://docs.mongodb.org/manual/,Databases,/`), or similar queries of sub-trees, where the node might be in the middle of the indexed string, the query must inspect the entire index.

  For these queries an index *may* provide some performance improvement *if* the index is significantly smaller than the entire collection.

Model Tree Structures with Nested Sets

On this page

- Overview (page 271)
- Pattern (page 271)

Overview

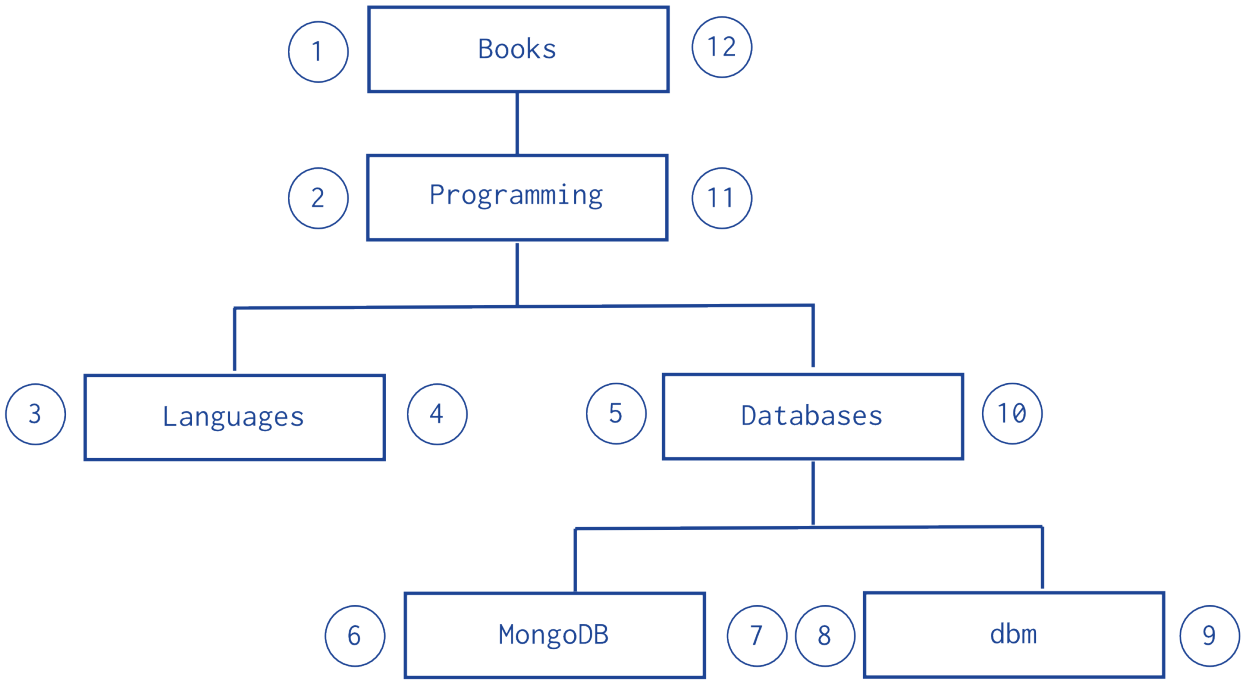
Data in MongoDB has a *flexible schema*. *Collections* do not enforce *document* structure. Decisions that affect how you model data can affect application performance and database capacity. See *Data Modeling Concepts* (page 252) for a full high level overview of data modeling in MongoDB.

This document describes a data model that describes a tree like structure that optimizes discovering subtrees at the expense of tree mutability.

Pattern

The *Nested Sets* pattern identifies each node in the tree as stops in a round-trip traversal of the tree. The application visits each node in the tree twice; first during the initial trip, and second during the return trip. The *Nested Sets* pattern stores each tree node in a document; in addition to the tree node, document stores the id of node's parent, the node's initial stop in the `left` field, and its return stop in the `right` field.

Consider the following hierarchy of categories:



The following example models the tree using *Nested Sets*:

```
db.categories.insert( { _id: "Books", parent: 0, left: 1, right: 12 } )
db.categories.insert( { _id: "Programming", parent: "Books", left: 2, right: 11 } )
db.categories.insert( { _id: "Languages", parent: "Programming", left: 3, right: 4 } )
db.categories.insert( { _id: "Databases", parent: "Programming", left: 5, right: 10 } )
db.categories.insert( { _id: "MongoDB", parent: "Databases", left: 6, right: 7 } )
db.categories.insert( { _id: "dbm", parent: "Databases", left: 8, right: 9 } )
```

You can query to retrieve the descendants of a node:

```
var databaseCategory = db.categories.findOne( { _id: "Databases" } );
db.categories.find( { left: { $gt: databaseCategory.left }, right: { $lt: databaseCategory.right } } )
```

The *Nested Sets* pattern provides a fast and efficient solution for finding subtrees but is inefficient for modifying the tree structure. As such, this pattern is best for static trees that do not change.

7.4.3 Model Specific Application Contexts

Model Data for Atomic Operations **(page 272)** Illustrates how embedding fields related to an atomic update within the same document ensures that the fields are in sync.

Model Data to Support Keyword Search **(page 273)** Describes one method for supporting keyword search by storing keywords in an array in the same document as the text field. Combined with a multi-key index, this pattern can support application's keyword search operations.

Model Monetary Data **(page 274)** Describes two methods to model monetary data in MongoDB.

Model Time Data **(page 276)** Describes how to deal with local time in MongoDB.

Model Data for Atomic Operations

On this page

- Pattern (page 272)

Pattern

In MongoDB, write operations, e.g. `db.collection.update()`, `db.collection.findAndModify()`, `db.collection.remove()`, are atomic on the level of a single document. For fields that must be updated together, embedding the fields within the same document ensures that the fields can be updated atomically.

For example, consider a situation where you need to maintain information on books, including the number of copies available for checkout as well as the current checkout information.

The available copies of the book and the checkout information should be in sync. As such, embedding the `available` field and the `checkout` field within the same document ensures that you can update the two fields atomically.

```
{
    _id: 123456789,
    title: "MongoDB: The Definitive Guide",
    author: [ "Kristina Chodorow", "Mike Dirolf" ],
    published_date: ISODate("2010-09-24"),
    pages: 216,
    language: "English",
    publisher_id: "oreilly",
    available: 3,
    checkout: [ { by: "joe", date: ISODate("2012-10-15") } ]
}
```

Then to update with new checkout information, you can use the `db.collection.update()` method to atomically update both the `available` field and the `checkout` field:

```
db.books.update (
   { _id: 123456789, available: { $gt: 0 } },
   {
     $inc: { available: -1 },
```

```
      $push: { checkout: { by: "abc", date: new Date() } }
   }
)
```

The operation returns a `WriteResult()` object that contains information on the status of the operation:

```
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
```

The `nMatched` field shows that `1` document matched the update condition, and `nModified` shows that the operation updated `1` document.

If no document matched the update condition, then `nMatched` and `nModified` would be `0` and would indicate that you could not check out the book.

Model Data to Support Keyword Search

On this page

- Pattern (page 273)
- Limitations of Keyword Indexes (page 274)

Note: Keyword search is *not* the same as text search or full text search, and does not provide stemming or other text-processing features. See the *Limitations of Keyword Indexes* (page 274) section for more information.

In 2.4, MongoDB provides a text search feature. See *Text Indexes* (page 533) for more information.

If your application needs to perform queries on the content of a field that holds text you can perform exact matches on the text or use `$regex` to use regular expression pattern matches. However, for many operations on text, these methods do not satisfy application requirements.

This pattern describes one method for supporting keyword search using MongoDB to support application search functionality, that uses keywords stored in an array in the same document as the text field. Combined with a *multi-key index* (page 525), this pattern can support application's keyword search operations.

Pattern

To add structures to your document to support keyword-based queries, create an array field in your documents and add the keywords as strings in the array. You can then create a *multi-key index* (page 525) on the array and create queries that select values from the array.

Example

Given a collection of library volumes that you want to provide topic-based search. For each volume, you add the array `topics`, and you add as many keywords as needed for a given volume.

For the `Moby-Dick` volume you might have the following document:

```
{ title : "Moby-Dick" ,
  author : "Herman Melville" ,
  published : 1851 ,
  ISBN : 0451526996 ,
  topics : [ "whaling" , "allegory" , "revenge" , "American" ,
    "novel" , "nautical" , "voyage" , "Cape Cod" ]
}
```

You then create a multi-key index on the `topics` array:

```
db.volumes.createIndex( { topics: 1 } )
```

The multi-key index creates separate index entries for each keyword in the `topics` array. For example the index contains one entry for `whaling` and another for `allegory`.

You then query based on the keywords. For example:

```
db.volumes.findOne( { topics : "voyage" }, { title: 1 } )
```

Note: An array with a large number of elements, such as one with several hundreds or thousands of keywords will incur greater indexing costs on insertion.

Limitations of Keyword Indexes

MongoDB can support keyword searches using specific data models and *multi-key indexes* (page 525); however, these keyword indexes are not sufficient or comparable to full-text products in the following respects:

- *Stemming*. Keyword queries in MongoDB can not parse keywords for root or related words.
- *Synonyms*. Keyword-based search features must provide support for synonym or related queries in the application layer.
- *Ranking*. The keyword look ups described in this document do not provide a way to weight results.
- *Asynchronous Indexing*. MongoDB builds indexes synchronously, which means that the indexes used for keyword indexes are always current and can operate in real-time. However, asynchronous bulk indexes may be more efficient for some kinds of content and workloads.

Model Monetary Data

On this page

- Overview (page 274)
- Use Cases for Exact Precision Model (page 275)
- Use Cases for Arbitrary Precision Model (page 275)
- Exact Precision (page 275)
- Arbitrary Precision (page 275)

Overview

MongoDB stores numeric data as either IEEE 754 standard 64-bit floating point numbers or as 32-bit or 64-bit signed integers. Applications that handle monetary data often require capturing fractional units of currency. However, arithmetic on floating point numbers, as implemented in modern hardware, often does not conform to requirements for monetary arithmetic. In addition, some fractional numeric quantities, such as one third and one tenth, have no exact representation in binary floating point numbers.

Note: Arithmetic mentioned on this page refers to server-side arithmetic performed by `mongod` or `mongos`, and not to client-side arithmetic.

This document describes two ways to model monetary data in MongoDB:

- *Exact Precision* (page 275) which multiplies the monetary value by a power of 10.
- *Arbitrary Precision* (page 275) which uses two fields for the value: one field to store the exact monetary value as a non-numeric and another field to store a floating point approximation of the value.

Use Cases for Exact Precision Model

If you regularly need to perform server-side arithmetic on monetary data, the exact precision model may be appropriate. For instance:

- If you need to query the database for exact, mathematically valid matches, use *Exact Precision* (page 275).
- If you need to be able to do server-side arithmetic, e.g., `$inc`, `$mul`, and `aggregation framework arithmetic`, use *Exact Precision* (page 275).

Use Cases for Arbitrary Precision Model

If there is no need to perform server-side arithmetic on monetary data, modeling monetary data using the arbitrary precision model may be suitable. For instance:

- If you need to handle arbitrary or unforeseen number of precision, see *Arbitrary Precision* (page 275).
- If server-side approximations are sufficient, possibly with client-side post-processing, see *Arbitrary Precision* (page 275).

Exact Precision

To model monetary data using the exact precision model:

1. Determine the maximum precision needed for the monetary value. For example, your application may require precision down to the tenth of one cent for monetary values in `USD` currency.
2. Convert the monetary value into an integer by multiplying the value by a power of 10 that ensures the maximum precision needed becomes the least significant digit of the integer. For example, if the required maximum precision is the tenth of one cent, multiply the monetary value by 1000.
3. Store the converted monetary value.

For example, the following scales `9.99 USD` by 1000 to preserve precision up to one tenth of a cent.

```
{ price: 9990, currency: "USD" }
```

The model assumes that for a given currency value:

- The scale factor is consistent for a currency; i.e. same scaling factor for a given currency.
- The scale factor is a constant and known property of the currency; i.e applications can determine the scale factor from the currency.

When using this model, applications must be consistent in performing the appropriate scaling of the values.

For use cases of this model, see *Use Cases for Exact Precision Model* (page 275).

Arbitrary Precision

To model monetary data using the arbitrary precision model, store the value in two fields:

1. In one field, encode the exact monetary value as a non-numeric data type; e.g., `BinData` or a `string`.

2. In the second field, store a double-precision floating point approximation of the exact value.

The following example uses the arbitrary precision model to store `9.99 USD` for the price and `0.25 USD` for the fee:

```
{
  price: { display: "9.99", approx: 9.9900000000000002, currency: "USD" },
  fee: { display: "0.25", approx: 0.2499999999999999, currency: "USD" }
}
```

With some care, applications can perform range and sort queries on the field with the numeric approximation. However, the use of the approximation field for the query and sort operations requires that applications perform client-side post-processing to decode the non-numeric representation of the exact value and then filter out the returned documents based on the exact monetary value.

For use cases of this model, see *Use Cases for Arbitrary Precision Model* (page 275).

Model Time Data

On this page

- Overview (page 276)
- Example (page 276)

Overview

MongoDB *stores times in UTC* (page 15) by default, and will convert any local time representations into this form. Applications that must operate or report on some unmodified local time value may store the time zone alongside the UTC timestamp, and compute the original local time in their application logic.

Example

In the MongoDB shell, you can store both the current date and the current client's offset from UTC.

```
var now = new Date();
db.data.save( { date: now,
                offset: now.getTimezoneOffset() } );
```

You can reconstruct the original local time by applying the saved offset:

```
var record = db.data.findOne();
var localNow = new Date( record.date.getTime() -  ( record.offset * 60000 ) );
```

7.5 Data Model Reference

Database References **(page 277)** Discusses manual references and DBRefs, which MongoDB can use to represent relationships between documents.

7.5.1 Database References

On this page

- Manual References (page 277)
- DBRefs (page 278)

MongoDB does not support joins. In MongoDB some data is *denormalized*, or stored with related data in *documents* to remove the need for joins. However, in some cases it makes sense to store related information in separate documents, typically in different collections or databases.

MongoDB applications use one of two methods for relating documents:

- *Manual references* (page 277) where you save the `_id` field of one document in another document as a reference. Then your application can run a second query to return the related data. These references are simple and sufficient for most use cases.
- *DBRefs* (page 278) are references from one document to another using the value of the first document's `_id` field, collection name, and, optionally, its database name. By including these names, DBRefs allow documents located in multiple collections to be more easily linked with documents from a single collection.

  To resolve DBRefs, your application must perform additional queries to return the referenced documents. Many `drivers` have helper methods that form the query for the DBRef automatically. The drivers [9] do not *automatically* resolve DBRefs into documents.

  DBRefs provide a common format and type to represent relationships among documents. The DBRef format also provides common semantics for representing links between documents if your database must interact with multiple frameworks and tools.

Unless you have a compelling reason to use DBRefs, use manual references instead.

Manual References

Background

Using manual references is the practice of including one *document's* `_id` field in another document. The application can then issue a second query to resolve the referenced fields as needed.

Process

Consider the following operation to insert two documents, using the `_id` field of the first document as a reference in the second document:

```
original_id = ObjectId()

db.places.insert({
    "_id": original_id,
    "name": "Broadway Center",
    "url": "bc.example.net"
})

db.people.insert({
    "name": "Erin",
```

[9] Some community supported drivers may have alternate behavior and may resolve a DBRef into a document automatically.

```
    "places_id": original_id,
    "url":  "bc.example.net/Erin"
})
```

Then, when a query returns the document from the `people` collection you can, if needed, make a second query for the document referenced by the `places_id` field in the `places` collection.

Use

For nearly every case where you want to store a relationship between two documents, use *manual references* (page 277). The references are simple to create and your application can resolve references as needed.

The only limitation of manual linking is that these references do not convey the database and collection names. If you have documents in a single collection that relate to documents in more than one collection, you may need to consider using DBRefs.

DBRefs

Background

DBRefs are a convention for representing a *document*, rather than a specific reference type. They include the name of the collection, and in some cases the database name, in addition to the value from the `_id` field.

Format

DBRefs have the following fields:

`$ref`
: The `$ref` field holds the name of the collection where the referenced document resides.

`$id`
: The `$id` field contains the value of the `_id` field in the referenced document.

`$db`
: *Optional.*

  Contains the name of the database where the referenced document resides.

  Only some drivers support `$db` references.

Example

DBRef documents resemble the following document:

```
{ "$ref" : <value>, "$id" : <value>, "$db" : <value> }
```

Consider a document from a collection that stored a DBRef in a `creator` field:

```
{
  "_id" : ObjectId("5126bbf64aed4daf9e2ab771"),
  // .. application fields
  "creator" : {
                  "$ref" : "creators",
                  "$id" : ObjectId("5126bc054aed4daf9e2ab772"),
                  "$db" : "users"
```

```
                }
}
```

The DBRef in this example points to a document in the `creators` collection of the `users` database that has `ObjectId("5126bc054aed4daf9e2ab772")` in its `_id` field.

Note: The order of fields in the DBRef matters, and you must use the above sequence when using a DBRef.

Driver Support for DBRefs

C	The C driver contains no support for DBRefs. You can traverse references manually.
C++	The C++ driver contains no support for DBRefs. You can traverse references manually.
C#	The C# driver supports DBRefs using the MongoDBRef[10] class and `FetchDBRef` and `FetchDBRefAs` methods.
Haskell	The Haskell driver contains no support for DBRefs. You can traverse references manually.
Java	The DBRef[11] class provides support for DBRefs from Java.
JavaScript	The `mongo` shell's `JavaScript` interface provides a DBRef.
Node.js	The Node.js driver supports DBRefs using the DBRef[12] class and the dereference[13] method.
Perl	The Perl driver supports DBRefs using the MongoDB::DBRef[14] class. You can traverse references manually.
PHP	The PHP driver supports DBRefs, including the optional `$db` reference, using the MongoDBRef[15] class.
Python	The Python driver supports DBRefs using the DBRef[16] class and the dereference[17] method.
Ruby	The Ruby driver supports DBRefs using the DBRef[18] class and the dereference[19] method.
Scala	The Scala driver contains no support for DBRefs. You can traverse references manually.

Use

In most cases you should use the *manual reference* (page 277) method for connecting two or more related documents. However, if you need to reference documents from multiple collections, consider using DBRefs.

[10] https://api.mongodb.org/csharp/current/html/T_MongoDB_Driver_MongoDBRef.htm
[11] https://api.mongodb.org/java/current/com/mongodb/DBRef.html
[12] http://mongodb.github.io/node-mongodb-native/api-bson-generated/db_ref.html
[13] http://mongodb.github.io/node-mongodb-native/api-generated/db.html#dereference
[14] https://metacpan.org/pod/MongoDB::DBRef
[15] http://www.php.net/manual/en/class.mongodbref.php/
[16] https://api.mongodb.org/python/current/api/bson/dbref.html
[17] https://api.mongodb.org/python/current/api/pymongo/database.html#pymongo.database.Database.deref eren ce
[18] https://api.mongodb.org/ruby/current/BSON/DBRef.html
[19] https://api.mongodb.org/ruby/current/Mongo/DB.html#dereference-instance_method

CHAPTER 8

Administration

The administration documentation addresses the ongoing operation and maintenance of MongoDB instances and deployments. This documentation includes both high level overviews of these concerns as well as tutorials that cover specific procedures and processes for operating MongoDB.

Administration Concepts **(page 281)** Core conceptual documentation of operational practices for managing MongoDB deployments and systems.

> ***MongoDB Backup Methods*** **(page 282)** Describes approaches and considerations for backing up a MongoDB database.
>
> ***Monitoring for MongoDB*** **(page 285)** An overview of monitoring tools, diagnostic strategies, and approaches to monitoring replica sets and sharded clusters.
>
> ***Production Notes*** **(page 296)** A collection of notes that describe best practices and considerations for the operations of MongoDB instances and deployments.
>
> Continue reading from *Administration Concepts* (page 281) for additional documentation of MongoDB administration.

Administration Tutorials **(page 318)** Tutorials that describe common administrative procedures and practices for operations for MongoDB instances and deployments.

> ***Configuration, Maintenance, and Analysis*** **(page 318)** Describes routine management operations, including configuration and performance analysis.
>
> ***Backup and Recovery*** **(page 343)** Outlines procedures for data backup and restoration with `mongod` instances and deployments.
>
> Continue reading from *Administration Tutorials* (page 318) for more tutorials of common MongoDB maintenance operations.

Administration Reference **(page 372)** Reference and documentation of internal mechanics of administrative features, systems and functions and operations.

See also:

The MongoDB Manual contains administrative documentation and tutorials though out several sections. See *Replica Set Tutorials* (page 665) and *Sharded Cluster Tutorials* (page 764) for additional tutorials and information.

8.1 Administration Concepts

The core administration documents address strategies and practices used in the operation of MongoDB systems and deployments.

Operational Strategies **(page 282)** Higher level documentation of key concepts for the operation and maintenance of MongoDB deployments.

MongoDB Backup Methods **(page 282)** Describes approaches and considerations for backing up a MongoDB database.

Monitoring for MongoDB **(page 285)** An overview of monitoring tools, diagnostic strategies, and approaches to monitoring replica sets and sharded clusters.

Run-time Database Configuration **(page 291)** Outlines common MongoDB configurations and examples of best-practice configurations for common use cases.

Continue reading from *Operational Strategies* (page 282) for additional documentation.

Data Management **(page 308)** Core documentation that addresses issues in data management, organization, maintenance, and lifecycle management.

Data Center Awareness **(page 308)** Presents the MongoDB features that allow application developers and database administrators to configure their deployments to be more data center aware or allow operational and location-based separation.

Optimization Strategies for MongoDB **(page 310)** Techniques for optimizing application performance with MongoDB.

Continue reading from *Optimization Strategies for MongoDB* (page 310) for additional documentation.

8.1.1 Operational Strategies

These documents address higher level strategies for common administrative tasks and requirements with respect to MongoDB deployments.

MongoDB Backup Methods **(page 282)** Describes approaches and considerations for backing up a MongoDB database.

Monitoring for MongoDB **(page 285)** An overview of monitoring tools, diagnostic strategies, and approaches to monitoring replica sets and sharded clusters.

Run-time Database Configuration **(page 291)** Outlines common MongoDB configurations and examples of best-practice configurations for common use cases.

Production Notes **(page 296)** A collection of notes that describe best practices and considerations for the operations of MongoDB instances and deployments.

MongoDB Backup Methods

On this page

- Backup by Copying Underlying Data Files (page 283)
- Backup with `mongodump` (page 283)
- MongoDB Cloud Manager Backup (page 284)
- Ops Manager Backup Software (page 284)
- Further Reading (page 285)
- Additional Resources (page 285)

When deploying MongoDB in production, you should have a strategy for capturing and restoring backups in the case of data loss events. There are several ways to back up MongoDB clusters:

- *Backup by Copying Underlying Data Files* (page 283)

- *Backup a Database with mongodump* (page 350)
- *MongoDB Cloud Manager Backup* (page 284)
- *Ops Manager Backup Software* (page 284)

Backup by Copying Underlying Data Files

You can create a backup by copying MongoDB's underlying data files.

If the volume where MongoDB stores data files supports point in time snapshots, you can use these snapshots to create backups of a MongoDB system at an exact moment in time.

File systems snapshots are an operating system volume manager feature, and are not specific to MongoDB. The mechanics of snapshots depend on the underlying storage system. For example, if you use Amazon's EBS storage system for EC2 supports snapshots. On Linux the LVM manager can create a snapshot.

To get a correct snapshot of a running `mongod` process, you must have journaling enabled and the journal must reside on the same logical volume as the other MongoDB data files. Without journaling enabled, there is no guarantee that the snapshot will be consistent or valid.

To get a consistent snapshot of a sharded system, you must disable the balancer and capture a snapshot from every shard and a config server at approximately the same moment in time.

If your storage system does not support snapshots, you can copy the files directly using `cp`, `rsync`, or a similar tool. Since copying multiple files is not an atomic operation, you must stop all writes to the `mongod` before copying the files. Otherwise, you will copy the files in an invalid state.

Backups produced by copying the underlying data do not support point in time recovery for replica sets and are difficult to manage for larger sharded clusters. Additionally, these backups are larger because they include the indexes and duplicate underlying storage padding and fragmentation. `mongodump`, by contrast, creates smaller backups.

For more information, see the *Backup and Restore with Filesystem Snapshots* (page 343) and *Backup a Sharded Cluster with Filesystem Snapshots* (page 356) for complete instructions on using LVM to create snapshots. Also see Back up and Restore Processes for MongoDB on Amazon EC2[1].

Backup with `mongodump`

The `mongodump` tool reads data from a MongoDB database and creates high fidelity BSON files. The `mongorestore` tool can populate a MongoDB database with the data from these BSON files.

Use Cases `mongodump` and `mongorestore` are simple and efficient for backing up small MongoDB deployments, for partial backup and restores based on a query, syncing from production to staging or development environments, or changing the storage engine of a standalone.

However, these tools can be problematic for capturing backups of larger systems, sharded clusters, or replica sets. For alternatives, see *MongoDB Cloud Manager Backup* (page 284) or *Ops Manager Backup Software* (page 284).

Data Exclusion `mongodump` excludes the content of the `local` database in its output.

`mongodump` only captures the documents in the database in its backup data and does not include index data. `mongorestore` or `mongod` must then rebuild the indexes after restoring data.

[1] https://docs.mongodb.org/ecosystem/tutorial/backup-and-restore-mongodb-on-amazon-ec2

Data Compression Handling When run against a `mongod` instance that uses the *WiredTiger* (page 595) storage engine, `mongodump` outputs uncompressed data.

Performance `mongodump` can adversely affect the performance of the `mongod`. If your data is larger than system memory, the `mongodump` will push the working set out of memory.

If applications modify data while `mongodump` is creating a backup, `mongodump` will compete for resources with those applications.

To mitigate the impact of `mongodump` on the performance of the replica set, use `mongodump` to capture backups from a *secondary* (page 628) member of a replica set.

Applications can continue to modify data while `mongodump` captures the output. For replica sets, `mongodump` provides the `--oplog` option to include in its output *oplog* entries that occur during the `mongodump` operation. This allows the corresponding `mongorestore` operation to replay the captured oplog. To restore a backup created with `--oplog`, use `mongorestore` with the `--oplogReplay` option.

However, for replica sets, consider *MongoDB Cloud Manager Backup* (page 284) or *Ops Manager Backup Software* (page 284).

See *Back Up and Restore with MongoDB Tools* (page 349), *Backup a Small Sharded Cluster with mongodump* (page 355), and *Backup a Sharded Cluster with Database Dumps* (page 359) for more information.

MongoDB Cloud Manager Backup

The MongoDB Cloud Manager[2] supports the backing up and restoring of MongoDB deployments.

MongoDB Cloud Manager continually backs up MongoDB replica sets and sharded clusters by reading the oplog data from your MongoDB deployment.

MongoDB Cloud Manager Backup offers point in time recovery of MongoDB replica sets and a consistent snapshot of sharded clusters.

MongoDB Cloud Manager achieves point in time recovery by storing oplog data so that it can create a restore for any moment in time in the last 24 hours for a particular replica set or sharded cluster. Sharded cluster snapshots are difficult to achieve with other MongoDB backup methods.

To restore a MongoDB deployment from an MongoDB Cloud Manager Backup snapshot, you download a compressed archive of your MongoDB data files and distribute those files before restarting the `mongod` processes.

To get started with MongoDB Cloud Manager Backup, sign up for MongoDB Cloud Manager[3]. For documentation on MongoDB Cloud Manager, see the MongoDB Cloud Manager documentation[4].

Ops Manager Backup Software

MongoDB Subscribers can install and run the same core software that powers *MongoDB Cloud Manager Backup* (page 284) on their own infrastructure. Ops Manager, an on-premise solution, has similar functionality to the cloud version and is available with Enterprise Advanced subscriptions.

For more information about Ops Manager, see the MongoDB Enterprise Advanced[5] page and the Ops Manager Manual[6].

[2] https://cloud.mongodb.com/?jmp=docs
[3] https://cloud.mongodb.com/?jmp=docs
[4] https://docs.cloud.mongodb.com/
[5] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[6] https://docs.opsmanager.mongodb.com/current/

Further Reading

Backup and Restore with Filesystem Snapshots **(page 343)** An outline of procedures for creating MongoDB data set backups using system-level file snapshot tool, such as *LVM* or native storage appliance tools.

Restore a Replica Set from MongoDB Backups **(page 348)** Describes procedure for restoring a replica set from an archived backup such as a `mongodump` or MongoDB Cloud Manager[7] Backup file.

Back Up and Restore with MongoDB Tools **(page 349)** Describes a procedure for exporting the contents of a database to either a binary dump or a textual exchange format, and for importing these files into a database.

Backup and Restore Sharded Clusters **(page 355)** Detailed procedures and considerations for backing up sharded clusters and single shards.

Recover Data after an Unexpected Shutdown **(page 366)** Recover data from MongoDB data files that were not properly closed or have an invalid state.

Additional Resources

- Backup and it's Role in Disaster Recovery White Paper[8]
- Backup vs. Replication: Why Do You Need Both?[9]
- MongoDB Production Readiness Consulting Package[10]

Monitoring for MongoDB

On this page

- Monitoring Strategies (page 286)
- MongoDB Reporting Tools (page 286)
- Process Logging (page 288)
- Diagnosing Performance Issues (page 289)
- Replication and Monitoring (page 289)
- Sharding and Monitoring (page 290)
- Additional Resources (page 291)

Monitoring is a critical component of all database administration. A firm grasp of MongoDB's reporting will allow you to assess the state of your database and maintain your deployment without crisis. Additionally, a sense of MongoDB's normal operational parameters will allow you to diagnose problems before they escalate to failures.

This document presents an overview of the available monitoring utilities and the reporting statistics available in MongoDB. It also introduces diagnostic strategies and suggestions for monitoring replica sets and sharded clusters.

Note: MongoDB Cloud Manager[11], a hosted service, and Ops Manager[12], an on-premise solution, provide monitoring, backup, and automation of MongoDB instances. See the MongoDB Cloud Manager documentation[13] and Ops Manager documentation[14] for more information.

[7] https://cloud.mongodb.com/?jmp=docs
[8] https://www.mongodb.com/lp/white-paper/backup-disaster-recovery?jmp=docs
[9] http://www.mongodb.com/blog/post/backup-vs-replication-why-do-you-need-both?jmp=docs
[10] https://www.mongodb.com/products/consulting?jmp=docs#s_product_readiness
[11] https://cloud.mongodb.com/?jmp=docs
[12] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[13] https://docs.cloud.mongodb.com/
[14] https://docs.opsmanager.mongodb.com?jmp=docs

Monitoring Strategies

There are three methods for collecting data about the state of a running MongoDB instance:

- First, there is a set of utilities distributed with MongoDB that provides real-time reporting of database activities.
- Second, `database commands` return statistics regarding the current database state with greater fidelity.
- Third, MongoDB Cloud Manager[15], a hosted service, and Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[16], provide monitoring to collect data from running MongoDB deployments as well as providing visualization and alerts based on that data.

Each strategy can help answer different questions and is useful in different contexts. These methods are complementary.

MongoDB Reporting Tools

This section provides an overview of the reporting methods distributed with MongoDB. It also offers examples of the kinds of questions that each method is best suited to help you address.

Utilities The MongoDB distribution includes a number of utilities that quickly return statistics about instances' performance and activity. Typically, these are most useful for diagnosing issues and assessing normal operation.

`mongostat` `mongostat` captures and returns the counts of database operations by type (e.g. insert, query, update, delete, etc.). These counts report on the load distribution on the server.

Use `mongostat` to understand the distribution of operation types and to inform capacity planning. See the `mongostat manual` for details.

`mongotop` `mongotop` tracks and reports the current read and write activity of a MongoDB instance, and reports these statistics on a per collection basis.

Use `mongotop` to check if your database activity and use match your expectations. See the `mongotop manual` for details.

HTTP Console Deprecated since version 3.2: HTTP interface for MongoDB

MongoDB provides a web interface that exposes diagnostic and monitoring information in a simple web page. The web interface is accessible at `localhost:<port>`, where the `<port>` number is **1000** more than the `mongod` port .

For example, if a locally running `mongod` is using the default port `27017`, access the HTTP console at `http://localhost:28017`.

Commands MongoDB includes a number of commands that report on the state of the database.

These data may provide a finer level of granularity than the utilities discussed above. Consider using their output in scripts and programs to develop custom alerts, or to modify the behavior of your application in response to the activity of your instance. The `db.currentOp` method is another useful tool for identifying the database instance's in-progress operations.

[15] https://cloud.mongodb.com/?jmp=docs
[16] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

serverStatus The `serverStatus` command, or `db.serverStatus()` from the shell, returns a general overview of the status of the database, detailing disk usage, memory use, connection, journaling, and index access. The command returns quickly and does not impact MongoDB performance.

`serverStatus` outputs an account of the state of a MongoDB instance. This command is rarely run directly. In most cases, the data is more meaningful when aggregated, as one would see with monitoring tools including MongoDB Cloud Manager[17] and Ops Manager[18]. Nevertheless, all administrators should be familiar with the data provided by `serverStatus`.

dbStats The `dbStats` command, or `db.stats()` from the shell, returns a document that addresses storage use and data volumes. The `dbStats` reflect the amount of storage used, the quantity of data contained in the database, and object, collection, and index counters.

Use this data to monitor the state and storage capacity of a specific database. This output also allows you to compare use between databases and to determine the average *document* size in a database.

collStats The `collStats` or `db.collection.stats()` from the shell that provides statistics that resemble `dbStats` on the collection level, including a count of the objects in the collection, the size of the collection, the amount of disk space used by the collection, and information about its indexes.

replSetGetStatus The `replSetGetStatus` command (`rs.status()` from the shell) returns an overview of your replica set's status. The `replSetGetStatus` document details the state and configuration of the replica set and statistics about its members.

Use this data to ensure that replication is properly configured, and to check the connections between the current host and the other members of the replica set.

Third Party Tools A number of third party monitoring tools have support for MongoDB, either directly, or through their own plugins.

Self Hosted Monitoring Tools These are monitoring tools that you must install, configure and maintain on your own servers. Most are open source.

Tool	Plugin	Description
Ganglia[19]	mongodb-ganglia[20]	Python script to report operations per second, memory usage, btree statistics, master/slave status and current connections.
Ganglia	gmond_python_modules[21]	Parses output from the `serverStatus` and `replSetGetStatus` commands.
Motop[22]	*None*	Realtime monitoring tool for MongoDB servers. Shows current operations ordered by durations every second.
mtop[23]	*None*	A top like tool.
Munin[24]	mongo-munin[25]	Retrieves server statistics.
Munin	mongomon[26]	Retrieves collection statistics (sizes, index sizes, and each (configured) collection count for one DB).
Munin	munin-plugins Ubuntu PPA[27]	Some additional munin plugins not in the main distribution.
Nagios[28]	nagios-plugin-mongodb[29]	A simple Nagios check script, written in Python.

[17] https://cloud.mongodb.com/?jmp=docs
[18] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[19] http://sourceforge.net/apps/trac/ganglia/wiki

Also consider dex[30], an index and query analyzing tool for MongoDB that compares MongoDB log files and indexes to make indexing recommendations.

See also:

Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[31].

Hosted (SaaS) Monitoring Tools These are monitoring tools provided as a hosted service, usually through a paid subscription.

Name	Notes
MongoDB Cloud Manager[32]	MongoDB Cloud Manager is a cloud-based suite of services for managing MongoDB deployments. MongoDB Cloud Manager provides monitoring, backup, and automation functionality. For an on-premise solution, see also Ops Manager, available in MongoDB Enterprise Advanced[33].
Scout[34]	Several plugins, including MongoDB Monitoring[35], MongoDB Slow Queries[36], and MongoDB Replica Set Monitoring[37].
Server Density[38]	Dashboard for MongoDB[39], MongoDB specific alerts, replication failover timeline and iPhone, iPad and Android mobile apps.
Application Performance Management[40]	IBM has an Application Performance Management SaaS offering that includes monitor for MongoDB and other applications and middleware.
New Relic[41]	New Relic offers full support for application performance management. In addition, New Relic Plugins and Insights enable you to view monitoring metrics from Cloud Manager in New Relic.
Datadog[42]	Infrastructure monitoring[43] to visualize the performance of your MongoDB deployments.

Process Logging

During normal operation, `mongod` and `mongos` instances report a live account of all server activity and operations to either standard output or a log file. The following runtime settings control these options.

- `quiet`. Limits the amount of information written to the log or output.

[20] https://github.com/quiiver/mongodb-ganglia
[21] https://github.com/ganglia/gmond_python_modules
[22] https://github.com/tart/motop
[23] https://github.com/beaufour/mtop
[24] http://munin-monitoring.org/
[25] https://github.com/erh/mongo-munin
[26] https://github.com/pcdummy/mongomon
[27] https://launchpad.net/ chris-lea/+archive/munin-plugins
[28] http://www.nagios.org/
[29] https://github.com/mzupan/nagios-plugin-mongodb
[30] https://github.com/mongolab/dex
[31] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[32] https://cloud.mongodb.com/?jmp=docs
[33] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[34] http://scoutapp.com
[35] https://scoutapp.com/plugin_urls/391-mongodb-monitoring
[36] http://scoutapp.com/plugin_urls/291-mongodb-slow-queries
[37] http://scoutapp.com/plugin_urls/2251-mongodb-replica-set-monitoring
[38] http://www.serverdensity.com
[39] http://www.serverdensity.com/mongodb-monitoring/
[40] http://ibmserviceengage.com
[41] http://newrelic.com/
[42] https://www.datadoghq.com/
[43] http://docs.datadoghq.com/integrations/mongodb/

- `verbosity`. Increases the amount of information written to the log or output. You can also modify the logging verbosity during runtime with the `logLevel` parameter or the `db.setLogLevel()` method in the shell.
- `path`. Enables logging to a file, rather than the standard output. You must specify the full path to the log file when adjusting this setting.
- `logAppend`. Adds information to a log file instead of overwriting the file.

Note: You can specify these configuration operations as the command line arguments to `mongod` or `mongos`

For example:

```
mongod -v --logpath /var/log/mongodb/server1.log --logappend
```

Starts a `mongod` instance in `verbose` mode, appending data to the log file at `/var/log/mongodb/server1.log/`.

The following *database commands* also affect logging:

- `getLog`. Displays recent messages from the `mongod` process log.
- `logRotate`. Rotates the log files for `mongod` processes only. See *Rotate Log Files* (page 330).

Diagnosing Performance Issues

As you develop and operate applications with MongoDB, you may want to analyze the performance of the database as the application. *Analyzing MongoDB Performance* (page 310) discusses some of the operational factors that can influence performance.

Replication and Monitoring

Beyond the basic monitoring requirements for any MongoDB instance, for replica sets, administrators must monitor *replication lag*. "Replication lag" refers to the amount of time that it takes to copy (i.e. replicate) a write operation on the *primary* to a *secondary*. Some small delay period may be acceptable, but two significant problems emerge as replication lag grows:

- First, operations that occurred during the period of lag are not replicated to one or more secondaries. If you're using replication to ensure data persistence, exceptionally long delays may impact the integrity of your data set.
- Second, if the replication lag exceeds the length of the operation log (*oplog*) then MongoDB will have to perform an initial sync on the secondary, copying all data from the *primary* and rebuilding all indexes. This is uncommon under normal circumstances, but if you configure the oplog to be smaller than the default, the issue can arise.

  Note: The size of the oplog is only configurable during the first run using the `--oplogSize` argument to the `mongod` command, or preferably, the `oplogSizeMB` setting in the MongoDB configuration file. If you do not specify this on the command line before running with the `--replSet` option, `mongod` will create a default sized oplog.

  By default, the oplog is 5 percent of total available disk space on 64-bit systems. For more information about changing the oplog size, see the *Change the Size of the Oplog* (page 693)

For causes of replication lag, see *Replication Lag* (page 712).

Replication issues are most often the result of network connectivity issues between members, or the result of a *primary* that does not have the resources to support application and replication traffic. To check the status of a replica, use the `replSetGetStatus` or the following helper in the shell:

```
rs.status()
```

The `replSetGetStatus` reference provides a more in-depth overview view of this output. In general, watch the value of `optimeDate`, and pay particular attention to the time difference between the *primary* and the *secondary* members.

Sharding and Monitoring

In most cases, the components of *sharded clusters* benefit from the same monitoring and analysis as all other MongoDB instances. In addition, clusters require further monitoring to ensure that data is effectively distributed among nodes and that sharding operations are functioning appropriately.

See also:

See the *Sharding Concepts* (page 739) documentation for more information.

Config Servers The *config database* maintains a map identifying which documents are on which shards. The cluster updates this map as *chunks* move between shards. When a configuration server becomes inaccessible, certain sharding operations become unavailable, such as moving chunks and starting `mongos` instances. However, clusters remain accessible from already-running `mongos` instances.

Because inaccessible configuration servers can seriously impact the availability of a sharded cluster, you should monitor your configuration servers to ensure that the cluster remains well balanced and that `mongos` instances can restart.

MongoDB Cloud Manager[44] and Ops Manager[45] monitor config servers and can create notifications if a config server becomes inaccessible. See the MongoDB Cloud Manager documentation[46] and Ops Manager documentation[47] for more information.

Balancing and Chunk Distribution The most effective *sharded cluster* deployments evenly balance *chunks* among the shards. To facilitate this, MongoDB has a background *balancer* process that distributes data to ensure that chunks are always optimally distributed among the *shards*.

Issue the `db.printShardingStatus()` or `sh.status()` command to the `mongos` by way of the `mongo` shell. This returns an overview of the entire cluster including the database name, and a list of the chunks.

Stale Locks In nearly every case, all locks used by the balancer are automatically released when they become stale. However, because any long lasting lock can block future balancing, it's important to ensure that all locks are legitimate. To check the lock status of the database, connect to a `mongos` instance using the `mongo` shell. Issue the following command sequence to switch to the `config` database and display all outstanding locks on the shard database:

```
use config
db.locks.find()
```

For active deployments, the above query can provide insights. The balancing process, which originates on a randomly selected `mongos`, takes a special "balancer" lock that prevents other balancing activity from transpiring. Use the following command, also to the `config` database, to check the status of the "balancer" lock.

```
db.locks.find( { _id : "balancer" } )
```

If this lock exists, make sure that the balancer process is actively using this lock.

[44] https://cloud.mongodb.com/?jmp=docs
[45] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[46] https://docs.cloud.mongodb.com/
[47] https://docs.opsmanager.mongodb.com/current/application

Additional Resources

- MongoDB Production Readiness Consulting Package[48]

Run-time Database Configuration

On this page

- Configure the Database (page 291)
- Security Considerations (page 292)
- Replication and Sharding Configuration (page 293)
- Run Multiple Database Instances on the Same System (page 295)
- Diagnostic Configurations (page 295)

The `command line` and `configuration file` interfaces provide MongoDB administrators with a large number of options and settings for controlling the operation of the database system. This document provides an overview of common configurations and examples of best-practice configurations for common use cases.

While both interfaces provide access to the same collection of options and settings, this document primarily uses the configuration file interface. If you run MongoDB using a *init script* or if you installed from a package for your operating system, you likely already have a configuration file located at `/etc/mongod.conf`. Confirm this by checking the contents of the `/etc/init.d/mongod` or `/etc/rc.d/mongod` script to ensure that the init scripts start the `mongod` with the appropriate configuration file.

To start a MongoDB instance using this configuration file, issue a command in the following form:

```
mongod --config /etc/mongod.conf
mongod -f /etc/mongod.conf
```

Modify the values in the `/etc/mongod.conf` file on your system to control the configuration of your database instance.

Configure the Database

Consider the following basic configuration which uses the `YAML format`:

```
processManagement:
   fork: true
net:
   bindIp: 127.0.0.1
   port: 27017
storage:
   dbPath: /srv/mongodb
systemLog:
   destination: file
   path: "/var/log/mongodb/mongod.log"
   logAppend: true
storage:
   journal:
      enabled: true
```

Or, if using the older `.ini` configuration file format:

[48] https://www.mongodb.com/products/consulting?jmp=docs#s_product_readiness

```
fork = true
bind_ip = 127.0.0.1
port = 27017
quiet = true
dbpath = /srv/mongodb
logpath = /var/log/mongodb/mongod.log
logappend = true
journal = true
```

For most standalone servers, this is a sufficient base configuration. It makes several assumptions, but consider the following explanation:

- `fork` is `true`, which enables a *daemon* mode for `mongod`, which detaches (i.e. "forks") the MongoDB from the current session and allows you to run the database as a conventional server.
- `bindIp` is `127.0.0.1`, which forces the server to only listen for requests on the localhost IP. Only bind to secure interfaces that the application-level systems can access with access control provided by system network filtering (i.e. "*firewall*").

  New in version 2.6: `mongod` installed from official *.deb* (page 36) and *.rpm* (page 23) packages have the `bind_ip` configuration set to `127.0.0.1` by default.
- `port` is `27017`, which is the default MongoDB port for database instances. MongoDB can bind to any port. You can also filter access based on port using network filtering tools.

  Note: UNIX-like systems require superuser privileges to attach processes to ports lower than 1024.
- `quiet` is `true`. This disables all but the most critical entries in output/log file, and is *not* recommended for production systems. If you do set this option, you can use `setParameter` to modify this setting during run time.
- `dbPath` is `/srv/mongodb`, which specifies where MongoDB will store its data files. `/srv/mongodb` and `/var/lib/mongodb` are popular locations. The user account that `mongod` runs under will need read and write access to this directory.
- `systemLog.path` is `/var/log/mongodb/mongod.log` which is where `mongod` will write its output. If you do not set this value, `mongod` writes all output to standard output (e.g. `stdout`.)
- `logAppend` is `true`, which ensures that `mongod` does not overwrite an existing log file following the server start operation.
- `storage.journal.enabled` is `true`, which enables *journaling*. Journaling ensures single instance write-durability. 64-bit builds of `mongod` enable journaling by default. Thus, this setting may be redundant.

Given the default configuration, some of these values may be redundant. However, in many situations explicitly stating the configuration increases overall system intelligibility.

Security Considerations

The following collection of configuration options are useful for limiting access to a `mongod` instance. Consider the following settings, shown in both `YAML` and older configuration file format:

In `YAML format`

```
security:
   authorization: enabled
net:
   bindIp: 127.0.0.1,10.8.0.10,192.168.4.24
```

Or, if using the older older configuration file format[49]:

```
bind_ip = 127.0.0.1,10.8.0.10,192.168.4.24
auth = true
```

Consider the following explanation for these configuration decisions:

- "`bindIp`" has three values: `127.0.0.1`, the localhost interface; `10.8.0.10`, a private IP address typically used for local networks and VPN interfaces; and `192.168.4.24`, a private network interface typically used for local networks.

  Because production MongoDB instances need to be accessible from multiple database servers, it is important to bind MongoDB to multiple interfaces that are accessible from your application servers. At the same time it's important to limit these interfaces to interfaces controlled and protected at the network layer.

- "`authorization`" is `true` enables the authorization system within MongoDB. If enabled you will need to log in by connecting over the `localhost` interface for the first time to create user credentials.

See also:

Security (page 391)

Replication and Sharding Configuration

Replication Configuration *Replica set* configuration is straightforward, and only requires that the `replSetName` have a value that is consistent among all members of the set. Consider the following:

In `YAML format`

```
replication:
   replSetName: set0
```

Or, if using the older configuration file format[50]:

```
replSet = set0
```

Use descriptive names for sets. Once configured, use the `mongo` shell to add hosts to the replica set.

See also:

Replica set reconfiguration.

To enable authentication for the *replica set*, add the following `keyFile` option:

In `YAML format`

```
security:
   keyFile: /srv/mongodb/keyfile
```

Or, if using the older configuration file format[51]:

```
keyFile = /srv/mongodb/keyfile
```

Setting `keyFile` enables authentication and specifies a key file for the replica set member use to when authenticating to each other. The content of the key file is arbitrary, but must be the same on all members of the *replica set* and `mongos` instances that connect to the set. The keyfile must be less than one kilobyte in size and may only contain characters in the base64 set and the file must not have group or "world" permissions on UNIX systems.

See also:

[49] https://docs.mongodb.org/v2.4/reference/configuration-options
[50] https://docs.mongodb.org/v2.4/reference/configuration-options
[51] https://docs.mongodb.org/v2.4/reference/configuration-options

The *Replica Set Security* (page 423) section for information on configuring authentication with replica sets.

The *Replication* (page 623) document for more information on replication in MongoDB and replica set configuration in general.

Sharding Configuration Sharding requires a number of `mongod` instances with different configurations. The config servers store the cluster's metadata, while the cluster distributes data among one or more shard servers.

Note: *Config servers* are not *replica sets*.

To set up one or three "config server" instances as *normal* (page 291) `mongod` instances, and then add the following configuration option:

In `YAML format`

```
sharding:
   clusterRole: configsvr
net:
   bindIp: 10.8.0.12
   port: 27001
```

Or, if using the older configuration file format[52]:

```
configsvr = true

bind_ip = 10.8.0.12
port = 27001
```

This creates a config server running on the private IP address `10.8.0.12` on port `27001`. Make sure that there are no port conflicts, and that your config server is accessible from all of your `mongos` and `mongod` instances.

To set up shards, configure two or more `mongod` instance using your *base configuration* (page 291), with the `shardsvr` value for the `sharding.clusterRole` setting:

```
sharding:
   clusterRole: shardsvr
```

Or, if using the older configuration file format[53]:

```
shardsvr = true
```

Finally, to establish the cluster, configure at least one `mongos` process with the following settings:

In `YAML format`:

```
sharding:
   configDB: 10.8.0.12:27001
   chunkSize: 64
```

Or, if using the older configuration file format[54]:

```
configdb = 10.8.0.12:27001
chunkSize = 64
```

Important: Always use 3 config servers in production environments.

[52] https://docs.mongodb.org/v2.4/reference/configuration-options
[53] https://docs.mongodb.org/v2.4/reference/configuration-options
[54] https://docs.mongodb.org/v2.4/reference/configuration-options

You can specify multiple `configDB` instances by specifying hostnames and ports in the form of a comma separated list.

In general, avoid modifying the `chunkSize` from the default value of 64, [55] and ensure this setting is consistent among all `mongos` instances.

See also:

The *Sharding* (page 733) section of the manual for more information on sharding and cluster configuration.

Run Multiple Database Instances on the Same System

In many cases running multiple instances of `mongod` on a single system is not recommended. On some types of deployments [56] and for testing purposes you may need to run more than one `mongod` on a single system.

In these cases, use a *base configuration* (page 291) for each instance, but consider the following configuration values:

In `YAML format`:

```
storage:
   dbPath: /srv/mongodb/db0/
processManagement:
   pidFilePath: /srv/mongodb/db0.pid
```

Or, if using the older configuration file format[57]:

```
dbpath = /srv/mongodb/db0/
pidfilepath = /srv/mongodb/db0.pid
```

The `dbPath` value controls the location of the `mongod` instance's data directory. Ensure that each database has a distinct and well labeled data directory. The `pidFilePath` controls where `mongod` process places it's *process id* file. As this tracks the specific `mongod` file, it is crucial that file be unique and well labeled to make it easy to start and stop these processes.

Create additional *init scripts* and/or adjust your existing MongoDB configuration and init script as needed to control these processes.

Diagnostic Configurations

The following configuration options control various `mongod` behaviors for diagnostic purposes:

- `operationProfiling.mode` sets the *database profiler* (page 312) level. The profiler is not active by default because of the possible impact on the profiler itself on performance. Unless this setting is on, queries are not profiled.
- `operationProfiling.slowOpThresholdMs` configures the threshold which determines whether a query is "slow" for the purpose of the logging system and the *profiler* (page 312). The default value is 100 milliseconds. Set a lower value if the database profiler does not return useful results or a higher value to only log the longest running queries.
- `systemLog.verbosity` controls the amount of logging output that `mongod` write to the log. Only use this option if you are experiencing an issue that is not reflected in the normal logging level.

[55] *Chunk* size is 64 megabytes by default, which provides the ideal balance between the most even distribution of data, for which smaller chunk sizes are best, and minimizing chunk migration, for which larger chunk sizes are optimal.

[56] Single-tenant systems with *SSD* or other high performance disks may provide acceptable performance levels for multiple `mongod` instances. Additionally, you may find that multiple databases with small working sets may function acceptably on a single system.

[57] https://docs.mongodb.org/v2.4/reference/configuration-options

Changed in version 3.0: You can also specify verbosity level for specific components using the `systemLog.component.<name>.verbosity` setting. For the available components, see `component verbosity settings`.

For more information, see also *Database Profiling* (page 312) and *Analyzing MongoDB Performance* (page 310).

Production Notes

On this page

- MongoDB Binaries (page 296)
- MongoDB `dbPath` (page 297)
- Concurrency (page 297)
- Data Consistency (page 298)
- Networking (page 298)
- Hardware Considerations (page 299)
- Architecture (page 302)
- Compression (page 302)
- Platform Specific Considerations (page 303)
- Performance Monitoring (page 307)
- Backups (page 307)
- Additional Resources (page 307)

This page details system configurations that affect MongoDB, especially when running in production.

Note: MongoDB Cloud Manager[58], a hosted service, and Ops Manager[59], an on-premise solution, provide monitoring, backup, and automation of MongoDB instances. See the MongoDB Cloud Manager documentation[60] and Ops Manager documentation[61] for more information.

MongoDB Binaries

Supported Platforms MongoDB provides builds for the following supported platforms. For running **in production**, refer to the *Recommended Platforms* (page 297) for operating system recommendations.

Platform	3.2	3.0	2.6	2.4	2.2
Amazon Linux	Y	Y	Y	Y	Y
Debian 7	Y	Y	Y	Y	Y
Fedora 8+			Y	Y	Y
RHEL/CentOS 6.2+	Y	Y	Y	Y	Y
RHEL/CentOS 7.0+	Y	Y	Y		
SLES 11	Y	Y	Y	Y	Y
SLES 12	Y				
Solaris 64-bit	Y	Y	Y	Y	Y
Ubuntu 12.04	Y	Y	Y	Y	Y
Ubuntu 14.04	Y	Y	Y		
Microsoft Azure	Y	Y	Y	Y	Y
Windows Vista/Server 2008R2/2012+	Y	Y	Y	Y	Y
OSX 10.7+	Y	Y	Y	Y	

[58] https://cloud.mongodb.com/?jmp=docs
[59] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[60] https://docs.cloud.mongodb.com/
[61] https://docs.opsmanager.mongodb.com?jmp=docs

Changed in version 3.2: MongoDB can now use the *WiredTiger storage engine* (page 595) on all supported platforms.

Recommended Platforms While MongoDB supports a variety of platorms, the following operating systems are recommended for production use:

- Amazon Linux
- Debian 7.1
- RHEL / CentOS 6.2+
- SLES 11+
- Ubuntu LTS 12.04
- Ubuntu LTS 14.04
- Windows Server 2012 & 2012 R2

See also:

Platform Specific Considerations (page 303)

Use the Latest Stable Packages Be sure you have the latest stable release.

All MongoDB releases are available on the Downloads[62] page. The Downloads[63] page is a good place to verify the current stable release, even if you are installing via a package manager.

Use 64-bit Builds Always use 64-bit builds for production.

Important: Starting in MongoDB 3.2, 32-bit binaries are deprecated and will be unavailable in future releases.

Although the 32-bit builds exist for Linux and Windows, they are **unsuitable** for production deployments. 32-bit builds also do **not** support the WiredTiger storage engine. For more information, see the *32-bit limitations page* (page 21)

MongoDB `dbPath`

The files in the `dbPath` directory must correspond to the configured *storage engine*. `mongod` will not start if `dbPath` contains data files created by a storage engine other than the one specified by `--storageEngine`.

Changed in version 3.2: As of MongoDB 3.2, MongoDB uses the *WiredTiger* (page 595) storage engine by default.

Changed in version 3.0: MongoDB includes support for two storage engines: *MMAPv1* (page 603), the storage engine available in previous versions of MongoDB, and *WiredTiger* (page 595).

`mongod` must possess read and write permissions for the specified `dbPath`.

Concurrency

MMAPv1 Changed in version 3.0: Beginning with MongoDB 3.0, *MMAPv1* (page 603) provides *collection-level locking*: All collections have a unique readers-writer lock that allows multiple clients to modify documents in different collections at the same time.

For MongoDB versions 2.2 through 2.6 series, each database has a readers-writer lock that allows concurrent read access to a database, but gives exclusive access to a single write operation per database. See the *Concurrency* (page 835)

[62] http://www.mongodb.org/downloads
[63] http://www.mongodb.org/downloads

page for more information. In earlier versions of MongoDB, all write operations contended for a single readers-writer lock for the entire `mongod` instance.

WiredTiger *WiredTiger* (page 595) supports concurrent access by readers and writers to the documents in a collection. Clients can read documents while write operations are in progress, and multiple threads can modify different documents in a collection at the same time.

See also:

Allocate Sufficient RAM and CPU (page 299) provides information about how WiredTiger takes advantage of multiple CPU cores and how to improve operation throughput.

Data Consistency

Journaling MongoDB uses *write ahead logging* to an on-disk *journal*. Journaling guarantees that MongoDB can quickly recover *write operations* (page 114) that were written to the journal but not written to data files in cases where `mongod` terminated due to a crash or other serious failure.

Leave journaling enabled in order to ensure that `mongod` will be able to recover its data files and keep the data files in a valid state following a crash. See *Journaling* (page 606) for more information.

Read Concern New in version 3.2.

If using `"majority"` (page 182) *read concern* (page 181), use `{ w: "majority" }` (page 180) *write concern* (page 179) for write operations to ensure that a single thread can read its own writes.

To use a *read concern* level of `"majority"` (page 182), you must use the WiredTiger storage engine and start the `mongod` instances with the `--enableMajorityReadConcern` command line option (or the `replication.enableMajorityReadConcern` setting if using a configuration file).

Only replica sets using `protocol version 1` (page 718) support `"majority"` (page 182) read concern. Replica sets running protocol version 0 do not support `"majority"` (page 182) read concern.

Write Concern *Write concern* (page 179) describes the level of acknowledgement requested from MongoDB for write operations. The level of the write concerns affects how quickly the write operation returns. When write operations have a *weak* write concern, they return quickly. With *stronger* write concerns, clients must wait after sending a write operation until MongoDB confirms the write operation at the requested write concern level. With insufficient write concerns, write operations may appear to a client to have succeeded, but may not persist in some cases of server failure.

See the *Write Concern* (page 179) document for more information about choosing an appropriate write concern level for your deployment.

Networking

Use Trusted Networking Environments Always run MongoDB in a *trusted environment*, with network rules that prevent access from *all* unknown machines, systems, and networks. As with any sensitive system that is dependent on network access, your MongoDB deployment should only be accessible to specific systems that require access, such as application servers, monitoring services, and other MongoDB components.

Important: By default, *authorization* (page 433) is not enabled, and `mongod` assumes a trusted environment. Enable `authorization` mode as needed. For more information on authentication mechanisms supported in MongoDB as well as authorization in MongoDB, see *Authentication* (page 393) and *Role-Based Access Control* (page 433).

For additional information and considerations on security, refer to the documents in the *Security Section* (page 391), specifically:

- *Security Checklist* (page 391)
- *MongoDB Configuration Hardening* (page 473)
- *Hardening Network Infrastructure* (page 474)

For Windows users, consider the Windows Server Technet Article on TCP Configuration[64] when deploying MongoDB on Windows.

Disable HTTP Interface MongoDB provides an HTTP interface to check the status of the server and, optionally, run queries. The HTTP interface is disabled by default. Do not enable the HTTP interface in production environments.

Deprecated since version 3.2: HTTP interface for MongoDB

See *HTTP Status Interface* (page 473).

Manage Connection Pool Sizes Avoid overloading the connection resources of a `mongod` or `mongos` instance by adjusting the connection pool size to suit your use case. Start at 110-115% of the typical number of current database requests, and modify the connection pool size as needed. Refer to the *connection-pool-options* for adjusting the connection pool size.

The `connPoolStats` command returns information regarding the number of open connections to the current database for `mongos` and `mongod` instances in sharded clusters.

See also *Allocate Sufficient RAM and CPU* (page 299).

Hardware Considerations

MongoDB is designed specifically with commodity hardware in mind and has few hardware requirements or limitations. MongoDB's core components run on little-endian hardware, primarily x86/x86_64 processors. Client libraries (i.e. drivers) can run on big or little endian systems.

Allocate Sufficient RAM and CPU

MMAPv1 Due to its concurrency model, the *MMAPv1* (page 603) storage engine does not require many CPU cores. As such, increasing the number of cores can improve performance but does not provide significant return.

At a minimum, ensure that your `mongod` or `mongos` has access to two real cores or one physical CPU.

Increasing the amount of RAM accessible to MongoDB may help reduce the frequency of page faults.

WiredTiger The *WiredTiger* (page 595) storage engine is multithreaded and can take advantage of additional CPU cores. Specifically, the total number of active threads (i.e. concurrent operations) relative to the number of available CPUs can impact performance:

- Throughput *increases* as the number of concurrent active operations increases up to the number of CPUs.
- Throughput *decreases* as the number of concurrent active operations exceeds the number of CPUs by some threshold amount.

[64] http://technet.microsoft.com/en-us/library/dd349797.aspx

The threshold depends on your application. You can determine the optimum number of concurrent active operations for your application by experimenting and measuring throughput. The output from `mongostat` provides statistics on the number of active reads/writes in the (`ar|aw`) column.

With WiredTiger, MongoDB utilizes both the WiredTiger cache and the filesystem cache.

Changed in version 3.2: Starting in MongoDB 3.2, the WiredTiger cache, by default, will use the larger of either:

- 60% of RAM minus 1 GB, or
- 1 GB.

For systems with up to 10 GB of RAM, the new default setting is less than or equal to the 3.0 default setting (For MongoDB 3.0, the WiredTiger cache uses either 1 GB or half of the installed physical RAM, whichever is larger).

For systems with more than 10 GB of RAM, the new default setting is greater than the 3.0 setting.

Via the filesystem cache, MongoDB automatically uses all free memory that is not used by the WiredTiger cache or by other processes. Data in the filesystem cache is compressed.

To adjust the size of the WiredTiger cache, see `storage.wiredTiger.engineConfig.cacheSizeGB` and *`--wiredTigerCacheSizeGB`*. Avoid increasing the WiredTiger cache size above its default value.

Note: The `storage.wiredTiger.engineConfig.cacheSizeGB` only limits the size of the WiredTiger cache, not the total amount of memory used by `mongod`. The WiredTiger cache is only one component of the RAM used by MongoDB. MongoDB also automatically uses all free memory on the machine via the filesystem cache (data in the filesystem cache is compressed).

In addition, the operating system will use any free RAM to buffer filesystem blocks.

To accommodate the additional consumers of RAM, you may have to decrease WiredTiger cache size.

The default WiredTiger cache size value assumes that there is a single `mongod` instance per machine. If a single machine contains multiple MongoDB instances, then you should decrease the setting to accommodate the other `mongod` instances.

If you run `mongod` in a container (e.g. `lxc`, `cgroups`, Docker, etc.) that does *not* have access to all of the RAM available in a system, you must set `storage.wiredTiger.engineConfig.cacheSizeGB` to a value less than the amount of RAM available in the container. The exact amount depends on the other processes running in the container.

To view statistics on the cache and eviction rate, see the `wiredTiger.cache` field returned from the `serverStatus` command.

See also:

Concurrency (page 297)

Use Solid State Disks (SSDs) MongoDB has good results and a good price-performance ratio with SATA SSD (Solid State Disk).

Use SSD if available and economical. Spinning disks can be performant, but SSDs' capacity for random I/O operations works well with the update model of MMAPv1.

Commodity (SATA) spinning drives are often a good option, as the random I/O performance increase with more expensive spinning drives is not that dramatic (only on the order of 2x). Using SSDs or increasing RAM may be more effective in increasing I/O throughput.

MongoDB and NUMA Hardware Running MongoDB on a system with Non-Uniform Access Memory (NUMA) can cause a number of operational problems, including slow performance for periods of time and high system process usage.

When running MongoDB servers and clients on NUMA hardware, you should configure a memory interleave policy so that the host behaves in a non-NUMA fashion. MongoDB checks NUMA settings on start up when deployed on Linux (since version 2.0) and Windows (since version 2.6) machines. If the NUMA configuration may degrade performance, MongoDB prints a warning.

See also:

- The MySQL "swap insanity" problem and the effects of NUMA[65] post, which describes the effects of NUMA on databases. The post introduces NUMA and its goals, and illustrates how these goals are not compatible with production databases. Although the blog post addresses the impact of NUMA for MySQL, the issues for MongoDB are similar.
- NUMA: An Overview[66].

Configuring NUMA on Windows On Windows, memory interleaving must be enabled through the machine's BIOS. Consult your system documentation for details.

Configuring NUMA on Linux When running MongoDB on Linux, you should disable *zone reclaim* in the `sysctl` settings using one of the following commands:

```
echo 0 | sudo tee /proc/sys/vm/zone_reclaim_mode

sudo sysctl -w vm.zone_reclaim_mode=0
```

Then, you should use the `numactl` command to start the MongoDB programs (`mongod`, including the *config servers* (page 742); `mongos`; and clients) in the following manner:

```
numactl --interleave=all <path> <options>
```

where `<path>` is the path to the program you are starting, and `<options>` are any optional arguments to pass to the program.

To fully disable NUMA behavior, you must perform both operations. For more information, see the Documentation for /proc/sys/vm/*[67].

Disk and Storage Systems

Swap Assign swap space for your systems. Allocating swap space can avoid issues with memory contention and can prevent the OOM Killer on Linux systems from killing `mongod`.

For the MMAPv1 storage engine, the method `mongod` uses to map files to memory ensures that the operating system will never store MongoDB data in swap space. On Windows systems, using MMAPv1 requires extra swap space due to commitment limits. For details, see *MongoDB on Windows* (page 304).

For the WiredTiger storage engine, given sufficient memory pressure, WiredTiger may store data in swap space.

[65] http://jcole.us/blog/archives/2010/09/28/mysql-swap-insanity-and-the-numa-architecture/
[66] https://queue.acm.org/detail.cfm?id=2513149
[67] http://www.kernel.org/doc/Documentation/sysctl/vm.txt

RAID Most MongoDB deployments should use disks backed by RAID-10.

RAID-5 and RAID-6 do not typically provide sufficient performance to support a MongoDB deployment.

Avoid RAID-0 with MongoDB deployments. While RAID-0 provides good write performance, it also provides limited availability and can lead to reduced performance on read operations, particularly when using Amazon's EBS volumes.

Remote Filesystems With the MMAPv1 storage engine, the Network File System protocol (NFS) is not recommended as you may see performance problems when both the data files and the journal files are hosted on NFS. You may experience better performance if you place the journal on local or `iscsi` volumes.

With the WiredTiger storage engine, WiredTiger objects may be stored on remote file systems if the remote file system conforms to ISO/IEC 9945-1:1996 (POSIX.1). Because remote file systems are often slower than local file systems, using a remote file system for storage may degrade performance.

If you decide to use NFS, add the following NFS options to your `/etc/fstab` file: `bg`, `nolock`, and `noatime`.

Separate Components onto Different Storage Devices For improved performance, consider separating your database's data, journal, and logs onto different storage devices, based on your application's access and write pattern.

For the WiredTiger storage engine, you can also store the indexes on a different storage device. See `storage.wiredTiger.engineConfig.directoryForIndexes`.

Note: Using different storage devices will affect your ability to create snapshot-style backups of your data, since the files will be on different devices and volumes.

Scheduling for Virtual Devices Local block devices attached to virtual machine instances via the hypervisor should use a *noop* scheduler for best performance. The *noop* scheduler allows the operating system to defer I/O scheduling to the underlying hypervisor.

Architecture

Replica Sets See the *Replica Set Architectures* (page 636) document for an overview of architectural considerations for replica set deployments.

Sharded Clusters See the *Sharded Cluster Production Architecture* (page 745) document for an overview of recommended sharded cluster architectures for production deployments.

See also:

Design Notes (page 315)

Compression

WiredTiger can compress collection data using either *snappy* or *zlib* compression library. *snappy* provides a lower compression rate but has little performance cost, whereas `zlib` provides better compression rate but has a higher performance cost.

By default, WiredTiger uses *snappy* compression library. To change the compression setting, see `storage.wiredTiger.collectionConfig.blockCompressor`.

WiredTiger uses *prefix compression* on all indexes by default.

Platform Specific Considerations

Note: MongoDB uses the GNU C Library[68] (glibc) if available on a system. MongoDB requires version at least `glibc-2.12-1.2.el6` to avoid a known bug with earlier versions. For best results use at least version 2.13.

MongoDB on Linux

Kernel and File Systems When running MongoDB in production on Linux, you should use Linux kernel version 2.6.36 or later, with either the XFS or EXT4 filesystem. If possible, use XFS as it generally performs better with MongoDB.

With the WiredTiger storage engine in particular, use of XFS is **strongly recommended** to avoid performance issues that have been observed when using EXT4 with WiredTiger.

- If you use the XFS file system, use at least version `2.6.25` of the Linux Kernel.
- If you use the EXT4 file system, use at least version `2.6.28` of the Linux Kernel.
- On Red Hat Enterprise Linux and CentOS, use at least version `2.6.18-194` of the Linux kernel.

`fsync()` on Directories

Important: MongoDB requires a filesystem that supports `fsync()` *on directories*. For example, HGFS and Virtual Box's shared folders do *not* support this operation.

Recommended Configuration For the WiredTiger and MMAPv1 storage engines, consider the following recommendations:

- Turn off `atime` for the storage volume containing the *database files*.
- Set the file descriptor limit, `-n`, and the user process limit (ulimit), `-u`, above 20,000, according to the suggestions in the *ulimit* (page 372) document. A low ulimit will affect MongoDB when under heavy use and can produce errors and lead to failed connections to MongoDB processes and loss of service.
- Disable Transparent Huge Pages, as MongoDB performs better with normal (4096 bytes) virtual memory pages. See *Transparent Huge Pages Settings* (page 319).
- Disable NUMA in your BIOS. If that is not possible, see *MongoDB on NUMA Hardware* (page 301).
- Configure SELinux on Red Hat. For more information, see *Configure SELinux for MongoDB* (page 25) and *Configure SELinux for MongoDB Enterprise* (page 52).

For the MMAPv1 storage engine:

- Ensure that readahead settings for the block devices that store the database files are appropriate. For random access use patterns, set low readahead values. A readahead of 32 (16 kB) often works well.

  For a standard block device, you can run `sudo blockdev --report` to get the readahead settings and `sudo blockdev --setra <value> <device>` to change the readahead settings. Refer to your specific operating system manual for more information.

For **all** MongoDB deployments:

- Use the Network Time Protocol (NTP) to synchronize time among your hosts. This is especially important in sharded clusters.

[68] http://www.gnu.org/software/libc/

MongoDB and TLS/SSL Libraries On Linux platforms, you may observe one of the following statements in the MongoDB log:

```
<path to SSL libs>/libssl.so.<version>: no version information available (required by /usr/bin/mongod
<path to SSL libs>/libcrypto.so.<version>: no version information available (required by /usr/bin/mon
```

These warnings indicate that the system's TLS/SSL libraries are different from the TLS/SSL libraries that the `mongod` was compiled against. Typically these messages do not require intervention; however, you can use the following operations to determine the symbol versions that `mongod` expects:

```
objdump -T <path to mongod>/mongod | grep " SSL_"
objdump -T <path to mongod>/mongod | grep " CRYPTO_"
```

These operations will return output that resembles one the of the following lines:

```
0000000000000000      DF *UND*       0000000000000000  libssl.so.10 SSL_write
0000000000000000      DF *UND*       0000000000000000  OPENSSL_1.0.0 SSL_write
```

The last two strings in this output are the symbol version and symbol name. Compare these values with the values returned by the following operations to detect symbol version mismatches:

```
objdump -T <path to TLS/SSL libs>/libssl.so.1*
objdump -T <path to TLS/SSL libs>/libcrypto.so.1*
```

This procedure is neither exact nor exhaustive: many symbols used by `mongod` from the `libcrypto` library do not begin with `CRYPTO_`.

MongoDB on Windows

MongoDB 3.0 Using WiredTiger For MongoDB instances using the WiredTiger storage engine, performance on Windows is comparable to performance on Linux.

MongoDB Using MMAPv1

Install Hotfix for MongoDB 2.6.6 and Later Microsoft has released a hotfix for Windows 7 and Windows Server 2008 R2, KB2731284[69], that repairs a bug in these operating systems' use of memory-mapped files that adversely affects the performance of MongoDB using the MMAPv1 storage engine.

Install this hotfix to obtain significant performance improvements on MongoDB 2.6.6 and later releases in the 2.6 series, which use MMAPv1 exclusively, and on 3.0 and later when using MMAPv1 as the storage engine.

Configure Windows Page File For MMAPv1 Configure the page file such that the minimum and maximum page file size are equal and at least 32 GB. Use a multiple of this size if, during peak usage, you expect concurrent writes to many databases or collections. However, the page file size does not need to exceed the maximum size of the database.

A large page file is needed as Windows requires enough space to accommodate all regions of memory mapped files made writable during peak usage, regardless of whether writes actually occur.

The page file is not used for database storage and will not receive writes during normal MongoDB operation. As such, the page file will not affect performance, but it must exist and be large enough to accommodate Windows' commitment rules during peak database use.

Note: Dynamic page file sizing is too slow to accommodate the rapidly fluctuating commit charge of an active

[69] http://support.microsoft.com/kb/2731284

MongoDB deployment. This can result in transient overcommitment situations that may lead to abrupt server shutdown with a VirtualProtect error 1455.

MongoDB on Virtual Environments This section describes considerations when running MongoDB in some of the more common virtual environments.

For all platforms, consider *Scheduling for Virtual Devices* (page 302).

EC2 MongoDB is compatible with EC2. MongoDB Cloud Manager[70] provides integration with Amazon Web Services (AWS) and lets you deploy new EC2 instances directly from MongoDB Cloud Manager. See Configure AWS Integration[71] for more details.

Azure For all MongoDB deployments using Azure, you **must** mount the volume that hosts the `mongod` instance's `dbPath` with the *Host Cache Preference* `READ/WRITE`.

This applies to all Azure deployments, using any guest operating system.

If your volumes have inappropriate cache settings, MongoDB may eventually shut down with the following error:

```
[DataFileSync] FlushViewOfFile for <data file> failed with error 1 ...
[DataFileSync] Fatal Assertion 16387
```

These shut downs do not produce data loss when `storage.journal.enabled` is set to `true`. You can safely restart `mongod` at any time following this event.

The performance characteristics of MongoDB may change with `READ/WRITE` caching enabled.

The TCP keepalive on the Azure load balancer is 240 seconds by default, which can cause it to silently drop connections if the TCP keepalive on your Azure systems is greater than this value. You should set `tcp_keepalive_time` to 120 to ameliorate this problem.

On Linux systems:

- To view the keep alive setting, you can use one of the following commands:

  ```
  sysctl net.ipv4.tcp_keepalive_time
  ```

  Or:

  ```
  cat /proc/sys/net/ipv4/tcp_keepalive_time
  ```

  The value is measured in seconds.

- To change the `tcp_keepalive_time` value, you can use one of the following command:

  ```
  sudo sysctl -w net.ipv4.tcp_keepalive_time=<value>
  ```

  Or:

  ```
  echo <value> | sudo tee /proc/sys/net/ipv4/tcp_keepalive_time
  ```

  These operations do not persist across system reboots. To persist the setting, add the following line to `/etc/sysctl.conf`:

  ```
  net.ipv4.tcp_keepalive_time = <value>
  ```

[70] https://cloud.mongodb.com/?jmp=docs
[71] https://docs.cloud.mongodb.com/tutorial/configure-aws-settings/

On Linux, `mongod` and `mongos` processes limit the keepalive to a maximum of 300 seconds (5 minutes) on their own sockets by overriding keepalive values greater than 5 minutes.

For Windows systems:

- To view the keep alive setting, issue the following command:

  ```
  reg query HKLM\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters /v KeepAliveTime
  ```

  The registry value is not present by default. The system default, used if the value is absent, is 7200000 *milliseconds* or `0x6ddd00` in hexadecimal.

- To change the `KeepAliveTime` value, use the following command in an Administrator *Command Prompt*, where `<value>` is expressed in hexadecimal (e.g. `0x0124c0` is 120000):

  ```
  reg add HKLM\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\ /v KeepAliveTime /d <value>
  ```

  Windows users should consider the Windows Server Technet Article on KeepAliveTime[72] for more information on setting keep alive for MongoDB deployments on Windows systems.

VMWare MongoDB is compatible with VMWare.

VMWare supports *memory overcommitment*, where you can assign more memory to your virtual machines than the physical machine has available. When memory is overcommitted, the hypervisor reallocates memory between the virtual machines. VMWare's balloon driver (`vmmemctl`) reclaims the pages that are considered least valuable. The balloon driver resides inside the guest operating system. When the balloon driver expands, it may induce the guest operating system to reclaim memory from guest applications, which can interfere with MongoDB's memory management and affect MongoDB's performance.

You can disable the balloon driver and VMWare's memory overcommitment feature to mitigate these problems. However, disabling the balloon driver can cause the hypervisor to use its swap, as there is no other available mechanism to perform the memory reclamation. Accessing data in swap is much slower than accessing data in memory, which can in turn affect performance. Instead of disabling the balloon driver and memory overcommitment features, map and reserve the full amount of memory for the virtual machine running MongoDB. This ensures that the balloon will not be inflated in the local operating system if there is memory pressure in the hypervisor due to an overcommitted configuration.

When using MongoDB with VMWare, ensure that the CPU reservation does not exceed more than 2 virtual CPUs per physical core.

Disable VMWare's Migration with vMotion ("live migration"). The live migration of a virtual machine can cause performance problems and affect *replica set* (page 644) and *sharded cluster high availability* (page 750) mechanisms.

It is possible to clone a virtual machine running MongoDB. You might use this function to spin up a new virtual host to add as a member of a replica set. If you clone a VM with journaling enabled, the clone snapshot will be valid. If not using journaling, first stop `mongod`, then clone the VM, and finally, restart `mongod`.

KVM MongoDB is compatible with KVM.

KVM supports *memory overcommitment*, where you can assign more memory to your virtual machines than the physical machine has available. When memory is overcommitted, the hypervisor reallocates memory between the virtual machines. KVM's balloon driver reclaims the pages that are considered least valuable. The balloon driver resides inside the guest operating system. When the balloon driver expands, it may induce the guest operating system to reclaim memory from guest applications, which can interfere with MongoDB's memory management and affect MongoDB's performance.

[72] https://technet.microsoft.com/en-us/library/cc957549.aspx

You can disable the balloon driver and KVM's memory overcommitment feature to mitigate these problems. However, disabling the balloon driver can cause the hypervisor to use its swap, as there is no other available mechanism to perform the memory reclamation. Accessing data in swap is much slower than accessing data in memory, which can in turn affect performance. Instead of disabling the balloon driver and memory overcommitment features, map and reserve the full amount of memory for the virtual machine running MongoDB. This ensures that the balloon will not be inflated in the local operating system if there is memory pressure in the hypervisor due to an overcommitted configuration.

When using MongoDB with KVM, ensure that the CPU reservation does not exceed more than 2 virtual CPUs per physical core.

Performance Monitoring

iostat On Linux, use the `iostat` command to check if disk I/O is a bottleneck for your database. Specify a number of seconds when running iostat to avoid displaying stats covering the time since server boot.

For example, the following command will display extended statistics and the time for each displayed report, with traffic in MB/s, at one second intervals:

```
iostat -xmt 1
```

Key fields from `iostat`:

- `%util`: this is the most useful field for a quick check, it indicates what percent of the time the device/drive is in use.
- `avgrq-sz`: average request size. Smaller number for this value reflect more random IO operations.

bwm-ng bwm-ng[73] is a command-line tool for monitoring network use. If you suspect a network-based bottleneck, you may use `bwm-ng` to begin your diagnostic process.

Backups

To make backups of your MongoDB database, please refer to *MongoDB Backup Methods Overview* (page 282).

Additional Resources

- Blog Post: Capacity Planning and Hardware Provisioning for MongoDB In Ten Minutes[74]
- Whitepaper: MongoDB Multi-Data Center Deployments[75]
- Whitepaper: Security Architecture[76]
- Whitepaper: MongoDB Architecture Guide[77]
- Presentation: MongoDB Administration 101[78]
- MongoDB Production Readiness Consulting Package[79]

[73] http://www.gropp.org/?id=projects&sub=bwm-ng
[74] https://www.mongodb.com/blog/post/capacity-planning-and-hardware-provisioning-mongodb-ten-minutes?jmp=docs
[75] http://www.mongodb.com/lp/white-paper/multi-dc?jmp=docs
[76] https://www.mongodb.com/lp/white-paper/mongodb-security-architecture?jmp=docs
[77] https://www.mongodb.com/lp/whitepaper/architecture-guide?jmp=docs
[78] http://www.mongodb.com/presentations/webinar-mongodb-administration-101?jmp=docs
[79] https://www.mongodb.com/products/consulting?jmp=docs#s_product_readiness

8.1.2 Data Management

These document introduce data management practices and strategies for MongoDB deployments, including strategies for managing multi-data center deployments, managing larger file stores, and data lifecycle tools.

Data Center Awareness **(page 308)** Presents the MongoDB features that allow application developers and database administrators to configure their deployments to be more data center aware or allow operational and location-based separation.

Data Center Awareness

On this page

- Further Reading (page 309)
- Additional Resource (page 309)

MongoDB provides a number of features that allow application developers and database administrators to customize the behavior of a *sharded cluster* or *replica set* deployment so that MongoDB may be *more* "data center aware," or allow operational and location-based separation.

MongoDB also supports segregation based on functional parameters, to ensure that certain `mongod` instances are only used for reporting workloads or that certain high-frequency portions of a sharded collection only exist on specific shards.

The following documents, *found either in this section or other sections of this manual*, provide information on customizing a deployment for operation- and location-based separation:

Operational Segregation in MongoDB Deployments **(page 308)** MongoDB lets you specify that certain application operations use certain `mongod` instances.

Tag Aware Sharding **(page 756)** Tags associate specific ranges of *shard key* values with specific shards for use in managing deployment patterns.

Manage Shard Tags **(page 816)** Use tags to associate specific ranges of shard key values with specific shards.

Operational Segregation in MongoDB Deployments

On this page

- Operational Overview (page 308)
- Additional Resource (page 309)

Operational Overview MongoDB includes a number of features that allow database administrators and developers to segregate application operations to MongoDB deployments by functional or geographical groupings.

This capability provides "data center awareness," which allows applications to target MongoDB deployments with consideration of the physical location of the `mongod` instances. MongoDB supports segmentation of operations across different dimensions, which may include multiple data centers and geographical regions in multi-data center deployments, racks, networks, or power circuits in single data center deployments.

MongoDB also supports segregation of database operations based on functional or operational parameters, to ensure that certain `mongod` instances are only used for reporting workloads or that certain high-frequency portions of a sharded collection only exist on specific shards.

Specifically, with MongoDB, you can:

- ensure write operations propagate to specific members of a replica set, or to specific members of replica sets.
- ensure that specific members of a replica set respond to queries.
- ensure that specific ranges of your *shard key* balance onto and reside on specific *shards*.
- combine the above features in a single distributed deployment, on a per-operation (for read and write operations) and collection (for chunk distribution in sharded clusters distribution) basis.

For full documentation of these features, see the following documentation in the MongoDB Manual:

- *Read Preferences* (page 651), which controls how drivers help applications target read operations to members of a replica set.
- *Write Concerns* (page 179), which controls how MongoDB ensures that write operations propagate to members of a replica set.
- *Replica Set Tags* (page 700), which control how applications create and interact with custom groupings of replica set members to create custom application-specific read preferences and write concerns.
- *Tag Aware Sharding* (page 756), which allows MongoDB administrators to define an application-specific balancing policy, to control how documents belonging to specific ranges of a shard key distribute to shards in the *sharded cluster*.

See also:

Before adding operational segregation features to your application and MongoDB deployment, become familiar with all documentation of *replication* (page 623), and *sharding* (page 733).

Additional Resource

- Whitepaper: MongoDB Multi-Data Center Deployments[80]
- Webinar: Multi-Data Center Deployment[81]

Further Reading

- The *Write Concern* (page 179) and *Read Preference* (page 651) documents, which address capabilities related to data center awareness.
- *Deploy a Geographically Redundant Replica Set* (page 672).

Additional Resource

- Whitepaper: MongoDB Multi-Data Center Deployments[82]
- Webinar: Multi-Data Center Deployment[83]

[80] http://www.mongodb.com/lp/white-paper/multi-dc?jmp=docs
[81] https://www.mongodb.com/presentations/webinar-multi-data-center-deployment?jmp=docs
[82] http://www.mongodb.com/lp/white-paper/multi-dc?jmp=docs
[83] https://www.mongodb.com/presentations/webinar-multi-data-center-deployment?jmp=docs

8.1.3 Optimization Strategies for MongoDB

There are many factors that can affect database performance and responsiveness including index use, query structure, data models and application design, as well as operational factors such as architecture and system configuration.

This section describes techniques for optimizing application performance with MongoDB.

Analyzing MongoDB Performance **(page 310)** Discusses some of the factors that can influence MongoDB's performance.

Evaluate Performance of Current Operations **(page 313)** MongoDB provides introspection tools that describe the query execution process, to allow users to test queries and build more efficient queries.

Optimize Query Performance **(page 314)** Introduces the use of *projections* (page 102) to reduce the amount of data MongoDB sends to clients.

Design Notes **(page 315)** A collection of notes related to the architecture, design, and administration of MongoDB-based applications.

Analyzing MongoDB Performance

On this page

- Locking Performance (page 310)
- Memory and the MMAPv1 Storage Engine (page 311)
- Number of Connections (page 311)
- Database Profiling (page 312)
- Additional Resources (page 313)

As you develop and operate applications with MongoDB, you may need to analyze the performance of the application and its database. When you encounter degraded performance, it is often a function of database access strategies, hardware availability, and the number of open database connections.

Some users may experience performance limitations as a result of inadequate or inappropriate indexing strategies, or as a consequence of poor schema design patterns. *Locking Performance* (page 310) discusses how these can impact MongoDB's internal locking.

Performance issues may indicate that the database is operating at capacity and that it is time to add additional capacity to the database. In particular, the application's *working set* should fit in the available physical memory. See *Memory and the MMAPv1 Storage Engine* (page 311) for more information on the working set.

In some cases performance issues may be temporary and related to abnormal traffic load. As discussed in *Number of Connections* (page 311), scaling can help relax excessive traffic.

Database Profiling (page 312) can help you to understand what operations are causing degradation.

Locking Performance

MongoDB uses a locking system to ensure data set consistency. If certain operations are long-running or a queue forms, performance will degrade as requests and operations wait for the lock.

Lock-related slowdowns can be intermittent. To see if the lock has been affecting your performance, refer to the *server-status-locks* section and the *globalLock* section of the `serverStatus` output.

Dividing `locks.timeAcquiringMicros` by `locks.acquireWaitCount` can give an approximate average wait time for a particular lock mode.

`locks.deadlockCount` provide the number of times the lock acquisitions encountered deadlocks.

If `globalLock.currentQueue.total` is consistently high, then there is a chance that a large number of requests are waiting for a lock. This indicates a possible concurrency issue that may be affecting performance.

If `globalLock.totalTime` is high relative to `uptime`, the database has existed in a lock state for a significant amount of time.

Long queries can result from ineffective use of indexes; non-optimal schema design; poor query structure; system architecture issues; or insufficient RAM resulting in *page faults* (page 311) and disk reads.

Memory and the MMAPv1 Storage Engine

Memory Use With the *MMAPv1* (page 603) storage engine, MongoDB uses memory-mapped files to store data. Given a data set of sufficient size, the `mongod` process will allocate all available memory on the system for its use.

While this is intentional and aids performance, the memory mapped files make it difficult to determine if the amount of RAM is sufficient for the data set.

The *memory usage statuses* metrics of the `serverStatus` output can provide insight into MongoDB's memory use.

The `mem.resident` field provides the amount of resident memory in use. If this exceeds the amount of system memory *and* there is a significant amount of data on disk that isn't in RAM, you may have exceeded the capacity of your system.

You can inspect `mem.mapped` to check the amount of mapped memory that `mongod` is using. If this value is greater than the amount of system memory, some operations will require a *page faults* to read data from disk.

Page Faults With the MMAPv1 storage engine, page faults can occur as MongoDB reads from or writes data to parts of its data files that are not currently located in physical memory. In contrast, operating system page faults happen when physical memory is exhausted and pages of physical memory are swapped to disk.

MongoDB reports its triggered page faults as the total number of *page faults* in one second. To check for page faults, see the `extra_info.page_faults` value in the `serverStatus` output.

Rapid increases in the MongoDB page fault counter may indicate that the server has too little physical memory. Page faults also can occur while accessing large data sets or scanning an entire collection.

A single page fault completes quickly and is not problematic. However, in aggregate, large volumes of page faults typically indicate that MongoDB is reading too much data from disk.

MongoDB can often "yield" read locks after a page fault, allowing other database processes to read while `mongod` loads the next page into memory. Yielding the read lock following a page fault improves concurrency, and also improves overall throughput in high volume systems.

Increasing the amount of RAM accessible to MongoDB may help reduce the frequency of page faults. If this is not possible, you may want to consider deploying a *sharded cluster* or adding *shards* to your deployment to distribute load among `mongod` instances.

See *What are page faults?* (page 854) for more information.

Number of Connections

In some cases, the number of connections between the applications and the database can overwhelm the ability of the server to handle requests. The following fields in the `serverStatus` document can provide insight:

- `globalLock.activeClients` contains a counter of the total number of clients with active operations in progress or queued.

- `connections` is a container for the following two fields:
  - `connections.current` the total number of current clients that connect to the database instance.
  - `connections.available` the total number of unused connections available for new clients.

If there are numerous concurrent application requests, the database may have trouble keeping up with demand. If this is the case, then you will need to increase the capacity of your deployment.

For read-heavy applications, increase the size of your *replica set* and distribute read operations to *secondary* members.

For write-heavy applications, deploy *sharding* and add one or more *shards* to a *sharded cluster* to distribute load among `mongod` instances.

Spikes in the number of connections can also be the result of application or driver errors. All of the officially supported MongoDB drivers implement connection pooling, which allows clients to use and reuse connections more efficiently. Extremely high numbers of connections, particularly without corresponding workload is often indicative of a driver or other configuration error.

Unless constrained by system-wide limits, MongoDB has no limit on incoming connections. On Unix-based systems, you can modify system limits using the `ulimit` command, or by editing your system's `/etc/sysctl` file. See *UNIX ulimit Settings* (page 372) for more information.

Database Profiling

MongoDB's "Profiler" is a database profiling system that can help identify inefficient queries and operations.

The following profiling levels are available:

Level	Setting
0	Off. No profiling
1	On. Only includes *"slow"* operations
2	On. Includes *all* operations

Enable the profiler by setting the `profile` value using the following command in the `mongo` shell:

```
db.setProfilingLevel(1)
```

The `slowOpThresholdMs` setting defines what constitutes a "slow" operation. To set the threshold above which the profiler considers operations "slow" (and thus, included in the level `1` profiling data), you can configure `slowOpThresholdMs` at runtime as an argument to the `db.setProfilingLevel()` operation.

See

The documentation of `db.setProfilingLevel()` for more information.

By default, `mongod` records all "slow" queries to its `log`, as defined by `slowOpThresholdMs`.

Note: Because the database profiler can negatively impact performance, only enable profiling for strategic intervals and as minimally as possible on production systems.

You may enable profiling on a per-`mongod` basis. This setting will not propagate across a *replica set* or *sharded cluster*.

You can view the output of the profiler in the `system.profile` collection of your database by issuing the `show profile` command in the `mongo` shell, or with the following operation:

```
db.system.profile.find( { millis : { $gt : 100 } } )
```

This returns all operations that lasted longer than 100 milliseconds. Ensure that the value specified here (`100`, in this example) is above the `slowOpThresholdMs` threshold.

You must use the `$query` operator to access the `query` field of documents within `system.profile`.

Additional Resources

- MongoDB Ops Optimization Consulting Package[84]

Evaluate Performance of Current Operations

On this page

- Use the Database Profiler to Evaluate Operations Against the Database (page 313)
- Use `db.currentOp()` to Evaluate `mongod` Operations (page 313)
- Use `explain` to Evaluate Query Performance (page 313)
- Additional Resources (page 314)

The following sections describe techniques for evaluating operational performance.

Use the Database Profiler to Evaluate Operations Against the Database

MongoDB provides a database profiler that shows performance characteristics of each operation against the database. Use the profiler to locate any queries or write operations that are running slow. You can use this information, for example, to determine what indexes to create.

For more information, see *Database Profiling* (page 312).

Use `db.currentOp()` to Evaluate `mongod` Operations

The `db.currentOp()` method reports on current operations running on a `mongod` instance.

Use `explain` to Evaluate Query Performance

The `cursor.explain()` and `db.collection.explain()` methods return information on a query execution, such as the index MongoDB selected to fulfill the query and execution statistics. You can run the methods in *queryPlanner* mode, *executionStats* mode, or *allPlansExecution* mode to control the amount of information returned.

Example

To use `cursor.explain()` on a query for documents matching the expression `{ a: 1 }`, in the collection named `records`, use an operation that resembles the following in the `mongo` shell:

```
db.records.find( { a: 1 } ).explain("executionStats")
```

For more information, see `https://docs.mongodb.org/manual/reference/explain-results`, `cursor.explain()`, `db.collection.explain()`, and *Analyze Query Performance* (page 159).

[84] https://www.mongodb.com/products/consulting?jmp=docs#ops_optimization

Additional Resources

- MongoDB Performance Evaluation and Tuning Consulting Package[85]

Optimize Query Performance

On this page

- Create Indexes to Support Queries (page 314)
- Limit the Number of Query Results to Reduce Network Demand (page 315)
- Use Projections to Return Only Necessary Data (page 315)
- Use `$hint` to Select a Particular Index (page 315)
- Use the Increment Operator to Perform Operations Server-Side (page 315)
- Additional Resources (page 315)

Create Indexes to Support Queries

For commonly issued queries, create *indexes* (page 515). If a query searches multiple fields, create a *compound index* (page 522). Scanning an index is much faster than scanning a collection. The indexes structures are smaller than the documents reference, and store references in order.

Example

If you have a `posts` collection containing blog posts, and if you regularly issue a query that sorts on the `author_name` field, then you can optimize the query by creating an index on the `author_name` field:

```
db.posts.createIndex( { author_name : 1 } )
```

Indexes also improve efficiency on queries that routinely sort on a given field.

Example

If you regularly issue a query that sorts on the `timestamp` field, then you can optimize the query by creating an index on the `timestamp` field:

Creating this index:

```
db.posts.createIndex( { timestamp : 1 } )
```

Optimizes this query:

```
db.posts.find().sort( { timestamp : -1 } )
```

Because MongoDB can read indexes in both ascending and descending order, the direction of a single-key index does not matter.

Indexes support queries, update operations, and some phases of the *aggregation pipeline* (page 200).

Index keys that are of the `BinData` type are more efficiently stored in the index if:

- the binary subtype value is in the range of 0-7 or 128-135, and
- the length of the byte array is: 0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 20, 24, or 32.

[85] https://www.mongodb.com/products/consulting?jmp=docs#performance_evaluation

Limit the Number of Query Results to Reduce Network Demand

MongoDB *cursors* return results in groups of multiple documents. If you know the number of results you want, you can reduce the demand on network resources by issuing the `limit()` method.

This is typically used in conjunction with sort operations. For example, if you need only 10 results from your query to the `posts` collection, you would issue the following command:

```
db.posts.find().sort( { timestamp : -1 } ).limit(10)
```

For more information on limiting results, see `limit()`

Use Projections to Return Only Necessary Data

When you need only a subset of fields from documents, you can achieve better performance by returning only the fields you need:

For example, if in your query to the `posts` collection, you need only the `timestamp`, `title`, `author`, and `abstract` fields, you would issue the following command:

```
db.posts.find( {}, { timestamp : 1 , title : 1 , author : 1 , abstract : 1} ).sort( { timestamp : -1
```

For more information on using projections, see *Limit Fields to Return from a Query* (page 153).

Use `$hint` to Select a Particular Index

In most cases the *query optimizer* (page 108) selects the optimal index for a specific operation; however, you can force MongoDB to use a specific index using the `hint()` method. Use `hint()` to support performance testing, or on some queries where you must select a field or field included in several indexes.

Use the Increment Operator to Perform Operations Server-Side

Use MongoDB's `$inc` operator to increment or decrement values in documents. The operator increments the value of the field on the server side, as an alternative to selecting a document, making simple modifications in the client and then writing the entire document to the server. The `$inc` operator can also help avoid race conditions, which would result when two application instances queried for a document, manually incremented a field, and saved the entire document back at the same time.

Additional Resources

- MongoDB Performance Evaluation and Tuning Consulting Package[86]

Design Notes

[86] https://www.mongodb.com/products/consulting?jmp=docs#performance_evaluation

On this page

- Schema Considerations (page 316)
- General Considerations (page 316)
- Replica Set Considerations (page 317)
- Sharding Considerations (page 317)
- Analyze Performance (page 318)
- Additional Resources (page 318)

This page details features of MongoDB that may be important to keep in mind when developing applications.

Schema Considerations

Dynamic Schema Data in MongoDB has a *dynamic schema*. *Collections* do not enforce *document* structure. This facilitates iterative development and polymorphism. Nevertheless, collections often hold documents with highly homogeneous structures. See *Data Modeling Concepts* (page 252) for more information.

Some operational considerations include:

- the exact set of collections to be used;
- the indexes to be used: with the exception of the `_id` index, all indexes must be created explicitly;
- shard key declarations: choosing a good shard key is very important as the shard key cannot be changed once set.

Avoid importing unmodified data directly from a relational database. In general, you will want to "roll up" certain data into richer documents that take advantage of MongoDB's support for embedded documents and nested arrays.

Case Sensitive Strings MongoDB strings are case sensitive. So a search for `"joe"` will not find `"Joe"`.

Consider:

- storing data in a normalized case format, or
- using regular expressions ending with the `i` option, and/or
- using `$toLower` or `$toUpper` in the *aggregation framework* (page 199).

Type Sensitive Fields MongoDB data is stored in the BSON format, a binary encoded serialization of JSON-like documents. BSON encodes additional type information. See bsonspec.org[87] for more information.

Consider the following document which has a field `x` with the *string* value `"123"`:

```
{ x : "123" }
```

Then the following query which looks for a *number* value `123` will **not** return that document:

```
db.mycollection.find( { x : 123 } )
```

General Considerations

By Default, Updates Affect one Document To update multiple documents that meet your query criteria, set the `update multi` option to `true` or `1`. See: *Update Multiple Documents* (page 121).

[87] http://bsonspec.org/#/specification

Prior to MongoDB 2.2, you would specify the `upsert` and `multi` options in the `update` method as positional boolean options. See: the `update` method reference documentation.

BSON Document Size Limit The `BSON Document Size` limit is currently set at 16 MB per document. If you require larger documents, use *GridFS* (page 611).

No Fully Generalized Transactions MongoDB does not have *fully generalized transactions* (page 125). If you model your data using rich documents that closely resemble your application's objects, each logical object will be in one MongoDB document. MongoDB allows you to modify a document in a single atomic operation. These kinds of data modification pattern covers most common uses of transactions in other systems.

Replica Set Considerations

Use an Odd Number of Replica Set Members *Replica sets* (page 623) perform consensus elections. To ensure that elections will proceed successfully, either use an odd number of members, typically three, or else use an *arbiter* to ensure an odd number of votes.

Keep Replica Set Members Up-to-Date MongoDB replica sets support *automatic failover* (page 644). It is important for your secondaries to be up-to-date. There are various strategies for assessing consistency:

1. Use monitoring tools to alert you to lag events. See *Monitoring for MongoDB* (page 285) for a detailed discussion of MongoDB's monitoring options.
2. Specify appropriate write concern.
3. If your application requires *manual* fail over, you can configure your secondaries as *priority 0* (page 631). Priority 0 secondaries require manual action for a failover. This may be practical for a small replica set, but large deployments should fail over automatically.

See also:

replica set rollbacks (page 647).

Sharding Considerations

- Pick your shard keys carefully. You cannot choose a new shard key for a collection that is already sharded.
- Shard key values are immutable.
- When enabling sharding on an *existing collection*, MongoDB imposes a maximum size on those collections to ensure that it is possible to create chunks. For a detailed explanation of this limit, see: `<sharding-existing-collection-data-size>`.

  To shard large amounts of data, create a new empty sharded collection, and ingest the data from the source collection using an application level import operation.
- Unique indexes are not enforced across shards except for the shard key itself. See *Enforce Unique Keys for Sharded Collections* (page 818).
- Consider *pre-splitting* (page 808) an empty sharded collection before a massive bulk import.

Analyze Performance

As you develop and operate applications with MongoDB, you may want to analyze the performance of the database as the application. *Analyzing MongoDB Performance* (page 310) discusses some of the operational factors that can influence performance.

Additional Resources

- MongoDB Ops Optimization Consulting Package[88]

8.2 Administration Tutorials

The administration tutorials provide specific step-by-step instructions for performing common MongoDB setup, maintenance, and configuration operations.

Configuration, Maintenance, and Analysis **(page 318)** Describes routine management operations, including configuration and performance analysis.

> ***Manage mongod Processes*** **(page 323)** Start, configure, and manage running `mongod` process.
>
> ***Rotate Log Files*** **(page 330)** Archive the current log files and start new ones.
>
> Continue reading from *Configuration, Maintenance, and Analysis* (page 318) for additional tutorials of fundamental MongoDB maintenance procedures.

Backup and Recovery **(page 343)** Outlines procedures for data backup and restoration with `mongod` instances and deployments.

> ***Backup and Restore with Filesystem Snapshots*** **(page 343)** An outline of procedures for creating MongoDB data set backups using system-level file snapshot tool, such as *LVM* or native storage appliance tools.
>
> ***Backup and Restore Sharded Clusters*** **(page 355)** Detailed procedures and considerations for backing up sharded clusters and single shards.
>
> ***Recover Data after an Unexpected Shutdown*** **(page 366)** Recover data from MongoDB data files that were not properly closed or have an invalid state.
>
> Continue reading from *Backup and Recovery* (page 343) for additional tutorials of MongoDB backup and recovery procedures.

MongoDB Tutorials **(page 369)** A complete list of tutorials in the MongoDB Manual that address MongoDB operation and use.

8.2.1 Configuration, Maintenance, and Analysis

The following tutorials describe routine management operations, including configuration and performance analysis:

Disable Transparent Huge Pages (THP) **(page 319)** Describes Transparent Huge Pages (THP) and provides detailed instructions on disabling them.

Use Database Commands **(page 322)** The process for running database commands that provide basic database operations.

Manage mongod Processes **(page 323)** Start, configure, and manage running `mongod` process.

[88] https://www.mongodb.com/products/consulting?jmp=docs#ops_optimization

Terminate Running Operations **(page 325)** Stop in progress MongoDB client operations using `db.killOp()` and `maxTimeMS()`.

Analyze Performance of Database Operations **(page 326)** Collect data that introspects the performance of query and update operations on a `mongod` instance.

Rotate Log Files **(page 330)** Archive the current log files and start new ones.

Manage Journaling **(page 332)** Describes the procedures for configuring and managing MongoDB's journaling system, which allows MongoDB to provide crash resiliency and durability.

Store a JavaScript Function on the Server **(page 334)** Describes how to store JavaScript functions on a MongoDB server.

Upgrade to the Latest Revision of MongoDB **(page 335)** Introduces the basic process for upgrading a MongoDB deployment between different minor release versions.

Monitor MongoDB With SNMP on Linux **(page 338)** The SNMP extension, available in MongoDB Enterprise, allows MongoDB to provide database metrics via SNMP.

Monitor MongoDB Windows with SNMP **(page 340)** The SNMP extension, available in the Windows build of MongoDB Enterprise, allows MongoDB to provide database metrics via SNMP.

Troubleshoot SNMP **(page 342)** Outlines common errors and diagnostic processes useful for deploying MongoDB Enterprise with SNMP support.

Disable Transparent Huge Pages (THP)

On this page

- Init Script (page 319)
- Using `tuned` and `ktune` (page 320)
- Test Your Changes (page 321)

Transparent Huge Pages (THP) is a Linux memory management system that reduces the overhead of Translation Lookaside Buffer (TLB) lookups on machines with large amounts of memory by using larger memory pages.

However, database workloads often perform poorly with THP, because they tend to have sparse rather than contiguous memory access patterns. You should disable THP on Linux machines to ensure best performance with MongoDB.

Init Script

Important: If you are using `tuned` or `ktune` (for example, if you are running Red Hat or CentOS 6+), you must additionally configure them so that THP is not re-enabled. See *Using tuned and ktune* (page 320).

Step 1: Create the `init.d` script. Create the following file at `/etc/init.d/disable-transparent-hugepages`:

```
#!/bin/sh
### BEGIN INIT INFO
# Provides:          disable-transparent-hugepages
# Required-Start:    $local_fs
# Required-Stop:
# X-Start-Before:    mongod mongodb-mms-automation-agent
# Default-Start:     2 3 4 5
# Default-Stop:      0 1 6
```

```
# Short-Description: Disable Linux transparent huge pages
# Description:       Disable Linux transparent huge pages, to improve
#                    database performance.
### END INIT INFO

case $1 in
  start)
    if [ -d /sys/kernel/mm/transparent_hugepage ]; then
      thp_path=/sys/kernel/mm/transparent_hugepage
    elif [ -d /sys/kernel/mm/redhat_transparent_hugepage ]; then
      thp_path=/sys/kernel/mm/redhat_transparent_hugepage
    else
      return 0
    fi

    echo 'never' > ${thp_path}/enabled
    echo 'never' > ${thp_path}/defrag

    unset thp_path
    ;;
esac
```

Step 2: Make it executable. Run the following command to ensure that the init script can be used:

```
sudo chmod 755 /etc/init.d/disable-transparent-hugepages
```

Step 3: Configure your operating system to run it on boot. Use the appropriate command to configure the new init script on your Linux distribution.

Distribution	Command
Ubuntu and Debian	`sudo update-rc.d disable-transparent-hugepages def`
SUSE	`sudo insserv /etc/init.d/disable-transparent-hugep`
Red Hat, CentOS, Amazon Linux, and derivatives	`sudo chkconfig --add disable-transparent-hugepages`

Step 4: Override tuned and ktune, if applicable If you are using `tuned` or `ktune` (for example, if you are running Red Hat or CentOS 6+) you must now configure them to preserve the above settings.

Using `tuned` and `ktune`

Important: If using `tuned` or `ktune`, you must perform this step in addition to installing the init script.

`tuned` and `ktune` are dynamic kernel tuning tools available on Red Hat and CentOS that can disable transparent huge pages.

To disable transparent huge pages in `tuned` or `ktune`, you need to edit or create a new profile that sets THP to `never`.

Red Hat/CentOS 6

Step 1: Create a new profile. Create a new profile from an existing default profile by copying the relevant directory. In the example we use the `default` profile as the base and call our new profile `no-thp`.

```
sudo cp -r /etc/tune-profiles/default /etc/tune-profiles/no-thp
```

Step 2: Edit `ktune.sh`. Edit `/etc/tune-profiles/no-thp/ktune.sh` and add the following:

```
set_transparent_hugepages never
```

to the `start()` block of the file, before the `return 0` statement.

Step 3: Enable the new profile. Finally, enable the new profile by issuing:

```
sudo tuned-adm profile no-thp
```

Red Hat/CentOS 7

Step 1: Create a new profile. Create a new `tuned` profile directory:

```
sudo mkdir /etc/tuned/no-thp
```

Step 2: Edit `tuned.conf`. Create and edit `/etc/tuned/no-thp/tuned.conf` so that it contains the following:

```
[main]
include=virtual-guest

[vm]
transparent_hugepages=never
```

Step 3: Enable the new profile. Finally, enable the new profile by issuing:

```
sudo tuned-adm profile no-thp
```

Test Your Changes

You can check the status of THP support by issuing the following commands:

```
cat /sys/kernel/mm/transparent_hugepage/enabled
cat /sys/kernel/mm/transparent_hugepage/defrag
```

On Red Hat Enterprise Linux, CentOS, and potentially other Red Hat-based derivatives, you may instead need to use the following:

```
cat /sys/kernel/mm/redhat_transparent_hugepage/enabled
cat /sys/kernel/mm/redhat_transparent_hugepage/defrag
```

For both files, the correct output resembles:

```
always madvise [never]
```

Use Database Commands

On this page

- Database Command Form (page 322)
- Issue Commands (page 322)
- admin Database Commands (page 322)
- Command Responses (page 322)

The MongoDB command interface provides access to all *non CRUD* database operations. Fetching server stats, initializing a replica set, and running a map-reduce job are all accomplished with commands.

See https://docs.mongodb.org/manual/reference/command for list of all commands sorted by function.

Database Command Form

You specify a command first by constructing a standard *BSON* document whose first key is the name of the command. For example, specify the isMaster command using the following *BSON* document:

```
{ isMaster: 1 }
```

Issue Commands

The mongo shell provides a helper method for running commands called db.runCommand(). The following operation in mongo runs the above command:

```
db.runCommand( { isMaster: 1 } )
```

Many drivers provide an equivalent for the db.runCommand() method. Internally, running commands with db.runCommand() is equivalent to a special query against the *$cmd* collection.

Many common commands have their own shell helpers or wrappers in the mongo shell and drivers, such as the db.isMaster() method in the mongo JavaScript shell.

You can use the maxTimeMS option to specify a time limit for the execution of a command, see *Terminate a Command* (page 326) for more information on operation termination.

admin Database Commands

You must run some commands on the *admin database*. Normally, these operations resemble the followings:

```
use admin
db.runCommand( {buildInfo: 1} )
```

However, there's also a command helper that automatically runs the command in the context of the admin database:

```
db._adminCommand( {buildInfo: 1} )
```

Command Responses

All commands return, at minimum, a document with an ok field indicating whether the command has succeeded:

```
{ 'ok': 1 }
```

Failed commands return the `ok` field with a value of `0`.

Manage `mongod` Processes

On this page

- Start `mongod` Processes (page 323)
- Stop `mongod` Processes (page 324)
- Stop a Replica Set (page 324)

MongoDB runs as a standard program. You can start MongoDB from a command line by issuing the `mongod` command and specifying options. For a list of options, see the `mongod` reference. MongoDB can also run as a Windows service. For details, see *Configure a Windows Service for MongoDB Community Edition* (page 47). To install MongoDB, see *Install MongoDB* (page 21).

The following examples assume the directory containing the `mongod` process is in your system paths. The `mongod` process is the primary database process that runs on an individual server. `mongos` provides a coherent MongoDB interface equivalent to a `mongod` from the perspective of a client. The `mongo` binary provides the administrative shell.

This document discusses the `mongod` process; however, some portions of this document may be applicable to `mongos` instances.

Start `mongod` Processes

By default, MongoDB stores data in the `/data/db` directory. On Windows, MongoDB stores data in `C:\data\db`. On all platforms, MongoDB listens for connections from clients on port `27017`.

To start MongoDB using all defaults, issue the following command at the system shell:

```
mongod
```

Specify a Data Directory If you want `mongod` to store data files at a path *other than* `/data/db` you can specify a `dbPath`. The `dbPath` must exist before you start `mongod`. If it does not exist, create the directory and the permissions so that `mongod` can read and write data to this path. For more information on permissions, see the *security operations documentation* (page 391).

To specify a `dbPath` for `mongod` to use as a data directory, use the *`--dbpath`* option. The following invocation will start a `mongod` instance and store data in the `/srv/mongodb` path

```
mongod --dbpath /srv/mongodb/
```

Specify a TCP Port Only a single process can listen for connections on a network interface at a time. If you run multiple `mongod` processes on a single machine, or have other processes that must use this port, you must assign each a different port to listen on for client connections.

To specify a port to `mongod`, use the *`--port`* option on the command line. The following command starts `mongod` listening on port `12345`:

```
mongod --port 12345
```

Use the default port number when possible, to avoid confusion.

Start `mongod` as a Daemon To run a `mongod` process as a daemon (i.e. `fork`), *and* write its output to a log file, use the `--fork` and `--logpath` options. You must create the log directory; however, `mongod` will create the log file if it does not exist.

The following command starts `mongod` as a daemon and records log output to `/var/log/mongodb.log`.

```
mongod --fork --logpath /var/log/mongodb.log
```

Additional Configuration Options For an overview of common configurations and deployments for common use cases, see *Run-time Database Configuration* (page 291).

Stop `mongod` Processes

In a clean shutdown a `mongod` completes all pending operations, flushes all data to data files, and closes all data files. Other shutdowns are *unclean* and can compromise the validity of the data files.

To ensure a clean shutdown, always shutdown `mongod` instances using one of the following methods:

Use `shutdownServer()` Shut down the `mongod` from the `mongo` shell using the `db.shutdownServer()` method as follows:

```
use admin
db.shutdownServer()
```

Calling the same method from a *init script* accomplishes the same result.

For systems with `authorization` enabled, users may only issue `db.shutdownServer()` when authenticated to the `admin` database or via the localhost interface on systems without authentication enabled.

Use `--shutdown` From the Linux command line, shut down the `mongod` using the `--shutdown` option in the following command:

```
mongod --shutdown
```

Use `CTRL-C` When running the `mongod` instance in interactive mode (i.e. without `--fork`), issue `Control-C` to perform a clean shutdown.

Use `kill` From the Linux command line, shut down a specific `mongod` instance using the following command:

```
kill <mongod process ID>
```

> **Warning:** Never use `kill -9` (i.e. `SIGKILL`) to terminate a mongod instance.

Stop a Replica Set

Procedure If the `mongod` is the *primary* in a *replica set*, the shutdown process for this `mongod` instance has the following steps:

1. Check how up-to-date the *secondaries* are.
2. If no secondary is within 10 seconds of the primary, `mongod` will return a message that it will not shut down. You can pass the `shutdown` command a `timeoutSecs` argument to wait for a secondary to catch up.

3. If there is a secondary within 10 seconds of the primary, the primary will step down and wait for the secondary to catch up.
4. After 60 seconds or once the secondary has caught up, the primary will shut down.

Force Replica Set Shutdown If there is no up-to-date secondary and you want the primary to shut down, issue the `shutdown` command with the `force` argument, as in the following `mongo` shell operation:

```
db.adminCommand({shutdown : 1, force : true})
```

To keep checking the secondaries for a specified number of seconds if none are immediately up-to-date, issue `shutdown` with the `timeoutSecs` argument. MongoDB will keep checking the secondaries for the specified number of seconds if none are immediately up-to-date. If any of the secondaries catch up within the allotted time, the primary will shut down. If no secondaries catch up, it will not shut down.

The following command issues `shutdown` with `timeoutSecs` set to `5`:

```
db.adminCommand({shutdown : 1, timeoutSecs : 5})
```

Alternately you can use the `timeoutSecs` argument with the `db.shutdownServer()` method:

```
db.shutdownServer({timeoutSecs : 5})
```

Terminate Running Operations

On this page

- Overview (page 325)
- Available Procedures (page 325)

Overview

MongoDB provides two facilitates to terminate running operations: `maxTimeMS()` and `db.killOp()`. Use these operations as needed to control the behavior of operations in a MongoDB deployment.

Available Procedures

`maxTimeMS` New in version 2.6.

The `maxTimeMS()` method sets a time limit for an operation. When the operation reaches the specified time limit, MongoDB interrupts the operation at the next *interrupt point*.

Terminate a Query From the `mongo` shell, use the following method to set a time limit of 30 milliseconds for this query:

```
db.location.find( { "town": { "$regex": "(Pine Lumber)",
                              "$options": 'i' } } ).maxTimeMS(30)
```

Terminate a Command Consider a potentially long running operation using `distinct` to return each distinct‘‘collection‘‘ field that has a `city` key:

```
db.runCommand( { distinct: "collection",
                 key: "city" } )
```

You can add the `maxTimeMS` field to the command document to set a time limit of 45 milliseconds for the operation:

```
db.runCommand( { distinct: "collection",
                 key: "city",
                 maxTimeMS: 45 } )
```

`db.getLastError()` and `db.getLastErrorObj()` will return errors for interrupted options:

```
{ "n" : 0,
  "connectionId" : 1,
  "err" : "operation exceeded time limit",
  "ok" : 1 }
```

`killOp` The `db.killOp()` method interrupts a running operation at the next *interrupt point*. `db.killOp()` identifies the target operation by operation ID.

```
db.killOp(<opId>)
```

Warning: Terminate running operations with extreme caution. Only use `db.killOp()` to terminate operations initiated by clients and *do not* terminate internal database operations.

Related

To return a list of running operations see `db.currentOp()`.

Analyze Performance of Database Operations

On this page

- Profiling Levels (page 327)
- Enable Database Profiling and Set the Profiling Level (page 327)
- View Profiler Data (page 328)
- Profiler Overhead (page 329)
- Additional Resources (page 330)

The database profiler collects fine grained data about MongoDB write operations, cursors, database commands on a running `mongod` instance. You can enable profiling on a per-database or per-instance basis. The *profiling level* (page 327) is also configurable when enabling profiling. The profiler is *off* by default.

The database profiler writes all the data it collects to the `system.profile` (page 377) collection, which is a *capped collection* (page 6). See *Database Profiler Output* (page 378) for overview of the data in the `system.profile` (page 377) documents created by the profiler.

This document outlines a number of key administration options for the database profiler. For additional related information, consider the following resources:

- *Database Profiler Output* (page 378)
- `Profile Command`

- db.currentOp()

Profiling Levels

The following profiling levels are available:

- 0 - the profiler is off, does not collect any data. mongod always writes operations longer than the slowOpThresholdMs threshold to its log. This is the default profiler level.
- 1 - collects profiling data for slow operations only. By default slow operations are those slower than 100 milliseconds.

  You can modify the threshold for "slow" operations with the slowOpThresholdMs runtime option or the setParameter command. See the *Specify the Threshold for Slow Operations* (page 327) section for more information.
- 2 - collects profiling data for all database operations.

Enable Database Profiling and Set the Profiling Level

You can enable database profiling from the mongo shell or through a driver using the profile command. This section will describe how to do so from the mongo shell. See your driver documentation if you want to control the profiler from within your application.

When you enable profiling, you also set the *profiling level* (page 327). The profiler records data in the system.profile (page 377) collection. MongoDB creates the system.profile (page 377) collection in a database after you enable profiling for that database.

To enable profiling and set the profiling level, use the db.setProfilingLevel() helper in the mongo shell, passing the profiling level as a parameter. For example, to enable profiling for all database operations, consider the following operation in the mongo shell:

```
db.setProfilingLevel(2)
```

The shell returns a document showing the *previous* level of profiling. The "ok" : 1 key-value pair indicates the operation succeeded:

```
{ "was" : 0, "slowms" : 100, "ok" : 1 }
```

To verify the new setting, see the *Check Profiling Level* (page 328) section.

Specify the Threshold for Slow Operations The threshold for slow operations applies to the entire mongod instance. When you change the threshold, you change it for all databases on the instance.

Important: Changing the slow operation threshold for the database profiler also affects the profiling subsystem's slow operation threshold for the entire mongod instance. Always set the threshold to the highest useful value.

By default the slow operation threshold is 100 milliseconds. Databases with a profiling level of 1 will log operations slower than 100 milliseconds.

To change the threshold, pass two parameters to the db.setProfilingLevel() helper in the mongo shell. The first parameter sets the profiling level for the current database, and the second sets the default slow operation threshold *for the entire* mongod *instance*.

For example, the following command sets the profiling level for the current database to 0, which disables profiling, and sets the slow-operation threshold for the mongod instance to 20 milliseconds. Any database on the instance with a profiling level of 1 will use this threshold:

```
db.setProfilingLevel(0,20)
```

Check Profiling Level To view the *profiling level* (page 327), issue the following from the `mongo` shell:

```
db.getProfilingStatus()
```

The shell returns a document similar to the following:

```
{ "was" : 0, "slowms" : 100 }
```

The `was` field indicates the current level of profiling.

The `slowms` field indicates how long an operation must exist in milliseconds for an operation to pass the "slow" threshold. MongoDB will log operations that take longer than the threshold if the profiling level is `1`. This document returns the profiling level in the `was` field. For an explanation of profiling levels, see *Profiling Levels* (page 327).

To return only the profiling level, use the `db.getProfilingLevel()` helper in the `mongo` as in the following:

```
db.getProfilingLevel()
```

Disable Profiling To disable profiling, use the following helper in the `mongo` shell:

```
db.setProfilingLevel(0)
```

Enable Profiling for an Entire `mongod` Instance For development purposes in testing environments, you can enable database profiling for an entire `mongod` instance. The profiling level applies to all databases provided by the `mongod` instance.

To enable profiling for a `mongod` instance, pass the following parameters to `mongod` at startup or within the `configuration file`:

```
mongod --profile=1 --slowms=15
```

This sets the profiling level to `1`, which collects profiling data for slow operations only, and defines slow operations as those that last longer than `15` milliseconds.

See also:

`mode` and `slowOpThresholdMs`.

Database Profiling and Sharding You *cannot* enable profiling on a `mongos` instance. To enable profiling in a shard cluster, you must enable profiling for each `mongod` instance in the cluster.

View Profiler Data

The database profiler logs information about database operations in the `system.profile` (page 377) collection.

To view profiling information, query the `system.profile` (page 377) collection. You can use `$comment` to add data to the query document to make it easier to analyze data from the profiler. To view example queries, see *Profiler Overhead* (page 329).

For an explanation of the output data, see *Database Profiler Output* (page 378).

Example Profiler Data Queries This section displays example queries to the `system.profile` (page 377) collection. For an explanation of the query output, see *Database Profiler Output* (page 378).

To return the most recent 10 log entries in the `system.profile` (page 377) collection, run a query similar to the following:

```
db.system.profile.find().limit(10).sort( { ts : -1 } ).pretty()
```

To return all operations except command operations (*$cmd*), run a query similar to the following:

```
db.system.profile.find( { op: { $ne : 'command' } } ).pretty()
```

To return operations for a particular collection, run a query similar to the following. This example returns operations in the `mydb` database's `test` collection:

```
db.system.profile.find( { ns : 'mydb.test' } ).pretty()
```

To return operations slower than `5` milliseconds, run a query similar to the following:

```
db.system.profile.find( { millis : { $gt : 5 } } ).pretty()
```

To return information from a certain time range, run a query similar to the following:

```
db.system.profile.find(
                        {
                         ts : {
                                $gt : new ISODate("2012-12-09T03:00:00Z") ,
                                $lt : new ISODate("2012-12-09T03:40:00Z")
                               }
                        }
                       ).pretty()
```

The following example looks at the time range, suppresses the `user` field from the output to make it easier to read, and sorts the results by how long each operation took to run:

```
db.system.profile.find(
                        {
                          ts : {
                                 $gt : new ISODate("2011-07-12T03:00:00Z") ,
                                 $lt : new ISODate("2011-07-12T03:40:00Z")
                                }
                        },
                        { user : 0 }
                       ).sort( { millis : -1 } )
```

Show the Five Most Recent Events On a database that has profiling enabled, the `show profile` helper in the `mongo` shell displays the 5 most recent operations that took at least 1 millisecond to execute. Issue `show profile` from the `mongo` shell, as follows:

```
show profile
```

Profiler Overhead

When enabled, profiling has a minor effect on performance. The `system.profile` (page 377) collection is a *capped collection* with a default size of 1 megabyte. A collection of this size can typically store several thousand profile documents, but some application may use more or less profiling data per operation.

Change Size of `system.profile` Collection on the Primary To change the size of the `system.profile` (page 377) collection, you must:

1. Disable profiling.
2. Drop the `system.profile` (page 377) collection.
3. Create a new `system.profile` (page 377) collection.
4. Re-enable profiling.

For example, to create a new `system.profile` (page 377) collections that's `4000000` bytes, use the following sequence of operations in the `mongo` shell:

```
db.setProfilingLevel(0)

db.system.profile.drop()

db.createCollection( "system.profile", { capped: true, size:4000000 } )

db.setProfilingLevel(1)
```

Change Size of `system.profile` Collection on a Secondary To change the size of the `system.profile` (page 377) collection on a *secondary*, you must stop the secondary, run it as a standalone, and then perform the steps above. When done, restart the standalone as a member of the replica set. For more information, see *Perform Maintenance on Replica Set Members* (page 695).

Additional Resources

- MongoDB Performance Evaluation and Tuning Consulting Package[89]

Rotate Log Files

On this page

- Overview (page 330)
- Default Log Rotation Behavior (page 331)
- Log Rotation with `--logRotate reopen` (page 331)
- Syslog Log Rotation (page 332)
- Forcing a Log Rotation with `SIGUSR1` (page 332)

Overview

When used with the *`--logpath`* option or `systemLog.path` setting, `mongod` and `mongos` instances report a live account of all activity and operations to a log file. When reporting activity data to a log file, by default, MongoDB only rotates logs in response to the `logRotate` command, or when the `mongod` or `mongos` process receives a `SIGUSR1` signal from the operating system.

MongoDB's standard log rotation approach archives the current log file and starts a new one. To do this, the `mongod` or `mongos` instance renames the current log file by appending a UTC timestamp to the filename, in *ISODate* format. It then opens a new log file, closes the old log file, and sends all new log entries to the new log file.

[89] https://www.mongodb.com/products/consulting?jmp=docs#performance_evaluation

You can also configure MongoDB to support the Linux/Unix logrotate utility by setting `systemLog.logRotate` or *`--logRotate`* to `reopen`. With `reopen`, `mongod` or `mongos` closes the log file, and then reopens a log file with the same name, expecting that another process renamed the file prior to rotation.

Finally, you can configure `mongod` to send log data to the `syslog`. using the *`--syslog`* option. In this case, you can take advantage of alternate logrotation tools.

See also:

For information on logging, see the *Process Logging* (page 288) section.

Default Log Rotation Behavior

By default, MongoDB uses the *`--logRotate rename`* behavior. With `rename`, `mongod` or `mongos` renames the current log file by appending a UTC timestamp to the filename, opens a new log file, closes the old log file, and sends all new log entries to the new log file.

Step 1: Start a `mongod` instance.

```
mongod -v --logpath /var/log/mongodb/server1.log
```

You can also explicitly specify *`logRotate --rename`*.

Step 2: List the log files In a separate terminal, list the matching files:

```
ls /var/log/mongodb/server1.log*
```

The results should include one log file, `server1.log`.

Step 3: Rotate the log file. Rotate the log file by issuing the `logRotate` command from the `admin` database in a `mongo` shell:

```
use admin
db.runCommand( { logRotate : 1 } )
```

Step 4: View the new log files List the new log files to view the newly-created log:

```
ls /var/log/mongodb/server1.log*
```

There should be two log files listed: `server1.log`, which is the log file that `mongod` or `mongos` made when it reopened the log file, and `server1.log.<timestamp>`, the renamed original log file.

Rotating log files does not modify the "old" rotated log files. When you rotate a log, you rename the `server1.log` file to include the timestamp, and a new, empty `server1.log` file receives all new log input.

Log Rotation with `--logRotate reopen`

New in version 3.0.0.

Log rotation with *`--logRotate reopen`* closes and opens the log file following the typical Linux/Unix log rotate behavior.

Step 1: Start a `mongod` instance, specifying the `reopen` *`--logRotate`* behavior.

```
mongod -v --logpath /var/log/mongodb/server1.log --logRotate reopen --logappend
```

You must use the *`--logappend`* option with *`--logRotate reopen`*.

Step 2: List the log files In a separate terminal, list the matching files:

```
ls /var/log/mongodb/server1.log*
```

The results should include one log file, `server1.log`.

Step 3: Rotate the log file. Rotate the log file by issuing the `logRotate` command from the `admin` database in a `mongo` shell:

```
use admin
db.runCommand( { logRotate : 1 } )
```

You should rename the log file using an external process, following the typical Linux/Unix log rotate behavior.

Syslog Log Rotation

With syslog log rotation, `mongod` sends log data to the syslog rather than writing it to a file.

Step 1: Start a `mongod` instance with the *`--syslog`* option

```
mongod --syslog
```

Do not include *`--logpath`*. Since *`--syslog`* tells `mongod` to send log data to the syslog, specifying a *`--logpath`* will causes an error.

To specify the facility level used when logging messages to the syslog, use the *`--syslogFacility`* option or `systemLog.syslogFacility` configuration setting.

Step 2: Rotate the log. Store and rotate the log output using your systems default log rotation mechanism.

Forcing a Log Rotation with `SIGUSR1`

For Linux and Unix-based systems, you can use the `SIGUSR1` signal to rotate the logs for a single process, as in the following:

```
kill -SIGUSR1 <mongod process id>
```

Manage Journaling

On this page

- Procedures (page 333)

MongoDB uses *write ahead logging* to an on-disk *journal* to guarantee *write operation* (page 114) durability. The MMAPv1 storage engine also requires the *journal* in order to provide crash resiliency.

The WiredTiger storage engine does not require journaling to guarantee a consistent state after a crash. The database will be restored to the last consistent *checkpoint* (page 596) during recovery. However, if MongoDB exits unexpectedly in between checkpoints, journaling is required to recover writes that occurred after the last checkpoint.

With journaling enabled, if `mongod` stops unexpectedly, the program can recover everything written to the journal. MongoDB will re-apply the write operations on restart and maintain a consistent state. By default, the greatest extent of lost writes, i.e., those not made to the journal, are those made in the last 100 milliseconds, plus the time it takes to perform the actual journal writes. See `commitIntervalMs` for more information on the default.

Procedures

Enable Journaling For 64-bit builds of `mongod`, journaling is enabled by default.

To enable journaling, start `mongod` with the `--journal` command line option.

Disable Journaling

> **Warning:** Do not disable journaling on production systems. When using the MMAPv1 storage engine *withou*
> journal, if your `mongod` instance stops without shutting down cleanly unexpectedly for any reason, (e.g. pow
> failure) and you are not running with journaling, then you must recover from an unaffected *replica set* member
> backup, as described in *repair* (page 366).

To disable journaling, start `mongod` with the `--nojournal` command line option.

Get Commit Acknowledgment You can get commit acknowledgment with the *Write Concern* (page 179) and the `j` (page 181) option. For details, see *Write Concern* (page 179).

Avoid Preallocation Lag for MMAPv1 With the *MMAPv1 storage engine* (page 603), MongoDB may preallocate journal files if the `mongod` process determines that it is more efficient to preallocate journal files than create new journal files as needed.

Depending on your filesystem, you might experience a preallocation lag the first time you start a `mongod` instance with journaling enabled. The amount of time required to pre-allocate files might last several minutes; during this time, you will not be able to connect to the database. This is a one-time preallocation and does not occur with future invocations.

To avoid *preallocation lag* (page 608), you can preallocate files in the journal directory by copying them from another instance of `mongod`.

Preallocated files do not contain data. It is safe to later remove them. But if you restart `mongod` with journaling, `mongod` will create them again.

Example

The following sequence preallocates journal files for an instance of `mongod` running on port `27017` with a database path of `/data/db`.

For demonstration purposes, the sequence starts by creating a set of journal files in the usual way.

1. Create a temporary directory into which to create a set of journal files:

   ```
   mkdir ~/tmpDbpath
   ```

2. Create a set of journal files by staring a `mongod` instance that uses the temporary directory:

```
mongod --port 10000 --dbpath ~/tmpDbpath --journal
```

3. When you see the following log output, indicating `mongod` has the files, press CONTROL+C to stop the `mongod` instance:

   ```
   [initandlisten] waiting for connections on port 10000
   ```

4. Preallocate journal files for the new instance of `mongod` by moving the journal files from the data directory of the existing instance to the data directory of the new instance:

   ```
   mv ~/tmpDbpath/journal /data/db/
   ```

5. Start the new `mongod` instance:

   ```
   mongod --port 27017 --dbpath /data/db --journal
   ```

Monitor Journal Status Use the following commands and methods to monitor journal status:

- `serverStatus`

  The `serverStatus` command returns database status information that is useful for assessing performance.

- `journalLatencyTest`

  Use `journalLatencyTest` to measure how long it takes on your volume to write to the disk in an append-only fashion. You can run this command on an idle system to get a baseline sync time for journaling. You can also run this command on a busy system to see the sync time on a busy system, which may be higher if the journal directory is on the same volume as the data files.

  The `journalLatencyTest` command also provides a way to check if your disk drive is buffering writes in its local cache. If the number is very low (i.e., less than 2 milliseconds) and the drive is non-SSD, the drive is probably buffering writes. In that case, enable cache write-through for the device in your operating system, unless you have a disk controller card with battery backed RAM.

Change the Group Commit Interval for MMAPv1 For the *MMAPv1 storage engine* (page 603), you can set the group commit interval using the `--journalCommitInterval` command line option. The allowed range is `2` to `300` milliseconds.

Lower values increase the durability of the journal at the expense of disk performance.

Recover Data After Unexpected Shutdown On a restart after a crash, MongoDB replays all journal files in the journal directory before the server becomes available. If MongoDB must replay journal files, `mongod` notes these events in the log output.

There is no reason to run `repairDatabase` in these situations.

Store a JavaScript Function on the Server

Note: Do not store application logic in the database. There are performance limitations to running JavaScript inside of MongoDB. Application code also is typically most effective when it shares version control with the application itself.

There is a special system collection named `system.js` that can store JavaScript functions for reuse.

To store a function, you can use the `db.collection.save()`, as in the following examples:

```
db.system.js.save(
   {
     _id: "echoFunction",
     value : function(x) { return x; }
   }
)

db.system.js.save(
   {
     _id : "myAddFunction" ,
     value : function (x, y){ return x + y; }
   }
);
```

- The `_id` field holds the name of the function and is unique per database.
- The `value` field holds the function definition.

Once you save a function in the `system.js` collection, you can use the function from any JavaScript context; e.g. `$where` operator, `mapReduce` command or `db.collection.mapReduce()`.

In the `mongo` shell, you can use `db.loadServerScripts()` to load all the scripts saved in the `system.js` collection for the current database. Once loaded, you can invoke the functions directly in the shell, as in the following example:

```
db.loadServerScripts();

echoFunction(3);

myAddFunction(3, 5);
```

Upgrade to the Latest Revision of MongoDB

On this page

- Before Upgrading (page 335)
- Upgrade Procedure (page 336)
- Upgrade a MongoDB Instance (page 336)
- Replace the Existing Binaries (page 336)
- Upgrade Sharded Clusters (page 337)
- Upgrade Replica Sets (page 337)
- Additional Resources (page 338)

Revisions provide security patches, bug fixes, and new or changed features that do not contain any backward breaking changes. Always upgrade to the latest revision in your release series. The third number in the *MongoDB version number* (page 1070) indicates the revision.

Before Upgrading

- Ensure you have an up-to-date backup of your data set. See *MongoDB Backup Methods* (page 282).
- Consult the following documents for any special considerations or compatibility issues specific to your MongoDB release:
  - The release notes, located at *Release Notes* (page 865).

  - The documentation for your driver. See Drivers[90] and Driver Compatibility[91] pages for more information.
- If your installation includes *replica sets*, plan the upgrade during a predefined maintenance window.
- Before you upgrade a production environment, use the procedures in this document to upgrade a *staging* environment that reproduces your production environment, to ensure that your production configuration is compatible with all changes.

Upgrade Procedure

Important: Always backup all of your data before upgrading MongoDB.

Upgrade each `mongod` and `mongos` binary separately, using the procedure described here. When upgrading a binary, use the procedure *Upgrade a MongoDB Instance* (page 336).

Follow this upgrade procedure:

1. For deployments that use authentication, first upgrade all of your MongoDB `drivers`. To upgrade, see the documentation for your driver as well as the Driver Compatibility[92] page.
2. Upgrade sharded clusters, as described in *Upgrade Sharded Clusters* (page 337).
3. Upgrade any standalone instances. See *Upgrade a MongoDB Instance* (page 336).
4. Upgrade any replica sets that are not part of a sharded cluster, as described in *Upgrade Replica Sets* (page 337).

Upgrade a MongoDB Instance

To upgrade a `mongod` or `mongos` instance, use one of the following approaches:

- Upgrade the instance using the operating system's package management tool and the official MongoDB packages. This is the preferred approach. See *Install MongoDB* (page 21).
- Upgrade the instance by replacing the existing binaries with new binaries. See *Replace the Existing Binaries* (page 336).

Replace the Existing Binaries

Important: Always backup all of your data before upgrading MongoDB.

This section describes how to upgrade MongoDB by replacing the existing binaries. The preferred approach to an upgrade is to use the operating system's package management tool and the official MongoDB packages, as described in *Install MongoDB* (page 21).

To upgrade a `mongod` or `mongos` instance by replacing the existing binaries:

1. Download the binaries for the latest MongoDB revision from the MongoDB Download Page[93] and store the binaries in a temporary location. The binaries download as compressed files that uncompress to the directory structure used by the MongoDB installation.
2. Shutdown the instance.
3. Replace the existing MongoDB binaries with the downloaded binaries.

[90] https://docs.mongodb.org/ecosystem/drivers
[91] https://docs.mongodb.org/ecosystem/drivers/driver-compatibility-reference
[92] https://docs.mongodb.org/ecosystem/drivers/driver-compatibility-reference
[93] http://downloads.mongodb.org/

4. Restart the instance.

Upgrade Sharded Clusters

To upgrade a sharded cluster:

1. Disable the cluster's balancer, as described in *Disable the Balancer* (page 802).
2. Upgrade each `mongos` instance by following the instructions below in *Upgrade a MongoDB Instance* (page 336). You can upgrade the `mongos` instances in any order.
3. Upgrade each `mongod` *config server* (page 742) individually starting with the last config server listed in your *`mongos --configdb`* string and working backward. To keep the cluster online, make sure at least one config server is always running. For each config server upgrade, follow the instructions below in *Upgrade a MongoDB Instance* (page 336)

   Example

   Given the following config string:

   ```
   mongos --configdb cfg0.example.net:27019,cfg1.example.net:27019,cfg2.example.net:27019
   ```

   You would upgrade the config servers in the following order:

   (a) cfg2.example.net

   (b) cfg1.example.net

   (c) cfg0.example.net

   ---

4. Upgrade each shard.
   - If a shard is a replica set, upgrade the shard using the procedure below titled *Upgrade Replica Sets* (page 337).
   - If a shard is a standalone instance, upgrade the shard using the procedure below titled *Upgrade a MongoDB Instance* (page 336).
5. Re-enable the balancer, as described in *Enable the Balancer* (page 803).

Upgrade Replica Sets

To upgrade a replica set, upgrade each member individually, starting with the *secondaries* and finishing with the *primary*. Plan the upgrade during a predefined maintenance window.

Upgrade Secondaries Upgrade each secondary separately as follows:

1. Upgrade the secondary's `mongod` binary by following the instructions below in *Upgrade a MongoDB Instance* (page 336).
2. After upgrading a secondary, wait for the secondary to recover to the `SECONDARY` state before upgrading the next instance. To check the member's state, issue `rs.status()` in the `mongo` shell.

   The secondary may briefly go into `STARTUP2` or `RECOVERING`. This is normal. Make sure to wait for the secondary to fully recover to `SECONDARY` before you continue the upgrade.

Upgrade the Primary

1. Step down the primary to initiate the normal *failover* (page 644) procedure. Using one of the following:
   - The `rs.stepDown()` helper in the `mongo` shell.
   - The `replSetStepDown` database command.

   During failover, the set cannot accept writes. Typically this takes 10-20 seconds. Plan the upgrade during a predefined maintenance window.

   Note: Stepping down the primary is preferable to directly *shutting down* the primary. Stepping down expedites the failover procedure.

2. Once the primary has stepped down, call the `rs.status()` method from the `mongo` shell until you see that another member has assumed the `PRIMARY` state.
3. Shut down the original primary and upgrade its instance by following the instructions below in *Upgrade a MongoDB Instance* (page 336).

Additional Resources

- Getting ready for MongoDB 3.2? Get our help.[94]

Monitor MongoDB With SNMP on Linux

On this page

- Overview (page 338)
- Considerations (page 339)
- Configuration Files (page 339)
- Procedure (page 339)
- Optional: Run MongoDB as SNMP Master (page 340)

Enterprise Feature

SNMP is only available in MongoDB Enterprise[95].

Overview

MongoDB Enterprise can provide database metrics via SNMP, in support of centralized data collection and aggregation. This procedure explains the setup and configuration of a `mongod` instance as an SNMP subagent, as well as initializing and testing of SNMP support with MongoDB Enterprise.

See also:

Troubleshoot SNMP (page 342) and *Monitor MongoDB Windows with SNMP* (page 340) for complete instructions on using MongoDB with SNMP on Windows systems.

[94] https://www.mongodb.com/contact/mongodb-3-2-upgrade-services?jmp=docs
[95] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Considerations

Only `mongod` instances provide SNMP support. `mongos` and the other MongoDB binaries do not support SNMP.

Configuration Files

Changed in version 2.6.

MongoDB Enterprise contains the following configuration files to support SNMP:

- `MONGOD-MIB.txt`:

  The management information base (MIB) file that defines MongoDB's SNMP output.

- `mongod.conf.subagent`:

  The configuration file to run `mongod` as the SNMP subagent. This file sets SNMP run-time configuration options, including the `AgentX` socket to connect to the SNMP master.

- `mongod.conf.master`:

  The configuration file to run `mongod` as the SNMP master. This file sets SNMP run-time configuration options.

Procedure

Step 1: Copy configuration files. Use the following sequence of commands to move the SNMP configuration files to the SNMP service configuration directory.

First, create the SNMP configuration directory if needed and then, from the installation directory, copy the configuration files to the SNMP service configuration directory:

```
mkdir -p /etc/snmp/
cp MONGOD-MIB.txt /usr/share/snmp/mibs/MONGOD-MIB.txt
cp mongod.conf.subagent /etc/snmp/mongod.conf
```

The configuration filename is tool-dependent. For example, when using `net-snmp` the configuration file is `snmpd.conf`.

By default SNMP uses UNIX domain for communication between the agent (i.e. `snmpd` or the master) and sub-agent (i.e. MongoDB).

Ensure that the `agentXAddress` specified in the SNMP configuration file for MongoDB matches the `agentXAddress` in the SNMP master configuration file.

Step 2: Start MongoDB. Start `mongod` with the `snmp-subagent` to send data to the SNMP master.

```
mongod --snmp-subagent
```

Step 3: Confirm SNMP data retrieval. Use `snmpwalk` to collect data from `mongod`:

Connect an SNMP client to verify the ability to collect SNMP data from MongoDB.

Install the net-snmp[96] package to access the `snmpwalk` client. `net-snmp` provides the `snmpwalk` SNMP client.

```
snmpwalk -m /usr/share/snmp/mibs/MONGOD-MIB.txt -v 2c -c mongodb 127.0.0.1:<port> 1.3.6.1.4.1.34601
```

[96] http://www.net-snmp.org/

<port> refers to the port defined by the SNMP master, *not* the primary port used by mongod for client communication.

Optional: Run MongoDB as SNMP Master

You can run mongod with the snmp-master option for testing purposes. To do this, use the SNMP master configuration file instead of the subagent configuration file. From the directory containing the unpacked MongoDB installation files:

```
cp mongod.conf.master /etc/snmp/mongod.conf
```

Additionally, start mongod with the snmp-master option, as in the following:

```
mongod --snmp-master
```

Monitor MongoDB Windows with SNMP

On this page

- Overview (page 340)
- Considerations (page 340)
- Configuration Files (page 341)
- Procedure (page 341)
- Optional: Run MongoDB as SNMP Master (page 342)

New in version 2.6.

Enterprise Feature

SNMP is only available in MongoDB Enterprise[97].

Overview

MongoDB Enterprise can provide database metrics via SNMP, in support of centralized data collection and aggregation. This procedure explains the setup and configuration of a mongod.exe instance as an SNMP subagent, as well as initializing and testing of SNMP support with MongoDB Enterprise.

See also:

Monitor MongoDB With SNMP on Linux (page 338) and *Troubleshoot SNMP* (page 342) for more information.

Considerations

Only mongod.exe instances provide SNMP support. mongos.exe and the other MongoDB binaries do not support SNMP.

[97] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Configuration Files

Changed in version 2.6.

MongoDB Enterprise contains the following configuration files to support SNMP:

- `MONGOD-MIB.txt`:

  The management information base (MIB) file that defines MongoDB's SNMP output.

- `mongod.conf.subagent`:

  The configuration file to run `mongod.exe` as the SNMP subagent. This file sets SNMP run-time configuration options, including the `AgentX` socket to connect to the SNMP master.

- `mongod.conf.master`:

  The configuration file to run `mongod.exe` as the SNMP master. This file sets SNMP run-time configuration options.

Procedure

Step 1: Copy configuration files. Use the following sequence of commands to move the SNMP configuration files to the SNMP service configuration directory.

First, create the SNMP configuration directory if needed and then, from the installation directory, copy the configuration files to the SNMP service configuration directory:

```
md C:\snmp\etc\config
copy MONGOD-MIB.txt C:\snmp\etc\config\MONGOD-MIB.txt
copy mongod.conf.subagent C:\snmp\etc\config\mongod.conf
```

The configuration filename is tool-dependent. For example, when using `net-snmp` the configuration file is `snmpd.conf`.

Edit the configuration file to ensure that the communication between the agent (i.e. `snmpd` or the master) and subagent (i.e. MongoDB) uses TCP.

Ensure that the `agentXAddress` specified in the SNMP configuration file for MongoDB matches the `agentXAddress` in the SNMP master configuration file.

Step 2: Start MongoDB. Start `mongod.exe` with the `snmp-subagent` to send data to the SNMP master.

```
mongod.exe --snmp-subagent
```

Step 3: Confirm SNMP data retrieval. Use `snmpwalk` to collect data from `mongod.exe`:

Connect an SNMP client to verify the ability to collect SNMP data from MongoDB.

Install the net-snmp[98] package to access the `snmpwalk` client. `net-snmp` provides the `snmpwalk` SNMP client.

```
snmpwalk -m C:\snmp\etc\config\MONGOD-MIB.txt -v 2c -c mongodb 127.0.0.1:<port> 1.3.6.1.4.1.34601
```

`<port>` refers to the port defined by the SNMP master, *not* the primary `port` used by `mongod.exe` for client communication.

[98] http://www.net-snmp.org/

Optional: Run MongoDB as SNMP Master

You can run `mongod.exe` with the `snmp-master` option for testing purposes. To do this, use the SNMP master configuration file instead of the subagent configuration file. From the directory containing the unpacked MongoDB installation files:

```
copy mongod.conf.master C:\snmp\etc\config\mongod.conf
```

Additionally, start `mongod.exe` with the `snmp-master` option, as in the following:

```
mongod.exe --snmp-master
```

Troubleshoot SNMP

On this page

- Overview (page 342)
- Issues (page 342)

New in version 2.6.

Enterprise Feature

SNMP is only available in MongoDB Enterprise.

Overview

MongoDB Enterprise can provide database metrics via SNMP, in support of centralized data collection and aggregation. This document identifies common problems you may encounter when deploying MongoDB Enterprise with SNMP as well as possible solutions for these issues.

See *Monitor MongoDB With SNMP on Linux* (page 338) and *Monitor MongoDB Windows with SNMP* (page 340) for complete installation instructions.

Issues

Failed to Connect The following in the `mongod` logfile:

```
Warning: Failed to connect to the agentx master agent
```

AgentX is the SNMP agent extensibility protocol defined in Internet RFC 2741[99]. It explains how to define additional data to monitor over SNMP. When MongoDB fails to connect to the agentx master agent, use the following procedure to ensure that the SNMP subagent can connect properly to the SNMP master.

1. Make sure the master agent is running.
2. Compare the SNMP master's configuration file with the subagent configuration file. Ensure that the agentx socket definition is the same between the two.
3. Check the SNMP configuration files to see if they specify using UNIX Domain Sockets. If so, confirm that the `mongod` has appropriate permissions to open a UNIX domain socket.

[99] http://www.ietf.org/rfc/rfc2741.txt

Error Parsing Command Line One of the following errors at the command line:

```
Error parsing command line: unknown option snmp-master
try 'mongod --help' for more information

Error parsing command line: unknown option snmp-subagent
try 'mongod --help' for more information
```

`mongod` binaries that are not part of the Enterprise Edition produce this error. *Install the Enterprise Edition* (page 49) and attempt to start `mongod` again.

Other MongoDB binaries, including `mongos` will produce this error if you attempt to star them with `snmp-master` or `snmp-subagent`. Only `mongod` supports SNMP.

Error Starting `SNMPAgent` The following line in the log file indicates that `mongod` cannot read the `mongod.conf` file:

```
[SNMPAgent] warning: error starting SNMPAgent as master err:1
```

If running on Linux, ensure `mongod.conf` exists in the `/etc/snmp` directory, and ensure that the `mongod` UNIX user has permission to read the `mongod.conf` file.

If running on Windows, ensure `mongod.conf` exists in `C:\snmp\etc\config`.

8.2.2 Backup and Recovery

The following tutorials describe backup and restoration for a `mongod` instance:

Backup and Restore with Filesystem Snapshots **(page 343)** An outline of procedures for creating MongoDB data set backups using system-level file snapshot tool, such as *LVM* or native storage appliance tools.

Restore a Replica Set from MongoDB Backups **(page 348)** Describes procedure for restoring a replica set from an archived backup such as a `mongodump` or MongoDB Cloud Manager[100] Backup file.

Back Up and Restore with MongoDB Tools **(page 349)** Describes a procedure for exporting the contents of a database to either a binary dump or a textual exchange format, and for importing these files into a database.

Backup and Restore Sharded Clusters **(page 355)** Detailed procedures and considerations for backing up sharded clusters and single shards.

Recover Data after an Unexpected Shutdown **(page 366)** Recover data from MongoDB data files that were not properly closed or have an invalid state.

Backup and Restore with Filesystem Snapshots

On this page

- Snapshots Overview (page 344)
- Back up and Restore Using LVM on Linux (page 345)
- Back up Instances with Journal Files on Separate Volume or without Journaling (page 347)
- Additional Resources (page 347)

[100] https://cloud.mongodb.com/?jmp=docs

This document describes a procedure for creating backups of MongoDB systems using system-level tools, such as *LVM* or storage appliance, as well as the corresponding restoration strategies.

These filesystem snapshots, or "block-level" backup methods, use system level tools to create copies of the device that holds MongoDB's data files. These methods complete quickly and work reliably, but require additional system configuration outside of MongoDB.

Changed in version 3.2: Starting in MongoDB 3.2, the data files as well as the journal files can reside on separate volumes to create volume-level backup of MongoDB instances using the *WiredTiger* (page 595) storage engine. With previous versions, for the purpose of volume-level backup of MongoDB instances using WiredTiger, the data files and the journal must reside on a single volume.

See also:

MongoDB Backup Methods (page 282) and *Back Up and Restore with MongoDB Tools* (page 349).

Snapshots Overview

Snapshots work by creating pointers between the live data and a special snapshot volume. These pointers are theoretically equivalent to "hard links." As the working data diverges from the snapshot, the snapshot process uses a copy-on-write strategy. As a result the snapshot only stores modified data.

After making the snapshot, you mount the snapshot image on your file system and copy data from the snapshot. The resulting backup contains a full copy of all data.

Considerations

Valid Database at the Time of Snapshot The database must be valid when the snapshot takes place. This means that all writes accepted by the database need to be fully written to disk: either to the *journal* or to data files.

If all writes are not on disk when the backup occurs, the backup will not reflect these changes.

For the *MMAPv1 storage engine* (page 603), if writes are *in progress* when the backup occurs, the data files will reflect an inconsistent state. With *journaling* (page 607), all data-file states resulting from in-progress writes are recoverable; without journaling, you must flush all pending writes to disk before running the backup operation and must ensure that no writes occur during the entire backup procedure. If you do use journaling, the journal **must** reside on the same volume as the data.

For the *WiredTiger storage engine* (page 595), the data files reflect a consistent state as of the last *checkpoint* (page 596), which occurs with every 2 GB of data or every minute.

Entire Disk Image Snapshots create an image of an entire disk image. Unless you need to back up your entire system, consider isolating your MongoDB data files, journal (if applicable), and configuration on one logical disk that doesn't contain any other data.

Alternately, store all MongoDB data files on a dedicated device so that you can make backups without duplicating extraneous data.

Site Failure Precaution Ensure that you copy data from snapshots onto other systems. This ensures that data is safe from site failures.

No Incremental Backups This tutorial does not include procedures for incremental backups. Although different snapshots methods provide different capability, the LVM method outlined below does not provide any capacity for capturing incremental backups.

Snapshots With Journaling If your `mongod` instance has journaling enabled, then you can use any kind of file system or volume/block level snapshot tool to create backups.

If you manage your own infrastructure on a Linux-based system, configure your system with *LVM* to provide your disk packages and provide snapshot capability. You can also use LVM-based setups *within* a cloud/virtualized environment.

Note: Running *LVM* provides additional flexibility and enables the possibility of using snapshots to back up MongoDB.

Snapshots with Amazon EBS in a RAID 10 Configuration If your deployment depends on Amazon's Elastic Block Storage (EBS) with RAID configured within your instance, it is impossible to get a consistent state across all disks using the platform's snapshot tool. As an alternative, you can do one of the following:

- Flush all writes to disk and create a write lock to ensure consistent state during the backup process.

  If you choose this option see *Back up Instances with Journal Files on Separate Volume or without Journaling* (page 347).

- Configure *LVM* to run and hold your MongoDB data files on top of the RAID within your system.

  If you choose this option, perform the LVM backup operation described in *Create a Snapshot* (page 345).

Back up and Restore Using LVM on Linux

This section provides an overview of a simple backup process using *LVM* on a Linux system. While the tools, commands, and paths may be (slightly) different on your system the following steps provide a high level overview of the backup operation.

Note: Only use the following procedure as a guideline for a backup system and infrastructure. Production backup systems must consider a number of application specific requirements and factors unique to specific environments.

Create a Snapshot Changed in version 3.2: Starting in MongoDB 3.2, for the purpose of volume-level backup of MongoDB instances using WiredTiger, the data files and the journal are no longer required to reside on a single volume.

To create a snapshot with *LVM*, issue a command as root in the following format:

```
lvcreate --size 100M --snapshot --name mdb-snap01 /dev/vg0/mongodb
```

This command creates an *LVM* snapshot (with the `--snapshot` option) named `mdb-snap01` of the `mongodb` volume in the `vg0` volume group.

This example creates a snapshot named `mdb-snap01` located at `/dev/vg0/mdb-snap01`. The location and paths to your systems volume groups and devices may vary slightly depending on your operating system's *LVM* configuration.

The snapshot has a cap of at 100 megabytes, because of the parameter `--size 100M`. This size does not reflect the total amount of the data on the disk, but rather the quantity of differences between the current state of `/dev/vg0/mongodb` and the creation of the snapshot (i.e. `/dev/vg0/mdb-snap01`.)

Warning: Ensure that you create snapshots with enough space to account for data growth, particularly for the period of time that it takes to copy data out of the system or to a temporary image.
If your snapshot runs out of space, the snapshot image becomes unusable. Discard this logical volume and create another.

The snapshot will exist when the command returns. You can restore directly from the snapshot at any time or by creating a new logical volume and restoring from this snapshot to the alternate image.

While snapshots are great for creating high quality backups very quickly, they are not ideal as a format for storing backup data. Snapshots typically depend and reside on the same storage infrastructure as the original disk images. Therefore, it's crucial that you archive these snapshots and store them elsewhere.

Archive a Snapshot After creating a snapshot, mount the snapshot and copy the data to separate storage. Your system might try to compress the backup images as you move them offline. Alternatively, take a block level copy of the snapshot image, such as with the following procedure:

```
umount /dev/vg0/mdb-snap01
dd if=/dev/vg0/mdb-snap01 | gzip > mdb-snap01.gz
```

The above command sequence does the following:

- Ensures that the `/dev/vg0/mdb-snap01` device is not mounted. Never take a block level copy of a filesystem or filesystem snapshot that is mounted.
- Performs a block level copy of the entire snapshot image using the `dd` command and compresses the result in a gzipped file in the current working directory.

  Warning: This command will create a large `gz` file in your current working directory. Make sure that you run this command in a file system that has enough free space.

Restore a Snapshot To restore a snapshot created with the above method, issue the following sequence of commands:

```
lvcreate --size 1G --name mdb-new vg0
gzip -d -c mdb-snap01.gz | dd of=/dev/vg0/mdb-new
mount /dev/vg0/mdb-new /srv/mongodb
```

The above sequence does the following:

- Creates a new logical volume named `mdb-new`, in the `/dev/vg0` volume group. The path to the new device will be `/dev/vg0/mdb-new`.

  Warning: This volume will have a maximum size of 1 gigabyte. The original file system must have had a total size of 1 gigabyte or smaller, or else the restoration will fail.
  Change `1G` to your desired volume size.

- Uncompresses and unarchives the `mdb-snap01.gz` into the `mdb-new` disk image.
- Mounts the `mdb-new` disk image to the `/srv/mongodb` directory. Modify the mount point to correspond to your MongoDB data file location, or other location as needed.

Note: The restored snapshot will have a stale `mongod.lock` file. If you do not remove this file from the snapshot, and MongoDB may assume that the stale lock file indicates an unclean shutdown. If you're running with `storage.journal.enabled` enabled, and you *do not* use `db.fsyncLock()`, you do not need to remove the `mongod.lock` file. If you use `db.fsyncLock()` you will need to remove the lock.

Restore Directly from a Snapshot To restore a backup without writing to a compressed `gz` file, use the following sequence of commands:

```
umount /dev/vg0/mdb-snap01
lvcreate --size 1G --name mdb-new vg0
dd if=/dev/vg0/mdb-snap01 of=/dev/vg0/mdb-new
mount /dev/vg0/mdb-new /srv/mongodb
```

Remote Backup Storage You can implement off-system backups using the *combined process* (page 346) and SSH.

This sequence is identical to procedures explained above, except that it archives and compresses the backup on a remote system using SSH.

Consider the following procedure:

```
umount /dev/vg0/mdb-snap01
dd if=/dev/vg0/mdb-snap01 | ssh username@example.com gzip > /opt/backup/mdb-snap01.gz
lvcreate --size 1G --name mdb-new vg0
ssh username@example.com gzip -d -c /opt/backup/mdb-snap01.gz | dd of=/dev/vg0/mdb-new
mount /dev/vg0/mdb-new /srv/mongodb
```

Back up Instances with Journal Files on Separate Volume or without Journaling

Changed in version 3.2: Starting in MongoDB 3.2, for the purpose of volume-level backup of MongoDB instances using WiredTiger, the data files and the journal are no longer required to reside on a single volume.

If your `mongod` instance is either running without journaling or has the journal files on a separate volume, you must flush all writes to disk and lock the database to prevent writes during the backup process. If you have a *replica set* configuration, then for your backup use a *secondary* which is not receiving reads (i.e. *hidden member*).

Important: In the following procedure to create backups, you **must** issue the `db.fsyncLock()` and `db.fsyncUnlock()` operations on the same connection. The client that issues `db.fsyncLock()` is solely responsible for issuing a `db.fsyncUnlock()` operation and must be able to handle potential error conditions so that it can perform the `db.fsyncUnlock()` before terminating the connection.

Step 1: Flush writes to disk and lock the database to prevent further writes. To flush writes to disk and to "lock" the database, issue the `db.fsyncLock()` method in the `mongo` shell:

```
db.fsyncLock();
```

Step 2: Perform the backup operation described in *Create a Snapshot*.

Step 3: After the snapshot completes, unlock the database. To unlock the database after the snapshot has completed, use the following command in the `mongo` shell:

```
db.fsyncUnlock();
```

Additional Resources

See also MongoDB Cloud Manager[101] for seamless automation, backup, and monitoring.

[101] https://cloud.mongodb.com/?jmp=docs

Restore a Replica Set from MongoDB Backups

On this page

- Restore Database into a Single Node Replica Set (page 348)
- Add Members to the Replica Set (page 348)

This procedure outlines the process for taking MongoDB data and restoring that data into a new *replica set*. Use this approach for seeding test deployments from production backups as well as part of disaster recovery.

You *cannot* restore a single data set to three new `mongod` instances and *then* create a replica set. In this situation MongoDB will force the secondaries to perform an initial sync. The procedures in this document describe the correct and efficient ways to deploy a replica set.

You can also use `mongorestore` to restore database files using data created with `mongodump`. See *Back Up and Restore with MongoDB Tools* (page 349) for more information.

Restore Database into a Single Node Replica Set

Step 1: Obtain backup MongoDB Database files. The backup files may come from a *file system snapshot* (page 343). The MongoDB Cloud Manager[102] produces MongoDB database files for stored snapshots[103] and point in time snapshots[104]. For Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[105], see also the Ops Manager Backup overview[106].

Step 2: Start a `mongod` using data files from the backup as the data path. Start a `mongod` instance for a new single-node replica set. Specify the path to the backup data files with `--dbpath` option and the replica set name with the `--replSet` option. For *config server replica set (CSRS)* (page 743), include the *`--configsvr`* option.

```
mongod --dbpath /data/db --replSet <replName>
```

Step 3: Connect a `mongo` shell to the `mongod` instance. For example, to connect to a `mongod` running on localhost on the default port of `27017`, simply issue:

```
mongo
```

Step 4: Initiate the new replica set. Use `rs.initiate()` on *one and only one* member of the replica set:

```
rs.initiate()
```

MongoDB initiates a set that consists of the current member and that uses the default replica set configuration.

Add Members to the Replica Set

MongoDB provides two options for restoring secondary members of a replica set:

- Manually copy the database files to each data directory.

[102] https://cloud.mongodb.com/?jmp=docs
[103] https://docs.cloud.mongodb.com/tutorial/restore-from-snapshot/
[104] https://docs.cloud.mongodb.com/tutorial/restore-from-point-in-time-snapshot/
[105] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[106] https://docs.opsmanager.mongodb.com/current/core/backup-overview

- Allow *initial sync* (page 658) to distribute data automatically.

The following sections outlines both approaches.

Note: If your database is large, initial sync can take a long time to complete. For large databases, it might be preferable to copy the database files onto each host.

Copy Database Files and Restart `mongod` Instance Use the following sequence of operations to "seed" additional members of the replica set with the restored data by copying MongoDB data files directly.

Step 1: Shut down the `mongod` instance that you restored. Use `--shutdown` or `db.shutdownServer()` to ensure a clean shut down.

Step 2: Copy the primary's data directory to each secondary. Copy the *primary's* data directory into the `dbPath` of the other members of the replica set. The `dbPath` is `/data/db` by default.

Step 3: Start the `mongod` instance that you restored.

Step 4: Add the secondaries to the replica set. In a `mongo` shell connected to the *primary*, add the *secondaries* to the replica set using `rs.add()`. See *Deploy a Replica Set* (page 667) for more information about deploying a replica set.

Update Secondaries using Initial Sync Use the following sequence of operations to "seed" additional members of the replica set with the restored data using the default *initial sync* operation.

Step 1: Ensure that the data directories on the prospective replica set members are empty.

Step 2: Add each prospective member to the replica set. When you add a member to the replica set, *Initial Sync* (page 658) copies the data from the *primary* to the new member.

Back Up and Restore with MongoDB Tools

On this page

- Binary BSON Dumps (page 350)
- Human Intelligible Import/Export Formats (page 352)

This document describes the process for creating backups and restoring data using the utilities provided with MongoDB.

Because all of these tools primarily operate by interacting with a running `mongod` instance, they can impact the performance of your running database.

Not only do they create traffic for a running database instance, they also force the database to read all data through memory. When MongoDB reads infrequently used data, it can supplant more frequently accessed data, causing a deterioration in performance for the database's regular workload.

No matter how you decide to import or export your data, consider the following guidelines:

- Label files so that you can identify the contents of the export or backup as well as the point in time the export/backup reflect.
- Do not create or apply exports if the backup process itself will have an adverse effect on a production system.
- Make sure that the backups reflect a consistent data state. Export or backup processes can impact data integrity (i.e. type fidelity) and consistency if updates continue during the backup process.
- Test backups and exports by restoring and importing to ensure that the backups are useful.

See also:

MongoDB Backup Methods (page 282) or MongoDB Cloud Manager Backup documentation[107] for more information on backing up MongoDB instances. Additionally, consider the following references for the MongoDB import/export tools:

- `mongoimport`
- `mongoexport`
- `mongorestore`
- `mongodump`

Binary BSON Dumps

The `mongorestore` and `mongodump` utilities work with *BSON* (page 12) data dumps, and are useful for creating backups of small deployments. For resilient and non-disruptive backups, use a file system or block-level disk snapshot function, such as the methods described in the *MongoDB Backup Methods* (page 282) document.

Use these tools for backups if other backup methods, such as the MongoDB Cloud Manager[108] or *file system snapshots* (page 343) are unavailable.

Backup a Database with `mongodump`

Exclude `local` Database `mongodump` excludes the content of the `local` database in its output.

Required Access To run `mongodump` against a MongoDB deployment that has *access control* (page 433) enabled, you must have privileges that grant `find` (page 501) action for each database to back up. The built-in `backup` (page 491) role provides the required privileges to perform backup of any and all databases.

Changed in version 3.2.1: The `backup` (page 491) role provides additional privileges to back up the `system.profile` (page 377) collections that exist when running with *database profiling* (page 312). Previously, users required an additional `read` access on this collection.

Basic `mongodump` Operations The `mongodump` utility backs up data by connecting to a running `mongod` or `mongos` instance.

The utility can create a backup for an entire server, database or collection, or can use a query to backup just part of a collection.

When you run `mongodump` without any arguments, the command connects to the MongoDB instance on the local system (e.g. `127.0.0.1` or `localhost`) on port `27017` and creates a database backup named `dump/` in the current directory.

[107] https://docs.cloud.mongodb.com/tutorial/nav/backup-use/
[108] https://cloud.mongodb.com/?jmp=docs

To backup data from a mongod or mongos instance running on the same machine and on the default port of 27017, use the following command:

```
mongodump
```

The data format used by mongodump from version 2.2 or later is *incompatible* with earlier versions of mongod. Do not use recent versions of mongodump to back up older data stores.

You can also specify the *--host* and *--port* of the MongoDB instance that the mongodump should connect to. For example:

```
mongodump --host mongodb.example.net --port 27017
```

mongodump will write *BSON* files that hold a copy of data accessible via the mongod listening on port 27017 of the mongodb.example.net host. See *Create Backups from Non-Local mongod Instances* (page 351) for more information.

To specify a different output directory, you can use the *--out or -o* option:

```
mongodump --out /data/backup/
```

To limit the amount of data included in the database dump, you can specify *--db* and *--collection* as options to mongodump. For example:

```
mongodump --collection myCollection --db test
```

This operation creates a dump of the collection named myCollection from the database test in a dump/ subdirectory of the current working directory.

mongodump overwrites output files if they exist in the backup data folder. Before running the mongodump command multiple times, either ensure that you no longer need the files in the output folder (the default is the dump/ folder) or rename the folders or files.

Point in Time Operation Using Oplogs Use the *--oplog* option with mongodump to collect the *oplog* entries to build a point-in-time snapshot of a database within a replica set. With *--oplog*, mongodump copies all the data from the source database as well as all of the *oplog* entries from the beginning to the end of the backup procedure. This operation, in conjunction with *mongorestore --oplogReplay*, allows you to restore a backup that reflects the specific moment in time that corresponds to when mongodump completed creating the dump file.

Create Backups from Non-Local mongod Instances The *--host* and *--port* options for mongodump allow you to connect to and backup from a remote host. Consider the following example:

```
mongodump --host mongodb1.example.net --port 3017 --username user --password pass --out /opt/backup/m
```

On any mongodump command you may, as above, specify username and password credentials to specify database authentication.

Restore a Database with mongorestore

Access Control To restore data to a MongoDB deployment that has *access control* (page 433) enabled, the restore (page 492) role provides access to restore any database if the backup data does not include system.profile (page 377) collection data.

If the backup data includes system.profile (page 377) collection data and the target database does not contain the system.profile (page 377) collection, mongorestore attempts to create the collection even though the program does not actually restore system.profile documents. As such, the user requires additional privileges to

perform createCollection (page 501) and convertToCapped (page 504) actions on the system.profile (page 377) collection for a database.

If running mongorestore with *--oplogReplay*, additional privilege *user-defined role* (page 442) that has anyAction (page 505) on *anyResource* (page 500) and grant only to users who must run mongorestore with *--oplogReplay*.

Basic mongorestore Operations The mongorestore utility restores a binary backup created by mongodump. By default, mongorestore looks for a database backup in the dump/ directory.

The mongorestore utility restores data by connecting to a running mongod or mongos directly.

mongorestore can restore either an entire database backup or a subset of the backup.

To use mongorestore to connect to an active mongod or mongos, use a command with the following prototype form:

```
mongorestore --port <port number> <path to the backup>
```

Consider the following example:

```
mongorestore dump-2013-10-25/
```

Here, mongorestore imports the database backup in the dump-2013-10-25 directory to the mongod instance running on the localhost interface.

Restore Point in Time Oplog Backup If you created your database dump using the *--oplog* option to ensure a point-in-time snapshot, call mongorestore with the *--oplogReplay* option, as in the following example:

```
mongorestore --oplogReplay
```

You may also consider using the *mongorestore --objcheck* option to check the integrity of objects while inserting them into the database, or you may consider the *mongorestore --drop* option to drop each collection from the database before restoring from backups.

Restore Backups to Non-Local mongod Instances By default, mongorestore connects to a MongoDB instance running on the localhost interface (e.g. 127.0.0.1) and on the default port (27017). If you want to restore to a different host or port, use the *--host* and *--port* options.

Consider the following example:

```
mongorestore --host mongodb1.example.net --port 3017 --username user --password pass /opt/backup/mong
```

As above, you may specify username and password connections if your mongod requires authentication.

Human Intelligible Import/Export Formats

MongoDB's mongoimport and mongoexport tools allow you to work with your data in a human-readable *Extended JSON* (page 16) or *CSV* format. This is useful for simple ingestion to or from a third-party system, and when you want to backup or export a small subset of your data. For more complex data migration tasks, you may want to write your own import and export scripts using a client *driver* to interact with the database.

The examples in this section use the MongoDB tools mongoimport and mongoexport. These tools may also be useful for importing data into a MongoDB database from third party applications.

If you want to simply copy a database or collection from one instance to another, consider using the `copydb`, `clone`, or `cloneCollection` commands, which may be more suited to this task. The `mongo` shell provides the `db.copyDatabase()` method.

Warning: Avoid using `mongoimport` and `mongoexport` for full instance production backups. They do not reliably preserve all rich *BSON* data types, because *JSON* can only represent a subset of the types supported by BSON. Use `mongodump` and `mongorestore` as described in *MongoDB Backup Methods* (page 282) for this kind of functionality.

Collection Export with `mongoexport`

Export in CSV Format Changed in version 3.0.0: `mongoexport` removed the `--csv` option. Use the `--type=csv` option to specify CSV format for the output.

In the following example, `mongoexport` exports data from the collection `contacts` collection in the `users` database in *CSV* format to the file `/opt/backups/contacts.csv`.

The `mongod` instance that `mongoexport` connects to is running on the localhost port number `27017`.

When you export in CSV format, you must specify the fields in the documents to export. The operation specifies the `name` and `address` fields to export.

```
mongoexport --db users --collection contacts --type=csv --fields name,address --out /opt/backups/cont
```

For CSV exports only, you can also specify the fields in a file containing the line-separated list of fields to export. The file must have only one field per line.

For example, you can specify the `name` and `address` fields in a file `fields.txt`:

```
name
address
```

Then, using the `--fieldFile` option, specify the fields to export with the file:

```
mongoexport --db users --collection contacts --type=csv --fieldFile fields.txt --out /opt/backups/co
```

Changed in version 3.0.0: `mongoexport` removed the `--csv` option and replaced with the `--type` option.

Export in JSON Format This example creates an export of the `contacts` collection from the MongoDB instance running on the localhost port number `27017`. This writes the export to the `contacts.json` file in *JSON* format.

```
mongoexport --db sales --collection contacts --out contacts.json
```

Export from Remote Host Running with Authentication The following example exports the `contacts` collection from the `marketing` database, which requires authentication.

This data resides on the MongoDB instance located on the host `mongodb1.example.net` running on port `37017`, which requires the username `user` and the password `pass`.

```
mongoexport --host mongodb1.example.net --port 37017 --username user --password pass --collection co
```

Export Query Results You can export only the results of a query by supplying a query filter with the `--query` option, and limit the results to a single database using the "`--db`" option.

For instance, this command returns all documents in the `sales` database's `contacts` collection that contain a field named `field` with a value of `1`.

```
mongoexport --db sales --collection contacts --query '{"field": 1}'
```

You must enclose the query in single quotes (e.g. `'`) to ensure that it does not interact with your shell environment.

Collection Import with `mongoimport`

Simple Usage `mongoimport` restores a database from a backup taken with `mongoexport`. Most of the arguments to `mongoexport` also exist for `mongoimport`.

In the following example, `mongoimport` imports the data in the *JSON* data from the `contacts.json` file into the collection `contacts` in the `users` database.

```
mongoimport --db users --collection contacts --file contacts.json
```

Import `JSON` to Remote Host Running with Authentication In the following example, `mongoimport` imports data from the file `/opt/backups/mdb1-examplenet.json` into the `contacts` collection within the database `marketing` on a remote MongoDB database with authentication enabled.

`mongoimport` connects to the `mongod` instance running on the host `mongodb1.example.net` over port `37017`. It authenticates with the username `user` and the password `pass`.

```
mongoimport --host mongodb1.example.net --port 37017 --username user --password pass --collection co
```

`CSV` Import In the following example, `mongoimport` imports the *csv* formatted data in the `/opt/backups/contacts.csv` file into the collection `contacts` in the `users` database on the MongoDB instance running on the localhost port numbered `27017`.

Specifying *`--headerline`* instructs `mongoimport` to determine the name of the fields using the first line in the CSV file.

```
mongoimport --db users --collection contacts --type csv --headerline --file /opt/backups/contacts.csv
```

`mongoimport` uses the input file name, without the extension, as the collection name if `-c` or `--collection` is unspecified. The following example is therefore equivalent:

```
mongoimport --db users --type csv --headerline --file /opt/backups/contacts.csv
```

Use the "*`--ignoreBlanks`*" option to ignore blank fields. For *CSV* and *TSV* imports, this option provides the desired functionality in most cases because it avoids inserting fields with null values into your collection.

Additional Resources

- Backup and its Role in Disaster Recovery White Paper[109]
- Cloud Backup through MongoDB Cloud Manager[110]
- Blog Post: Backup vs. Replication, Why you Need Both[111]
- Backup Service with Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[112]

[109] https://www.mongodb.com/lp/white-paper/backup-disaster-recovery?jmp=docs
[110] https://cloud.mongodb.com/?jmp=docs
[111] http://www.mongodb.com/blog/post/backup-vs-replication-why-do-you-need-both?jmp=docs
[112] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

Backup and Restore Sharded Clusters

The following tutorials describe backup and restoration for sharded clusters:

Backup a Small Sharded Cluster with mongodump **(page 355)** If your *sharded cluster* holds a small data set, you can use `mongodump` to capture the entire backup in a reasonable amount of time.

Backup a Sharded Cluster with Filesystem Snapshots **(page 356)** Use file system snapshots back up each component in the sharded cluster individually. The procedure involves stopping the cluster balancer. If your system configuration allows file system backups, this might be more efficient than using MongoDB tools.

Backup a Sharded Cluster with Database Dumps **(page 359)** Create backups using `mongodump` to back up each component in the cluster individually.

Schedule Backup Window for Sharded Clusters **(page 362)** Limit the operation of the cluster balancer to provide a window for regular backup operations.

Restore a Single Shard **(page 362)** An outline of the procedure and consideration for restoring a single shard from a backup.

Restore a Sharded Cluster **(page 363)** An outline of the procedure and consideration for restoring an *entire* sharded cluster from backup.

Backup a Small Sharded Cluster with `mongodump`

On this page

- Overview (page 355)
- Considerations (page 355)
- Procedure (page 356)
- Additional Resources (page 356)

Overview If your *sharded cluster* holds a small data set, you can connect to a `mongos` using `mongodump`. You can create backups of your MongoDB cluster, if your backup infrastructure can capture the entire backup in a reasonable amount of time and if you have a storage system that can hold the complete MongoDB data set.

See *MongoDB Backup Methods* (page 282) and *Backup and Restore Sharded Clusters* (page 355) for complete information on backups in MongoDB and backups of sharded clusters in particular.

Important: By default `mongodump` issue its queries to the non-primary nodes.

Considerations If you use `mongodump` without specifying a database or collection, `mongodump` will capture collection data *and* the cluster meta-data from the *config servers* (page 742).

You cannot use the `--oplog` option for `mongodump` when capturing data from `mongos`. As a result, if you need to capture a backup that reflects a single moment in time, you must stop all writes to the cluster for the duration of the backup operation.

To run `mongodump` against a MongoDB deployment that has *access control* (page 433) enabled, you must have privileges that grant `find` (page 501) action for each database to back up. The built-in `backup` (page 491) role provides the required privileges to perform backup of any and all databases.

Changed in version 3.2.1: The `backup` (page 491) role provides additional privileges to back up the `system.profile` (page 377) collections that exist when running with *database profiling* (page 312). Previously, users required an additional `read` access on this collection.

Procedure

Capture Data You can perform a backup of a *sharded cluster* by connecting `mongodump` to a `mongos`. Use the following operation at your system's prompt:

```
mongodump --host mongos3.example.net --port 27017
```

`mongodump` will write *BSON* files that hold a copy of data stored in the *sharded cluster* accessible via the `mongos` listening on port `27017` of the `mongos3.example.net` host.

Restore Data Backups created with `mongodump` do not reflect the chunks or the distribution of data in the sharded collection or collections. Like all `mongodump` output, these backups contain separate directories for each database and *BSON* files for each collection in that database.

You can restore `mongodump` output to any MongoDB instance, including a standalone, a *replica set*, or a new *sharded cluster*. When restoring data to sharded cluster, you must deploy and configure sharding before restoring data from the backup. See *Deploy a Sharded Cluster* (page 765) for more information.

Additional Resources See also MongoDB Cloud Manager[113] for seamless automation, backup, and monitoring.

Backup a Sharded Cluster with Filesystem Snapshots

On this page

- Overview (page 356)
- Considerations (page 356)
- Procedure (page 357)
- Additional Resources (page 359)

Changed in version 3.2: Starting in MongoDB 3.2, the procedure can be used with the *MMAPv1* (page 603) and the *WiredTiger* (page 595) storage engines. With previous versions of MongoDB, the procedure applied to *MMAPv1* (page 603) only.

Overview This document describes a procedure for taking a backup of all components of a sharded cluster. This procedure uses file system snapshots to capture a copy of the `mongod` instance. An alternate procedure uses `mongodump` to create binary database dumps when file-system snapshots are not available. See *Backup a Sharded Cluster with Database Dumps* (page 359) for the alternate procedure.

Important: To capture a point-in-time backup from a sharded cluster you **must** stop *all* writes to the cluster. On a running production system, you can only capture an *approximation* of point-in-time snapshot.

For more information on backups in MongoDB and backups of sharded clusters in particular, see *MongoDB Backup Methods* (page 282) and *Backup and Restore Sharded Clusters* (page 355).

Considerations

[113] https://cloud.mongodb.com/?jmp=docs

Balancer It is *essential* that you stop the *balancer* (page 758) before capturing a backup.

If the balancer is active while you capture backups, the backup artifacts may be incomplete and/or have duplicate data, as *chunks* may migrate while recording backups.

Precision In this procedure, you will stop the cluster balancer and take a backup up of the *config database*, and then take backups of each shard in the cluster using a file-system snapshot tool. If you need an exact moment-in-time snapshot of the system, you will need to stop all application writes before taking the filesystem snapshots; otherwise the snapshot will only approximate a moment in time.

For approximate point-in-time snapshots, you can minimize the impact on the cluster by taking the backup from a secondary member of each replica set shard.

Consistency If the journal and data files are on the same logical volume, you can use a single point-in-time snapshot to capture a consistent copy of the data files.

If the journal and data files are on different file systems, you must use `db.fsyncLock()` and `db.fsyncUnlock()` to ensure that the data files do not change, providing consistency for the purposes of creating backups.

Procedure

Step 1: Disable the balancer. To disable the *balancer* (page 758), connect the `mongo` shell to a `mongos` instance and run `sh.stopBalancer()` in the `config` database.

```
use config
sh.stopBalancer()
```

For more information, see the *Disable the Balancer* (page 802) procedure.

Step 2: If necessary, lock one secondary member of each replica set. If your secondary does not have journaling enabled *or* its journal and data files are on different volumes, you **must** lock the secondary's `mongod` instance before capturing a backup.

If your secondary has journaling enabled and its journal and data files are on the same volume, you may skip this step.

Important: If your deployment requires this step, you must perform it on one secondary of each shard and, if your sharded cluster uses a replica set for the config servers, one secondary of the *config server replica set (CSRS)* (page 743).

Ensure that the *oplog* has sufficient capacity to allow these secondaries to catch up to the state of the primaries after finishing the backup procedure. See *Oplog Size* (page 657) for more information.

Lock shard replica set secondary. For each shard replica set in the sharded cluster, connect a `mongo` shell to the secondary member's `mongod` instance and run `db.fsyncLock()`.

```
db.fsyncLock()
```

When calling `db.fsyncLock()`, ensure that the connection is kept open to allow a subsequent call to `db.fsyncUnlock()`.

Lock config server replica set secondary. If locking a secondary of the CSRS, confirm that the member has replicated data up to some control point. To verify, first connect a `mongo` shell to the CSRS primary and perform a write operation with `"majority"` (page 180) write concern on a control collection:

```
use config
db.BackupControl.findAndModify(
   {
     query: { _id: 'BackupControlDocument' },
     update: { $inc: { counter : 1 } },
     new: true,
     upsert: true,
     writeConcern: { w: 'majority', wtimeout: 15000 } }
   }
);
```

The operation should return the modified (or inserted) control document:

```
{ "_id" : "BackupControlDocument", "counter" : 1 }
```

Query the CSRS secondary member for the returned control document. Connect a `mongo` shell to the CSRS secondary to lock and use `db.collection.find()` to query for the control document:

```
rs.slaveOk();
use config;

db.BackupControl.find(
   { "_id" : "BackupControlDocument", "counter" : 1 }
).readConcern('majority');
```

If the secondary member contains the latest control document, it is safe to lock the member. Otherwise, wait until the member contains the document or select a different secondary member that contains the latest control document.

To lock the secondary member, run `db.fsyncLock()` on the member:

```
db.fsyncLock()
```

When calling `db.fsyncLock()`, ensure that the connection is kept open to allow a subsequent call to `db.fsyncUnlock()`.

Step 3: Back up one of the config servers.

Note: Backing up a *config server* (page 742) backs up the sharded cluster's metadata. You only need to back up one config server, as they all hold the same data. If you are using CSRS config servers, perform this step against the locked config server.

To create a file-system snapshot of the config server, follow the procedure in *Create a Snapshot* (page 345).

Step 4: Back up a replica set member for each shard. If you locked a member of the replica set shards, perform this step against the locked secondary.

You may back up the shards in parallel. For each shard, create a snapshot, using the procedure in *Backup and Restore with Filesystem Snapshots* (page 343).

Step 5: Unlock all locked replica set members. If you locked any `mongod` instances to capture the backup, unlock them.

To unlock the replica set members, use `db.fsyncUnlock()` method in the `mongo` shell. For each locked member, use the same `mongo` shell used to lock the instance.

```
db.fsyncUnlock()
```

Step 6: Enable the balancer. To re-enable to balancer, connect the `mongo` shell to a `mongos` instance and run `sh.setBalancerState()`.

```
sh.setBalancerState(true)
```

Additional Resources See also MongoDB Cloud Manager[114] for seamless automation, backup, and monitoring.

Backup a Sharded Cluster with Database Dumps

On this page

- Overview (page 359)
- Prerequisites (page 359)
- Consideration (page 359)
- Procedure (page 360)
- Additional Resources (page 361)

Changed in version 3.2: Starting in MongoDB 3.2, the following procedure can be used with the *MMAPv1* (page 603) and the *WiredTiger* (page 595) storage engines. With previous versions of MongoDB, the procedure applied to *MMAPv1* (page 603) only.

Overview This document describes a procedure for taking a backup of all components of a sharded cluster. This procedure uses `mongodump` to create dumps of the `mongod` instance. An alternate procedure uses file system snapshots to capture the backup data, and may be more efficient in some situations if your system configuration allows file system backups.

For more information on backups in MongoDB and backups of sharded clusters in particular, see *MongoDB Backup Methods* (page 282) and *Backup and Restore Sharded Clusters* (page 355).

Prerequisites

Important: To capture a point-in-time backup from a sharded cluster you **must** stop *all* writes to the cluster. On a running production system, you can only capture an *approximation* of point-in-time snapshot.

Access Control The `backup` (page 491) role provides the required privileges to perform backup on a sharded cluster that has access control enabled.

Changed in version 3.2.1: The `backup` (page 491) role provides additional privileges to back up the `system.profile` (page 377) collections that exist when running with *database profiling* (page 312). Previously, users required an additional `read` access on this collection.

Consideration To create these backups of a sharded cluster, you will stop the cluster balancer and take a backup of the *config database*, and then take backups of each shard in the cluster using `mongodump` to capture the backup data. To capture a more exact moment-in-time snapshot of the system, you will need to stop all application writes before taking the filesystem snapshots; otherwise the snapshot will only approximate a moment in time.

[114] https://cloud.mongodb.com/?jmp=docs

For approximate point-in-time snapshots, you can minimize the impact on the cluster by taking the backup from a secondary member of each replica set shard.

Procedure

Step 1: Disable the balancer process. To disable the *balancer* (page 758), connect the `mongo` shell to a `mongos` instance and run `sh.stopBalancer()` in the `config` database.

```
use config
sh.stopBalancer()
```

For more information, see the *Disable the Balancer* (page 802) procedure.

> **Warning:** If you do not stop the balancer, the backup could have duplicate data or omit data as *chunks* migrate while recording backups.

Step 2: Lock one secondary member of each replica set. Lock a secondary member of each replica set in the sharded cluster, and, if your sharded cluster uses a replica set for the config servers, one secondary of the *config server replica set (CSRS)* (page 743).

Ensure that the *oplog* has sufficient capacity to allow these secondaries to catch up to the state of the primaries after finishing the backup procedure. See *Oplog Size* (page 657) for more information.

Lock shard replica set secondary. For each shard replica set in the sharded cluster, connect a `mongo` shell to the secondary member's `mongod` instance and run `db.fsyncLock()`.

```
db.fsyncLock()
```

When calling `db.fsyncLock()`, ensure that the connection is kept open to allow a subsequent call to `db.fsyncUnlock()`.

Lock config server replica set secondary. If locking a secondary of the CSRS, confirm that the member has replicated data up to some control point. To verify, first connect a `mongo` shell to the CSRS primary and perform a write operation with `"majority"` (page 180) write concern on a control collection:

```
use config
db.BackupControl.findAndModify(
   {
     query: { _id: 'BackupControlDocument' },
     update: { $inc: { counter : 1 } },
     new: true,
     upsert: true,
     writeConcern: { w: 'majority', wtimeout: 15000 } }
   }
);
```

The operation should return either the newly inserted document or the updated document:

```
{ "_id" : "BackupControlDocument", "counter" : 1 }
```

Query the CSRS secondary member for the returned control document. Connect a `mongo` shell to the CSRS secondary to lock and use `db.collection.find()` to query for the control document:

```
rs.slaveOk();
use config;

db.BackupControl.find(
   { "_id" : "BackupControlDocument", "counter" : 1 }
).readConcern('majority');
```

If the secondary member contains the latest control document, it is safe to lock the member. Otherwise, wait until the member contains the document or select a different secondary member that contains the latest control document.

To lock the secondary member, run `db.fsyncLock()` on the member:

```
db.fsyncLock()
```

When calling `db.fsyncLock()`, ensure that the connection is kept open to allow a subsequent call to `db.fsyncUnlock()`.

Step 3: Backup one config server. Run `mongodump` against a config server `mongod` instance to back up the cluster's metadata. You only need to back up one config server. If you are using CSRS config servers and locked a config server secondary in the previous step, perform this step against the locked config server.

Use `mongodump` with the *`--oplog`* option to backup one of the *config servers* (page 742).

```
mongodump --oplog
```

If your deployment uses CSRS config servers, unlock the config server node before proceeding to the next step. To unlock the CSRS member, use `db.fsyncUnlock()` method in the `mongo` shell used to lock the instance.

```
db.fsyncUnlock()
```

Step 4: Back up a replica set member for each shard. Back up the locked replica set members of the shards using `mongodump` with the *`--oplog`* option. You may back up the shards in parallel.

```
mongodump --oplog
```

Step 5: Unlock replica set members for each shard. To unlock the replica set members, use `db.fsyncUnlock()` method in the `mongo` shell. For each locked member, use the same `mongo` shell used to lock the instance.

```
db.fsyncUnlock()
```

Allow these members to catch up with the state of the primary.

Step 6: Re-enable the balancer process. To re-enable to balancer, connect the `mongo` shell to a `mongos` instance and run `sh.setBalancerState()`.

```
use config
sh.setBalancerState(true)
```

Additional Resources See also MongoDB Cloud Manager[115] for seamless automation, backup, and monitoring.

[115] https://cloud.mongodb.com/?jmp=docs

Schedule Backup Window for Sharded Clusters

On this page

- Overview (page 362)
- Procedure (page 362)

Overview In a *sharded cluster*, the balancer process is responsible for distributing sharded data around the cluster, so that each *shard* has roughly the same amount of data.

However, when creating backups from a sharded cluster it is important that you disable the balancer while taking backups to ensure that no chunk migrations affect the content of the backup captured by the backup procedure. Using the procedure outlined in the section *Disable the Balancer* (page 802) you can manually stop the balancer process temporarily. As an alternative you can use this procedure to define a balancing window so that the balancer is always disabled during your automated backup operation.

Procedure If you have an automated backup schedule, you can disable all balancing operations for a period of time. For instance, consider the following command:

```
use config
db.settings.update( { _id : "balancer" }, { $set : { activeWindow : { start : "06:00", stop : "23:00"
```

This operation configures the balancer to run between 6:00am and 11:00pm, server time. Schedule your backup operation to run *and complete* outside of this time. Ensure that the backup can complete outside the window when the balancer is running *and* that the balancer can effectively balance the collection among the shards in the window allotted to each.

Restore a Single Shard

On this page

- Overview (page 362)
- Procedure (page 362)

Overview Restoring a single shard from backup with other unaffected shards requires a number of special considerations and practices. This document outlines the additional tasks you must perform when restoring a single shard.

Consider the following resources on backups in general as well as backup and restoration of sharded clusters specifically:

- *Backup and Restore Sharded Clusters* (page 355)
- *Restore a Sharded Cluster* (page 363)
- *MongoDB Backup Methods* (page 282)

Procedure Always restore *sharded clusters* as a whole. When you restore a single shard, keep in mind that the *balancer* process might have moved *chunks* to or from this shard since the last backup. If that's the case, you must manually move those chunks, as described in this procedure.

Step 1: Restore the shard as you would any other `mongod` instance. See *MongoDB Backup Methods* (page 282) for overviews of these procedures.

Step 2: Manage the chunks. For all chunks that migrate away from this shard, you do not need to do anything at this time. You do not need to delete these documents from the shard because the chunks are automatically filtered out from queries by `mongos`. You can remove these documents from the shard, if you like, at your leisure.

For chunks that migrate to this shard after the most recent backup, you must manually recover the chunks using backups of other shards, or some other source. To determine what chunks have moved, view the `changelog` collection in the *Config Database* (page 823).

Restore a Sharded Cluster

On this page

- Overview (page 363)
- Procedures (page 363)

Overview You can restore a sharded cluster either from *snapshots* (page 343) or from *BSON database dumps* (page 359) created by the `mongodump` tool. This document describes procedures to

- *Restore a Sharded Cluster with Filesystem Snapshots* (page 363)
- *Restore a Sharded Cluster with Database Dumps* (page 365)

Procedures

Restore a Sharded Cluster with Filesystem Snapshots The following procedure outlines the steps to restore a sharded cluster from filesystem snapshots. To create filesystem snapshots of sharded clusters, see *Backup a Sharded Cluster with Filesystem Snapshots* (page 356).

Step 1: Shut down the entire cluster. Stop all `mongos` and `mongod` processes, including all shards *and* all config servers. To stop all members, connect to **each** member and issue following operations:

```
use admin
db.shutdownServer()
```

Step 2: Restore the data files. On each server, extract the data files to the location where the `mongod` instance will access them and restore the following:

- **Data files for each server in each shard.**

  For each shard replica set, restore all the members of the replica set. See *Restore a Replica Set from MongoDB Backups* (page 348).

- **Data files for each config server.**

  Changed in version 3.2: If restoring to a *config server replica set (CSRS)* (page 743), restore the members of the replica set. See *Restore a Replica Set from MongoDB Backups* (page 348).

  Else, if restoring to 3 mirrored config servers, restore the files on each config server `mongod` instance as you would a standalone node.

See also:

Restore a Snapshot (page 346).

Step 3: Restart the config servers.

- If restoring to a *config server replica set* (page 743), restart each member of the CSRS.

```
mongod --configsvr --replSet <CSRS name> --dbpath <config dbpath> --port 27019
```

- Or, if restoring to a three mirrored `mongod` instances, start exactly three `mongod` config server instances.

```
mongod --configsvr --dbpath <config dbpath> --port 27019
```

Step 4: Start one `mongos` instance.

- If using a *CSRS* (page 743) deployment, start `mongos` with the `--configdb` option set to the replica set name and seed list of the CSRS started in the step Restart the config servers. (page **??**)
- Or, if using three mirrored config servers, start `mongos` with the `--configdb` option set to the hostnames (and port numbers) of the config servers started in the step Restart the config servers. (page **??**)

Step 5: If shard hostnames have changed, update the config database. If shard hostnames have changed, connect a `mongo` shell to the `mongos` instance and update the `shards` (page 828) collection in the *Config Database* (page 823) to reflect the new hostnames.

Step 6: Clear per-shard sharding recovery information. If the backup data was from a deployment using *CSRS* (page 743), clear out the no longer applicable recovery information on each shard. For each shard:

1. Restart the replica set members for the shard with the `recoverShardingState` parameter set to `false`. Include additional options as required for your specific configuration.

```
mongod --setParameter=recoverShardingState=false --replSet <replSetName>
```

2. Connect `mongo` shell to the primary of the replica set and delete from the `admin.system.version` collection the document where `_id` equals `minOpTimeRecovery` id. Use write concern `"majority"`.

```
use admin
db.system.version.remove(
   { _id: "minOpTimeRecovery" },
   { writeConcern: { w: "majority" } }
)
```

3. Shut down the replica set members for the shard.

Step 7: Restart all the shard `mongod` instances. Do not include the `recoverShardingState` parameter.

Step 8: Restart the other `mongos` instances.

- If using a CSRS deployment, specify for `--configdb` the config server replica set name and a seed list of the CSRS started in the step Restart the config servers. (page **??**)
- Or, if using three mirrored config servers, specify for `--configdb` the hostnames (and port numbers) of the config servers started in the step Restart the config servers. (page **??**) All `mongos` must specify the same `--configdb` string.

Step 9: Verify that the cluster is operational. Connect to a `mongos` instance from a `mongo` shell and use the `db.printShardingStatus()` method to ensure that the cluster is operational.

```
db.printShardingStatus()
show collections
```

Restore a Sharded Cluster with Database Dumps The following procedure outlines the steps to restore a sharded cluster from the BSON database dumps created by `mongodump`. For information on using `mongodump` to backup sharded clusters, see *Backup a Sharded Cluster with Database Dumps* (page 359).

Changed in version 3.0: `mongorestore` requires a running MongoDB instances. Earlier versions of `mongorestore` did not require a running MongoDB instances and instead used the `--dbpath` option. For instructions specific to your version of `mongorestore`, refer to the appropriate version of the manual.

Step 1: Deploy a new replica set for each shard. For each shard, deploy a new replica set:

1. Start a new `mongod` for each member of the replica set. Include any other configuration as appropriate.
2. Connect a `mongo` to *one* of the `mongod` instances. In the `mongo` shell:
   (a) Run `rs.initiate()`.
   (b) Use `rs.add()` to add the other members of the replica set.

For detailed instructions on deploying a replica set, see *Deploy a Replica Set* (page 667).

Step 2: Deploy new config servers. To deploy config servers as replica set (CSRS), see *Deploy the Config Server Replica Set* (page 766).

To deploy config servers as 3 mirrored `mongod` instances, see *Start 3 Mirrored Config Servers (Deprecated)* (page 769).

Step 3: Start the `mongos` instances. Start the `mongos` instances, specifying the new config servers with `--configdb`. Include any other configuration as appropriate.

For sharded clusters with CSRS, see *Start the mongos Instances* (page 767).

For sharded clusters with 3 mirrored config servers, see *Start the mongos Instances (Deprecated)* (page 770).

Step 4: Add shards to the cluster. Connect a `mongo` shell to a `mongos` instance. Use `sh.addShard()` to add each replica sets as a shard.

For detailed instructions in adding shards to the cluster, see *Add Shards to the Cluster* (page 767).

Step 5: Shut down the `mongos` instances. Once the new sharded cluster is up, shut down all `mongos` instances.

Step 6: Restore the shard data. For each shard, use `mongorestore` to restore the data dump to the primary's data directory. Include the `--drop` option to drop the collections before restoring and, because the *backup procedure* (page 359) included the `--oplog` option, include the `--oplogReplay` option for `mongorestore`.

For example, on the primary for ShardA, run the `mongorestore`. Specify any other configuration as appropriate.

```
mongorestore --drop --oplogReplay /data/dump/shardA
```

After you have finished restoring all the shards, shut down all shard instances.

Step 7: Restore the config server data.

```
mongorestore --drop --oplogReplay /data/dump/configData
```

Step 8: Start one `mongos` instance.

- If using a *CSRS* (page 743) deployment, start `mongos` with the `--configdb` option set to the replica set name and seed list of the CSRS started in the step Deploy new config servers. (page **??**)
- Or, if using three mirrored config servers, start `mongos` with the `--configdb` option set to the hostnames (and port numbers) of the config servers started in the step Deploy new config servers. (page **??**)

Step 9: If shard hostnames have changed, update the config database. If shard hostnames have changed, connect a `mongo` shell to the `mongos` instance and update the `shards` (page 828) collection in the *Config Database* (page 823) to reflect the new hostnames.

Step 10: Restart all the shard `mongod` instances. Do not include the `recoverShardingState` parameter.

Step 11: Restart the other `mongos` instances.

- If using a CSRS deployment, specify for `--configdb` the config server replica set name and a seed list of the CSRS started in the step Deploy new config servers. (page **??**)
- Or, if using three mirrored config servers, specify for `--configdb` the hostnames (and port numbers) of the config servers started in the step Deploy new config servers. (page **??**) All `mongos` must specify the same `--configdb` string.

Step 12: Verify that the cluster is operational. Connect to a `mongos` instance from a `mongo` shell and use the `db.printShardingStatus()` method to ensure that the cluster is operational.

```
db.printShardingStatus()
show collections
```

See also:

MongoDB Backup Methods (page 282), *Backup and Restore Sharded Clusters* (page 355)

Recover Data after an Unexpected Shutdown

On this page

- Process (page 367)
- Procedures (page 368)
- `mongod.lock` (page 369)

If MongoDB does not shutdown cleanly, the on-disk representation of the data files will likely reflect an inconsistent state which could lead to data corruption. [116]

[116] You can also use the `db.collection.validate()` method to test the integrity of a single collection. However, this process is time consuming, and without journaling you can safely assume that the data is in an invalid state and you should either run the repair operation or resync from an intact member of the replica set.

To prevent data inconsistency and corruption, always shut down the database cleanly and use the *durability journaling*. MongoDB writes data to the journal, by default, every 100 milliseconds, such that MongoDB can always recover to a consistent state even in the case of an unclean shutdown due to power loss or other system failure.

If you are *not* running as part of a *replica set* **and** do *not* have journaling enabled, use the following procedure to recover data that may be in an inconsistent state. If you are running as part of a replica set, you should *always* restore from a backup or restart the `mongod` instance with an empty `dbPath` and allow MongoDB to perform an initial sync to restore the data.

To ensure a clean shut down, use one of the following methods:

- `db.shutdownServer()` from the `mongo` shell,
- Your system's *init script*,
- "Control-C" when running `mongod` in interactive mode,
- `kill $(pidof mongod)`; or `kill -2 $(pidof mongod)`,
- On Linux, the *`mongod --shutdown`* option.

See also:

The *Administration* (page 281) documents, including *Replica Set Syncing* (page 656), and the documentation on the *`--repair`* `repairPath` and `storage.journal.enabled` settings.

Process

Indications When you are aware of a `mongod` instance running without journaling that stops unexpectedly **and** you're not running with replication, you should always run the repair operation before starting MongoDB again. If you're using replication, then restore from a backup and allow replication to perform an initial *sync* (page 656) to restore data.

If the `mongod.lock` file in the data directory specified by `dbPath`, `/data/db` by default, is *not* a zero-byte file, then `mongod` will refuse to start, and you will find a message that contains the following line in your MongoDB log our output:

```
Unclean shutdown detected.
```

This indicates that you need to run `mongod` with the *`--repair`* option. If you run repair when the `mongodb.lock` file exists in your `dbPath`, or the optional *`--repairpath`*, you will see a message that contains the following line:

```
old lock file: /data/db/mongod.lock. probably means unclean shutdown
```

If you see this message, as a last resort you may remove the lockfile **and** run the repair operation before starting the database normally, as in the following procedure:

Overview

> **Warning:** Recovering a member of a replica set.
> Do not use this procedure to recover a member of a *replica set*. Instead you should either restore from a *backup* (page 282) or perform an initial sync using data from an intact member of the set, as described in *Resync a Member of a Replica Set* (page 699).

There are two processes to repair data files that result from an unexpected shutdown:

- Use the *`--repair`* option in conjunction with the *`--repairpath`* option. `mongod` will read the existing data files, and write the existing data to new data files.

  You do not need to remove the `mongod.lock` file before using this procedure.

- Use the *--repair* option. mongod will read the existing data files, write the existing data to new files and replace the existing, possibly corrupt, files with new files.

  You must remove the mongod.lock file before using this procedure.

Note: *--repair* functionality is also available in the shell with the db.repairDatabase() helper for the repairDatabase command.

Procedures

Important: Always Run mongod as the same user to avoid changing the permissions of the MongoDB data files.

Repair Data Files and Preserve Original Files To repair your data files using the *--repairpath* option to preserve the original data files unmodified.

Step 1: Start mongod using the option to replace the original files with the repaired files. Start the mongod instance using the *--repair* option **and** the *--repairpath* option. Issue a command similar to the following:

```
mongod --dbpath /data/db --repair --repairpath /data/db0
```

When this completes, the new repaired data files will be in the /data/db0 directory.

Step 2: Start mongod with the new data directory. Start mongod using the following invocation to point the dbPath at /data/db0:

```
mongod --dbpath /data/db0
```

Once you confirm that the data files are operational you may delete or archive the old data files in the /data/db directory. You may also wish to move the repaired files to the old database location or update the dbPath to indicate the new location.

Repair Data Files without Preserving Original Files To repair your data files without preserving the original files, do not use the *--repairpath* option, as in the following procedure:

> **Warning:** After you remove the mongod.lock file you *must* run the *--repair* process before using your database.

Step 1: Remove the stale lock file. For example:

```
rm /data/db/mongod.lock
```

Replace /data/db with your dbPath where your MongoDB instance's data files reside.

Step 2: Start mongod using the option to replace the original files with the repaired files. Start the mongod instance using the *--repair* option, which replaces the original data files with the repaired data files. Issue a command similar to the following:

```
mongod --dbpath /data/db --repair
```

When this completes, the repaired data files will replace the original data files in the /data/db directory.

Step 3: Start `mongod` as usual. Start `mongod` using the following invocation to point the `dbPath` at `/data/db`:

```
mongod --dbpath /data/db
```

`mongod.lock`

In normal operation, you should **never** remove the `mongod.lock` file and start `mongod`. Instead consider the one of the above methods to recover the database and remove the lock files. In dire situations you can remove the lockfile, and start the database using the possibly corrupt files, and attempt to recover data from the database; however, it's impossible to predict the state of the database in these situations.

If you are not running with journaling, and your database shuts down unexpectedly for *any* reason, you should always proceed *as if* your database is in an inconsistent and likely corrupt state. If at all possible restore from *backup* (page 282) or, if running as a *replica set*, restore by performing an initial sync using data from an intact member of the set, as described in *Resync a Member of a Replica Set* (page 699).

8.2.3 MongoDB Tutorials

This page lists the tutorials available as part of the `MongoDB Manual`. In addition to these tutorial in the manual, MongoDB provides *Getting Started Guides* in various driver editions. If there is a process or pattern that you would like to see included here, please open a Jira Case[117].

Installation

- *Install MongoDB Community Edition From Tarball* (page 39)
- *Install MongoDB Community Edition on Red Hat Enterprise or CentOS Linux* (page 23)
- *Install MongoDB Community Edition on Debian* (page 36)
- *Install MongoDB Community Edition on Ubuntu* (page 33)
- *Install MongoDB Community Edition on Amazon Linux* (page 30)
- *Install MongoDB Community Edition on SUSE* (page 27)
- *Install MongoDB Community Edition on OS X* (page 42)
- *Install MongoDB Community Edition on Windows* (page 44)

Administration

Replica Sets

- *Deploy a Replica Set* (page 667)
- *Convert a Standalone to a Replica Set* (page 678)
- *Add Members to a Replica Set* (page 679)
- *Remove Members from Replica Set* (page 682)
- *Replace a Replica Set Member* (page 684)
- *Adjust Priority for Replica Set Member* (page 685)

[117] https://jira.mongodb.org/browse/DOCS

- *Resync a Member of a Replica Set* (page 699)
- *Deploy a Geographically Redundant Replica Set* (page 672)
- *Change the Size of the Oplog* (page 693)
- *Force a Member to Become Primary* (page 697)
- *Change Hostnames in a Replica Set* (page 706)
- *Add an Arbiter to Replica Set* (page 677)
- *Convert a Secondary to an Arbiter* (page 691)
- *Configure a Secondary's Sync Target* (page 710)
- *Configure a Delayed Replica Set Member* (page 689)
- *Configure a Hidden Replica Set Member* (page 687)
- *Configure Non-Voting Replica Set Member* (page 690)
- *Prevent Secondary from Becoming Primary* (page 686)
- *Configure Replica Set Tag Sets* (page 700)
- *Manage Chained Replication* (page 705)
- *Reconfigure a Replica Set with Unavailable Members* (page 704)
- *Recover Data after an Unexpected Shutdown* (page 366)
- *Troubleshoot Replica Sets* (page 711)

Sharding

- *Deploy a Sharded Cluster* (page 765)
- *Convert a Replica Set to a Sharded Cluster* (page 775)
- *Add Shards to a Cluster* (page 773)
- *Remove Shards from an Existing Sharded Cluster* (page 805)
- *Migrate Config Servers with the Same Hostname* (page 792)
- *Migrate Config Servers with Different Hostnames* (page 793)
- *Replace a Config Server* (page 791)
- *Migrate a Sharded Cluster to Different Hardware* (page 794)
- *Backup Cluster Metadata* (page 797)
- *Backup a Small Sharded Cluster with mongodump* (page 355)
- *Backup a Sharded Cluster with Filesystem Snapshots* (page 356)
- *Backup a Sharded Cluster with Database Dumps* (page 359)
- *Restore a Single Shard* (page 362)
- *Restore a Sharded Cluster* (page 363)
- *Schedule Backup Window for Sharded Clusters* (page 362)
- *Manage Shard Tags* (page 816)

Basic Operations

- *Use Database Commands* (page 322)
- *Recover Data after an Unexpected Shutdown* (page 366)
- *Expire Data from Collections by Setting TTL* (page 567)
- *Analyze Performance of Database Operations* (page 326)
- *Rotate Log Files* (page 330)
- *Manage mongod Processes* (page 323)
- *Back Up and Restore with MongoDB Tools* (page 349)
- *Backup and Restore with Filesystem Snapshots* (page 343)

Security

- *Configure Linux iptables Firewall for MongoDB* (page 475)
- *Configure Windows netsh Firewall for MongoDB* (page 479)
- *Enable Client Access Control* (page 435)
- *Enable Internal Authentication* (page 425)
- *Manage Users and Roles* (page 441)
- *Configure MongoDB with Kerberos Authentication on Linux* (page 409)
- *Create a Vulnerability Report* (page 512)

Development Patterns

- *Perform Two Phase Commits* (page 164)
- *Create an Auto-Incrementing Sequence Field* (page 173)
- *Enforce Unique Keys for Sharded Collections* (page 818)
- *Aggregation with the Zip Code Data Set* (page 206)
- *Aggregation with User Preference Data* (page 209)
- *Model Data to Support Keyword Search* (page 273)
- *Limit Number of Elements in an Array after an Update* (page 156)
- *Perform Incremental Map-Reduce* (page 219)
- *Troubleshoot the Map Function* (page 221)
- *Troubleshoot the Reduce Function* (page 222)
- *Store a JavaScript Function on the Server* (page 334)

Text Search Patterns

- *Specify a Language for Text Index* (page 538)
- *Specify Name for text Index* (page 540)
- *Control Search Results with Weights* (page 541)
- *Limit the Number of Entries Scanned* (page 542)

Data Modeling Patterns

- *Model One-to-One Relationships with Embedded Documents* (page 259)
- *Model One-to-Many Relationships with Embedded Documents* (page 260)
- *Model One-to-Many Relationships with Document References* (page 261)
- *Model Data for Atomic Operations* (page 272)
- *Model Tree Structures with Parent References* (page 264)
- *Model Tree Structures with Child References* (page 266)
- *Model Tree Structures with Materialized Paths* (page 269)
- *Model Tree Structures with Nested Sets* (page 270)

See also:

The MongoDB Manual contains administrative documentation and tutorials though out several sections. See *Replica Set Tutorials* (page 665) and *Sharded Cluster Tutorials* (page 764) for additional tutorials and information.

8.3 Administration Reference

UNIX ulimit Settings **(page 372)** Describes user resources limits (i.e. `ulimit`) and introduces the considerations and optimal configurations for systems that run MongoDB deployments.

System Collections **(page 376)** Introduces the internal collections that MongoDB uses to track per-database metadata, including indexes, collections, and authentication credentials.

Database Profiler Output **(page 378)** Describes the data collected by MongoDB's operation profiler, which introspects operations and reports data for analysis on performance and behavior.

Server-side JavaScript **(page 383)** Describes MongoDB's support for executing JavaScript code for server-side operations.

Exit Codes and Statuses **(page 385)** Lists the unique codes returned by `mongos` and `mongod` processes upon exit.

8.3.1 UNIX `ulimit` Settings

On this page

- Resource Utilization (page 373)
- Review and Set Resource Limits (page 374)

Most UNIX-like operating systems, including Linux and OS X, provide ways to limit and control the usage of system resources such as threads, files, and network connections on a per-process and per-user basis. These "ulimits" prevent single users from using too many system resources. Sometimes, these limits have low default values that can cause a number of issues in the course of normal MongoDB operation.

Note: Red Hat Enterprise Linux and CentOS 6 place a max process limitation of 1024 which overrides `ulimit` settings. Create a file named `/etc/security/limits.d/99-mongodb-nproc.conf` with new `soft nproc` and `hard nproc` values to increase the process limit. See `/etc/security/limits.d/90-nproc.conf` file as an example.

Resource Utilization

`mongod` and `mongos` each use threads and file descriptors to track connections and manage internal operations. This section outlines the general resource utilization patterns for MongoDB. Use these figures in combination with the actual information about your deployment and its use to determine ideal `ulimit` settings.

Generally, all `mongod` and `mongos` instances:

- track each incoming connection with a file descriptor *and* a thread.
- track each internal thread or *pthread* as a system process.

mongod

- 1 file descriptor for each data file in use by the `mongod` instance.
- 1 file descriptor for each journal file used by the `mongod` instance when `storage.journal.enabled` is `true`.
- In replica sets, each `mongod` maintains a connection to all other members of the set.

`mongod` uses background threads for a number of internal processes, including *TTL collections* (page 567), replication, and replica set health checks, which may require a small number of additional resources.

mongos

In addition to the threads and file descriptors for client connections, `mongos` must maintain connects to all config servers and all shards, which includes all members of all replica sets.

For `mongos`, consider the following behaviors:

- `mongos` instances maintain a connection pool to each shard so that the `mongos` can reuse connections and quickly fulfill requests without needing to create new connections.
- You can limit the number of incoming connections using the `maxIncomingConnections` run-time option. By restricting the number of incoming connections you can prevent a cascade effect where the `mongos` creates too many connections on the `mongod` instances.

  Note: Changed in version 2.6: MongoDB removed the upward limit on the `maxIncomingConnections` setting.

Review and Set Resource Limits

ulimit

You can use the `ulimit` command at the system prompt to check system limits, as in the following example:

```
$ ulimit -a
-t: cpu time (seconds)         unlimited
-f: file size (blocks)         unlimited
-d: data seg size (kbytes)     unlimited
-s: stack size (kbytes)        8192
-c: core file size (blocks)    0
-m: resident set size (kbytes) unlimited
-u: processes                  192276
-n: file descriptors           21000
-l: locked-in-memory size (kb) 40000
-v: address space (kb)         unlimited
-x: file locks                 unlimited
-i: pending signals            192276
-q: bytes in POSIX msg queues  819200
-e: max nice                   30
-r: max rt priority            65
-N 15:                         unlimited
```

`ulimit` refers to the per-*user* limitations for various resources. Therefore, if your `mongod` instance executes as a user that is also running multiple processes, or multiple `mongod` processes, you might see contention for these resources. Also, be aware that the `processes` value (i.e. `-u`) refers to the combined number of distinct processes and sub-process threads.

You can change `ulimit` settings by issuing a command in the following form:

```
ulimit -n <value>
```

There are both "hard" and the "soft" `ulimits` that affect MongoDB's performance. The "hard" `ulimit` refers to the maximum number of processes that a user can have active at any time. This is the ceiling: no non-root process can increase the "hard" `ulimit`. In contrast, the "soft" `ulimit` is the limit that is actually enforced for a session or process, but any process can increase it up to "hard" `ulimit` maximum.

A low "soft" `ulimit` can cause `can't create new thread, closing connection` errors if the number of connections grows too high. For this reason, it is extremely important to set *both* `ulimit` values to the recommended values.

`ulimit` will modify both "hard" and "soft" values unless the `-H` or `-S` modifiers are specified when modifying limit values.

For many distributions of Linux you can change values by substituting the `-n` option for any possible value in the output of `ulimit -a`. On OS X, use the `launchctl limit` command. See your operating system documentation for the precise procedure for changing system limits on running systems.

After changing the `ulimit` settings, you *must* restart the process to take advantage of the modified settings. You can use the `/proc` file system to see the current limitations on a running process.

Depending on your system's configuration, and default settings, any change to system limits made using `ulimit` may revert following system a system restart. Check your distribution and operating system documentation for more information.

Note: SUSE Linux Enterprise Server and potentially other SUSE distributions ship with virtual memory address space limited to 8 GB by default. You *must* adjust this in order to prevent virtual memory allocation failures as the database grows.

The SLES packages for MongoDB adjust these limits in the default scripts, but you will need to make this change manually if you are using custom scripts and/or the tarball release rather than the SLES packages.

Recommended `ulimit` Settings

Every deployment may have unique requirements and settings; however, the following thresholds and settings are particularly important for `mongod` and `mongos` deployments:

- `-f` (file size): `unlimited`
- `-t` (cpu time): `unlimited`
- `-v` (virtual memory): `unlimited` [118]
- `-n` (open files): `64000`
- `-m` (memory size): `unlimited` [1] [119]
- `-u` (processes/threads): `64000`

Always remember to restart your `mongod` and `mongos` instances after changing the `ulimit` settings to ensure that the changes take effect.

Linux distributions using Upstart

For Linux distributions that use Upstart, you can specify limits within service scripts if you start `mongod` and/or `mongos` instances as Upstart services. You can do this by using `limit` stanzas[120].

Specify the *Recommended ulimit Settings* (page 375), as in the following example:

```
limit fsize unlimited unlimited  # (file size)
limit cpu unlimited unlimited    # (cpu time)
limit as unlimited unlimited     # (virtual memory size)
limit nofile 64000 64000         # (open files)
limit nproc 64000 64000          # (processes/threads)
```

Each `limit` stanza sets the "soft" limit to the first value specified and the "hard" limit to the second.

After changing `limit` stanzas, ensure that the changes take effect by restarting the application services, using the following form:

```
restart <service name>
```

Linux distributions using `systemd`

For Linux distributions that use `systemd`, you can specify limits within the `[Service]` sections of service scripts if you start `mongod` and/or `mongos` instances as `systemd` services. You can do this by using resource limit directives[121].

Specify the *Recommended ulimit Settings* (page 375), as in the following example:

[118] If you limit virtual or resident memory size on a system running MongoDB the operating system will refuse to honor additional allocation requests.

[119] The `-m` parameter to `ulimit` has no effect on Linux systems with kernel versions more recent than 2.4.30. You may omit `-m` if you wish.

[120] http://upstart.ubuntu.com/wiki/Stanzas#limit

[121] http://www.freedesktop.org/software/systemd/man/systemd.exec.html#LimitCPU=

```
[Service]
# Other directives omitted
# (file size)
LimitFSIZE=infinity
# (cpu time)
LimitCPU=infinity
# (virtual memory size)
LimitAS=infinity
# (open files)
LimitNOFILE=64000
# (processes/threads)
LimitNPROC=64000
```

Each `systemd` limit directive sets both the "hard" and "soft" limits to the value specified.

After changing `limit` stanzas, ensure that the changes take effect by restarting the application services, using the following form:

```
systemctl restart <service name>
```

`/proc` File System

Note: This section applies only to Linux operating systems.

The `/proc` file-system stores the per-process limits in the file system object located at `/proc/<pid>/limits`, where `<pid>` is the process's *PID* or process identifier. You can use the following `bash` function to return the content of the `limits` object for a process or processes with a given name:

```
return-limits(){

     for process in $@; do
          process_pids=`ps -C $process -o pid --no-headers | cut -d " " -f 2`

          if [ -z $@ ]; then
             echo "[no $process running]"
          else
             for pid in $process_pids; do
                   echo "[$process #$pid -- limits]"
                   cat /proc/$pid/limits
             done
          fi

     done

}
```

You can copy and paste this function into a current shell session or load it as part of a script. Call the function with one the following invocations:

```
return-limits mongod
return-limits mongos
return-limits mongod mongos
```

8.3.2 System Collections

On this page

- Synopsis (page 377)
- Collections (page 377)

Synopsis

MongoDB stores system information in collections that use the `<database>.system.*` *namespace*, which MongoDB reserves for internal use. Do not create collections that begin with `system`.

MongoDB also stores some additional instance-local metadata in the *local database* (page 723), specifically for replication purposes.

Collections

System collections include these collections stored in the `admin` database:

`admin.system.`**`roles`**
: New in version 2.6.

  The `admin.system.roles` (page 377) collection stores custom roles that administrators create and assign to users to provide access to specific resources.

`admin.system.`**`users`**
: Changed in version 2.6.

  The `admin.system.users` (page 377) collection stores the user's authentication credentials as well as any roles assigned to the user. Users may define authorization roles in the `admin.system.roles` (page 377) collection.

`admin.system.`**`version`**
: New in version 2.6.

  Stores the schema version of the user credential documents.

System collections also include these collections stored directly in each database:

`<database>.system.`**`namespaces`**
: Deprecated since version 3.0: Access this data using `listCollections`.

  The `<database>.system.namespaces` (page 377) collection contains information about all of the database's collections.

`<database>.system.`**`indexes`**
: Deprecated since version 3.0: Access this data using `listIndexes`.

  The `<database>.system.indexes` (page 377) collection lists all the indexes in the database.

`<database>.system.`**`profile`**
: The `<database>.system.profile` (page 377) collection stores database profiling information. For information on profiling, see *Database Profiling* (page 312).

`<database>.system.`**`js`**
: The `<database>.system.js` (page 377) collection holds special JavaScript code for use in *server side JavaScript* (page 383). See *Store a JavaScript Function on the Server* (page 334) for more information.

8.3.3 Database Profiler Output

On this page

- Example `system.profile` Document (page 378)
- Output Reference (page 380)

The database profiler captures data information about read and write operations, cursor operations, and database commands. To configure the database profile and set the thresholds for capturing profile data, see the *Analyze Performance of Database Operations* (page 326) section.

The database profiler writes data in the `system.profile` (page 377) collection, which is a *capped collection*. To view the profiler's output, use normal MongoDB queries on the `system.profile` (page 377) collection.

Note: Because the database profiler writes data to the `system.profile` (page 377) collection in a database, the profiler will profile some write activity, even for databases that are otherwise read-only.

Example `system.profile` Document

The documents in the `system.profile` (page 377) collection have the following form. This example document reflects a find operation:

```
{
   "op" : "query",
   "ns" : "test.c",
   "query" : {
      "find" : "c",
      "filter" : {
         "a" : 1
      }
   },
   "keysExamined" : 2,
   "docsExamined" : 2,
   "cursorExhausted" : true,
   "keyUpdates" : 0,
   "writeConflicts" : 0,
   "numYield" : 0,
   "locks" : {
      "Global" : {
         "acquireCount" : {
            "r" : NumberLong(2)
         }
      },
      "Database" : {
         "acquireCount" : {
            "r" : NumberLong(1)
         }
      },
      "Collection" : {
         "acquireCount" : {
            "r" : NumberLong(1)
         }
      }
   },
   "nreturned" : 2,
```

```
    "responseLength" : 108,
    "millis" : 0,
    "execStats" : {
        "stage" : "FETCH",
        "nReturned" : 2,
        "executionTimeMillisEstimate" : 0,
        "works" : 3,
        "advanced" : 2,
        "needTime" : 0,
        "needYield" : 0,
        "saveState" : 0,
        "restoreState" : 0,
        "isEOF" : 1,
        "invalidates" : 0,
        "docsExamined" : 2,
        "alreadyHasObj" : 0,
        "inputStage" : {
            "stage" : "IXSCAN",
            "nReturned" : 2,
            "executionTimeMillisEstimate" : 0,
            "works" : 3,
            "advanced" : 2,
            "needTime" : 0,
            "needYield" : 0,
            "saveState" : 0,
            "restoreState" : 0,
            "isEOF" : 1,
            "invalidates" : 0,
            "keyPattern" : {
                "a" : 1
            },
            "indexName" : "a_1",
            "isMultiKey" : false,
            "isUnique" : false,
            "isSparse" : false,
            "isPartial" : false,
            "indexVersion" : 1,
            "direction" : "forward",
            "indexBounds" : {
                "a" : [
                    "[1.0, 1.0]"
                ]
            },
            "keysExamined" : 2,
            "dupsTested" : 0,
            "dupsDropped" : 0,
            "seenInvalidated" : 0
        }
    },
    "ts" : ISODate("2015-09-03T15:26:14.948Z"),
    "client" : "127.0.0.1",
    "allUsers" : [ ],
    "user" : ""
}
```

Output Reference

For any single operation, the documents created by the database profiler will include a subset of the following fields. The precise selection of fields in these documents depends on the type of operation.

Changed in version 3.2.0: `system.profile.query.skip` replaces the `system.profile.ntoskip` field.

Changed in version 3.2.0: The information in the `system.profile.ntoreturn` field has been replaced by two separate fields, `system.profile.query.limit` and `system.profile.query.batchSize`. Older drivers or older versions of the `mongo` shell may still use `ntoreturn`; this will appear as `system.profile.query.ntoreturn`.

Note: For the output specific to the version of your MongoDB, refer to the appropriate version of the MongoDB Manual.

`system.profile.`**`op`**
The type of operation. The possible values are:

- `insert`
- `query`
- `update`
- `remove`
- `getmore`
- `command`

`system.profile.`**`ns`**
The *namespace* the operation targets. Namespaces in MongoDB take the form of the *database*, followed by a dot (`.`), followed by the name of the *collection*.

`system.profile.`**`query`**
The *query document* (page 140) used, or for an insert operation, the inserted document. If the document exceeds 50 kilobytes, the value is a string summary of the object. If the string summary exceeds 50 kilobytes, the string summary is truncated, denoted with an ellipsis (`...`) at the end of the string.

Changed in version 3.0.4: For `"getmore"` (page 380) operations on cursors returned from a `db.collection.find()` or a `db.collection.aggregate()`, the `query` (page 380) field contains respectively the query predicate or the issued `aggregate` command document. For details on the `aggregate` command document, see the `aggregate` reference page.

`system.profile.`**`command`**
The command operation. If the command document exceeds 50 kilobytes, the value is a string summary of the object. If the string summary exceeds 50 kilobytes, the string summary is truncated, denoted with an ellipsis (`...`) at the end of the string.

`system.profile.`**`updateobj`**
The `<update>` document passed in during an *update* (page 148) operation. If the document exceeds 50 kilobytes, the value is a string summary of the object. If the string summary exceeds 50 kilobytes, the string summary is truncated, denoted with an ellipsis (`...`) at the end of the string.

`system.profile.`**`cursorid`**
The ID of the cursor accessed by a `query` and `getmore` operations.

`system.profile.`**`keysExamined`**
Changed in version 3.2.0: Renamed from `system.profile.nscanned`.

The number of *index* (page 515) keys that MongoDB scanned in order to carry out the operation.

In general, if keysExamined (page 380) is much higher than nreturned (page 382), the database is scanning many index keys to find the result documents. Consider creating or adjusting indexes to improve query performance..

system.profile.**docsExamined**
Changed in version 3.2.0: Renamed from system.profile.nscannedObjects.

The number of documents in the collection that MongoDB scanned in order to carry out the operation.

system.profile.**moved**
Changed in version 3.0.0: Only appears when using the MMAPv1 storage engine.

This field appears with a value of true when an update operation moved one or more documents to a new location on disk. If the operation did not result in a move, this field does not appear. Operations that result in a move take more time than in-place updates and typically occur as a result of document growth.

system.profile.**nmoved**
Changed in version 3.0.0: Only appears when using the MMAPv1 storage engine.

The number of documents the operation moved on disk. This field appears only if the operation resulted in a move. The field's implicit value is zero, and the field is present only when non-zero.

system.profile.**hasSortStage**
Changed in version 3.2.0: Renamed from system.profile.scanAndOrder.

hasSortStage (page 381) is a boolean that is true when a query **cannot** use the ordering in the index to return the requested sorted results; i.e. MongoDB must sort the documents after it receives the documents from a cursor. The field only appears when the value is true.

system.profile.**ndeleted**
The number of documents deleted by the operation.

system.profile.**ninserted**
The number of documents inserted by the operation.

system.profile.**nMatched**
New in version 2.6.

The number of documents that match the system.profile.query (page 380) condition for the update operation.

system.profile.**nModified**
New in version 2.6.

The number of documents modified by the update operation.

system.profile.**upsert**
A boolean that indicates the update operation's upsert option value. Only appears if upsert is true.

system.profile.**keyUpdates**
The number of *index* (page 515) keys the update changed in the operation. Changing an index key carries a small performance cost because the database must remove the old key and inserts a new key into the B-tree index.

system.profile.**writeConflicts**
New in version 3.0.0.

The number of conflicts encountered during the write operation; e.g. an update operation attempts to modify the same document as another update operation. See also *write conflict*.

system.profile.**numYield**
The number of times the operation yielded to allow other operations to complete. Typically, operations yield when they need access to data that MongoDB has not yet fully read into memory. This allows other operations

that have data in memory to complete while MongoDB reads in data for the yielding operation. For more information, see *the FAQ on when operations yield* (page 837).

system.profile.**locks**

New in version 3.0.0: `locks` (page 382) replaces the `lockStats` field.

The `system.profile.locks` (page 382) provides information for various *lock types and lock modes* (page 836) held during the operation.

The possible lock types are:

Lock Type	Description
`Global`	Represents global lock.
`MMAPV1Journal`	Represents MMAPv1 storage engine specific lock to synchronize journal writes; for non-MMAPv1 storage engines, the mode for `MMAPV1Journal` is empty.
`Database`	Represents database lock.
`Collection`	Represents collection lock.
`Metadata`	Represents metadata lock.
`oplog`	Represents lock on the *oplog*.

The possible locking modes for the lock types are as follows:

Lock Mode	Description
`R`	Represents Shared (S) lock.
`W`	Represents Exclusive (X) lock.
`r`	Represents Intent Shared (IS) lock.
`w`	Represents Intent Exclusive (IX) lock.

The returned lock information for the various lock types include:

system.profile.locks.**acquireCount**

Number of times the operation acquired the lock in the specified mode.

system.profile.locks.**acquireWaitCount**

Number of times the operation had to wait for the `acquireCount` (page 382) lock acquisitions because the locks were held in a conflicting mode. `acquireWaitCount` (page 382) is less than or equal to `acquireCount` (page 382).

system.profile.locks.**timeAcquiringMicros**

Cumulative time in microseconds that the operation had to wait to acquire the locks.

`timeAcquiringMicros` (page 382) divided by `acquireWaitCount` (page 382) gives an approximate average wait time for the particular lock mode.

system.profile.locks.**deadlockCount**

Number of times the operation encountered deadlocks while waiting for lock acquisitions.

For more information on lock modes, see *What type of locking does MongoDB use?* (page 836).

system.profile.**nreturned**

The number of documents returned by the operation.

system.profile.**responseLength**

The length in bytes of the operation's result document. A large `responseLength` (page 382) can affect performance. To limit the size of the result document for a query operation, you can use any of the following:

- *Projections* (page 153)
- `The limit() method`
- `The batchSize() method`

Note: When MongoDB writes query profile information to the log, the `responseLength` (page 382) value is in a field named `reslen`.

`system.profile.`**`millis`**
The time in milliseconds from the perspective of the `mongod` from the beginning of the operation to the end of the operation.

`system.profile.`**`execStats`**
Changed in version 3.0.

A document that contains the execution statistics of the query operation. For other operations, the value is an empty document.

The `system.profile.execStats` (page 383) presents the statistics as a tree; each node provides the statistics for the operation executed during that stage of the query operation.

Note: The following fields list for `execStats` (page 383) is not meant to be exhaustive as the returned fields vary per stage.

`system.profile.execStats.`**`stage`**
New in version 3.0: `stage` (page 383) replaces the `type` field.

The descriptive name for the operation performed as part of the query execution; e.g.

- `COLLSCAN` for a collection scan
- `IXSCAN` for scanning index keys
- `FETCH` for retrieving documents

`system.profile.execStats.`**`inputStages`**
New in version 3.0: `inputStages` (page 383) replaces the `children` field.

An array that contains statistics for the operations that are the input stages of the current stage.

`system.profile.`**`ts`**
The timestamp of the operation.

`system.profile.`**`client`**
The IP address or hostname of the client connection where the operation originates.

For some operations, such as `db.eval()`, the client is `0.0.0.0:0` instead of an actual client.

`system.profile.`**`allUsers`**
An array of authenticated user information (user name and database) for the session. See also *Users* (page 394).

`system.profile.`**`user`**
The authenticated user who ran the operation. If the operation was not run by an authenticated user, this field's value is an empty string.

8.3.4 Server-side JavaScript

On this page

- Overview (page 384)
- Running `.js` files via a `mongo` shell Instance on the Server (page 384)
- Concurrency (page 384)
- Disable Server-Side Execution of JavaScript (page 384)

Overview

MongoDB provides the following commands, methods, and operator that perform server-side execution of JavaScript code:

- `mapReduce` and the corresponding `mongo` shell method `db.collection.mapReduce()`. `mapReduce` operations *map*, or associate, values to keys, and for keys with multiple values, *reduce* the values for each key to a single object. For more information, see *Map-Reduce* (page 214).
- `$where` operator that evaluates a JavaScript expression or a function in order to query for documents.

You can also specify a JavaScript file to the `mongo` shell to run on the server. For more information, see *Running .js files via a mongo shell Instance on the Server* (page 384)

JavaScript in MongoDB

Although these methods use JavaScript, most interactions with MongoDB do not use JavaScript but use an `idiomatic driver` in the language of the interacting application.

You can also disable server-side execution of JavaScript. For details, see *Disable Server-Side Execution of JavaScript* (page 384).

Running `.js` files via a `mongo` shell Instance on the Server

You can specify a JavaScript (`.js`) file to a `mongo` shell instance to execute the file on the server. This is a good technique for performing batch administrative work. When you run `mongo` shell on the server, connecting via the localhost interface, the connection is fast with low latency.

For more information, see *Write Scripts for the mongo Shell* (page 84).

Concurrency

Changed in version 3.2: MongoDB 3.2 uses SpiderMonkey as the JavaScript engine for the `mongo` shell. For information on this change, see *JavaScript Changes in MongoDB 3.2* (page 899).

Refer to the individual method or operator documentation for any concurrency information. See also the *concurrency table* (page 837).

Disable Server-Side Execution of JavaScript

You can disable all server-side execution of JavaScript, by passing the `--noscripting` option on the command line or setting `security.javascriptEnabled` in a configuration file.

See also:

Store a JavaScript Function on the Server (page 334)

8.3.5 Exit Codes and Statuses

MongoDB will return one of the following codes and statuses when exiting. Use this guide to interpret logs and when troubleshooting issues with `mongod` and `mongos` instances.

0

Returned by MongoDB applications upon successful exit.

2

The specified options are in error or are incompatible with other options.

3

Returned by `mongod` if there is a mismatch between hostnames specified on the command line and in the `local.sources` (page 725) collection. `mongod` may also return this status if *oplog* collection in the `local` database is not readable.

4

The version of the database is different from the version supported by the `mongod` (or `mongod.exe`) instance. The instance exits cleanly. Restart `mongod` with the `--upgrade` option to upgrade the database to the version supported by this `mongod` instance.

5

Returned by `mongod` if a `moveChunk` operation fails to confirm a commit.

12

Returned by the `mongod.exe` process on Windows when it receives a Control-C, Close, Break or Shutdown event.

14

Returned by MongoDB applications which encounter an unrecoverable error, an uncaught exception or uncaught signal. The system exits without performing a clean shut down.

20

Message: `ERROR: wsastartup failed <reason>`

Returned by MongoDB applications on Windows following an error in the WSAStartup function.

Message: `NT Service Error`

Returned by MongoDB applications for Windows due to failures installing, starting or removing the NT Service for the application.

45

Returned when a MongoDB application cannot open a file or cannot obtain a lock on a file.

47

MongoDB applications exit cleanly following a large clock skew (32768 milliseconds) event.

48

`mongod` exits cleanly if the server socket closes. The server socket is on port `27017` by default, or as specified to the `--port` run-time option.

49

Returned by `mongod.exe` or `mongos.exe` on Windows when either receives a shutdown message from the *Windows Service Control Manager*.

100

Returned by `mongod` when the process throws an uncaught exception.

8.4 Production Checklist

On this page

- Additional Resources (page 390)

The following checklists provide recommendations that will help you avoid issues in your production MongoDB deployment.

8.4.1 Operations Checklist

On this page

- Filesystem (page 386)
- Replication (page 386)
- Sharding (page 387)
- Journaling: MMAPv1 Storage Engine (page 387)
- Hardware (page 387)
- Deployments to Cloud Hardware (page 387)
- Operating System Configuration (page 388)
- Backups (page 388)
- Monitoring (page 389)
- Load Balancing (page 389)

The following checklist, along with the *Development* (page 389) list, provides recommendations to help you avoid issues in your production MongoDB deployment.

Filesystem

- Align your disk partitions with your RAID configuration.
- Avoid using NFS drives for your `dbPath`. Using NFS drives can result in degraded and unstable performance. See: *Remote Filesystems* (page 302) for more information.
  - VMWare users should use VMWare virtual drives over NFS.
- Linux/Unix: format your drives into XFS or EXT4. If possible, use XFS as it generally performs better with MongoDB.
  - With the WiredTiger storage engine, use of XFS is **strongly recommended** to avoid performance issues found when using EXT4 with WiredTiger.
  - If using RAID, you may need to configure XFS with your RAID geometry.
- Windows: use the NTFS file system. **Do not** use any FAT file system (i.e. FAT 16/32/exFAT).

Replication

- Verify that all non-hidden replica set members are identically provisioned in terms of their RAM, CPU, disk, network setup, etc.
- *Configure the oplog size* (page 693) to suit your use case:

  - The replication oplog window should cover normal maintenance and downtime windows to avoid the need for a full resync.
  - The replication oplog window should cover the time needed to restore a replica set member, either by an initial sync or by restoring from the last backup.
- Ensure that your replica set includes at least three data-bearing nodes with `w:majority` *write concern* (page 179). Three data-bearing nodes are required for replica set-wide data durability.
- Use hostnames when configuring replica set members, rather than IP addresses.
- Ensure full bidirectional network connectivity between all `mongod` instances.
- Ensure that each host can resolve itself.
- Ensure that your replica set contains an odd number of voting members.
- Ensure that `mongod` instances have `0` or `1` votes.
- For high availability, deploy your replica set into a *minimum* of three data centers.

Sharding

- Place your *config servers* (page 742) on dedicated hardware for optimal performance in large clusters. Ensure that the hardware has enough RAM to hold the data files entirely in memory and that it has dedicated storage.
- Use NTP to synchronize the clocks on all components of your sharded cluster.
- Ensure full bidirectional network connectivity between `mongod`, `mongos` and config servers.
- Use CNAMEs to identify your config servers to the cluster so that you can rename and renumber your config servers without downtime.

Journaling: MMAPv1 Storage Engine

- Ensure that all instances use *journaling* (page 606).
- Place the journal on its own low-latency disk for write-intensive workloads. Note that this will affect snapshot-style backups as the files constituting the state of the database will reside on separate volumes.

Hardware

- Use RAID10 and SSD drives for optimal performance.
- SAN and Virtualization:
  - Ensure that each `mongod` has provisioned IOPS for its `dbPath`, or has its own physical drive or LUN.
  - Avoid dynamic memory features, such as memory ballooning, when running in virtual environments.
  - Avoid placing all replica set members on the same SAN, as the SAN can be a single point of failure.

Deployments to Cloud Hardware

- Windows Azure: Adjust the TCP keepalive (`tcp_keepalive_time`) to 100-120. The default TTL for TCP connections on Windows Azure load balancers is too slow for MongoDB's connection pooling behavior.

- Use MongoDB version 2.6.4 or later on systems with high-latency storage, such as Windows Azure, as these versions include performance improvements for those systems. See: Azure Deployment Recommendations[122] for more information.

Operating System Configuration

Linux

- Turn off transparent hugepages and defrag. See *Transparent Huge Pages Settings* (page 319) for more information.
- *Adjust the readahead settings* (page 303) on the devices storing your database files to suit your use case. If your working set is bigger that the available RAM, and the document access pattern is random, consider lowering the readahead to 32 or 16. Evaluate different settings to find an optimal value that maximizes the resident memory and lowers the number of page faults.
- Use the `noop` or `deadline` disk schedulers for SSD drives.
- Use the `noop` disk scheduler for virtualized drives in guest VMs.
- Disable NUMA or set vm.zone_reclaim_mode to 0 and run `mongod` instances with node interleaving. See: *MongoDB and NUMA Hardware* (page 301) for more information.
- Adjust the `ulimit` values on your hardware to suit your use case. If multiple `mongod` or `mongos` instances are running under the same user, scale the `ulimit` values accordingly. See: *UNIX ulimit Settings* (page 372) for more information.
- Use `noatime` for the `dbPath` mount point.
- Configure sufficient file handles (`fs.file-max`), kernel pid limit (`kernel.pid_max`), and maximum threads per process (`kernel.threads-max`) for your deployment. For large systems, the following values provide a good starting point:
  - `fs.file-max` value of 98000,
  - `kernel.pid_max` value of 64000, and
  - `kernel.threads-max` value of 64000
- Ensure that your system has swap space configured. Refer to your operating system's documentation for details on appropriate sizing.
- Ensure that the system default TCP keepalive is set correctly. A value of 300 often provides better performance for replica sets and sharded clusters. See: *Does TCP keepalive time affect MongoDB Deployments?* (page 857) in the Frequently Asked Questions for more information.

Windows

- Consider disabling NTFS "last access time" updates. This is analogous to disabling `atime` on Unix-like systems.

Backups

- Schedule periodic tests of your back up and restore process to have time estimates on hand, and to verify its functionality.

[122] https://docs.mongodb.org/ecosystem/platforms/windows-azure

Monitoring

- Use MongoDB Cloud Manager[123] or Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[124] or another monitoring system to monitor key database metrics and set up alerts for them. Include alerts for the following metrics:
  - lock percent (for the *MMAPv1 storage engine* (page 603))
  - replication lag
  - replication oplog window
  - assertions
  - queues
  - page faults
- Monitor hardware statistics for your servers. In particular, pay attention to the disk use, CPU, and available disk space.

  In the absence of disk space monitoring, or as a precaution:
  - Create a dummy 4 GB file on the `storage.dbPath` drive to ensure available space if the disk becomes full.
  - A combination of `cron+df` can alert when disk space hits a high-water mark, if no other monitoring tool is available.

Load Balancing

- Configure load balancers to enable "sticky sessions" or "client affinity", with a sufficient timeout for existing connections.
- Avoid placing load balancers between MongoDB cluster or replica set components.

8.4.2 Development

On this page

- Data Durability (page 389)
- Schema Design (page 390)
- Replication (page 390)
- Sharding (page 390)
- Drivers (page 390)

The following checklist, along with the *Operations Checklist* (page 386), provides recommendations to help you avoid issues in your production MongoDB deployment.

Data Durability

- Ensure that your replica set includes at least three data-bearing nodes with `w:majority` *write concern* (page 179). Three data-bearing nodes are required for replica-set wide data durability.

[123] https://cloud.mongodb.com/?jmp=docs
[124] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

- Ensure that all instances use *journaling* (page 606).

Schema Design

- Ensure that your schema design does not rely on indexed arrays that grow in length without bound. Typically, best performance can be achieved when such indexed arrays have fewer than 1000 elements.

Replication

- Do not use secondary reads to scale overall read throughput. See: Can I use more replica nodes to scale[125] for an overview of read scaling. For information about secondary reads, see: *Read Preference* (page 651).

Sharding

- Ensure that your shard key distributes the load evenly on your shards. See: *Considerations for Selecting Shard Keys* (page 771) for more information.
- Use *targeted queries* (page 752) for workloads that need to scale with the number of shards.
- Always read from primary nodes for non-targeted queries that may be sensitive to stale or orphaned data[126].
- *Pre-split and manually balance chunks* (page 808) when inserting large data sets into a new non-hashed sharded collection. Pre-splitting and manually balancing enables the insert load to be distributed among the shards, increasing performance for the initial load.

Drivers

- Make use of connection pooling. Most MongoDB drivers support connection pooling. Adjust the connection pool size to suit your use case, beginning at 110-115% of the typical number of concurrent database requests.
- Ensure that your applications handle transient write and read errors during replica set elections.
- Ensure that your applications handle failed requests and retry them if applicable. Drivers **do not** automatically retry failed requests.
- Use exponential backoff logic for database request retries.
- Use `cursor.maxTimeMS()` for reads and *wtimeout* (page 181) for writes if you need to cap execution time for database operations.

8.4.3 Additional Resources

- MongoDB Production Readiness Consulting Package[127]
- MongoDB Ops Optimization Consulting Package[128]

[125] http://askasya.com/post/canreplicashelpscaling
[126] http://blog.mongodb.org/post/74730554385/background-indexing-on-secondaries-and-orphaned
[127] https://www.mongodb.com/products/consulting?jmp=docs#s_product_readiness
[128] https://www.mongodb.com/products/consulting?jmp=docs#ops_optimization

CHAPTER 9

Security

On this page

- Additional Resources (page 514)

Maintaining a secure MongoDB deployment requires administrators to implement controls to ensure that users and applications have access to only the data that they require. MongoDB provides features that allow administrators to implement these controls and restrictions for any MongoDB deployment.

If you are already familiar with security and MongoDB security practices, consider the *Security Checklist* (page 391) for a collection of recommended actions to protect a MongoDB deployment.

9.1 Security Checklist

On this page

- Enable Access Control and Enforce Authentication (page 391)
- Configure Role-Based Access Control (page 392)
- Encrypt Communication (page 392)
- Limit Network Exposure (page 392)
- Audit System Activity (page 392)
- Encrypt and Protect Data (page 392)
- Run MongoDB with a Dedicated User (page 392)
- Run MongoDB with Secure Configuration Options (page 393)
- Request a Security Technical Implementation Guide (where applicable) (page 393)
- Consider Security Standards Compliance (page 393)

This documents provides a list of security measures that you should implement to protect your MongoDB installation.

9.1.1 Enable Access Control and Enforce Authentication

Enable access control and specify the authentication mechanism. You can use the default MongoDB authentication mechanism or an existing external framework. Authentication requires that all clients and servers provide valid credentials before they can connect to the system. In clustered deployments, enable authentication for each MongoDB server.

See *Authentication* (page 393) and *Enable Client Access Control* (page 435).

9.1.2 Configure Role-Based Access Control

Create a user administrator **first**, then create additional users. Create a unique MongoDB user for each person and application that accesses the system.

Create roles that define the exact access a set of users needs. Follow a principle of least privilege. Then create users and assign them only the roles they need to perform their operations. A user can be a person or a client application.

See *Role-Based Access Control* (page 433) and *Manage Users and Roles* (page 441), .

9.1.3 Encrypt Communication

Configure MongoDB to use TLS/SSL for all incoming and outgoing connections. Use TLS/SSL to encrypt communication between `mongod` and `mongos` components of a MongoDB client as well as between all applications and MongoDB.

See *Configure mongod and mongos for TLS/SSL* (page 451).

9.1.4 Limit Network Exposure

Ensure that MongoDB runs in a trusted network environment and limit the interfaces on which MongoDB instances listen for incoming connections. Allow only trusted clients to access the network interfaces and ports on which MongoDB instances are available.

See *Security Hardening* (page 472) and the `bindIp` setting.

9.1.5 Audit System Activity

Track access and changes to database configurations and data. MongoDB Enterprise[1] includes a system auditing facility that can record system events (e.g. user operations, connection events) on a MongoDB instance. These audit records permit forensic analysis and allow administrators to verify proper controls.

See *Auditing* (page 466) and *Configure Auditing* (page 467).

9.1.6 Encrypt and Protect Data

Encrypt MongoDB data on each host using file-system, device, or physical encryption. Protect MongoDB data using file-system permissions. MongoDB data includes data files, configuration files, auditing logs, and key files.

9.1.7 Run MongoDB with a Dedicated User

Run MongoDB processes with a dedicated operating system user account. Ensure that the account has permissions to access data but no unnecessary permissions.

See *Install MongoDB* (page 21) for more information on running MongoDB.

[1] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

9.1.8 Run MongoDB with Secure Configuration Options

MongoDB supports the execution of JavaScript code for certain server-side operations: `mapReduce`, `group`, and `$where`. If you do not use these operations, disable server-side scripting by using the `--noscripting` option on the command line.

Use only the MongoDB wire protocol on production deployments. Do **not** enable the following, all of which enable the web server interface: `net.http.enabled`, `net.http.JSONPEnabled`, and `net.http.RESTInterfaceEnabled`. Leave these *disabled*, unless required for backwards compatibility.

Deprecated since version 3.2: HTTP interface for MongoDB

Keep input validation enabled. MongoDB enables input validation by default through the `wireObjectCheck` setting. This ensures that all documents stored by the `mongod` instance are valid *BSON*.

See *Security Hardening* (page 472) for more information on hardening MongoDB configuration.

9.1.9 Request a Security Technical Implementation Guide (where applicable)

The Security Technical Implementation Guide (STIG) contains security guidelines for deployments within the United States Department of Defense. MongoDB Inc. provides its STIG, upon request, for situations where it is required. Please request a copy[2] for more information.

9.1.10 Consider Security Standards Compliance

For applications requiring HIPAA or PCI-DSS compliance, please refer to the MongoDB Security Reference Architecture[3] to learn more about how you can use the key security capabilities to build compliant application infrastructure.

9.2 Authentication

On this page

- Authentication Methods (page 393)
- Authentication Mechanisms (page 394)
- Internal Authentication (page 394)
- Authentication on Sharded Clusters (page 394)

Authentication is the process of verifying the identity of a client. When access control, i.e. *authorization* (page 433), is enabled, MongoDB requires all clients to authenticate themselves in order to determine their access.

Although authentication and *authorization* (page 433) are closely connected, authentication is distinct from authorization. Authentication verifies the identity of a user; authorization determines the verified user's access to resources and operations.

9.2.1 Authentication Methods

To authenticate a *user* (page 394), MongoDB provides the `db.auth()` method.

For the `mongo` shell and the MongoDB tools, you can also authenticate a user by passing in the user authentication information from the command line.

[2] http://www.mongodb.com/lp/contact/stig-requests
[3] http://info.mongodb.com/rs/mongodb/images/MongoDB_Security_Architecture_WP.pdf

9.2.2 Authentication Mechanisms

MongoDB supports a number of *authentication mechanisms* (page 398) that clients can use to verify their identity. These mechanisms allow MongoDB to integrate into your existing authentication system.

MongoDB supports multiple authentication mechanisms:

- *SCRAM-SHA-1* (page 399)
- *MongoDB Challenge and Response (MONGODB-CR)* (page 400)

  Changed in version 3.0: New challenge-response users created in 3.0 will use `SCRAM-SHA-1`. If using 2.6 user data, MongoDB 3.0 will continue to use the `MONGODB-CR`.
- *x.509 Certificate Authentication* (page 401).

In addition to supporting the aforementioned mechanisms, MongoDB Enterprise also supports the following mechanisms:

- *LDAP proxy authentication* (page 406), and
- *Kerberos authentication* (page 406).

9.2.3 Internal Authentication

In addition to verifying the identity of a client, MongoDB can require members of replica sets and sharded clusters to *authenticate their membership* (page 423) to their respective replica set or sharded cluster. See *Internal Authentication* (page 423) for more information.

9.2.4 Authentication on Sharded Clusters

In sharded clusters, clients generally authenticate directly to the `mongos` instances. However, some maintenance operations may require authenticating directly to a specific shard. For more information on authentication and sharded clusters, see *Sharded Cluster Users* (page 395).

Users

On this page

- User Management Interface (page 394)
- Authentication Database (page 395)
- Authenticate a User (page 395)
- Centralized User Data (page 395)
- Sharded Cluster Users (page 395)
- Localhost Exception (page 396)

To authenticate a client in MongoDB, you must add a corresponding user to MongoDB.

User Management Interface

To add a user, MongoDB provides the `db.createUser()` method. When adding a user, you can assign *roles* (page 433) to the user in order to grant privileges.

Note: The first user created in the database should be a user administrator who has the privileges to manage other users. See *Enable Client Access Control* (page 435).

You can also update existing users, such as to change password and grant or revoke roles. For a full list of user management methods, see *user-management-methods*.

Authentication Database

When adding a user, you create the user in a specific database. This database is the authentication database for the user.

A user can have privileges across different databases; i.e. a user's privileges are not limited to the authentication database. By assigning to the user roles in other databases, a user created in one database can have permissions to act on other databases. For more information on roles, see *Role-Based Access Control* (page 433).

The user's name and authentication database serve as a unique identifier for that user. That is, if two users have the same name but are created in different databases, they are two separate users. If you intend to have a single user with permissions on multiple databases, create a single user with roles in the applicable databases instead of creating the user multiple times in different databases.

Authenticate a User

To authenticate a user, either

- Use the command line authentication options (e.g. `-u`, `-p`, `--authenticationDatabase`) when connecting to the `mongod` or `mongos` instance, or
- Connect first to the `mongod` or `mongos` instance, and then run the `authenticate` command or the `db.auth()` method against the authentication database.

To authenticate, the client must authenticate the user against the user's *authentication database*.

For instance, if using the `mongo` shell as a client, you can specify the authentication database for the user with the `--authenticationDatabase` option.

Centralized User Data

Changed in version 2.6.

MongoDB stores all user information, including `name` (page 497), `password` (page 497), and the `user's authentication database` (page 497), in the *system.users* (page 497) collection in the `admin` database.

Do not access this collection directly but instead use the *user management commands*.

Sharded Cluster Users

To create users for a sharded cluster, connect to the `mongos` instance and add the users. Clients then authenticate these users through the `mongos` instances.

Changed in version 2.6: MongoDB stores these sharded cluster user data in the `admin` database of the *config servers*. Previously, the credentials for authenticating to a database on a sharded cluster resided on the *primary shard* (page 742) for that database.

Shard Local Users However, some maintenance operations, such as `cleanupOrphaned`, `compact`, `rs.reconfig()`, require direct connections to specific shards in a sharded cluster. To perform these operations, you must connect directly to the shard and authenticate as a *shard local* administrative user.

To create a *shard local* administrative user, connect directly to the shard and create the user. MongoDB stores *shard local* users in the `admin` database of the shard itself.

These *shard local* users are completely independent from the users added to the sharded cluster via `mongos`. *Shard local* users are local to the shard and are inaccessible by `mongos`.

Direct connections to a shard should only be for shard-specific maintenance and configuration. In general, clients should connect to the sharded cluster through the `mongos`.

Localhost Exception

The localhost exception allows you to enable access control and then create the first user in the system. With the localhost exception, after you enable access control, connect to the localhost interface and create the first user in the `admin` database. The first user must have privileges to create other users, such as a user with the `userAdmin` (page 488) or `userAdminAnyDatabase` (page 493) role.

Changed in version 3.0: The localhost exception changed so that these connections *only* have access to create the first user on the `admin` database. In previous versions, connections that gained access using the localhost exception had unrestricted access to the MongoDB instance.

The localhost exception applies only when there are no users created in the MongoDB instance.

In the case of a sharded cluster, the localhost exception applies to each shard individually as well as to the cluster as a whole. Once you create a sharded cluster and add a user administrator through the `mongos` instance, you must still prevent unauthorized access to the individual shards. Follow one of the following steps for each shard in your cluster:

- Create an administrative user, or
- Disable the localhost exception at startup. To disable the localhost exception, set the `enableLocalhostAuthBypass` parameter to `0`.

Add Users

On this page

- Overview (page 396)
- Prerequisites (page 397)
- Examples (page 397)
  - Username/Password Authentication (page 397)
  - Kerberos Authentication (page 397)
  - LDAP Authentication (page 398)
  - x.509 Client Certificate Authentication (page 398)

Overview MongoDB employs role-based access control (RBAC) to determine access for users. A user is granted one or more *roles* (page 434) that determine the user's access or privileges to MongoDB *resources* (page 498) and the *actions* (page 500) that user can perform. A user should have only the minimal set of privileges required to ensure a system of *least privilege*.

Each application and user of a MongoDB system should map to a distinct user. This *access isolation* facilitates access revocation and ongoing user maintenance.

Prerequisites If you have enabled access control for your deployment, you can use the *localhost exception* (page 396) to create the first user in the system. This first user must have privileges to create other users. As of MongoDB 3.0, with the localhost exception, you can only create users on the `admin` database. Once you create the first user, you must authenticate as that user to add subsequent users. *Enable Client Access Control* (page 435) provides more detail about adding users when enabling access control for a deployment.

For routine user creation, you must possess the following permissions:

- To create a new user in a database, you must have the `createUser` (page 501) *action* (page 500) on that *database resource* (page 499).
- To grant roles to a user, you must have the `grantRole` (page 502) *action* (page 500) on the role's database.

The `userAdmin` (page 488) and `userAdminAnyDatabase` (page 493) built-in roles provide `createUser` (page 501) and `grantRole` (page 502) actions on their respective *resources* (page 498).

Examples To create a user in a MongoDB deployment, you connect to the deployment, and then use the `db.createUser()` method or `createUser` command to add the user.

Username/Password Authentication The following operation creates a user in the `reporting` database with the specified name, password, and roles.

```
use reporting
db.createUser(
  {
    user: "reportsUser",
    pwd: "12345678",
    roles: [
       { role: "read", db: "reporting" },
       { role: "read", db: "products" },
       { role: "read", db: "sales" },
       { role: "readWrite", db: "accounts" }
    ]
  }
)
```

Enable Client Access Control (page 435) provides more details about enforcing authentication for your MongoDB deployment.

Kerberos Authentication Users that will authenticate to MongoDB using an external authentication mechanism, such as Kerberos, must be created in the `$external` database, which allows `mongos` or `mongod` to consult an external source for authentication.

For Kerberos authentication, you must add the Kerberos principal as the username. You do not need to specify a password.

The following operation adds the Kerberos principal `reportingapp@EXAMPLE.NET` with read-only access to the `records` database.

```
use $external
db.createUser(
    {
      user: "reportingapp@EXAMPLE.NET",
      roles: [
         { role: "read", db: "records" }
      ]
    }
)
```

Configure MongoDB with Kerberos Authentication on Linux (page 409) and *Configure MongoDB with Kerberos Authentication on Windows* (page 412) provide more details about setting up Kerberos authentication for your MongoDB deployment.

LDAP Authentication Users that will authenticate to MongoDB using an external authentication mechanism, such as LDAP, must be created in the `$external` database, which allows `mongos` or `mongod` to consult an external source for authentication.

For LDAP authentication, you must specify a username. You do not need to specify the password, as that is handled by the LDAP service.

The following operation adds the `reporting` user with read-only access to the `records` database.

```
use $external
db.createUser(
    {
      user: "reporting",
      roles: [
         { role: "read", db: "records" }
      ]
    }
)
```

Authenticate Using SASL and LDAP with ActiveDirectory (page 417) and *Authenticate Using SASL and LDAP with OpenLDAP* (page 420) provide more detail about using authenticating using LDAP.

x.509 Client Certificate Authentication Users that will authenticate to MongoDB using an external authentication mechanism, such as x.509 Client Certificate Authentication, must be created in the `$external` database, which allows `mongos` or `mongod` to consult an external source for authentication.

For x.509 Client Certificate authentication, you must add the value of the `subject` from the client certificate as a MongoDB user. Each unique x.509 client certificate corresponds to a single MongoDB user. You do not need to specify a password.

The following operation adds the client certificate subject `CN=myName,OU=myOrgUnit,O=myOrg,L=myLocality,ST=myStat` user with read-only access to the `records` database.

```
use $external
db.createUser(
    {
      user: "CN=myName,OU=myOrgUnit,O=myOrg,L=myLocality,ST=myState,C=myCountry",
      roles: [
         { role: "read", db: "records" }
      ]
    }
)
```

Use x.509 Certificates to Authenticate Clients (page 403) provides details about setting up x.509 Client Certificate authentication for your MongoDB deployment.

Authentication Mechanisms

On this page

- Default Authentication Mechanism (page 399)
- Specify Authentication Mechanism (page 399)

MongoDB supports the following authentication mechanisms:

- *SCRAM-SHA-1* (page 399)
- *MongoDB Challenge and Response (MONGODB-CR)* (page 400)

  Changed in version 3.0: New challenge-response users created in 3.0 will use `SCRAM-SHA-1`. If using 2.6 user data, MongoDB 3.0 will continue to use the `MONGODB-CR`.
- *x.509 Certificate Authentication* (page 401).

In addition, MongoDB Enterprise also provides supports for additional mechanisms. See *Enterprise Authentication Mechanisms* (page 405) for additional mechanisms available in MongoDB Enterprise.

Default Authentication Mechanism

Changed in version 3.0.

MongoDB uses the *SCRAM-SHA-1* (page 399) as the default challenge and response authentication mechanism. Previous versions used *MONGODB-CR* (page 400) as the default.

Specify Authentication Mechanism

To specify the authentication mechanism to use, set the `authenticationMechanisms` parameter for `mongod` and `mongos`.

Clients specify the authentication mechanism in the `db.auth()` method. For the `mongo` shell and the MongoDB tools, you can also specify the authentication mechanism from the command line.

SCRAM-SHA-1

On this page

- `SCRAM-SHA-1` Advantages (page 400)
- `SCRAM-SHA-1` and `MongoDB-CR` User Credentials (page 400)
- Additional Information (page 400)

New in version 3.0.

`SCRAM-SHA-1` is the default authentication mechanism for MongoDB. `SCRAM-SHA-1` is an IETF standard, RFC 5802[4], that defines best practice methods for implementation of challenge-response mechanisms for authenticating users with passwords.

`SCRAM-SHA-1` verifies the supplied user credentials against the user's `name` (page 497), `password` (page 497) and `authentication database` (page 497). The authentication database is the database where the user was created, and together with the user's name, serves to identify the user.

Note: A driver upgrade is **necessary** to use the `SCRAM-SHA-1` authentication mechanism if your current driver version does not support `SCRAM-SHA-1`. See *required driver versions* (page 958) for details.

[4] https://tools.ietf.org/html/rfc5802

SCRAM-SHA-1 Advantages MongoDB's implementation of SCRAM-SHA-1 represents an improvement in security over the previously-used MONGODB-CR, providing:

- A tunable work factor (iterationCount),
- Per-user random salts rather than server-wide salts,
- A cryptographically stronger hash function (SHA-1 rather than MD5), and
- Authentication of the server to the client as well as the client to the server.

SCRAM-SHA-1 and MongoDB-CR User Credentials SCRAM-SHA-1 is the default mechanism for MongoDB versions beginning with the 3.0 series. However, if you are upgrading a MongoDB 2.6 instances that already have users credentials, MongoDB will continue to use MONGODB-CR for challenge-response authentication until you upgrade the authentication schema.

Even when using the MONGODB-CR authentication mechanism, clients and drivers that support MongoDB 3.0 features (see *Driver Compatibility Changes* (page 950)) will use the SCRAM communication protocol. That is, MONGODB-CR authentication mechanism also implies *SCRAM-SHA-1* (page 399).

For details on upgrading the authentication schema model to SCRAM-SHA-1, see *Upgrade to SCRAM-SHA-1* (page 957).

Warning: The procedure to upgrade to SCRAM-SHA-1 **discards** the MONGODB-CR credentials used by 2.6. As such, the procedure is **irreversible**, short of restoring from backups.
The procedure also disables MONGODB-CR as an authentication mechanism.

Additional Information

- Blog Post: Improved Password-Based Authentication in MongoDB 3.0: SCRAM Explained (Part 1)[5]
- Blog Post: Improved Password-Based Authentication in MongoDB 3.0: SCRAM Explained (Part 2)[6]

MONGODB-CR

On this page

- MONGODB-CR and SCRAM-SHA-1 (page 400)

MONGODB-CR is a challenge-response mechanism that authenticates users through passwords. MONGODB-CR verifies supplied user credentials against the user's name (page 497), password (page 497) and authentication database (page 497). The authentication database is the database where the user was created, and the user's database and the user's name together serve to identify the user.

MONGODB-CR and SCRAM-SHA-1 Changed in version 3.0.

MongoDB no longer defaults to MONGODB-CR and instead uses SCRAM-SHA-1 as the default authentication mechanism.

Even when using the MONGODB-CR authentication mechanism, clients and drivers that support MongoDB 3.0 features (see *Driver Compatibility Changes* (page 950)) will use the SCRAM communication protocol. That is, MONGODB-CR authentication mechanism also implies *SCRAM-SHA-1* (page 399).

[5] https://www.mongodb.com/blog/post/improved-password-based-authentication-mongodb-30-scram-explained-part-1?jmp=docs
[6] https://www.mongodb.com/blog/post/improved-password-based-authentication-mongodb-30-scram-explained-part-2?jmp=docs

x.509

On this page

- Certificate Authority (page 401)
- Client x.509 Certificates (page 401)
- Member x.509 Certificates (page 402)

New in version 2.6.

MongoDB supports x.509 certificate authentication for client authentication and internal authentication of the members of replica sets and sharded clusters.

x.509 certificate authentication requires a secure *TLS/SSL connection* (page 451).

Certificate Authority For production use, your MongoDB deployment should use valid certificates generated and signed by a single certificate authority. You or your organization can generate and maintain an independent certificate authority, or use certificates generated by a third-party SSL vendor. Obtaining and managing certificates is beyond the scope of this documentation.

Client x.509 Certificates To authenticate to servers, clients can use x.509 certificates instead of usernames and passwords.

Client Certificate Requirements The client certificate must have the following properties:

- A single Certificate Authority (CA) must issue the certificates for both the client and the server.
- Client certificates must contain the following fields:

```
keyUsage = digitalSignature
extendedKeyUsage = clientAuth
```

- Each unique MongoDB user must have a unique certificate.
- A client x.509 certificate's subject, which contains the Distinguished Name (`DN`), must **differ** from that of a *Member x.509 Certificate* (page 430). Specifically, the subjects must differ with regards to at least one of the following attributes: Organization (`O`), the Organizational Unit (`OU`) or the Domain Component (`DC`).

  Warning: If a client x.509 certificate's subject has the same `O`, `OU`, and `DC` combination as the *Member x.509 Certificate* (page 430), the client will be identified as a cluster member and granted full permission on the system.

MongoDB User and `$external` Database To authenticate with a client certificate, you must first add the value of the `subject` from the client certificate as a MongoDB user. Each unique x.509 client certificate corresponds to a single MongoDB user; i.e. you cannot use a single client certificate to authenticate more than one MongoDB user.

Add the user in the `$external` database; i.e. the *Authentication Database* (page 395) is the `$external` database

Authenticate To authenticate using x.509 client certificate, connect to MongoDB over TLS/SSL connection; i.e. include the `--ssl` and `--sslPEMKeyFile` command line options.

Then in the `$external` database, use `db.auth()` to authenticate the *user corresponding to the client certificate* (page 401).

For an example, see *Use x.509 Certificates to Authenticate Clients* (page 403)

Member x.509 Certificates For internal authentication, members of sharded clusters and replica sets can use x.509 certificates instead of keyfiles, which use *MONGODB-CR* (page 400) authentication mechanism.

Member Certificate Requirements The member certificate, used for internal authentication to verify membership to the sharded cluster or a replica set, must have the following properties:

- A single Certificate Authority (CA) must issue all the x.509 certificates for the members of a sharded cluster or a replica set.
- The Distinguished Name (`DN`), found in the member certificate's `subject`, must specify a non-empty value for *at least one* of the following attributes: Organization (`O`), the Organizational Unit (`OU`) or the Domain Component (`DC`).
- The Organization attributes (`O`'s), the Organizational Unit attributes (`OU`'s), and the Domain Components (`DC`'s) must match those from the certificates for the other cluster members. To match, the certificate must match all specifications of these attributes, or even the non-specification of these attributes. The order of the attributes does not matter.

  In the following example, the two `DN`'s contain matching specifications for `O`, `OU` as well as the non-specification of the `DC` attribute.

  ```
  CN=host1,OU=Dept1,O=MongoDB,ST=NY,C=US
  C=US, ST=CA, O=MongoDB, OU=Dept1, CN=host2
  ```

  However, the following two `DN`'s contain a mismatch for the `OU` attribute since one contains two `OU` specifications and the other, only one specification.

  ```
  CN=host1,OU=Dept1,OU=Sales,O=MongoDB
  CN=host2,OU=Dept1,O=MongoDB
  ```

- Either the Common Name (`CN`) or one of the Subject Alternative Name (`SAN`) entries must match the hostname of the server, used by the other members of the cluster.

  For example, the certificates for a cluster could have the following subjects:

  ```
  subject= CN=<myhostname1>,OU=Dept1,O=MongoDB,ST=NY,C=US
  subject= CN=<myhostname2>,OU=Dept1,O=MongoDB,ST=NY,C=US
  subject= CN=<myhostname3>,OU=Dept1,O=MongoDB,ST=NY,C=US
  ```

- *If* the certificate includes the Extended Key Usage (`extendedKeyUsage`) setting, the value must include `clientAuth` ("TLS Web Client Authentication").

  ```
  extendedKeyUsage = clientAuth
  ```

  You can also use a certificate that does not include the Extended Key Usage (EKU).

MongoDB Configuration To specify x.509 for internal authentication, in addition to the other SSL configurations appropriate for your deployment, for each member of the replica set or sharded cluster, include either:

- `security.clusterAuthMode` and `net.ssl.clusterFile` if using a `configuration file`, or
- `--clusterAuthMode` and `--sslClusterFile` command line options.

Member Certificate and `PEMKeyFile` To configure MongoDB for client certificate authentication, the `mongod` and `mongos` specify a `PEMKeyFile` to prove its identity to clients, either through `net.ssl.PEMKeyFile` setting in the configuration file or `--sslPEMKeyFile` command line option.

If no `clusterFile` certificate is specified for internal member authentication, MongoDB will attempt to use the `PEMKeyFile` certificate for member authentication. In order to use `PEMKeyFile` certificate for internal authentication as well as for client authentication, then the `PEMKeyFile` certificate must either:

- Omit `extendedKeyUsage` or
- Specify `extendedKeyUsage` values that include `clientAuth` in addition to `serverAuth`.

For an example of x.509 internal authentication, see *Use x.509 Certificate for Membership Authentication* (page 430).

Use x.509 Certificates to Authenticate Clients

On this page

- Prerequisites (page 403)
- Procedures (page 403)

New in version 2.6.

MongoDB supports x.509 certificate authentication for use with a secure *TLS/SSL connection* (page 451). The x.509 client authentication allows *clients to authenticate to servers with certificates* (page 403) rather than with a username and password.

To use x.509 authentication for the internal authentication of replica set/sharded cluster members, see *Use x.509 Certificate for Membership Authentication* (page 430).

Prerequisites

Important: A full description of TLS/SSL, PKI (Public Key Infrastructure) certificates, in particular x.509 certificates, and Certificate Authority is beyond the scope of this document. This tutorial assumes prior knowledge of TLS/SSL as well as access to valid x.509 certificates.

Certificate Authority For production use, your MongoDB deployment should use valid certificates generated and signed by a single certificate authority. You or your organization can generate and maintain an independent certificate authority, or use certificates generated by a third-party SSL vendor. Obtaining and managing certificates is beyond the scope of this documentation.

Client x.509 Certificate The client certificate must have the following properties:

- A single Certificate Authority (CA) must issue the certificates for both the client and the server.
- Client certificates must contain the following fields:

```
keyUsage = digitalSignature
extendedKeyUsage = clientAuth
```

- Each unique MongoDB user must have a unique certificate.
- A client x.509 certificate's subject, which contains the Distinguished Name (`DN`), must **differ** from that of a *Member x.509 Certificate* (page 430). Specifically, the subjects must differ with regards to at least one of the following attributes: Organization (`O`), the Organizational Unit (`OU`) or the Domain Component (`DC`).

  Warning: If a client x.509 certificate's subject has the same `O`, `OU`, and `DC` combination as the *Member x.509 Certificate* (page 430), the client will be identified as a cluster member and granted full permission on the system.

Procedures

Configure MongoDB Server

Use Command-line Options You can configure the MongoDB server from the command line, e.g.:

```
mongod --clusterAuthMode x509 --sslMode requireSSL --sslPEMKeyFile <path to SSL certificate and key P
```

> **Warning:** If the *--sslCAFile* option and its target file are not specified, x.509 client and member authentication will not function. `mongod`, and `mongos` in sharded systems, will not be able to verify the certificates of processes connecting to it against the trusted certificate authority (CA) that issued them, breaking the certificate chain.
> As of version 2.6.4, `mongod` will not start with x.509 authentication enabled if the CA file is not specified.

Use Configuration File You may also specify these options in the `configuration file`.

Starting in MongoDB 2.6, you can specify the configuration for MongoDB in `YAML format`, e.g.:

```
security:
   clusterAuthMode: x509
net:
   ssl:
      mode: requireSSL
      PEMKeyFile: <path to TLS/SSL certificate and key PEM file>
      CAFile: <path to root CA PEM file>
```

For backwards compatibility, you can also specify the configuration using the older configuration file format[7], e.g.:

```
clusterAuthMode = x509
sslMode = requireSSL
sslPEMKeyFile = <path to TLS/SSL certificate and key PEM file>
sslCAFile = <path to the root CA PEM file>
```

Include any additional options, TLS/SSL or otherwise, that are required for your specific configuration.

Add x.509 Certificate `subject` as a User To authenticate with a client certificate, you must first add the value of the `subject` from the client certificate as a MongoDB user. Each unique x.509 client certificate corresponds to a single MongoDB user; i.e. you cannot use a single client certificate to authenticate more than one MongoDB user.

Note: The RDNs in the `subject` string must be compatible with the RFC2253[8] standard.

---

1. You can retrieve the `RFC2253` formatted `subject` from the client certificate with the following command:

   ```
   openssl x509 -in <pathToClient PEM> -inform PEM -subject -nameopt RFC2253
   ```

   The command returns the `subject` string as well as certificate:

   ```
   subject= CN=myName,OU=myOrgUnit,O=myOrg,L=myLocality,ST=myState,C=myCountry
   -----BEGIN CERTIFICATE-----
   # ...
   -----END CERTIFICATE-----
   ```

2. Add the `RFC2253` compliant value of the `subject` as a user. Omit spaces as needed.

   For example, in the `mongo` shell, to add the user with both the `readWrite` role in the `test` database and the `userAdminAnyDatabase` role which is defined only in the `admin` database:

[7] https://docs.mongodb.org/v2.4/reference/configuration-options
[8] https://www.ietf.org/rfc/rfc2253.txt

```
db.getSiblingDB("$external").runCommand(
  {
    createUser: "CN=myName,OU=myOrgUnit,O=myOrg,L=myLocality,ST=myState,C=myCountry",
    roles: [
             { role: 'readWrite', db: 'test' },
             { role: 'userAdminAnyDatabase', db: 'admin' }
           ],
    writeConcern: { w: "majority" , wtimeout: 5000 }
  }
)
```

In the above example, to add the user with the `readWrite` role in the `test` database, the role specification document specified `'test'` in the `db` field. To add `userAdminAnyDatabase` role for the user, the above example specified `'admin'` in the `db` field.

Note: Some roles are defined only in the `admin` database, including: `clusterAdmin`, `readAnyDatabase`, `readWriteAnyDatabase`, `dbAdminAnyDatabase`, and `userAdminAnyDatabase`. To add a user with these roles, specify `'admin'` in the `db`.

See *Manage Users and Roles* (page 441) for details on adding a user with roles.

Authenticate with a x.509 Certificate To authenticate with a client certificate, you must first add a MongoDB user that corresponds to the client certificate. See *Add x.509 Certificate subject as a User* (page 404).

To authenticate, use the `db.auth()` method in the `$external` database, specifying `"MONGODB-X509"` for the `mechanism` field, and the *user that corresponds to the client certificate* (page 404) for the `user` field.

For example, if using the `mongo` shell,

1. Connect `mongo` shell to the `mongod` set up for SSL:

   ```
   mongo --ssl --sslPEMKeyFile <path to CA signed client PEM file> --sslCAFile <path to root CA PEM
   ```

2. To perform the authentication, use the `db.auth()` method in the `$external` database. For the `mechanism` field, specify `"MONGODB-X509"`, and for the `user` field, specify the user, or the `subject`, that corresponds to the client certificate.

   ```
   db.getSiblingDB("$external").auth(
     {
       mechanism: "MONGODB-X509",
       user: "CN=myName,OU=myOrgUnit,O=myOrg,L=myLocality,ST=myState,C=myCountry"
     }
   )
   ```

Enterprise Authentication Mechanisms

On this page

- Kerberos Authentication (page 406)
- LDAP Proxy Authority Authentication (page 406)

In addition to the authentication mechanisms offered, MongoDB Enterprise provides integration with the following authentication mechanisms.

Kerberos Authentication

MongoDB Enterprise[9] supports authentication using a Kerberos service. Kerberos is an industry standard authentication protocol for large client/server systems.

To use MongoDB with Kerberos, you must have a properly configured Kerberos deployment, configured *Kerberos service principals* (page 407) for MongoDB, and added *Kerberos user principal* (page 407) to MongoDB.

For more information on Kerberos and MongoDB, see:

- *Kerberos Authentication* (page 406),
- *Configure MongoDB with Kerberos Authentication on Linux* (page 409) and
- *Configure MongoDB with Kerberos Authentication on Windows* (page 412).

LDAP Proxy Authority Authentication

MongoDB Enterprise (excluding Windows version)[10] supports proxy authentication through a Lightweight Directory Access Protocol (LDAP) service.

LDAP support for user authentication requires proper configuration of the `saslauthd` daemon process as well as the MongoDB server.

For more information on LDAP and MongoDB, see

- *LDAP Proxy Authority Authentication* (page 416),
- *Authenticate Using SASL and LDAP with OpenLDAP* (page 420) and
- *Authenticate Using SASL and LDAP with ActiveDirectory* (page 417).

Kerberos Authentication

On this page

- Overview (page 406)
- Kerberos Components and MongoDB (page 406)
- Operational Considerations (page 408)
- Kerberized MongoDB Environments (page 408)
- Additional Resources (page 416)

New in version 2.4.

Overview MongoDB Enterprise provides support for Kerberos authentication of MongoDB clients to `mongod` and `mongos`. Kerberos is an industry standard authentication protocol for large client/server systems. Kerberos allows MongoDB and applications to take advantage of existing authentication infrastructure and processes.

Kerberos Components and MongoDB

[9] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs
[10] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Principals In a Kerberos-based system, every participant in the authenticated communication is known as a "principal", and every principal must have a unique name.

Principals belong to administrative units called *realms*. For each realm, the Kerberos Key Distribution Center (KDC) maintains a database of the realm's principal and the principals' associated "secret keys".

For a client-server authentication, the client requests from the KDC a "ticket" for access to a specific asset. KDC uses the client's secret and the server's secret to construct the ticket which allows the client and server to mutually authenticate each other, while keeping the secrets hidden.

For the configuration of MongoDB for Kerberos support, two kinds of principal names are of interest: *user principals* (page 407) and *service principals* (page 407).

User Principal To authenticate using Kerberos, you must add the Kerberos user principals to MongoDB to the `$external` database. User principal names have the form:

```
<username>@<KERBEROS REALM>
```

For every user you want to authenticate using Kerberos, you must create a corresponding user in MongoDB in the `$external` database.

For examples of adding a user to MongoDB as well as authenticating as that user, see *Configure MongoDB with Kerberos Authentication on Linux* (page 409) and *Configure MongoDB with Kerberos Authentication on Windows* (page 412).

See also:

Manage Users and Roles (page 441) for general information regarding creating and managing users in MongoDB.

Service Principal Every MongoDB `mongod` and `mongos` instance (or `mongod.exe` or `mongos.exe` on Windows) must have an associated service principal. Service principal names have the form:

```
<service>/<fully qualified domain name>@<KERBEROS REALM>
```

For MongoDB, the `<service>` defaults to `mongodb`. For example, if `m1.example.com` is a MongoDB server, and `example.com` maintains the `EXAMPLE.COM` Kerberos realm, then `m1` should have the service principal name `mongodb/m1.example.com@EXAMPLE.COM`.

To specify a different value for `<service>`, use `serviceName` during the start up of `mongod` or `mongos` (or `mongod.exe` or `mongos.exe`). `mongo` shell or other clients may also specify a different service principal name using `serviceName`.

Service principal names must be reachable over the network using the fully qualified domain name (FQDN) part of its service principal name.

By default, Kerberos attempts to identify hosts using the `/etc/krb5.conf` file before using DNS to resolve hosts.

On Windows, if running MongoDB as a service, see *Assign Service Principal Name to MongoDB Windows Service* (page 414).

Linux Keytab Files Linux systems can store Kerberos authentication keys for a *service principal* (page 407) in *keytab* files. Each Kerberized `mongod` and `mongos` instance running on Linux must have access to a keytab file containing keys for its *service principal* (page 407).

To keep keytab files secure, use file permissions that restrict access to only the user that runs the `mongod` or `mongos` process.

Tickets On Linux, MongoDB clients can use Kerberos's `kinit` program to initialize a credential cache for authenticating the user principal to servers.

Windows Active Directory Unlike on Linux systems, `mongod` and `mongos` instances running on Windows do not require access to keytab files. Instead, the `mongod` and `mongos` instances read their server credentials from a credential store specific to the operating system.

However, from the Windows Active Directory, you can export a keytab file for use on Linux systems. See Ktpass[11] for more information.

Authenticate With Kerberos To configure MongoDB for Kerberos support and authenticate, see *Configure MongoDB with Kerberos Authentication on Linux* (page 409) and *Configure MongoDB with Kerberos Authentication on Windows* (page 412).

Operational Considerations

The HTTP Console The MongoDB HTTP Console[12] interface does not support Kerberos authentication.

Deprecated since version 3.2: HTTP interface for MongoDB

DNS Each host that runs a `mongod` or `mongos` instance must have both `A` and `PTR` DNS records to provide forward and reverse lookup.

Without `A` and `PTR` DNS records, the host cannot resolve the components of the Kerberos domain or the Key Distribution Center (KDC).

System Time Synchronization To successfully authenticate, the system time for each `mongod` and `mongos` instance must be within 5 minutes of the system time of the other hosts in the Kerberos infrastructure.

Kerberized MongoDB Environments

Driver Support The following MongoDB drivers support Kerberos authentication:

- C[13]
- C++[14]
- Java[15]
- C#[16]
- Node.js[17]
- PHP[18]
- Python[19]

[11] http://technet.microsoft.com/en-us/library/cc753771.aspx
[12] https://docs.mongodb.org/ecosystem/tools/http-interfaces/#http-console
[13] https://api.mongodb.org/c/current/authentication.html#kerberos
[14] https://docs.mongodb.org/ecosystem/tutorial/authenticate-with-cpp-driver/
[15] https://docs.mongodb.org/ecosystem/tutorial/authenticate-with-java-driver/
[16] http://mongodb.github.io/mongo-csharp-driver/2.0/reference/driver/authentication/#gssapi-kerberos
[17] http://mongodb.github.io/node-mongodb-native/2.0/tutorials/enterprise_features/
[18] http://php.net/manual/en/mongoclient.construct.php
[19] http://api.mongodb.org/python/current/examples/authentication.html

- Ruby[20]

Use with Additional MongoDB Authentication Mechanism Although MongoDB supports the use of Kerberos authentication with other authentication mechanisms, only add the other mechanisms as necessary. See the `Incorporate Additional Authentication Mechanisms` section in *Configure MongoDB with Kerberos Authentication on Linux* (page 409) and *Configure MongoDB with Kerberos Authentication on Windows* (page 412) for details.

Configure MongoDB with Kerberos Authentication on Linux

On this page

- Overview (page 409)
- Prerequisites (page 409)
- Procedure (page 409)
- Additional Considerations (page 411)
- Additional Resources (page 412)

New in version 2.4.

Overview MongoDB Enterprise supports authentication using a *Kerberos service* (page 406). Kerberos is an industry standard authentication protocol for large client/server system.

Prerequisites Setting up and configuring a Kerberos deployment is beyond the scope of this document. This tutorial assumes you have configured a *Kerberos service principal* (page 407) for each `mongod` and `mongos` instance in your MongoDB deployment, and you have a valid *keytab file* (page 407) for for each `mongod` and `mongos` instance.

To verify MongoDB Enterprise binaries:

```
mongod --version
```

In the output from this command, look for the string `modules: subscription` or `modules: enterprise` to confirm your system has MongoDB Enterprise.

Procedure The following procedure outlines the steps to add a Kerberos user principal to MongoDB, configure a standalone `mongod` instance for Kerberos support, and connect using the `mongo` shell and authenticate the user principal.

Step 1: Start `mongod` without Kerberos. For the initial addition of Kerberos users, start `mongod` without Kerberos support.

If a Kerberos user is already in MongoDB and has the *privileges required to create a user*, you can start `mongod` with Kerberos support.

Step 2: Connect to `mongod`. Connect via the `mongo` shell to the `mongod` instance. If `mongod` has `--auth` enabled, ensure you connect with the *privileges required to create a user*.

[20] https://docs.mongodb.org/ecosystem/tutorial/ruby-driver-tutorial/#gssapi-kerberos-mechanism

Step 3: Add Kerberos Principal(s) to MongoDB. Add a Kerberos principal, `<username>@<KERBEROS REALM>` or `<username>/<instance>@<KERBEROS REALM>`, to MongoDB in the `$external` database. Specify the Kerberos realm in all uppercase. The `$external` database allows `mongod` to consult an external source (e.g. Kerberos) to authenticate. To specify the user's privileges, assign *roles* (page 433) to the user.

The following example adds the Kerberos principal `application/reporting@EXAMPLE.NET` with read-only access to the `records` database:

```
use $external
db.createUser(
   {
     user: "application/reporting@EXAMPLE.NET",
     roles: [ { role: "read", db: "records" } ]
   }
)
```

Add additional principals as needed. For every user you want to authenticate using Kerberos, you must create a corresponding user in MongoDB. For more information about creating and managing users, see `https://docs.mongodb.org/manual/reference/command/nav-user-management`.

Step 4: Start `mongod` with Kerberos support. To start `mongod` with Kerberos support, set the environmental variable `KRB5_KTNAME` to the path of the keytab file and the `mongod` parameter `authenticationMechanisms` to `GSSAPI` in the following form:

```
env KRB5_KTNAME=<path to keytab file> \
mongod \
--setParameter authenticationMechanisms=GSSAPI
<additional mongod options>
```

For example, the following starts a standalone `mongod` instance with Kerberos support:

```
env KRB5_KTNAME=/opt/mongodb/mongod.keytab \
/opt/mongodb/bin/mongod --auth \
--setParameter authenticationMechanisms=GSSAPI \
--dbpath /opt/mongodb/data
```

The path to your `mongod` as well as your *keytab file* (page 407) may differ. Modify or include additional `mongod` options as required for your configuration. The *keytab file* (page 407) must be only accessible to the owner of the `mongod` process.

With the official `.deb` or `.rpm` packages, you can set the `KRB5_KTNAME` in a environment settings file. See *KRB5_KTNAME* (page 411) for details.

Step 5: Connect `mongo` shell to `mongod` and authenticate. Connect the `mongo` shell client as the Kerberos principal `application/reporting@EXAMPLE.NET`. Before connecting, you must have used Kerberos's `kinit` program to get credentials for `application/reporting@EXAMPLE.NET`.

You can connect and authenticate from the command line.

```
mongo --authenticationMechanism=GSSAPI --authenticationDatabase='$external' \
--username application/reporting@EXAMPLE.NET
```

Or, alternatively, you can first connect `mongo` to the `mongod`, and then from the `mongo` shell, use the `db.auth()` method to authenticate in the `$external` database.

```
use $external
db.auth( { mechanism: "GSSAPI", user: "application/reporting@EXAMPLE.NET" } )
```

Additional Considerations

KRB5_KTNAME If you installed MongoDB Enterprise using one of the official `.deb` or `.rpm` packages, and you use the included init/upstart scripts to control the `mongod` instance, you can set the `KR5_KTNAME` variable in the default environment settings file instead of setting the variable each time.

For `.rpm` packages, the default environment settings file is `/etc/sysconfig/mongod`.

For `.deb` packages, the file is `/etc/default/mongodb`.

Set the `KRB5_KTNAME` value in a line that resembles the following:

```
export KRB5_KTNAME="<path to keytab>"
```

Configure `mongos` for Kerberos To start `mongos` with Kerberos support, set the environmental variable `KRB5_KTNAME` to the path of its *keytab file* (page 407) and the `mongos` parameter `authenticationMechanisms` to `GSSAPI` in the following form:

```
env KRB5_KTNAME=<path to keytab file> \
mongos \
--setParameter authenticationMechanisms=GSSAPI \
<additional mongos options>
```

For example, the following starts a `mongos` instance with Kerberos support:

```
env KRB5_KTNAME=/opt/mongodb/mongos.keytab \
mongos \
--setParameter authenticationMechanisms=GSSAPI \
--configdb shard0.example.net, shard1.example.net,shard2.example.net \
--keyFile /opt/mongodb/mongos.keyfile
```

The path to your `mongos` as well as your *keytab file* (page 407) may differ. The *keytab file* (page 407) must be only accessible to the owner of the `mongos` process.

Modify or include any additional `mongos` options as required for your configuration. For example, instead of using *`--keyFile`* for internal authentication of sharded cluster members, you can use *x.509 member authentication* (page 430) instead.

Use a Config File To configure `mongod` or `mongos` for Kerberos support using a `configuration file`, specify the `authenticationMechanisms` setting in the configuration file:

If using the `YAML configuration file format`:

```
setParameter:
   authenticationMechanisms: GSSAPI
```

Or, if using the older `.ini` configuration file format:

```
setParameter=authenticationMechanisms=GSSAPI
```

Modify or include any additional `mongod` options as required for your configuration. For example, if `/opt/mongodb/mongod.conf` contains the following configuration settings for a standalone `mongod`:

```
security:
   authorization: enabled
setParameter:
   authenticationMechanisms: GSSAPI
storage:
   dbPath: /opt/mongodb/data
```

Or, if using the older configuration file format[21]:

```
auth = true
setParameter=authenticationMechanisms=GSSAPI
dbpath=/opt/mongodb/data
```

To start `mongod` with Kerberos support, use the following form:

```
env KRB5_KTNAME=/opt/mongodb/mongod.keytab \
/opt/mongodb/bin/mongod --config /opt/mongodb/mongod.conf
```

The path to your `mongod`, *keytab file* (page 407), and configuration file may differ. The *keytab file* (page 407) must be only accessible to the owner of the `mongod` process.

Troubleshoot Kerberos Setup for MongoDB If you encounter problems when starting `mongod` or `mongos` with Kerberos authentication, see *Troubleshoot Kerberos Authentication* (page 414).

Incorporate Additional Authentication Mechanisms Kerberos authentication (*GSSAPI* (page 406) (Kerberos)) can work alongside MongoDB's challenge/response authentication mechanisms (*SCRAM-SHA-1* (page 399) and *MONGODB-CR* (page 400)), MongoDB's authentication mechanism for LDAP (*PLAIN* (page 406) (LDAP SASL)), and MongoDB's authentication mechanism for x.509 (*MONGODB-X509* (page 401)). Specify the mechanisms as follows:

```
--setParameter authenticationMechanisms=GSSAPI,SCRAM-SHA-1
```

Only add the other mechanisms if in use. This parameter setting does not affect MongoDB's internal authentication of cluster members.

Additional Resources

- MongoDB LDAP and Kerberos Authentication with Dell (Quest) Authentication Services[22]
- MongoDB with Red Hat Enterprise Linux Identity Management and Kerberos[23]

Configure MongoDB with Kerberos Authentication on Windows

On this page

- Overview (page 412)
- Prerequisites (page 412)
- Procedures (page 413)
- Additional Considerations (page 414)

New in version 2.6.

Overview MongoDB Enterprise supports authentication using a *Kerberos service* (page 406). Kerberos is an industry standard authentication protocol for large client/server system. Kerberos allows MongoDB and applications to take advantage of existing authentication infrastructure and processes.

Prerequisites Setting up and configuring a Kerberos deployment is beyond the scope of this document. This tutorial assumes have configured a *Kerberos service principal* (page 407) for each `mongod.exe` and `mongos.exe` instance.

[21] https://docs.mongodb.org/v2.4/reference/configuration-options
[22] https://www.mongodb.com/blog/post/mongodb-ldap-and-kerberos-authentication-dell-quest-authentication-services?jmp=docs
[23] http://docs.mongodb.org/ecosystem/tutorial/manage-red-hat-enterprise-linux-identity-management?jmp=docs

Procedures

Step 1: Start `mongod.exe` without Kerberos. For the initial addition of Kerberos users, start `mongod.exe` without Kerberos support.

If a Kerberos user is already in MongoDB and has the *privileges required to create a user*, you can start `mongod.exe` with Kerberos support.

Step 2: Connect to `mongod`. Connect via the `mongo.exe` shell to the `mongod.exe` instance. If `mongod.exe` has `--auth` enabled, ensure you connect with the *privileges required to create a user*.

Step 3: Add Kerberos Principal(s) to MongoDB. Add a Kerberos principal, `<username>@<KERBEROS REALM>`, to MongoDB in the `$external` database. Specify the Kerberos realm in **ALL UPPERCASE**. The `$external` database allows `mongod.exe` to consult an external source (e.g. Kerberos) to authenticate. To specify the user's privileges, assign *roles* (page 433) to the user.

The following example adds the Kerberos principal `reportingapp@EXAMPLE.NET` with read-only access to the `records` database:

```
use $external
db.createUser(
   {
     user: "reportingapp@EXAMPLE.NET",
     roles: [ { role: "read", db: "records" } ]
   }
)
```

Add additional principals as needed. For every user you want to authenticate using Kerberos, you must create a corresponding user in MongoDB. For more information about creating and managing users, see `https://docs.mongodb.org/manual/reference/command/nav-user-management`.

Step 4: Start `mongod.exe` with Kerberos support. You must start `mongod.exe` as the *service principal account* (page 414).

To start `mongod.exe` with Kerberos support, set the `mongod.exe` parameter `authenticationMechanisms` to `GSSAPI`:

```
mongod.exe --setParameter authenticationMechanisms=GSSAPI <additional mongod.exe options>
```

For example, the following starts a standalone `mongod.exe` instance with Kerberos support:

```
mongod.exe --auth --setParameter authenticationMechanisms=GSSAPI
```

Modify or include additional `mongod.exe` options as required for your configuration.

Step 5: Connect `mongo.exe` shell to `mongod.exe` and authenticate. Connect the `mongo.exe` shell client as the Kerberos principal `application@EXAMPLE.NET`.

You can connect and authenticate from the command line.

```
mongo.exe --authenticationMechanism=GSSAPI --authenticationDatabase='$external' \
--username reportingapp@EXAMPLE.NET
```

Or, alternatively, you can first connect `mongo.exe` to the `mongod.exe`, and then from the `mongo.exe` shell, use the `db.auth()` method to authenticate in the `$external` database.

```
use $external
db.auth( { mechanism: "GSSAPI", user: "reportingapp@EXAMPLE.NET" } )
```

Additional Considerations

Configure `mongos.exe` for Kerberos To start `mongos.exe` with Kerberos support, set the `mongos.exe` parameter `authenticationMechanisms` to `GSSAPI`. You must start `mongos.exe` as the *service principal account* (page 414).:

```
mongos.exe --setParameter authenticationMechanisms=GSSAPI <additional mongos options>
```

For example, the following starts a `mongos` instance with Kerberos support:

```
mongos.exe --setParameter authenticationMechanisms=GSSAPI --configdb shard0.example.net, shard1.examp
```

Modify or include any additional `mongos.exe` options as required for your configuration. For example, instead of using *`--keyFile`* for internal authentication of sharded cluster members, you can use *x.509 member authentication* (page 430) instead.

Assign Service Principal Name to MongoDB Windows Service Use `setspn.exe` to assign the service principal name (SPN) to the account running the `mongod.exe` and the `mongos.exe` service:

```
setspn.exe -A <service>/<fully qualified domain name> <service account name>
```

For example, if `mongod.exe` runs as a service named `mongodb` on `testserver.mongodb.com` with the service account name `mongodtest`, assign the SPN as follows:

```
setspn.exe -A mongodb/testserver.mongodb.com mongodtest
```

Incorporate Additional Authentication Mechanisms Kerberos authentication (*GSSAPI* (page 406) (Kerberos)) can work alongside MongoDB's challenge/response authentication mechanisms (*SCRAM-SHA-1* (page 399) and *MONGODB-CR* (page 400)), MongoDB's authentication mechanism for LDAP (*PLAIN* (page 406) (LDAP SASL)), and MongoDB's authentication mechanism for x.509 (*MONGODB-X509* (page 401)). Specify the mechanisms as follows:

```
--setParameter authenticationMechanisms=GSSAPI,SCRAM-SHA-1
```

Only add the other mechanisms if in use. This parameter setting does not affect MongoDB's internal authentication of cluster members.

Troubleshoot Kerberos Authentication

On this page

- Kerberos Configuration Checklist (page 415)
- Debug with More Verbose Logs on Linux (page 415)
- Common Error Messages (page 415)

New in version 2.4.

Kerberos Configuration Checklist If you have difficulty starting `mongod` or `mongos` with *Kerberos* (page 406), ensure that:

- The `mongod` and the `mongos` binaries are from MongoDB Enterprise.

  To verify MongoDB Enterprise binaries:

  ```
  mongod --version
  ```

  In the output from this command, look for the string `modules: subscription` or `modules: enterprise` to confirm your system has MongoDB Enterprise.

- You are not using the HTTP Console[24]. MongoDB Enterprise does not support Kerberos authentication over the HTTP Console interface.
- On Linux, either the service principal name (SPN) in the *keytab file* (page 407) matches the SPN for the `mongod` or `mongos` instance, or the `mongod` or the `mongos` instance use the `--setParameter saslHostName=<host name>` to match the name in the keytab file.
- The canonical system hostname of the system that runs the `mongod` or `mongos` instance is a resolvable, fully qualified domain for this host. You can test the system hostname resolution with the `hostname -f` command at the system prompt.
- Each host that runs a `mongod` or `mongos` instance has both the `A` and `PTR` DNS records to provide forward and reverse lookup. The records allow the host to resolve the components of the Kerberos infrastructure.
- Both the Kerberos Key Distribution Center (KDC) and the system running `mongod` instance or `mongos` must be able to resolve each other using DNS. By default, Kerberos attempts to resolve hosts using the content of the `/etc/kerb5.conf` before using DNS to resolve hosts.
- The time synchronization of the systems running `mongod` or the `mongos` instances and the Kerberos infrastructure are within the maximum time skew (default is 5 minutes) of each other. Time differences greater than the maximum time skew will prevent successful authentication.

Debug with More Verbose Logs on Linux If you still encounter problems with Kerberos on Linux, you can start both `mongod` and `mongo` (or another client) with the environment variable `KRB5_TRACE` set to different files to produce more verbose logging of the Kerberos process to help further troubleshooting. For example, the following starts a standalone `mongod` with `KRB5_TRACE` set:

```
env KRB5_KTNAME=/opt/mongodb/mongod.keytab \
    KRB5_TRACE=/opt/mongodb/log/mongodb-kerberos.log \
    /opt/mongodb/bin/mongod --dbpath /opt/mongodb/data \
    --fork --logpath /opt/mongodb/log/mongod.log \
    --auth --setParameter authenticationMechanisms=GSSAPI
```

Common Error Messages In some situations, MongoDB will return error messages from the GSSAPI interface if there is a problem with the Kerberos service. Some common error messages are:

`GSSAPI error in client while negotiating security context.` This error occurs on the client and reflects insufficient credentials or a malicious attempt to authenticate.

If you receive this error, ensure that you are using the correct credentials and the correct fully qualified domain name when connecting to the host.

`GSSAPI error acquiring credentials.` This error occurs during the start of the `mongod` or `mongos` and reflects improper configuration of the system hostname or a missing or incorrectly configured keytab file.

[24] https://docs.mongodb.org/ecosystem/tools/http-interface/#http-console

If you encounter this problem, consider the items in the *Kerberos Configuration Checklist* (page 415), in particular, whether the SPN in the *keytab file* (page 407) matches the SPN for the `mongod` or `mongos` instance.

To determine whether the SPNs match:

1. Examine the keytab file, with the following command:

   ```
   klist -k <keytab>
   ```

   Replace `<keytab>` with the path to your keytab file.

2. Check the configured hostname for your system, with the following command:

   ```
   hostname -f
   ```

   Ensure that this name matches the name in the keytab file, or start `mongod` or `mongos` with the `--setParameter saslHostName=<hostname>`.

See also:

- *Kerberos Authentication* (page 406)
- *Configure MongoDB with Kerberos Authentication on Linux* (page 409)
- *Configure MongoDB with Kerberos Authentication on Windows* (page 412)

Additional Resources

- MongoDB LDAP and Kerberos Authentication with Dell (Quest) Authentication Services[25]
- MongoDB with Red Hat Enterprise Linux Identity Management and Kerberos[26]

LDAP Proxy Authority Authentication

On this page

- Considerations (page 416)
- MongoDB Configuration (page 417)
- LDAP User (page 417)
- Additional Information (page 417)

MongoDB Enterprise[27] supports proxy authentication through a Lightweight Directory Access Protocol (LDAP) service.

Considerations MongoDB Enterprise for Windows does **not** include LDAP support for authentication. However, MongoDB Enterprise for Linux supports using LDAP authentication with an ActiveDirectory server.

MongoDB does **not** support LDAP authentication in mixed sharded cluster deployments that contain both version 2.4 and version 2.6 shards. See *Upgrade MongoDB to 2.6* (page 1010) for upgrade instructions.

Use secure encrypted or trusted connections between clients and the server, as well as between `saslauthd` and the LDAP server. The LDAP server uses the `SASL PLAIN` mechanism, sending and receiving data in **plain text**. You should use only a trusted channel such as a VPN, a connection encrypted with TLS/SSL, or a trusted wired network.

[25] https://www.mongodb.com/blog/post/mongodb-ldap-and-kerberos-authentication-dell-quest-authentication-services?jmp=docs
[26] http://docs.mongodb.org/ecosystem/tutorial/manage-red-hat-enterprise-linux-identity-management?jmp=docs
[27] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

MongoDB Configuration To configure the MongoDB server to use LDAP authentication mechanism, use the following command line options:

- `--auth` to enable access control,
- `--authenticationMechanisms` set to `PLAIN`, and
- `--saslauthdPath` parameter set to the path to the Unix-domain Socket of the saslauthd instance.

Or, if using the `YAML configuration file`, use the following settings:

- `security.authorization` set to `enabled`,
- `setParameter.authenticationMechanisms` set to `PLAIN`, and
- `setParameter.saslauthdPath` set to the path to the Unix-domain Socket of the saslauthd instance.

LDAP User In order to authenticate a user with the LDAP authentication mechanism, add a corresponding *user* (page 394) to the `$external` database. You do not need to save the user's password in MongoDB.

The `$external` database is the *authentication database* (page 395) for the LDAP user. To authenticate the LDAP user, you must authenticate against the `$external` database. When authenticating, specify `PLAIN` for the authentication mechanism .

LDAP authentication requires that MongoDB forward the user's password in plan text. As such, you must specify `digestPassword` set to false during authentication.

Additional Information For information on configuring MongoDB to use LDAP and authenticating users using LDAP, see:

- *Authenticate Using SASL and LDAP with OpenLDAP* (page 420) and
- *Authenticate Using SASL and LDAP with ActiveDirectory* (page 417).

Authenticate Using SASL and LDAP with ActiveDirectory	**On this page** • Considerations (page 417) • Configure `saslauthd` (page 417) • Configure MongoDB (page 418)

MongoDB Enterprise provides support for proxy authentication of users. This allows administrators to configure a MongoDB cluster to authenticate users by proxying authentication requests to a specified Lightweight Directory Access Protocol (LDAP) service.

Considerations MongoDB Enterprise for Windows does **not** include LDAP support for authentication. However, MongoDB Enterprise for Linux supports using LDAP authentication with an ActiveDirectory server.

MongoDB does **not** support LDAP authentication in mixed sharded cluster deployments that contain both version 2.4 and version 2.6 shards. See *Upgrade MongoDB to 2.6* (page 1010) for upgrade instructions.

Use secure encrypted or trusted connections between clients and the server, as well as between `saslauthd` and the LDAP server. The LDAP server uses the `SASL PLAIN` mechanism, sending and receiving data in **plain text**. You should use only a trusted channel such as a VPN, a connection encrypted with TLS/SSL, or a trusted wired network.

Configure `saslauthd` LDAP support for user authentication requires proper configuration of the `saslauthd` daemon process as well as the MongoDB server.

Step 1: Specify the mechanism. On systems that configure `saslauthd` with the `/etc/sysconfig/saslauthd` file, such as Red Hat Enterprise Linux, Fedora, CentOS, and Amazon Linux AMI, set the mechanism `MECH` to `ldap`:

```
MECH=ldap
```

On systems that configure `saslauthd` with the `/etc/default/saslauthd` file, such as Ubuntu, set the `MECHANISMS` option to `ldap`:

```
MECHANISMS="ldap"
```

Step 2: Adjust caching behavior. On certain Linux distributions, `saslauthd` starts with the caching of authentication credentials *enabled*. Until restarted or until the cache expires, `saslauthd` will not contact the LDAP server to re-authenticate users in its authentication cache. This allows `saslauthd` to successfully authenticate users in its cache, even in the LDAP server is down or if the cached users' credentials are revoked.

To set the expiration time (in seconds) for the authentication cache, see the -t option[28] of `saslauthd`.

Step 3: Configure LDAP Options with ActiveDirectory. If the `saslauthd.conf` file does not exist, create it. The `saslauthd.conf` file usually resides in the `/etc` folder. If specifying a different file path, see the -O option[29] of `saslauthd`.

To use with ActiveDirectory, start `saslauthd` with the following configuration options set in the `saslauthd.conf` file:

```
ldap_servers: <ldap uri>
ldap_use_sasl: yes
ldap_mech: DIGEST-MD5
ldap_auth_method: fastbind
```

For the `<ldap uri>`, specify the uri of the ldap server. For example, `ldap_servers: ldaps://ad.example.net`.

For more information on `saslauthd` configuration, see http://www.openldap.org/doc/admin24/guide.html#Configuringsaslauthd.

Step 4: Test the `saslauthd` configuration. Use `testsaslauthd` utility to test the `saslauthd` configuration. For example:

```
testsaslauthd -u testuser -p testpassword -f /var/run/saslauthd/mux
```

Note: `/var/run/saslauthd` directory must have permissions set to `755` for MongoDB to successfully authenticate.

Configure MongoDB

Step 1: Add user to MongoDB for authentication. Add the user to the `$external` database in MongoDB. To specify the user's privileges, assign *roles* (page 433) to the user.

For example, the following adds a user with read-only access to the `records` database.

[28] http://www.linuxcommand.org/man_pages/saslauthd8.html
[29] http://www.linuxcommand.org/man_pages/saslauthd8.html

```
db.getSiblingDB("$external").createUser(
    {
      user : <username>,
      roles: [ { role: "read", db: "records" } ]
    }
)
```

Add additional principals as needed. For more information about creating and managing users, see https://docs.mongodb.org/manual/reference/command/nav-user-management.

Step 2: Configure MongoDB server. To configure the MongoDB server to use the saslauthd instance for proxy authentication, start the mongod with the following options:

- *--auth*,
- authenticationMechanisms parameter set to PLAIN, and
- saslauthdPath parameter set to the path to the Unix-domain Socket of the saslauthd instance.

Configure the MongoDB server using either the command line option --setParameter or the configuration file. Specify additional configurations as appropriate for your configuration.

If you use the authorization option to enforce authentication, you will need privileges to create a user.

Use specific saslauthd socket path. For socket path of /<some>/<path>/saslauthd, set the saslauthdPath to /<some>/<path>/saslauthd/mux, as in the following command line example:

```
mongod --auth --setParameter saslauthdPath=/<some>/<path>/saslauthd/mux --setParameter authenticatio
```

Or if using a YAML format configuration file, specify the following settings in the file:

```
security:
   authorization: enabled

setParameter:
   saslauthdPath: /<some>/<path>/saslauthd/mux
   authenticationMechanisms: PLAIN
```

Or, if using the older configuration file format[30]:

```
auth=true
setParameter=saslauthdPath=/<some>/<path>/saslauthd/mux
setParameter=authenticationMechanisms=PLAIN
```

Use default Unix-domain socket path. To use the default Unix-domain socket path, set the saslauthdPath to the empty string "", as in the following command line example:

```
mongod --auth --setParameter saslauthdPath="" --setParameter authenticationMechanisms=PLAIN
```

Or if using a YAML format configuration file, specify the following settings in the file:

```
security:
   authorization: enabled

setParameter:
   saslauthdPath: ""
   authenticationMechanisms: PLAIN
```

[30] https://docs.mongodb.org/v2.4/reference/configuration-options

Or, if using the older configuration file format[31]:

```
auth=true
setParameter=saslauthdPath=""
setParameter=authenticationMechanisms=PLAIN
```

Step 3: Authenticate the user in the `mongo` shell. To perform the authentication in the `mongo` shell, use the `db.auth()` method in the `$external` database.

Specify the value `"PLAIN"` in the `mechanism` field, the user and password in the `user` and `pwd` fields respectively, and the value `false` in the `digestPassword` field. You **must** specify `false` for `digestPassword` since the server must receive an undigested password to forward on to `saslauthd`, as in the following example:

```
db.getSiblingDB("$external").auth(
   {
     mechanism: "PLAIN",
     user: <username>,
     pwd:  <cleartext password>,
     digestPassword: false
   }
)
```

The server forwards the password in plain text. In general, use only on a trusted channel (VPN, TLS/SSL, trusted wired network). See Considerations.

Authenticate Using SASL and LDAP with OpenLDAP

On this page

- Considerations (page 420)
- Configure `saslauthd` (page 420)
- Configure MongoDB (page 422)

MongoDB Enterprise provides support for proxy authentication of users. This allows administrators to configure a MongoDB cluster to authenticate users by proxying authentication requests to a specified Lightweight Directory Access Protocol (LDAP) service.

Considerations MongoDB Enterprise for Windows does **not** include LDAP support for authentication. However, MongoDB Enterprise for Linux supports using LDAP authentication with an ActiveDirectory server.

MongoDB does **not** support LDAP authentication in mixed sharded cluster deployments that contain both version 2.4 and version 2.6 shards. See *Upgrade MongoDB to 2.6* (page 1010) for upgrade instructions.

Use secure encrypted or trusted connections between clients and the server, as well as between `saslauthd` and the LDAP server. The LDAP server uses the `SASL PLAIN` mechanism, sending and receiving data in **plain text**. You should use only a trusted channel such as a VPN, a connection encrypted with TLS/SSL, or a trusted wired network.

Configure `saslauthd` LDAP support for user authentication requires proper configuration of the `saslauthd` daemon process as well as the MongoDB server.

Step 1: Specify the mechanism. On systems that configure `saslauthd` with the `/etc/sysconfig/saslauthd` file, such as Red Hat Enterprise Linux, Fedora, CentOS, and Amazon Linux AMI, set the mechanism `MECH` to `ldap`:

[31] https://docs.mongodb.org/v2.4/reference/configuration-options

```
MECH=ldap
```

On systems that configure `saslauthd` with the `/etc/default/saslauthd` file, such as Ubuntu, set the `MECHANISMS` option to `ldap`:

```
MECHANISMS="ldap"
```

Step 2: Adjust caching behavior. On certain Linux distributions, `saslauthd` starts with the caching of authentication credentials *enabled*. Until restarted or until the cache expires, `saslauthd` will not contact the LDAP server to re-authenticate users in its authentication cache. This allows `saslauthd` to successfully authenticate users in its cache, even in the LDAP server is down or if the cached users' credentials are revoked.

To set the expiration time (in seconds) for the authentication cache, see the -t option[32] of `saslauthd`.

Step 3: Configure LDAP Options with OpenLDAP. If the `saslauthd.conf` file does not exist, create it. The `saslauthd.conf` file usually resides in the `/etc` folder. If specifying a different file path, see the -O option[33] of `saslauthd`.

To connect to an OpenLDAP server, update the `saslauthd.conf` file with the following configuration options:

```
ldap_servers: <ldap uri>
ldap_search_base: <search base>
ldap_filter: <filter>
```

The `ldap_servers` specifies the uri of the LDAP server used for authentication. In general, for OpenLDAP installed on the local machine, you can specify the value `ldap://localhost:389` or if using LDAP over TLS/SSL, you can specify the value `ldaps://localhost:636`.

The `ldap_search_base` specifies distinguished name to which the search is relative. The search includes the base or objects below.

The `ldap_filter` specifies the search filter.

The values for these configuration options should correspond to the values specific for your test. For example, to filter on email, specify `ldap_filter: (mail=%n)` instead.

OpenLDAP Example A sample `saslauthd.conf` file for OpenLDAP includes the following content:

```
ldap_servers: ldaps://ad.example.net
ldap_search_base: ou=Users,dc=example,dc=com
ldap_filter: (uid=%u)
```

To use this sample OpenLDAP configuration, create users with a `uid` attribute (login name) and place under the `Users` organizational unit (`ou`) under the domain components (`dc`) `example` and `com`.

For more information on `saslauthd` configuration, see http://www.openldap.org/doc/admin24/guide.html#Configuringsaslauthd.

Step 4: Test the `saslauthd` configuration. Use `testsaslauthd` utility to test the `saslauthd` configuration. For example:

```
testsaslauthd -u testuser -p testpassword -f /var/run/saslauthd/mux
```

[32] http://www.linuxcommand.org/man_pages/saslauthd8.html
[33] http://www.linuxcommand.org/man_pages/saslauthd8.html

Note: `/var/run/saslauthd` directory must have permissions set to `755` for MongoDB to successfully authenticate.

Configure MongoDB

Step 1: Add user to MongoDB for authentication. Add the user to the `$external` database in MongoDB. To specify the user's privileges, assign *roles* (page 433) to the user.

For example, the following adds a user with read-only access to the `records` database.

```
db.getSiblingDB("$external").createUser(
    {
      user : <username>,
      roles: [ { role: "read", db: "records" } ]
    }
)
```

Add additional principals as needed. For more information about creating and managing users, see `https://docs.mongodb.org/manual/reference/command/nav-user-management`.

Step 2: Configure MongoDB server. To configure the MongoDB server to use the `saslauthd` instance for proxy authentication, start the `mongod` with the following options:

- `--auth`,
- `authenticationMechanisms` parameter set to `PLAIN`, and
- `saslauthdPath` parameter set to the path to the Unix-domain Socket of the `saslauthd` instance.

Configure the MongoDB server using either the command line option `--setParameter` or the `configuration file`. Specify additional configurations as appropriate for your configuration.

If you use the `authorization` option to enforce authentication, you will need privileges to create a user.

Use specific `saslauthd` socket path. For socket path of `/<some>/<path>/saslauthd`, set the `saslauthdPath` to `/<some>/<path>/saslauthd/mux`, as in the following command line example:

```
mongod --auth --setParameter saslauthdPath=/<some>/<path>/saslauthd/mux --setParameter authentication
```

Or if using a `YAML format configuration file`, specify the following settings in the file:

```
security:
   authorization: enabled

setParameter:
   saslauthdPath: /<some>/<path>/saslauthd/mux
   authenticationMechanisms: PLAIN
```

Or, if using the older configuration file format[34]:

```
auth=true
setParameter=saslauthdPath=/<some>/<path>/saslauthd/mux
setParameter=authenticationMechanisms=PLAIN
```

[34] https://docs.mongodb.org/v2.4/reference/configuration-options

Use default Unix-domain socket path. To use the default Unix-domain socket path, set the `saslauthdPath` to the empty string `""`, as in the following command line example:

```
mongod --auth --setParameter saslauthdPath="" --setParameter authenticationMechanisms=PLAIN
```

Or if using a `YAML format configuration file`, specify the following settings in the file:

```
security:
   authorization: enabled

setParameter:
   saslauthdPath: ""
   authenticationMechanisms: PLAIN
```

Or, if using the older configuration file format[35]:

```
auth=true
setParameter=saslauthdPath=""
setParameter=authenticationMechanisms=PLAIN
```

Step 3: Authenticate the user in the `mongo` shell. To perform the authentication in the `mongo` shell, use the `db.auth()` method in the `$external` database.

Specify the value `"PLAIN"` in the `mechanism` field, the user and password in the `user` and `pwd` fields respectively, and the value `false` in the `digestPassword` field. You **must** specify `false` for `digestPassword` since the server must receive an undigested password to forward on to `saslauthd`, as in the following example:

```
db.getSiblingDB("$external").auth(
   {
     mechanism: "PLAIN",
     user: <username>,
     pwd:  <cleartext password>,
     digestPassword: false
   }
)
```

The server forwards the password in plain text. In general, use only on a trusted channel (VPN, TLS/SSL, trusted wired network). See Considerations.

Internal Authentication

On this page

- Keyfiles (page 424)
- x.509 (page 424)

You can authenticate members of *replica sets* and *sharded clusters*. For the internal authentication of the members, MongoDB can use either keyfiles or *x.509* (page 401) certificates.

Note: Enabling internal authentication also enables *client authorization* (page 433).

[35] https://docs.mongodb.org/v2.4/reference/configuration-options

Keyfiles

Keyfiles use *SCRAM-SHA-1* (page 399) challenge and response authentication mechanism. The contents of the keyfiles serve as the shared password for the members. A key's length must be between 6 and 1024 characters and may only contain characters in the base64 set.

MongoDB strips whitespace characters (e.g. `x0d`, `x09`, and `x20`) for cross-platform convenience. As a result, the following operations produce identical keys:

```
echo -e "my secret key" > key1
echo -e "my secret key\n" > key2
echo -e "my    secret    key" > key3
echo -e "my\r\nsecret\r\nkey\r\n" > key4
```

On UNIX systems, the keyfile must not have group or world permissions. On Windows systems, keyfile permissions are not checked

The content of the keyfile must be the same on all `mongod` and `mongos` instances that connect to each other. You must store the keyfile on each member of the replica set or sharded clusters.

To specify the keyfile, use the `security.keyFile` setting or `--keyFile` command line option.

For an example of keyfile internal authentication, see *Enable Internal Authentication* (page 425).

x.509

Members of a replica set or sharded cluster can use x.509 certificates for internal authentication instead of using keyfiles. MongoDB supports x.509 certificate authentication for use with a secure TLS/SSL connection.

Member Certificate Requirements The member certificate, used for internal authentication to verify membership to the sharded cluster or a replica set, must have the following properties:

- A single Certificate Authority (CA) must issue all the x.509 certificates for the members of a sharded cluster or a replica set.
- The Distinguished Name (`DN`), found in the member certificate's `subject`, must specify a non-empty value for *at least one* of the following attributes: Organization (`O`), the Organizational Unit (`OU`) or the Domain Component (`DC`).
- The Organization attributes (`O`'s), the Organizational Unit attributes (`OU`'s), and the Domain Components (`DC`'s) must match those from the certificates for the other cluster members. To match, the certificate must match all specifications of these attributes, or even the non-specification of these attributes. The order of the attributes does not matter.

  In the following example, the two `DN`'s contain matching specifications for `O`, `OU` as well as the non-specification of the `DC` attribute.

  ```
  CN=host1,OU=Dept1,O=MongoDB,ST=NY,C=US
  C=US, ST=CA, O=MongoDB, OU=Dept1, CN=host2
  ```

  However, the following two `DN`'s contain a mismatch for the `OU` attribute since one contains two `OU` specifications and the other, only one specification.

  ```
  CN=host1,OU=Dept1,OU=Sales,O=MongoDB
  CN=host2,OU=Dept1,O=MongoDB
  ```

- Either the Common Name (`CN`) or one of the Subject Alternative Name (`SAN`) entries must match the hostname of the server, used by the other members of the cluster.

For example, the certificates for a cluster could have the following subjects:

```
subject= CN=<myhostname1>,OU=Dept1,O=MongoDB,ST=NY,C=US
subject= CN=<myhostname2>,OU=Dept1,O=MongoDB,ST=NY,C=US
subject= CN=<myhostname3>,OU=Dept1,O=MongoDB,ST=NY,C=US
```

- *If* the certificate includes the Extended Key Usage (`extendedKeyUsage`) setting, the value must include `clientAuth` ("TLS Web Client Authentication").

  ```
  extendedKeyUsage = clientAuth
  ```

  You can also use a certificate that does not include the Extended Key Usage (EKU).

MongoDB Configuration To specify x.509 for internal authentication, in addition to the other SSL configurations appropriate for your deployment, for each member of the replica set or sharded cluster, include either:

- `security.clusterAuthMode` and `net.ssl.clusterFile` if using a `configuration file`, or
- `--clusterAuthMode` and `--sslClusterFile` command line options.

Member Certificate and `PEMKeyFile` To configure MongoDB for client certificate authentication, the `mongod` and `mongos` specify a `PEMKeyFile` to prove its identity to clients, either through `net.ssl.PEMKeyFile` setting in the configuration file or `--sslPEMKeyFile` command line option.

If no `clusterFile` certificate is specified for internal member authentication, MongoDB will attempt to use the `PEMKeyFile` certificate for member authentication. In order to use `PEMKeyFile` certificate for internal authentication as well as for client authentication, then the `PEMKeyFile` certificate must either:

- Omit `extendedKeyUsage` or
- Specify `extendedKeyUsage` values that include `clientAuth` in addition to `serverAuth`.

For an example of x.509 internal authentication, see *Use x.509 Certificate for Membership Authentication* (page 430).

To upgrade from keyfile internal authentication to x.509 internal authentication, see *Upgrade from Keyfile Authentication to x.509 Authentication* (page 432).

Enable Internal Authentication

On this page

- Overview (page 425)
- Considerations (page 426)
- Procedures (page 426)
- x.509 Internal Authentication (page 430)

Overview When authentication is enabled on a replica set or a sharded cluster, members of the replica set or the sharded clusters must provide credentials to authenticate.

To enable authentication on a replica set or a sharded cluster, you must enable authentication individually for each member. For a sharded cluster, this means enabling authentication on each `mongos` and each `mongod`, including the config servers and each member of a shard's replica set.

The following tutorial uses a *keyfile* (page 424) to enable internal authentication. You can also use x.509 certificate for internal authentication. For details on using x.509, see *Use x.509 Certificate for Membership Authentication* (page 430).

Considerations

Access Control Enabling internal authentication enables *access control* (page 433). The following tutorial assumes *no* users have been created in the system before enabling internal authentication, and uses *Localhost Exception* (page 396) to add a user administrator after access control has been enabled.

If you prefer, you can create the users before enabling internal authentication.

Sharded Cluster It is not possible to convert an existing sharded cluster that does not enforce access control to require authentication without taking all components of the cluster offline for a short period of time.

For sharded clusters, the *Localhost Exception* (page 396) will apply to the individual shards unless you either create an administrative user or disable the localhost exception on each shard.

Procedures

Update Existing Deployment

Step 1: Create a keyfile. Create the *keyfile* (page 424) your deployment will use to authenticate to members to each other. You can generate a keyfile using any method you choose. Ensure that the password stored in the keyfile is both long and contains a high amount of randomness.

For example, the following operation uses `openssl` command to generate pseudo-random data to use for a keyfile:

```
openssl rand -base64 741 > /srv/mongodb/mongodb-keyfile
chmod 600 mongodb-keyfile
```

Step 2: Enable authentication for each member of the sharded cluster or replica set. For *each* `mongod` in the replica set or for *each* `mongos` and `mongod` in the sharded cluster, including all config servers and shards, specify the keyfile using either a configuration file or a command line option.

In a configuration file, set the `security.keyFile` option to the keyfile's path and then start the component, as in the following example:

```
security:
  keyFile: /srv/mongodb/keyfile
```

Include any other settings as appropriate for your deployment.

Or, when starting the component, specify the `--keyFile` option. For example, for a `mongod`

```
mongod --keyFile /srv/mongodb/mongodb-keyfile --dbpath <path to data>
```

Include any other options as appropriate for your deployment.

Enabling internal authentication enables *access control* (page 433).

Step 3: Connect to the MongoDB instance via the localhost exception. To add the first user using *Localhost Exception* (page 396):

- For a replica set, connect a `mongo` shell to the primary. Run the `mongo` shell from the same host as the primary.
- For a sharded cluster, connect a `mongo` shell to the `mongos`. Run the `mongo` shell from same host as the `mongos`.

Step 4: Add first user. Add a user with the `userAdminAnyDatabase` (page 493) role. For example, the following creates the user `myUserAdmin` on the `admin` database:

```
use admin
db.createUser(
  {
    user: "myUserAdmin",
    pwd: "abc123",
    roles: [ { role: "userAdminAnyDatabase", db: "admin" } ]
  }
)
```

After you create the user administrator, for a replica set, the *localhost exception* (page 396) is no longer available.

For sharded clusters, you must still prevent unauthorized access to the individual shards. Follow one of the following steps for each shard in your cluster:

- Create an administrative user, or
- Disable the *Localhost Exception* (page 396) at startup. To disable the localhost exception, set the `enableLocalhostAuthBypass` to 0.

Step 5: Authenticate as the user administrator. Either connect a new `mongo` shell to the MongoDB instance with the `-u <username>`, `-p <password>`, and the `--authenticationDatabase <database>`:

```
mongo --port 27017 -u "myUserAdmin" -p "abc123" --authenticationDatabase "admin"
```

The `mongo` shell executes a number of commands at start up. As a result, when you log in as the user administrator, you may see authentication errors from one or more commands. You may ignore these errors, which are expected, because the `userAdminAnyDatabase` (page 493) role does not have permissions to run some of the start up commands.

Or, in the `mongo` shell connected without authentication, switch to the authentication database, and use `db.auth()` method to authenticate:

```
use admin
db.auth("myUserAdmin", "abc123" )
```

Step 6: Create additional users as needed for your deployment.

Deploy New Replica Set with Access Control

Step 1: Start one member of the replica set. This `mongod` should *not* enable `auth`.

Step 2: Create administrative users. The following operations will create two users: a user administrator that will be able to create and modify users (`myUserAdmin`), and a `root` (page 494) user (`siteRootAdmin`) that you will use to complete the remainder of the tutorial:

```
use admin
db.createUser( {
    user: "myUserAdmin",
    pwd: "<password>",
    roles: [ { role: "userAdminAnyDatabase", db: "admin" } ]
  });
db.createUser( {
```

```
    user: "siteRootAdmin",
    pwd: "<password>",
    roles: [ { role: "root", db: "admin" } ]
  });
```

Step 3: Stop the `mongod` instance.

Step 4: Create the key file to be used by each member of the replica set. Create the key file your deployment will use to authenticate servers to each other.

To generate pseudo-random data to use for a `keyfile`, issue the following `openssl` command:

```
openssl rand -base64 741 > mongodb-keyfile
chmod 600 mongodb-keyfile
```

You may generate a key file using any method you choose. Always ensure that the password stored in the key file is both long and contains a high amount of entropy. Using `openssl` in this manner helps generate such a key.

Step 5: Copy the key file to each member of the replica set. Copy the `mongodb-keyfile` to all hosts where components of a MongoDB deployment run. Set the permissions of these files to `600` so that only the *owner* of the file can read or write this file to prevent other users on the system from accessing the shared secret.

Step 6: Start each member of the replica set with the appropriate options. For each member, start a `mongod` and specify the key file and the name of the replica set. Also specify other parameters as needed for your deployment. For replication-specific parameters, see *cli-mongod-replica-set* required by your deployment.

If your application connects to more than one replica set, each set should have a distinct name. Some drivers group replica set connections by replica set name.

The following example specifies parameters through the *`--keyFile`* and *`--replSet`* command-line options:

```
mongod --keyFile /mysecretdirectory/mongodb-keyfile --replSet "rs0"
```

The following example specifies parameters through a configuration file:

```
mongod --config $HOME/.mongodb/config
```

In production deployments, you can configure a *init script* to manage this process. Init scripts are beyond the scope of this document.

Step 7: Connect to the member of the replica set where you created the administrative users. Connect to the replica set member you started and authenticate as the `siteRootAdmin` user. From the `mongo` shell, use the following operation to authenticate:

```
use admin
db.auth("siteRootAdmin", "<password>");
```

Step 8: Initiate the replica set. Use `rs.initiate()` on *one and only one* member of the replica set:

```
rs.initiate()
```

MongoDB initiates a set that consists of the current member and that uses the default replica set configuration.

Step 9: Verify the initial replica set configuration. Use `rs.conf()` to display the *replica set configuration object* (page 717):

```
rs.conf()
```

The replica set configuration object resembles the following:

```
{
   "_id" : "rs0",
   "version" : 1,
   "members" : [
      {
         "_id" : 1,
         "host" : "mongodb0.example.net:27017"
      }
   ]
}
```

Step 10: Add the remaining members to the replica set. Add the remaining members with the `rs.add()` method. You must be connected to the *primary* to add members to a replica set.

`rs.add()` can, in some cases, trigger an election. If the *mongod* you are connected to becomes a *secondary*, you need to connect the *mongo* shell to the new primary to continue adding new replica set members. Use `rs.status()` to identify the primary in the replica set.

The following example adds two members:

```
rs.add("mongodb1.example.net")
rs.add("mongodb2.example.net")
```

When complete, you have a fully functional replica set. The new replica set will elect a *primary*.

Step 11: Check the status of the replica set. Use the `rs.status()` operation:

```
rs.status()
```

Step 12: Create additional users to address operational requirements. You can use *built-in roles* (page 485) to create common types of database users, such as the `dbOwner` (page 488) role to create a database administrator, the `readWrite` (page 486) role to create a user who can update data, or the `read` (page 486) role to create user who can search data but no more. You also can define *custom roles* (page 440).

For example, the following creates a database administrator for the `products` database:

```
use products
db.createUser(
  {
    user: "productsDBAdmin",
    pwd: "password",
    roles:
    [
      {
        role: "dbOwner",
        db: "products"
      }
    ]
  }
)
```

For an overview of roles and privileges, see *Role-Based Access Control* (page 433). For more information on adding users, see *Manage Users and Roles* (page 441).

x.509 Internal Authentication For details on using x.509 for internal authentication, see *Use x.509 Certificate for Membership Authentication* (page 430).

To upgrade from keyfile internal authentication to x.509 internal authentication, see *Upgrade from Keyfile Authentication to x.509 Authentication* (page 432).

Use x.509 Certificate for Membership Authentication

On this page

- Member x.509 Certificate (page 430)
- Configure Replica Set/Sharded Cluster (page 431)
- Additional Information (page 432)

New in version 2.6.

MongoDB supports x.509 certificate authentication for use with a secure *TLS/SSL connection* (page 451). Sharded cluster members and replica set members can use x.509 certificates to verify their membership to the cluster or the replica set instead of using *keyfiles* (page 424). The membership authentication is an internal process.

For client authentication with x.509, see *Use x.509 Certificates to Authenticate Clients* (page 403).

Important: A full description of TLS/SSL, PKI (Public Key Infrastructure) certificates, in particular x.509 certificates, and Certificate Authority is beyond the scope of this document. This tutorial assumes prior knowledge of TLS/SSL as well as access to valid x.509 certificates.

Member x.509 Certificate

Certificate Requirements The member certificate, used for internal authentication to verify membership to the sharded cluster or a replica set, must have the following properties:

- A single Certificate Authority (CA) must issue all the x.509 certificates for the members of a sharded cluster or a replica set.
- The Distinguished Name (`DN`), found in the member certificate's `subject`, must specify a non-empty value for *at least one* of the following attributes: Organization (`O`), the Organizational Unit (`OU`) or the Domain Component (`DC`).
- The Organization attributes (`O`'s), the Organizational Unit attributes (`OU`'s), and the Domain Components (`DC`'s) must match those from the certificates for the other cluster members. To match, the certificate must match all specifications of these attributes, or even the non-specification of these attributes. The order of the attributes does not matter.

  In the following example, the two `DN`'s contain matching specifications for `O`, `OU` as well as the non-specification of the `DC` attribute.

  ```
  CN=host1,OU=Dept1,O=MongoDB,ST=NY,C=US
  C=US, ST=CA, O=MongoDB, OU=Dept1, CN=host2
  ```

  However, the following two `DN`'s contain a mismatch for the `OU` attribute since one contains two `OU` specifications and the other, only one specification.

```
CN=host1,OU=Dept1,OU=Sales,O=MongoDB
CN=host2,OU=Dept1,O=MongoDB
```

- Either the Common Name (`CN`) or one of the Subject Alternative Name (`SAN`) entries must match the hostname of the server, used by the other members of the cluster.

  For example, the certificates for a cluster could have the following subjects:

  ```
  subject= CN=<myhostname1>,OU=Dept1,O=MongoDB,ST=NY,C=US
  subject= CN=<myhostname2>,OU=Dept1,O=MongoDB,ST=NY,C=US
  subject= CN=<myhostname3>,OU=Dept1,O=MongoDB,ST=NY,C=US
  ```

- *If* the certificate includes the Extended Key Usage (`extendedKeyUsage`) setting, the value must include `clientAuth` ("TLS Web Client Authentication").

  ```
  extendedKeyUsage = clientAuth
  ```

  You can also use a certificate that does not include the Extended Key Usage (EKU).

Member Certificate and `PEMKeyFile` To configure MongoDB for client certificate authentication, the `mongod` and `mongos` specify a `PEMKeyFile` to prove its identity to clients, either through `net.ssl.PEMKeyFile` setting in the configuration file or `--sslPEMKeyFile` command line option.

If no `clusterFile` certificate is specified for internal member authentication, MongoDB will attempt to use the `PEMKeyFile` certificate for member authentication. In order to use `PEMKeyFile` certificate for internal authentication as well as for client authentication, then the `PEMKeyFile` certificate must either:

- Omit `extendedKeyUsage` or
- Specify `extendedKeyUsage` values that include `clientAuth` in addition to `serverAuth`.

Configure Replica Set/Sharded Cluster

Use Command-line Options To specify the x.509 certificate for internal cluster member authentication, append the additional TLS/SSL options *`--clusterAuthMode`* and *`--sslClusterFile`*, as in the following example for a member of a replica set:

```
mongod --replSet <name> --sslMode requireSSL --clusterAuthMode x509 --sslClusterFile <path to members
```

Include any additional options, TLS/SSL or otherwise, that are required for your specific configuration. For instance, if the membership key is encrypted, set the *`--sslClusterPassword`* to the passphrase to decrypt the key or have MongoDB prompt for the passphrase. See *SSL Certificate Passphrase* (page 455) for details.

> **Warning:** If the *`--sslCAFile`* option and its target file are not specified, x.509 client and member authentication will not function. `mongod`, and `mongos` in sharded systems, will not be able to verify the certificates of processes connecting to it against the trusted certificate authority (CA) that issued them, breaking the certificate chain.
> As of version 2.6.4, `mongod` will not start with x.509 authentication enabled if the CA file is not specified.

Use Configuration File You can specify the configuration for MongoDB in a `YAML formatted` configuration file, as in the following example:

```
security:
   clusterAuthMode: x509
net:
   ssl:
      mode: requireSSL
      PEMKeyFile: <path to TLS/SSL certificate and key PEM file>
      CAFile: <path to root CA PEM file>
      clusterFile: <path to x.509 membership certificate and key PEM file>
```

See `security.clusterAuthMode`, `net.ssl.mode`, `net.ssl.PEMKeyFile`, `net.ssl.CAFile`, and `net.ssl.clusterFile` for more information on the settings.

Additional Information To upgrade from keyfile internal authentication to x.509 internal authentication, see *Upgrade from Keyfile Authentication to x.509 Authentication* (page 432).

Upgrade from Keyfile Authentication to x.509 Authentication

On this page

- Clusters Currently Using TLS/SSL (page 432)
- Clusters Currently Not Using TLS/SSL (page 433)

To upgrade clusters that are currently using *keyfile authentication* (page 424) to x.509 authentication, use the following rolling upgrade processes.

Clusters Currently Using TLS/SSL For clusters using TLS/SSL and keyfile authentication, to upgrade to x.509 cluster authentication, use the following rolling upgrade process:

1. For each node of a cluster, start the node with the option `--clusterAuthMode` set to `sendKeyFile` and the option `--sslClusterFile` set to the appropriate path of the node's certificate. Include other *TLS/SSL options* (page 451) as well as any other options that are required for your specific configuration. For example:

   ```
   mongod --replSet <name> --sslMode requireSSL --clusterAuthMode sendKeyFile --sslClusterFile <pat
   ```

   With this setting, each node continues to use its keyfile to authenticate itself as a member. However, each node can now accept either a keyfile or an x.509 certificate from other members to authenticate those members. Upgrade all nodes of the cluster to this setting.

2. Then, for each node of a cluster, connect to the node and use the `setParameter` command to update the `clusterAuthMode` to `sendX509`. [36] For example,

   ```
   db.getSiblingDB('admin').runCommand( { setParameter: 1, clusterAuthMode: "sendX509" } )
   ```

   With this setting, each node uses its x.509 certificate, specified with the `--sslClusterFile` option in the previous step, to authenticate itself as a member. However, each node continues to accept either a keyfile or an x.509 certificate from other members to authenticate those members. Upgrade all nodes of the cluster to this setting.

3. Optional but recommended. Finally, for each node of the cluster, connect to the node and use the `setParameter` command to update the `clusterAuthMode` to `x509` to only use the x.509 certificate for authentication. [1] For example:

   ```
   db.getSiblingDB('admin').runCommand( { setParameter: 1, clusterAuthMode: "x509" } )
   ```

[36] As an alternative to using the `setParameter` command, you can also restart the nodes with the appropriate TLS/SSL and x509 options and values.

4. After the upgrade of all nodes, edit the `configuration file` with the appropriate x.509 settings to ensure that upon subsequent restarts, the cluster uses x.509 authentication.

See *`--clusterAuthMode`* for the various modes and their descriptions.

Clusters Currently Not Using TLS/SSL For clusters using keyfile authentication but not TLS/SSL, to upgrade to x.509 authentication, use the following rolling upgrade process:

1. For each node of a cluster, start the node with the option *`--sslMode`* set to `allowSSL`, the option *`--clusterAuthMode`* set to `sendKeyFile` and the option *`--sslClusterFile`* set to the appropriate path of the node's certificate. Include other *TLS/SSL options* (page 451) as well as any other options that are required for your specific configuration. For example:

   ```
   mongod --replSet <name> --sslMode allowSSL --clusterAuthMode sendKeyFile --sslClusterFile <path
   ```

   The *`--sslMode allowSSL`* setting allows the node to accept both TLS/SSL and non-TLS/non-SSL incoming connections. Its outgoing connections do not use TLS/SSL.

   The *`--clusterAuthMode sendKeyFile`* setting allows each node continues to use its keyfile to authenticate itself as a member. However, each node can now accept either a keyfile or an x.509 certificate from other members to authenticate those members.

   Upgrade all nodes of the cluster to these settings.

2. Then, for each node of a cluster, connect to the node and use the `setParameter` command to update the `sslMode` to `preferSSL` and the `clusterAuthMode` to `sendX509`. [1] For example:

   ```
   db.getSiblingDB('admin').runCommand( { setParameter: 1, sslMode: "preferSSL", clusterAuthMode: "
   ```

   With the `sslMode` set to `preferSSL`, the node accepts both TLS/SSL and non-TLS/non-SSL incoming connections, and its outgoing connections use TLS/SSL.

   With the `clusterAuthMode` set to `sendX509`, each node uses its x.509 certificate, specified with the *`--sslClusterFile`* option in the previous step, to authenticate itself as a member. However, each node continues to accept either a keyfile or an x.509 certificate from other members to authenticate those members.

   Upgrade all nodes of the cluster to these settings.

3. Optional but recommended. Finally, for each node of the cluster, connect to the node and use the `setParameter` command to update the `sslMode` to `requireSSL` and the `clusterAuthMode` to `x509`. [1] For example:

   ```
   db.getSiblingDB('admin').runCommand( { setParameter: 1, sslMode: "requireSSL", clusterAuthMode:
   ```

   With the `sslMode` set to `requireSSL`, the node only uses TLS/SSLs connections.

   With the `clusterAuthMode` set to `x509`, the node only uses the x.509 certificate for authentication.

4. After the upgrade of all nodes, edit the `configuration file` with the appropriate TLS/SSL and x.509 settings to ensure that upon subsequent restarts, the cluster uses x.509 authentication.

See *`--clusterAuthMode`* for the various modes and their descriptions.

9.3 Role-Based Access Control

On this page

- Enable Access Control (page 434)
- Roles (page 434)
- Users and Roles (page 435)
- Built-In Roles and User-Defined Roles (page 435)

MongoDB employs Role-Based Access Control (RBAC) to govern access to a MongoDB system. A user is granted one or more *roles* (page 434) that determine the user's access to database resources and operations. Outside of role assignments, the user has no access to the system.

9.3.1 Enable Access Control

MongoDB does not enable access control by default. You can enable authorization using the `--auth` or the `security.authorization` setting. Enabling *internal authentication* (page 423) also enables client authorization.

Once access control is enabled, users must *authenticate* (page 393) themselves.

9.3.2 Roles

A role grants privileges to perform the specified *actions* (page 500) on *resource* (page 498). Each privilege is either specified explicitly in the role or inherited from another role or both.

Privileges

A privilege consists of a specified resource and the actions permitted on the resource.

A *resource* (page 498) is either a database, collection, set of collections, or the cluster. If the resource is the cluster, the affiliated actions affect the state of the system rather than a specific database or collection. For information on the resource documents, see *Resource Document* (page 498).

An *action* (page 500) specifies the operation allowed on the resource. For available actions see *Privilege Actions* (page 500).

Inherited Privileges

A role can include one or more existing roles in its definition, in which case the role inherits all the privileges of the included roles.

A role can inherit privileges from other roles in its database. A role created on the `admin` database can inherit privileges from roles in any database.

View Role's Privileges

You can view the privileges for a role by issuing the `rolesInfo` command with the `showPrivileges` and `showBuiltinRoles` fields both set to `true`.

9.3.3 Users and Roles

You can assign roles to users during the user creation. You can also update existing users to grant or revoke roles. For a full list of user management methods, see *user-management-methods*

A user assigned a role receives all the privileges of that role. A user can have multiple roles. By assigning to the user roles in various databases, a user created in one database can have permissions to act on other databases.

Note: The first user created in the database should be a user administrator who has the privileges to manage other users. See *Enable Client Access Control* (page 435).

9.3.4 Built-In Roles and User-Defined Roles

MongoDB provides *built-in roles* (page 438) that provide set of privileges commonly needed in a database system.

If these built-in-roles cannot provide the desired set of privileges, MongoDB provides methods to create and modify *user-defined roles* (page 440).

Enable Client Access Control

On this page

- Overview (page 435)
- Considerations (page 435)
- Procedures (page 436)
- Additional Information (page 438)

Overview

Enabling access control requires authentication of every user. Once authenticated, users only have the privileges as defined in the roles granted to the users.

To enable access control, use either the command line option `--auth` or `security.authorization` configuration file setting.

Note: The tutorial enables access control and uses the *default authentication mechanism* (page 399). To specify a different authentication mechanism, see *Authentication Mechanisms* (page 398).

You can also enable client access control by *enabling internal authentication* (page 425) of replica sets or sharded clusters. For instructions on enabling internal authentication, see *Enable Internal Authentication* (page 425).

Considerations

With access control enabled, ensure you have a user with `userAdmin` (page 488) or `userAdminAnyDatabase` (page 493) role in the `admin` database.

This tutorial assumes a *standalone* environment.

The *Enable Internal Authentication* (page 425) tutorial has steps specific to enabling access control on replica sets and sharded clusters.

You can create users before enabling access control or you can create users after enabling access control. If you enable access control before creating any user, MongoDB provides a *localhost exception* (page 396) which allows you to create a user administrator in the `admin` database. Once created, authenticate as the user administrator to create additional users as needed.

Procedures

Add Users Before Enabling Access Control The following procedure first adds a user administrator to a MongoDB instance running without access control and then enables access control.

Step 1: Start MongoDB without access control. For example, the following starts a standalone `mongod` instance without access control.

```
mongod --port 27017 --dbpath /data/db1
```

For details on starting a `mongod` or `mongos`, see *Manage mongod Processes* (page 323) or *Deploy a Sharded Cluster* (page 765).

Step 2: Connect to the instance. For example, connect a `mongo` shell to the instance.

```
mongo --port 27017
```

Specify additional command line options as appropriate to connect the `mongo` shell to your deployment, such as `--host`.

Step 3: Create the user administrator. Add a user with the `userAdminAnyDatabase` (page 493) role. For example, the following creates the user `myUserAdmin` on the `admin` database:

```
use admin
db.createUser(
  {
    user: "myUserAdmin",
    pwd: "abc123",
    roles: [ { role: "userAdminAnyDatabase", db: "admin" } ]
  }
)
```

Step 4: Re-start the MongoDB instance with access control. Re-start the `mongod` instance with the `--auth` command line option or, if using a configuration file, the `security.authorization` setting.

```
mongod --auth --port 27017 --dbpath /data/db1
```

Step 5: Authenticate as the user administrator. Either connect a new `mongo` shell to the MongoDB instance with the `-u <username>`, `-p <password>`, and the `--authenticationDatabase <database>`:

```
mongo --port 27017 -u "myUserAdmin" -p "abc123" --authenticationDatabase "admin"
```

The `mongo` shell executes a number of commands at start up. As a result, when you log in as the user administrator, you may see authentication errors from one or more commands. You may ignore these errors, which are expected, because the `userAdminAnyDatabase` (page 493) role does not have permissions to run some of the start up commands.

Or, in the mongo shell connected without authentication, switch to the authentication database, and use db.auth() method to authenticate:

```
use admin
db.auth("myUserAdmin", "abc123" )
```

Step 5: Create additional users as needed for your deployment. If you need to disable access control for any reason, restart the MongoDB instance without the --auth command line option, or if using a configuration file, the security.authorization setting.

Add Users After Enabling Access Control The following procedure first enables access control, and then uses *localhost exception* (page 396) to add a user administrator.

Step 1: Start the MongoDB instance with access control. Start the mongod instance with the --auth command line option or, if using a configuration file, the security.authorization setting.

```
mongod --auth --port 27017 --dbpath /data/db1
```

Step 2: Connect to the MongoDB instance via the localhost exception. To add the first user using *Localhost Exception* (page 396), connect a mongo shell to the mongod instance. Run the mongo shell from the same host as the mongod instance.

Step 3: Create the system user administrator. Add the user with the userAdminAnyDatabase (page 493) role, and only that role.

The following example creates the user myUserAdmin user on the admin database:

```
use admin
db.createUser(
  {
    user: "myUserAdmin",
    pwd: "abc123",
    roles: [ { role: "userAdminAnyDatabase", db: "admin" } ]
  }
)
```

After you create the user administrator, the *localhost exception* (page 396) is no longer available.

Step 4: Authenticate as the user administrator. Either connect a new mongo shell to the MongoDB instance with the -u <username>, -p <password>, and the --authenticationDatabase <database>:

```
mongo --port 27017 -u "myUserAdmin" -p "abc123" --authenticationDatabase "admin"
```

The mongo shell executes a number of commands at start up. As a result, when you log in as the user administrator, you may see authentication errors from one or more commands. You may ignore these errors, which are expected, because the userAdminAnyDatabase (page 493) role does not have permissions to run some of the start up commands.

Or, in the mongo shell connected without authentication, switch to the authentication database, and use db.auth() method to authenticate:

```
use admin
db.auth("myUserAdmin", "abc123" )
```

Step 5: Create additional users as needed for your deployment.

Additional Information

See also *Manage Users and Roles* (page 441).

Built-In Roles

On this page

- Database User Roles (page 438)
- Database Administration Roles (page 438)
- Cluster Administration Roles (page 439)
- Backup and Restoration Roles (page 439)
- All-Database Roles (page 439)
- Superuser Roles (page 440)
- Internal Role (page 440)

MongoDB provides built-in roles that provide the different levels of access commonly needed in a database system. Built-in *database user roles* (page 486) and *database administration roles* (page 487) roles exist in *each* database. The `admin` database contains additional roles.

This page provides a brief description of the built-in roles. For the specific privileges granted by each role, see the *Built-In Roles* (page 485) reference page.

Database User Roles

Every database includes the following roles:

Role	Short Description
`read` (page 486)	Provides the ability to read data on all *non*-system collections and on the following system collections: `system.indexes` (page 377), `system.js` (page 377), and `system.namespaces` (page 377) collections. For the specific privileges granted by the role, see `read` (page 486).
`readWrite` (page 486)	Provides all the privileges of the `read` (page 486) role and the ability to modify data on all *non*-system collections and the `system.js` (page 377) collection. For the specific privileges granted by the role, see `readWrite` (page 486).

Database Administration Roles

Every database includes the following database administration roles:

Role	Short Description
dbAdmin (page 487)	Provides the ability to perform administrative tasks such as schema-related tasks, indexing, gathering statistics. This role does not grant privileges for user and role management. For the specific privileges granted by the role, see dbAdmin (page 487).
dbOwner (page 488)	Provides the ability to perform any administrative action on the database. This role combines the privileges granted by the readWrite (page 486), dbAdmin (page 487) and userAdmin (page 488) roles.
userAdmin (page 488)	Provides the ability to create and modify roles and users on the current database. Since the userAdmin (page 488) role allows users to grant any privilege to any user, including themselves, the role also indirectly provides *superuser* (page 493) access to either the database or, if scoped to the admin database, the cluster. For the specific privileges granted by the role, see userAdmin (page 488).

Cluster Administration Roles

The admin database includes the following roles for administering the whole system rather than a specific database. These roles include but are not limited to *replica set* and *sharded cluster* administrative functions.

Role	Short Description
clusterAdmin (page 488)	Provides the greatest cluster-management access. This role combines the privileges granted by the clusterManager (page 489), clusterMonitor (page 490), and hostManager (page 490) roles. Additionally, the role provides the dropDatabase (page 504) action.
clusterManager (page 489)	Provides management and monitoring actions on the cluster. A user with this role can access the config and local databases, which are used in sharding and replication, respectively. For the specific privileges granted by the role, see clusterManager (page 489).
clusterMonitor (page 490)	Provides read-only access to monitoring tools, such as the MongoDB Cloud Manager[37] and Ops Manager[38] monitoring agent. For the specific privileges granted by the role, see clusterMonitor (page 490).
hostManager (page 490)	Provides the ability to monitor and manage servers. For the specific privileges granted by the role, see hostManager (page 490).

Backup and Restoration Roles

The admin database includes the following roles for backing up and restoring data:

Role	Short Description
backup (page 491)	Provides privileges needed to back up data. This role provides sufficient privileges to use the MongoDB Cloud Manager[39] backup agent, Ops Manager[40] backup agent, or to use mongodump. For the specific privileges granted by the role, see backup (page 491).
restore (page 492)	Provides privileges needed to restore data with mongorestore without the *--oplogReplay* option or without system.profile collection data. For the specific privileges granted by the role, see restore (page 492).

All-Database Roles

The admin database provides the following roles that apply to all databases in a mongod instance and are roughly equivalent to their single-database equivalents:

[37] https://cloud.mongodb.com/?jmp=docs
[38] https://docs.opsmanager.mongodb.com/current/
[39] https://cloud.mongodb.com/?jmp=docs
[40] https://docs.opsmanager.mongodb.com/current/

Role	Short Description
readAnyDatabase (page 493)	Provides the same read-only permissions as read (page 486), except it applies to *all* databases in the cluster. The role also provides the listDatabases (page 505) action on the cluster as a whole. For the specific privileges granted by the role, see readAnyDatabase (page 493).
readWriteAnyDatabase (page 493)	Provides the same read and write permissions as readWrite (page 486), except it applies to *all* databases in the cluster. The role also provides the listDatabases (page 505) action on the cluster as a whole. For the specific privileges granted by the role, see readWriteAnyDatabase (page 493).
userAdminAnyDatabase (page 493)	Provides the same access to user administration operations as userAdmin (page 488), except it applies to *all* databases in the cluster. Since the userAdminAnyDatabase (page 493) role allows users to grant any privilege to any user, including themselves, the role also indirectly provides *superuser* (page 493) access. For the specific privileges granted by the role, see userAdminAnyDatabase (page 493).
dbAdminAnyDatabase (page 493)	Provides the same access to database administration operations as dbAdmin (page 487), except it applies to *all* databases in the cluster. The role also provides the listDatabases (page 505) action on the cluster as a whole. For the specific privileges granted by the role, see dbAdminAnyDatabase (page 493).

Superuser Roles

The following role provides full privileges on all resources:

Role	Short Description
root (page 494)	Provides access to the operations and all the resources of the readWriteAnyDatabase (page 493), dbAdminAnyDatabase (page 493), userAdminAnyDatabase (page 493) and clusterAdmin (page 488) roles *combined*. For the specific privileges granted by the role, see root (page 494).

Internal Role

Role	Short Description
__system (page 494)	Provides privileges to take any action against any object in the database. **Do not** assign this role to user objects representing applications or human administrators, other than in exceptional circumstances. For more information, see root (page 494).

See also:

Built-In Roles (page 485)

User-Defined Roles

On this page

- Role Management Interface (page 441)
- Scope (page 441)
- Centralized Role Data (page 441)

New in version 2.6.

MongoDB provides a number of *built-in roles* (page 485). However, if these roles cannot describe the desired set of privileges, you can create new roles.

Role Management Interface

To add a role, MongoDB provides the `db.createRole()` method. MongoDB also provides methods to update existing user-defined roles. For a full list of role management methods, see *role-management-methods*.

Scope

When adding a role, you create the role in a specific database. MongoDB uses the combination of the database and the role name to uniquely define a role.

Except for roles created in the `admin` database, a role can only include privileges that apply to its database and can only inherit from other roles in its database.

A role created in the `admin` database can include privileges that apply to the `admin` database, other databases or to the *cluster* (page 500) resource, and can inherit from roles in other databases as well as the `admin` database.

Centralized Role Data

MongoDB stores all role information in the *system.roles* (page 494) collection in the `admin` database

Do not access this collection directly but instead use the *role management commands* to view and edit custom roles.

Manage Users and Roles

On this page

- Overview (page 441)
- Prerequisites (page 442)
- Create a User-Defined Role (page 442)
- Modify Access for an Existing User (page 443)
- Modify the Password for an Existing User (page 445)
- View a User's Roles (page 445)
- View a Role's Privileges (page 446)

Overview

Changed in version 2.6: MongoDB 2.6 introduces a new *authorization model* (page 433).

This tutorial provides examples for user and role management under the MongoDB's authorization model. *Add Users* (page 396) describes how to add a new user to MongoDB.

Prerequisites

Important: If you have *enabled access control* (page 435) for your deployment, you must authenticate as a user with the required privileges specified in each section. A user administrator with the `userAdminAnyDatabase` (page 493) role, or `userAdmin` (page 488) role in the specific databases, provides the required privileges to perform the operations listed in this tutorial. See *Enable Client Access Control* (page 435) for details on adding user administrator as the first user.

Create a User-Defined Role

Roles grant users access to MongoDB resources. MongoDB provides a number of *built-in roles* (page 485) that administrators can use to control access to a MongoDB system. However, if these roles cannot describe the desired set of privileges, you can create new roles in a particular database.

Except for roles created in the `admin` database, a role can only include privileges that apply to its database and can only inherit from other roles in its database.

A role created in the `admin` database can include privileges that apply to the `admin` database, other databases or to the *cluster* (page 500) resource, and can inherit from roles in other databases as well as the `admin` database.

To create a new role, use the `db.createRole()` method, specifying the privileges in the `privileges` array and the inherited roles in the `roles` array.

MongoDB uses the combination of the database name and the role name to uniquely define a role. Each role is scoped to the database in which you create the role, but MongoDB stores all role information in the `admin.system.roles` (page 377) collection in the `admin` database.

Prerequisites To create a role in a database, you must have:

- the `createRole` (page 501) *action* (page 500) on that *database resource* (page 499).
- the `grantRole` (page 502) *action* (page 500) on that database to specify privileges for the new role as well as to specify roles to inherit from.

Built-in roles `userAdmin` (page 488) and `userAdminAnyDatabase` (page 493) provide `createRole` (page 501) and `grantRole` (page 502) actions on their respective *resources* (page 498).

Create a Role to Manage Current Operations The following example creates a role named `manageOpRole` which provides only the privileges to run both `db.currentOp()` and `db.killOp()`.[41]

Step 1: Connect to MongoDB with the appropriate privileges. Connect to `mongod` or `mongos` with the privileges specified in the *Prerequisites* (page 442) section.

The following procedure uses the `myUserAdmin` created in *Enable Client Access Control* (page 435).

```
mongo --port 27017 -u myUserAdmin -p abc123 --authenticationDatabase admin
```

The `myUserAdmin` has privileges to create roles in the `admin` as well as other databases.

[41] The built-in role `clusterMonitor` (page 490) also provides the privilege to run `db.currentOp()` along with other privileges, and the built-in role `hostManager` (page 490) provides the privilege to run `db.killOp()` along with other privileges.

Step 2: Create a new role to manage current operations. `manageOpRole` has privileges that act on multiple databases as well as the *cluster resource* (page 500). As such, you must create the role in the `admin` database.

```
use admin
db.createRole(
   {
     role: "manageOpRole",
     privileges: [
       { resource: { cluster: true }, actions: [ "killop", "inprog" ] },
       { resource: { db: "", collection: "" }, actions: [ "killCursors" ] }
     ],
     roles: []
   }
)
```

The new role grants permissions to kill any operations.

> **Warning:** Terminate running operations with extreme caution. Only use `db.killOp()` to terminate operations initiated by clients and *do not* terminate internal database operations.

Create a Role to Run `mongostat` The following example creates a role named `mongostatRole` that provides only the privileges to run `mongostat`. [42]

Step 1: Connect to MongoDB with the appropriate privileges. Connect to `mongod` or `mongos` with the privileges specified in the *Prerequisites* (page 442) section.

The following procedure uses the `myUserAdmin` created in *Enable Client Access Control* (page 435).

```
mongo --port 27017 -u myUserAdmin -p abc123 --authenticationDatabase admin
```

The `myUserAdmin` has privileges to create roles in the `admin` as well as other databases.

Step 2: Create a new role to manage current operations. `mongostatRole` has privileges that act on the *cluster resource* (page 500). As such, you must create the role in the `admin` database.

```
use admin
db.createRole(
   {
     role: "mongostatRole",
     privileges: [
       { resource: { cluster: true }, actions: [ "serverStatus" ] }
     ],
     roles: []
   }
)
```

Modify Access for an Existing User

Prerequisites

- You must have the `grantRole` (page 502) *action* (page 500) on a database to grant a role on that database.
- You must have the `revokeRole` (page 502) *action* (page 500) on a database to revoke a role on that database.

[42] The built-in role `clusterMonitor` (page 490) also provides the privilege to run `mongostat` along with other privileges.

- To view a role's information, you must be either explicitly granted the role or must have the `viewRole` (page 502) *action* (page 500) on the role's database.

Procedure

Step 1: Connect to MongoDB with the appropriate privileges. Connect to `mongod` or `mongos` as a user with the privileges specified in the prerequisite section.

The following procedure uses the `myUserAdmin` created in *Enable Client Access Control* (page 435).

```
mongo --port 27017 -u myUserAdmin -p abc123 --authenticationDatabase admin
```

Step 2: Identify the user's roles and privileges. To display the roles and privileges of the user to be modified, use the `db.getUser()` and `db.getRole()` methods.

For example, to view roles for `reportsUser` created in *Examples* (page 397), issue:

```
use reporting
db.getUser("reportsUser")
```

To display the privileges granted to the user by the `readWrite` role on the `"accounts"` database, issue:

```
use accounts
db.getRole( "readWrite", { showPrivileges: true } )
```

Step 3: Identify the privileges to grant or revoke. If the user requires additional privileges, grant to the user the role, or roles, with the required set of privileges. If such a role does not exist, *create a new role* (page 442) with the appropriate set of privileges.

To revoke a subset of privileges provided by an existing role: revoke the original role and grant a role that contains only the required privileges. You may need to *create a new role* (page 442) if a role does not exist.

Step 4: Modify the user's access.

Revoke a Role Revoke a role with the `db.revokeRolesFromUser()` method. The following example operation removes the `readWrite` (page 486) role on the `accounts` database from the `reportsUser`:

```
use reporting
db.revokeRolesFromUser(
    "reportsUser",
    [
      { role: "readWrite", db: "accounts" }
    ]
)
```

Grant a Role Grant a role using the `db.grantRolesToUser()` method. For example, the following operation grants the `reportsUser` user the `read` (page 486) role on the `accounts` database:

```
use reporting
db.grantRolesToUser(
    "reportsUser",
    [
      { role: "read", db: "accounts" }
```

```
    ]
)
```

For sharded clusters, the changes to the user are instant on the `mongos` on which the command runs. However, for other `mongos` instances in the cluster, the user cache may wait up to 10 minutes to refresh. See `userCacheInvalidationIntervalSecs`.

Modify the Password for an Existing User

Prerequisites To modify the password of another user on a database, you must have the `changeAnyPassword` *action* (page 500) on that database.

Procedure

Step 1: Connect to MongoDB with the appropriate privileges. Connect to the `mongod` or `mongos` with the privileges specified in the *Prerequisites* (page 445) section.

The following procedure uses the `myUserAdmin` created in *Enable Client Access Control* (page 435).

```
mongo --port 27017 -u myUserAdmin -p abc123 --authenticationDatabase admin
```

Step 2: Change the password. Pass the user's username and the new password to the `db.changeUserPassword()` method.

The following operation changes the `reporting` user's password to `SOh3TbYhxuLiW8ypJPxmt1oOfL`:

```
db.changeUserPassword("reporting", "SOh3TbYhxuLiW8ypJPxmt1oOfL")
```

See also:

Change Your Password and Custom Data (page 447)

View a User's Roles

Prerequisites To view another user's information, you must have the `viewUser` (page 502) *action* (page 500) on the other user's database.

Users can view their own information.

Procedure

Step 1: Connect to MongoDB with the appropriate privileges. Connect to `mongod` or `mongos` as a user with the privileges specified in the prerequisite section.

The following procedure uses the `myUserAdmin` created in *Enable Client Access Control* (page 435).

```
mongo --port 27017 -u myUserAdmin -p abc123 --authenticationDatabase admin
```

Step 2: Identify the user's roles. Use the `usersInfo` command or `db.getUser()` method to display user information.

For example, to view roles for `reportsUser` created in *Examples* (page 397), issue:

```
use reporting
db.getUser("reportsUser")
```

In the returned document, the `roles` (page 497) field displays all roles for `reportsUser`:

```
...
"roles" : [
   { "role" : "readWrite", "db" : "accounts" },
   { "role" : "read", "db" : "reporting" },
   { "role" : "read", "db" : "products" },
   { "role" : "read", "db" : "sales" }
]
```

View a Role's Privileges

Prerequisites To view a role's information, you must be either explicitly granted the role or must have the `viewRole` (page 502) *action* (page 500) on the role's database.

Procedure

Step 1: Connect to MongoDB with the appropriate privileges. Connect to `mongod` or `mongos` as a user with the privileges specified in the prerequisite section.

The following procedure uses the `myUserAdmin` created in *Enable Client Access Control* (page 435).

```
mongo --port 27017 -u myUserAdmin -p abc123 --authenticationDatabase admin
```

Step 2: Identify the privileges granted by a role. For a given role, use the `db.getRole()` method, or the `rolesInfo` command, with the `showPrivileges` option:

For example, to view the privileges granted by `read` role on the `products` database, use the following operation, issue:

```
use products
db.getRole( "read", { showPrivileges: true } )
```

In the returned document, the `privileges` and `inheritedPrivileges` arrays. The `privileges` lists the privileges directly specified by the role and excludes those privileges inherited from other roles. The `inheritedPrivileges` lists all privileges granted by this role, both directly specified and inherited. If the role does not inherit from other roles, the two fields are the same.

```
...
"privileges" : [
  {
    "resource": { "db" : "products", "collection" : "" },
    "actions": [ "collStats","dbHash","dbStats","find","killCursors","planCacheRead" ]
  },
  {
    "resource" : { "db" : "products", "collection" : "system.js" },
    "actions": [ "collStats","dbHash","dbStats","find","killCursors","planCacheRead" ]
  }
```

```
],
"inheritedPrivileges" : [
  {
    "resource": { "db" : "products", "collection" : "" },
    "actions": [ "collStats","dbHash","dbStats","find","killCursors","planCacheRead" ]
  },
  {
    "resource" : { "db" : "products", "collection" : "system.js" },
    "actions": [ "collStats","dbHash","dbStats","find","killCursors","planCacheRead" ]
  }
]
```

Change Your Password and Custom Data

Changed in version 2.6.

On this page

- Overview (page 447)
- Considerations (page 447)
- Prerequisites (page 447)
- Procedure (page 448)

Overview

Users with appropriate privileges can change their own passwords and custom data. `Custom data` (page 498) stores optional user information.

Considerations

To generate a strong password for use in this procedure, you can use the `openssl` utility's `rand` command. For example, issue `openssl rand` with the following options to create a base64-encoded string of 48 pseudo-random bytes:

```
openssl rand -base64 48
```

Prerequisites

To modify your own password and custom data, you must have privileges that grant `changeOwnPassword` (page 501) and `changeOwnCustomData` (page 501) *actions* (page 500) respectively on the user's database.

Step 1: Connect as a user with privileges to manage users and roles. Connect to the `mongod` or `mongos` with privileges to manage users and roles, such as a user with `userAdminAnyDatabase` (page 493) role. The following procedure uses the `myUserAdmin` created in *Enable Client Access Control* (page 435).

```
mongo --port 27017 -u myUserAdmin -p abc123 --authenticationDatabase admin
```

Step 2: Create a role with appropriate privileges. In the `admin` database, `create` a new role with `changeOwnPassword` (page 501) and `changeOwnCustomData` (page 501).

```
use admin
db.createRole(
   { role: "changeOwnPasswordCustomDataRole",
     privileges: [
        {
          resource: { db: "", collection: ""},
          actions: [ "changeOwnPassword", "changeOwnCustomData" ]
        }
     ],
     roles: []
   }
)
```

Step 3: Add a user with this role. In the `test` database, `create` a new user with the created `"changeOwnPasswordCustomDataRole"` role. For example, the following operation creates a user with both the built-in role `readWrite` (page 486) and the user-created `"changeOwnPasswordCustomDataRole"`.

```
use test
db.createUser(
   {
     user:"user123",
     pwd:"12345678",
     roles:[ "readWrite", { role:"changeOwnPasswordCustomDataRole", db:"admin" } ]
   }
)
```

To grant an existing user the new role, use `db.grantRolesToUser()`.

Procedure

Step 1: Connect with the appropriate privileges. Connect to the `mongod` or `mongos` as a user with appropriate privileges.

For example, the following operation connects to MongoDB as `user123` created in the *Prerequisites* (page 447) section.

```
mongo --port 27017 -u user123 -p 12345678 --authenticationDatabase test
```

To check that you have the privileges specified in the *Prerequisites* (page 447) section as well as to see user information, use the `usersInfo` command with the `showPrivileges` option.

Step 2: Change your password and custom data. Use the `db.updateUser()` method to update the password and custom data.

For example, the following operation changes thw user's password to `KNlZmiaNUp0B` and custom data to `{ title: "Senior Manager" }`:

```
use test
db.updateUser(
   "user123",
   {
      pwd: "KNlZmiaNUp0B",
      customData: { title: "Senior Manager" }
```

```
    }
)
```

Collection-Level Access Control

On this page

- Privileges and Scope (page 449)
- Additional Information (page 449)

Collection-level access control allows administrators to grant users privileges that are scoped to specific collections.

Administrators can implement collection-level access control through *user-defined roles* (page 440). By creating a role with *privileges* (page 434) that are scoped to a specific collection in a particular database, administrators can provision users with roles that grant privileges on a collection level.

Privileges and Scope

A privilege consists of *actions* (page 500) and the *resources* (page 498) upon which the actions are permissible; i.e. the resources define the scope of the actions for that privilege.

By specifying both the database and the collection in the *resource document* (page 499) for a privilege, administrator can limit the privilege actions just to a specific collection in a specific database. Each privilege action in a role can be scoped to a different collection.

For example, a user defined role can contain the following privileges:

```
privileges: [
  { resource: { db: "products", collection: "inventory" }, actions: [ "find", "update", "insert" ] },
  { resource: { db: "products", collection: "orders" },  actions: [ "find" ] }
]
```

The first privilege scopes its actions to the `inventory` collection of the `products` database. The second privilege scopes its actions to the `orders` collection of the `products` database.

Additional Information

For more information on user-defined roles and MongoDB authorization model, see *Role-Based Access Control* (page 433). For a tutorial on creating user-defined roles, see *Manage Users and Roles* (page 441).

9.4 Encryption

On this page

- Transport Encryption (page 450)
- Encryption at Rest (page 450)

9.4.1 Transport Encryption

You can use TLS/SSL (Transport Layer Security/Secure Sockets Layer) to encrypt all of MongoDB's network traffic. TLS/SSL ensures that MongoDB network traffic is only readable by the intended client.

See *Transport Encryption* (page 450) for more information.

9.4.2 Encryption at Rest

There are two broad classes of approaches to encrypting data at rest with MongoDB: Application Level Encryption and Storage Encryption. You can use these solutions together or independently.

New in version 3.2: MongoDB Enterprise 3.2 introduces a native encryption option for the WiredTiger storage engine. This feature allows MongoDB to encrypt data files such that only parties with the decryption key can decode and read the data.

See *Encryption At Rest* (page 461) for more information.

Transport Encryption

On this page

- TLS/SSL (page 450)
- Certificates (page 450)
- Identity Verification (page 450)
- FIPS Mode (page 451)

TLS/SSL

MongoDB supports TLS/SSL (Transport Layer Security/Secure Sockets Layer) to encrypt all of MongoDB's network traffic. TLS/SSL ensures that MongoDB network traffic is only readable by the intended client.

MongoDB TLS/SSL implementation uses OpenSSL libraries. MongoDB's SSL encryption only allows use of strong SSL ciphers with a minimum of 128-bit key length for all connections.

Certificates

Before you can use SSL, you must have a `.pem` file containing a public key certificate and its associated private key.

MongoDB can use any valid SSL certificate issued by a certificate authority or a self-signed certificate. If you use a self-signed certificate, although the communications channel will be encrypted, there will be *no* validation of server identity. Although such a situation will prevent eavesdropping on the connection, it leaves you vulnerable to a man-in-the-middle attack. Using a certificate signed by a trusted certificate authority will permit MongoDB drivers to verify the server's identity.

For example, see *TLS/SSL Configuration for Clients* (page 455).

Identity Verification

In addition to encrypting connections, SSL allows for authentication using certificates, both for *client authentication* (page 393) and for *internal authentication* (page 423) of members of replica sets and sharded clusters.

For more information, see:

- *Configure mongod and mongos for TLS/SSL* (page 451)
- *TLS/SSL Configuration for Clients* (page 455)
- *Use x.509 Certificates to Authenticate Clients* (page 403)
- *Use x.509 Certificate for Membership Authentication* (page 430)

FIPS Mode

Enterprise Feature

Available in MongoDB Enterprise only.

The Federal Information Processing Standard (FIPS) is a U.S. government computer security standard used to certify software modules and libraries that encrypt and decrypt data securely. You can configure MongoDB to run with a FIPS 140-2 certified library for OpenSSL. Configure FIPS to run by default or as needed from the command line.

For an example, see *Configure MongoDB for FIPS* (page 459).

On this page

Configure `mongod` and `mongos` for TLS/SSL

- Overview (page 451)
- Prerequisites (page 451)
- Procedures (page 452)

Overview This document helps you to configure MongoDB to support TLS/SSL. MongoDB clients can use TLS/SSL to encrypt connections to `mongod` and `mongos` instances. MongoDB TLS/SSL implementation uses OpenSSL libraries.

Note: Although TLS is the successor to SSL, this page uses the more familiar term SSL to refer to TLS/SSL.

These instructions assume that you have already installed a build of MongoDB that includes SSL support and that your client driver supports SSL. For instructions on upgrading a cluster currently not using SSL to using SSL, see *Upgrade a Cluster to Use TLS/SSL* (page 458).

Changed in version 2.6: MongoDB's SSL encryption only allows use of strong SSL ciphers with a minimum of 128-bit key length for all connections.

Prerequisites

Important: A full description of TLS/SSL, PKI (Public Key Infrastructure) certificates, and Certificate Authority is beyond the scope of this document. This page assumes prior knowledge of TLS/SSL as well as access to valid certificates.

MongoDB Support New in version 3.0: Most MongoDB distributions now include support for SSL.

Certain distributions of MongoDB[43] do **not** contain support for SSL. To use SSL, be sure to choose a package that supports SSL. All MongoDB Enterprise[44] supported platforms include SSL support.

[43] http://www.mongodb.org/downloads?jmp=docs
[44] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Client Support See *TLS/SSL Configuration for Clients* (page 455) to learn about SSL support for Python, Java, Ruby, and other clients.

Certificate Authorities For production use, your MongoDB deployment should use valid certificates generated and signed by a single certificate authority. You or your organization can generate and maintain an independent certificate authority, or use certificates generated by a third-party SSL vendor. Obtaining and managing certificates is beyond the scope of this documentation.

`.pem` File Before you can use SSL, you must have a `.pem` file containing a public key certificate and its associated private key.

MongoDB can use any valid SSL certificate issued by a certificate authority, or a self-signed certificate. If you use a self-signed certificate, although the communications channel will be encrypted, there will be *no* validation of server identity. Although such a situation will prevent eavesdropping on the connection, it leaves you vulnerable to a man-in-the-middle attack. Using a certificate signed by a trusted certificate authority will permit MongoDB drivers to verify the server's identity.

In general, avoid using self-signed certificates unless the network is trusted.

Additionally, with regards to *authentication among replica set/sharded cluster members* (page 423), in order to minimize exposure of the private key and allow hostname validation, it is advisable to use different certificates on different servers.

For *testing* purposes, you can generate a self-signed certificate and private key on a Unix system with a command that resembles the following:

```
cd /etc/ssl/
openssl req -newkey rsa:2048 -new -x509 -days 365 -nodes -out mongodb-cert.crt -keyout mongodb-cert.k
```

This operation generates a new, self-signed certificate with no passphrase that is valid for 365 days. Once you have the certificate, concatenate the certificate and private key to a `.pem` file, as in the following example:

```
cat mongodb-cert.key mongodb-cert.crt > mongodb.pem
```

See also:

Use x.509 Certificates to Authenticate Clients (page 403)

Procedures

Set Up `mongod` and `mongos` with SSL Certificate and Key To use SSL in your MongoDB deployment, include the following run-time options with `mongod` and `mongos`:

- `net.ssl.mode` set to `requireSSL`. This setting restricts each server to use only SSL encrypted connections. You can also specify either the value `allowSSL` or `preferSSL` to set up the use of mixed SSL modes on a port. See `net.ssl.mode` for details.
- `PEMKeyfile` with the `.pem` file that contains the SSL certificate and key.

Consider the following syntax for `mongod`:

```
mongod --sslMode requireSSL --sslPEMKeyFile <pem>
```

For example, given an SSL certificate located at `/etc/ssl/mongodb.pem`, configure `mongod` to use SSL encryption for all connections with the following command:

```
mongod --sslMode requireSSL --sslPEMKeyFile /etc/ssl/mongodb.pem
```

Note:

- Specify <pem> with the full path name to the certificate.
- If the private key portion of the <pem> is encrypted, specify the passphrase. See *SSL Certificate Passphrase* (page 455).

You may also specify these options in the configuration file, as in the following examples:

If using the YAML configuration file format:

```
net:
   ssl:
      mode: requireSSL
      PEMKeyFile: /etc/ssl/mongodb.pem
```

Or, if using the older older configuration file format[45]:

```
sslMode = requireSSL
sslPEMKeyFile = /etc/ssl/mongodb.pem
```

To connect, to mongod and mongos instances using SSL, the mongo shell and MongoDB tools must include the --ssl option. See *TLS/SSL Configuration for Clients* (page 455) for more information on connecting to mongod and mongos running with SSL.

See also:

Upgrade a Cluster to Use TLS/SSL (page 458)

Set Up mongod and mongos with Certificate Validation To set up mongod or mongos for SSL encryption using an SSL certificate signed by a certificate authority, include the following run-time options during startup:

- net.ssl.mode set to requireSSL. This setting restricts each server to use only SSL encrypted connections. You can also specify either the value allowSSL or preferSSL to set up the use of mixed SSL modes on a port. See net.ssl.mode for details.
- PEMKeyfile with the name of the .pem file that contains the signed SSL certificate and key.
- CAFile with the name of the .pem file that contains the root certificate chain from the Certificate Authority.

Consider the following syntax for mongod:

```
mongod --sslMode requireSSL --sslPEMKeyFile <pem> --sslCAFile <ca>
```

For example, given a signed SSL certificate located at /etc/ssl/mongodb.pem and the certificate authority file at /etc/ssl/ca.pem, you can configure mongod for SSL encryption as follows:

```
mongod --sslMode requireSSL --sslPEMKeyFile /etc/ssl/mongodb.pem --sslCAFile /etc/ssl/ca.pem
```

Note:

- Specify the <pem> file and the <ca> file with either the full path name or the relative path name.
- If the <pem> is encrypted, specify the passphrase. See *SSL Certificate Passphrase* (page 455).

You may also specify these options in the configuration file, as in the following examples:

[45] https://docs.mongodb.org/v2.4/reference/configuration-options

If using the YAML configuration file format:

```
net:
   ssl:
      mode: requireSSL
      PEMKeyFile: /etc/ssl/mongodb.pem
      CAFile: /etc/ssl/ca.pem
```

Or, if using the older older configuration file format[46]:

```
sslMode = requireSSL
sslPEMKeyFile = /etc/ssl/mongodb.pem
sslCAFile = /etc/ssl/ca.pem
```

To connect, to mongod and mongos instances using SSL, the mongo tools must include the both the *--ssl* and *--sslPEMKeyFile* option. See *TLS/SSL Configuration for Clients* (page 455) for more information on connecting to mongod and mongos running with SSL.

See also:

Upgrade a Cluster to Use TLS/SSL (page 458)

Block Revoked Certificates for Clients To prevent clients with revoked certificates from connecting, include the sslCRLFile to specify a .pem file that contains revoked certificates.

For example, the following mongod with SSL configuration includes the sslCRLFile setting:

```
mongod --sslMode requireSSL --sslCRLFile /etc/ssl/ca-crl.pem --sslPEMKeyFile /etc/ssl/mongodb.pem --s
```

Clients with revoked certificates in the /etc/ssl/ca-crl.pem will not be able to connect to this mongod instance.

Validate Only if a Client Presents a Certificate In most cases it is important to ensure that clients present valid certificates. However, if you have clients that cannot present a client certificate, or are transitioning to using a certificate authority you may only want to validate certificates from clients that present a certificate.

If you want to bypass validation for clients that don't present certificates, include the allowConnectionsWithoutCertificates run-time option with mongod and mongos. If the client does not present a certificate, no validation occurs. These connections, though not validated, are still encrypted using SSL.

For example, consider the following mongod with an SSL configuration that includes the allowConnectionsWithoutCertificates setting:

```
mongod --sslMode requireSSL --sslAllowConnectionsWithoutCertificates --sslPEMKeyFile /etc/ssl/mongodb
```

Then, clients can connect either with the option *--ssl* and **no** certificate or with the option *--ssl* and a **valid** certificate. See *TLS/SSL Configuration for Clients* (page 455) for more information on SSL connections for clients.

Note: If the client presents a certificate, the certificate must be a valid certificate.

All connections, including those that have not presented certificates are encrypted using SSL.

---

[46] https://docs.mongodb.org/v2.4/reference/configuration-options

Disallow Protocols New in version 3.0.7.

To prevent MongoDB servers from accepting incoming connections that use specific protocols, include the `--sslDisabledProtocols` option, or if using the `configuration file` the `net.ssl.disabledProtocols` setting.

For example, the following configuration uses `--sslDisabledProtocols` option to prevent `mongod` from accepting incoming connections that use either `TLS1_0` or `TLS1_1`:

```
mongod --sslMode requireSSL --sslDisabledProtocols TLS1_0,TLS1_1 --sslPEMKeyFile /etc/ssl/mongodb.pem
```

If using the `YAML configuration file format`:

```
net:
   ssl:
      mode: requireSSL
      PEMKeyFile: /etc/ssl/mongodb.pem
      CAFile: /etc/ssl/ca.pem
      disabledProtocols: TLS1_0,TLS1_1
```

For more information, including the protocols recognized by the option, see `net.ssl.disabledProtocols` or the `--sslDisabledProtocols` option for `mongod` and `mongos`.

SSL Certificate Passphrase The PEM files for `PEMKeyfile` and `ClusterFile` may be encrypted. With encrypted PEM files, you must specify the passphrase at startup with a command-line or a configuration file option or enter the passphrase when prompted.

Changed in version 2.6: In previous versions, you can only specify the passphrase with a command-line or a configuration file option.

To specify the passphrase in clear text on the command line or in a configuration file, use the `PEMKeyPassword` and/or the `ClusterPassword` option.

To have MongoDB prompt for the passphrase at the start of `mongod` or `mongos` and avoid specifying the passphrase in clear text, omit the `PEMKeyPassword` and/or the `ClusterPassword` option. MongoDB will prompt for each passphrase as necessary.

Important: The passphrase prompt option is available if you run the MongoDB instance in the foreground with a connected terminal. If you run `mongod` or `mongos` in a non-interactive session (e.g. without a terminal or as a service on Windows), you cannot use the passphrase prompt option.

Run in FIPS Mode

Note: FIPS-compatible SSL is available only in MongoDB Enterprise[47]. See *Configure MongoDB for FIPS* (page 459) for more information.

See *Configure MongoDB for FIPS* (page 459) for more details.

TLS/SSL Configuration for Clients

On this page

- `mongo` Shell SSL Configuration (page 456)
- MongoDB Cloud Manager and Ops Manager Monitoring Agent (page 457)
- MongoDB Drivers (page 457)
- MongoDB Tools (page 458)

[47] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Clients must have support for TLS/SSL to work with a `mongod` or a `mongos` instance that has TLS/SSL support enabled.

Important: A full description of TLS/SSL, PKI (Public Key Infrastructure) certificates, and Certificate Authority is beyond the scope of this document. This page assumes prior knowledge of TLS/SSL as well as access to valid certificates.

Note: Although TLS is the successor to SSL, this page uses the more familiar term SSL to refer to TLS/SSL.

See also:

Configure mongod and mongos for TLS/SSL (page 451).

mongo Shell SSL Configuration For SSL connections, you must use the `mongo` shell built with SSL support or distributed with MongoDB Enterprise.

New in version 3.0: Most MongoDB distributions now include support for SSL.

The `mongo` shell provides various *mongo-shell-ssl* settings, including:

- `--ssl`
- `--sslPEMKeyFile` with the name of the `.pem` file that contains the SSL certificate and key.
- `--sslCAFile` with the name of the `.pem` file that contains the certificate from the Certificate Authority (CA).

  Changed in version 3.0: When running `mongo` with the `--ssl` option, you must include either `--sslCAFile` or `--sslAllowInvalidCertificates`.

  This restriction does not apply to the MongoDB tools. However, running the tools without `-sslCAFile` creates the same vulnerability to invalid certificates.

  Warning: For SSL connections (`--ssl`) to `mongod` and `mongos`, if the `mongo` shell (or *MongoDB tools* (page 458)) runs without the `--sslCAFile <CAFile>` option (i.e. specifies the `--sslAllowInvalidCertificates` instead), the `mongo` shell (or *MongoDB tools* (page 458)) will not attempt to validate the server certificates. This creates a vulnerability to expired `mongod` and `mongos` certificates as well as to foreign processes posing as valid `mongod` or `mongos` instances. Ensure that you *always* specify the CA file to validate the server certificates in cases where intrusion is a possibility.

- `--sslPEMKeyPassword` option if the client certificate-key file is encrypted.

For a complete list of the `mongo` shell's SSL settings, see *mongo-shell-ssl*.

Connect to MongoDB Instance with SSL Encryption To connect to a `mongod` or `mongos` instance that requires *only a SSL encryption mode* (page 452), start `mongo` shell with `--ssl` and include the `--sslCAFile` to validate the server certificates.

```
mongo --ssl --sslCAFile /etc/ssl/ca.pem
```

Changed in version 3.0: When running `mongo` with the `--ssl` option, you must include either `--sslCAFile` or `--sslAllowInvalidCertificates`.

This restriction does not apply to the MongoDB tools. However, running the tools without `-sslCAFile` creates the same vulnerability to invalid certificates.

Connect to MongoDB Instance that Requires Client Certificates To connect to a `mongod` or `mongos` that requires *CA-signed client certificates* (page 453), start the `mongo` shell with `--ssl`, the `--sslPEMKeyFile` option to specify the signed certificate-key file, and the `--sslCAFile` to validate the server certificates.

```
mongo --ssl --sslPEMKeyFile /etc/ssl/client.pem --sslCAFile /etc/ssl/ca.pem
```

Changed in version 3.0: When running `mongo` with the `--ssl` option, you must include either `--sslCAFile` or `--sslAllowInvalidCertificates`.

This restriction does not apply to the MongoDB tools. However, running the tools without `-sslCAFile` creates the same vulnerability to invalid certificates.

Connect to MongoDB Instance that Validates when Presented with a Certificate To connect to a `mongod` or `mongos` instance that *only requires valid certificates when the client presents a certificate* (page 454), start `mongo` shell either:

- with the `--ssl`, `--sslCAFile`, and **no** certificate or
- with the `--ssl`, `--sslCAFile`, and a **valid** signed certificate.

Changed in version 3.0: When running `mongo` with the `--ssl` option, you must include either `--sslCAFile` or `--sslAllowInvalidCertificates`.

This restriction does not apply to the MongoDB tools. However, running the tools without `-sslCAFile` creates the same vulnerability to invalid certificates.

For example, if `mongod` is running with weak certificate validation, both of the following `mongo` shell clients can connect to that `mongod`:

```
mongo --ssl --sslCAFile /etc/ssl/ca.pem
mongo --ssl --sslPEMKeyFile /etc/ssl/client.pem --sslCAFile /etc/ssl/ca.pem
```

Important: If the client presents a certificate, the certificate must be valid.

MongoDB Cloud Manager and Ops Manager Monitoring Agent The MongoDB Cloud Manager Monitoring agent will also have to connect via SSL in order to gather its statistics. Because the agent already utilizes SSL for its communications to the MongoDB Cloud Manager servers, this is just a matter of enabling SSL support in MongoDB Cloud Manager itself on a per host basis. å See the MongoDB Cloud Manager documentation[48] for more information about SSL configuration.

For Ops Manager, see Ops Manager documentation[49].

MongoDB Drivers The MongoDB Drivers support for connection to SSL enabled MongoDB. See:

- C Driver[50]
- C++ Driver[51]
- C# Driver[52]
- Java Driver[53]

[48] https://docs.cloud.mongodb.com/
[49] https://docs.opsmanager.mongodb.com/current/
[50] http://api.mongodb.org/c/current/advanced-connections.html
[51] https://github.com/mongodb/mongo-cxx-driver/wiki/Configuring%20the%20Legacy%20Driver
[52] http://mongodb.github.io/mongo-csharp-driver/2.0/reference/driver/ssl/
[53] http://mongodb.github.io/mongo-java-driver/3.0/driver/reference/connecting/ssl/

- Node.js Driver[54]
- Perl Driver[55]
- PHP Driver[56]
- Python Driver[57]
- Ruby Driver[58]
- Scala Driver[59]

MongoDB Tools Changed in version 2.6.

Various MongoDB utility programs supports SSL. These tools include:

- `mongodump`
- `mongoexport`
- `mongofiles`
- `mongoimport`
- `mongorestore`
- `mongostat`
- `mongotop`

To use SSL connections with these tools, use the same SSL options as the `mongo` shell. See *mongo Shell SSL Configuration* (page 456).

Upgrade a Cluster to Use TLS/SSL Changed in version 3.0: Most MongoDB distributions now include support for TLS/SSL. See *Configure mongod and mongos for TLS/SSL* (page 451) and *TLS/SSL Configuration for Clients* (page 455) for more information about TLS/SSL and MongoDB.

Important: A full description of TLS/SSL, PKI (Public Key Infrastructure) certificates, and Certificate Authority is beyond the scope of this document. This page assumes prior knowledge of TLS/SSL as well as access to valid certificates.

Changed in version 2.6.

The MongoDB server supports listening for both TLS/SSL encrypted and unencrypted connections on the same TCP port. This allows upgrades of MongoDB clusters to use TLS/SSL encrypted connections.

To upgrade from a MongoDB cluster using no TLS/SSL encryption to one using *only* TLS/SSL encryption, use the following rolling upgrade process:

1. For each node of a cluster, start the node with the option `--sslMode` set to `allowSSL`. The `--sslMode allowSSL` setting allows the node to accept both TLS/SSL and non-TLS/non-SSL incoming connections. Its connections to other servers do not use TLS/SSL. Include other *TLS/SSL options* (page 451) as well as any other options that are required for your specific configuration. For example:

   ```
   mongod --replSet <name> --sslMode allowSSL --sslPEMKeyFile <path to TLS/SSL Certificate and key
   ```

[54] http://mongodb.github.io/node-mongodb-native/2.0/tutorials/enterprise_features/
[55] https://metacpan.org/pod/MongoDB::MongoClient#ssl
[56] http://php.net/manual/en/mongo.connecting.ssl.php
[57] http://api.mongodb.org/python/current/examples/tls.html
[58] http://docs.mongodb.org/ecosystem/tutorial/ruby-driver-tutorial/#mongodb-x509-mechanism
[59] http://mongodb.github.io/mongo-scala-driver/1.1/reference/connecting/ssl/

Upgrade all nodes of the cluster to these settings.

You may also specify these options in the `configuration file`. If using a `YAML format configuration file`, specify the following settings in the file:

```
net:
   ssl:
      mode: <disabled|allowSSL|preferSSL|requireSSL>
      PEMKeyFile: <path to TLS/SSL certificate and key PEM file>
      CAFile: <path to root CA PEM file>
```

Or, if using the older configuration file format[60]:

```
sslMode = <disabled|allowSSL|preferSSL|requireSSL>
sslPEMKeyFile = <path to TLS/SSL certificate and key PEM file>
sslCAFile = <path to root CA PEM file>
```

2. Switch all clients to use TLS/SSL. See *TLS/SSL Configuration for Clients* (page 455).
3. For each node of a cluster, use the `setParameter` command to update the `sslMode` to `preferSSL`. [61] With `preferSSL` as its `net.ssl.mode`, the node accepts both TLS/SSL and non-TLS/non-SSL incoming connections, and its connections to other servers use TLS/SSL. For example:

   ```
   db.getSiblingDB('admin').runCommand( { setParameter: 1, sslMode: "preferSSL" } )
   ```

   Upgrade all nodes of the cluster to these settings.

   At this point, all connections should be using TLS/SSL.
4. For each node of the cluster, use the `setParameter` command to update the `sslMode` to `requireSSL`. [1] With `requireSSL` as its `net.ssl.mode`, the node will reject any non-TLS/non-SSL connections. For example:

   ```
   db.getSiblingDB('admin').runCommand( { setParameter: 1, sslMode: "requireSSL" } )
   ```
5. After the upgrade of all nodes, edit the `configuration file` with the appropriate TLS/SSL settings to ensure that upon subsequent restarts, the cluster uses TLS/SSL.

Configure MongoDB for FIPS

On this page

- Overview (page 459)
- Prerequisites (page 460)
- Considerations (page 460)
- Procedure (page 460)

New in version 2.6.

Overview The Federal Information Processing Standard (FIPS) is a U.S. government computer security standard used to certify software modules and libraries that encrypt and decrypt data securely. You can configure MongoDB to run with a FIPS 140-2 certified library for OpenSSL. Configure FIPS to run by default or as needed from the command line.

[60] https://docs.mongodb.org/v2.4/reference/configuration-options

[61] As an alternative to using the `setParameter` command, you can also restart the nodes with the appropriate TLS/SSL options and values.

Prerequisites

Important: A full description of FIPS and TLS/SSL is beyond the scope of this document. This tutorial assumes prior knowledge of FIPS and TLS/SSL.

Only the MongoDB Enterprise[62] version supports FIPS mode. See *Install MongoDB Enterprise* (page 49) to download and install MongoDB Enterprise[63] to use FIPS mode.

Your system must have an OpenSSL library configured with the FIPS 140-2 module. At the command line, type `openssl version` to confirm your OpenSSL software includes FIPS support.

For Red Hat Enterprise Linux 6.x (RHEL 6.x) or its derivatives such as CentOS 6.x, the OpenSSL toolkit must be at least `openssl-1.0.1e-16.el6_5` to use FIPS mode. To upgrade the toolkit for these platforms, issue the following command:

```
sudo yum update openssl
```

Some versions of Linux periodically execute a process to *prelink* dynamic libraries with pre-assigned addresses. This process modifies the OpenSSL libraries, specifically `libcrypto`. The OpenSSL FIPS mode will subsequently fail the signature check performed upon startup to ensure `libcrypto` has not been modified since compilation.

To configure the Linux prelink process to not prelink `libcrypto`:

```
sudo bash -c "echo '-b /usr/lib64/libcrypto.so.*' >>/etc/prelink.conf.d/openssl-prelink.conf"
```

Considerations FIPS is property of the encryption system and not the access control system. However, if your environment requires FIPS compliant encryption *and* access control, you must ensure that the access control system uses only FIPS-compliant encryption.

MongoDB's FIPS support covers the way that MongoDB uses OpenSSL for network encryption and X509 authentication. If you use Kerberos or LDAP Proxy authentication, you muse ensure that these external mechanisms are FIPS-compliant. `MONGODB-CR` authentication is *not* FIPS compliant.

Procedure

Configure MongoDB to use TLS/SSL See *Configure mongod and mongos for TLS/SSL* (page 451) for details about configuring OpenSSL.

Run `mongod` or `mongos` instance in FIPS mode Perform these steps after you *Configure mongod and mongos for TLS/SSL* (page 451).

Step 1: Change configuration file. To configure your `mongod` or `mongos` instance to use FIPS mode, shut down the instance and update the configuration file with the following setting:

```
net:
   ssl:
      FIPSMode: true
```

[62] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs
[63] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

Step 2: Start `mongod` or `mongos` instance with configuration file. For example, run this command to start the `mongod` instance with its configuration file:

```
mongod --config /etc/mongod.conf
```

Confirm FIPS mode is running Check the server log file for a message FIPS is active:

```
FIPS 140-2 mode activated
```

Encryption At Rest

On this page

- Encrypted Storage Engine (page 461)
- Application Level Encryption (page 462)

Encryption at rest, when used in conjunction with transport encryption and good security policies that protect relevant accounts, passwords, and encryption keys, can help ensure compliance with security and privacy standards, including HIPAA, PCI-DSS, and FERPA.

Encrypted Storage Engine

New in version 3.2.

Enterprise Feature

Available in MongoDB Enterprise only.

Important: Available for the WiredTiger Storage Engine only.

MongoDB Enterprise 3.2 introduces a native encryption option for the WiredTiger storage engine. This feature allows MongoDB to encrypt data files such that only parties with the decryption key can decode and read the data.

Encryption Process If encryption is enabled, the default encryption mode that MongoDB Enterprise uses is the `AES256-CBC` (or 256-bit Advanced Encryption Standard in Cipher Block Chaining mode) via OpenSSL. AES-256 uses a symmetric key; i.e. the same key to encrypt and decrypt text. MongoDB Enterprise also supports authenticated encryption `AES256-GCM` (or 256-bit Advanced Encryption Standard in Galois/Counter Mode). FIPS mode encryption is also available.

The data encryption includes:

- Generating a master key.
- Generating keys for each database.
- Encrypting data with the database keys.
- Encrypting the database keys with the master key.

The encryption occur transparently in the storage layer; i.e. all data files are fully encrypted from a filesystem perspective, and data only exists in an unencrypted state in memory and during transmission.

To encrypt all of MongoDB's network traffic, you can use TLS/SSL (Transport Layer Security/Secure Sockets Layer). See *Configure mongod and mongos for TLS/SSL* (page 451) and *TLS/SSL Configuration for Clients* (page 455).

Key Management

Important: Secure management of the encryption keys is critical.

The database keys are internal to the server and are only paged to disk in an encrypted format. MongoDB never pages the master key to disk under any circumstances.

Only the master key is external to the server (i.e. kept separate from the data and the database keys), and requires external management. To manage the master key, MongoDB's encrypted storage engine supports two key management options:

- Integration with a third party key management appliance via the Key Management Interoperability Protocol (KMIP). **Recommended**
- Local key management via a keyfile.

To configure MongoDB for encryption and use one of the two key management options, see *Configure Encryption* (page 462).

Encryption and Replication Encryption is not a part of replication:

- Master keys and database keys are not replicated, and
- Data is not natively encrypted over the wire.

Although you could reuse the same key for the nodes, MongoDB recommends the use of individual keys for each node as well as the use of transport encryption.

For details, see *Rotate Encryption Keys* (page 465).

Application Level Encryption

Application Level Encryption provides encryption on a per-field or per-document basis within the application layer. To encrypt document or field level data, write custom encryption and decryption routines or use a commercial solution.

For a list of MongoDB's certified partners, refer to the Partners List[64]. To view security partners, select "Security" from the *Technology* filter, and "Certified" from the *Certified* filter.

Configure Encryption	**On this page** • Overview (page 462) • Key Manager (page 463) • Local Key Management (page 464)

New in version 3.2.

Overview

Enterprise Feature

Available in MongoDB Enterprise only.

Important: Available for the WiredTiger Storage Engine Only.

[64] https://www.mongodb.com/partners/list

MongoDB Enterprise 3.2 introduces a native encryption option for the WiredTiger storage engine. With storage encryption, the secure management of the encryption keys is critical.

Only the master key is external to the server and requires external management. To manage the master key, MongoDB's encrypted storage engine supports two key management options:

- Integration with a third party key management appliance via the Key Management Interoperability Protocol (KMIP). **Recommended**
- Use of local key management via a keyfile.

The following tutorial outlines the procedures to configure MongoDB for encryption and key management.

Key Manager MongoDB Enterprise supports secure transfer of keys with compatible key management appliances. Using a key manager allows for the keys to be stored in the key manager.

MongoDB Enterprise supports secure transfer of keys with Key Management Interoperability Protocol (KMIP) compliant key management appliances. Any appliance vendor that provides support for KMIP is expected to be compatible.

For a list of MongoDB's certified partners, refer to the Partners List[65]. To view security partners, select "Security" from the *Technology* filter, and "Certified" from the *Certified* filter.

Recommended

Using a key manager meets regulatory key management guidelines, such as HIPAA, PCI-DSS, and FERPA, and is recommended over the local key management.

Prerequisites

- Your key manager must support the KMIP communication protocol.
- To authenticate MongoDB to a KMIP server, you must have a valid certificate issued by the key management appliance.

Encrypt Using a New Key To create a new key, connect `mongod` to the key manager by starting `mongod` with the following options:

- `--enableEncryption`,
- `--kmipServerName <KMIP Server Hostname>`,
- `--kmipServerCAFile <path to KMIP Server's CA File>`, and
- `--kmipClientCertificateFile <path to valid client certificate>`.

Include any other options specific to your configuration.

```
mongod --enableEncryption --kmipServerName <KMIP Server HostName> \
  --kmipServerCAFile ca.pem --kmipClientCertificateFile client.pem
```

This operation creates a new master key in your key manager for use by the `mongod` to wrap the keys `mongod` generates for each database.

To verify that the key creation and usage was successful, check the log file. If successful, the process will log the following messages:

```
[initandlisten] Created KMIP key with id: <UID>
[initandlisten] Encryption key manager initialized using master key with id: <UID>
```

[65] https://www.mongodb.com/partners/list

See also:

encryption-key-management-options

Encrypt Using an Existing Key You can use an existing master key created and managed by your KMIP. To use an existing key, connect `mongod` to the key manager by starting `mongod` with the following options:

- `--enableEncryption`,
- `--kmipServerName <KMIP Server Hostname`,
- `--kmipServerCAFile <path to KMIP Server's CA File>`,
- `--kmipClientCertificateFile <path to valid client certificate>`, and
- `--kmipKeyIdentifier <UID>`.

Include any other options specific to your configuration.

```
mongod --enableEncryption --kmipServerName <KMIP Server HostName> \
  --kmipServerCAFile ca.pem --kmipClientCertificateFile client.pem \
  --kmipKeyIdentifier <UID>
```

Important: If data is already encrypted with a key, you must specify that key's `<UID>` for the `--kmipKeyIdentifier` option. Otherwise, MongoDB will not start and log an error.

See also:

encryption-key-management-options

Local Key Management

Important: Using the keyfile method does not meet most regulatory key management guidelines and requires users to securely manage their own keys.

The safe management of the keyfile is critical.

To encrypt using a keyfile, you must have a base64 encoded keyfile that contains a 16 or 32 character string. The keyfile must only be accessible by the owner of the `mongod` process.

1. Create the base64 encoded keyfile with the 16 or 32 character string. You can generate the encoded keyfile using any method you prefer. For example,

   ```
   openssl rand -base64 32 > mongodb-keyfile
   ```

2. Update the file permissions.

   ```
   chmod 600 mongodb-keyfile
   ```

3. To use the key file, start `mongod` with the following options:

   - `--enableEncryption`,
   - `--encryptionKeyFile <path to keyfile>`,

   ```
   mongod --enableEncryption --encryptionKeyFile  mongodb-keyfile
   ```

4. Verify if the encryption key manager successfully initialized with the keyfile. If the operation was successful, the process will log the following message:

   ```
   [initandlisten] Encryption key manager initialized with key file: <path to keyfile>
   ```

See also:

encryption-key-management-options

Rotate Encryption Keys

On this page

- Rotate a Member of Replica Set (page 465)
- KMIP Master Key Rotation (page 465)

Most regulatory requirements mandate that a managed key used to decrypt sensitive data must be rotated out and replaced with a new key once a year.

MongoDB provides two options for key rotation. You can rotate out the binary with a new instance that uses a new key. Or, if you are using a KMIP server for key management, you can rotate the master key.

Rotate a Member of Replica Set For a replica set, to rotate out a member:

1. Start a new `mongod` instance, configured to use a new key. Include the `--replSet` option with the name of the replica set as well as any other options specific to your configuration, such as `--dbpath`.

   ```
   mongod --replSet myReplSet --enableEncryption --kmipServerName
   <KMIP Server HostName> \ --kmipServerCAFile ca.pem
   --kmipClientCertificateFile client.pem
   ```

2. Connect a `mongo` shell to the replica set's primary.

3. Add the instance to the replica set.

   ```
   rs.add("<hostname>:<port>")
   ```

   During the initial sync process, the re-encryption of the data with an entirely new set of database keys as well as a new system key occurs.

4. Once the new node completes its initial sync process, remove the old node from the replica set and delete all its data. For instructions, see *Remove Members from Replica Set* (page 682)

KMIP Master Key Rotation If you are using a KMIP server for key management, you can rotate the master key, the only externally managed key. With the new master key, the internal keystore will be re-encrypted but the database keys will be otherwise left unchanged. This obviates the need to re-encrypt the entire data set.

1. Rotate the master key for the *secondary* (page **??**) members of the replica set one at a time.

   (a) Restart the secondary, including the *`--kmipRotateMasterKey`* parameter. Include any other options specific to your configuration. If the member already includes the *`--kmipKeyIdentifier`* option, either update the *`--kmipKeyIdentifier`* option with the new key to use or omit to request a new key from the KMIP server:

   ```
   mongod --enableEncryption --kmipRotateMasterKey \
     --kmipServerName <KMIP Server HostName> \
     --kmipServerCAFile ca.pem --kmipClientCertificateFile client.pem
   ```

   If using a configuration file, include the `security.kmip.rotateMasterKey`.

   (b) Upon successful completion of the master key rotation and re-encryption of the database keystore, the `mongod` will exit.

(c) Restart the secondary without the *--kmipRotateMasterKey* parameter. Include any other options specific to your configuration.

```
mongod --enableEncryption --kmipServerName <KMIP Server HostName> \
  --kmipServerCAFile ca.pem --kmipClientCertificateFile client.pem
```

If using a configuration file, remove the `security.kmip.rotateMasterKey` setting.

2. Step down the replica set primary.

   Connect a `mongo` shell to the primary and use `rs.stepDown()` to step down the primary and force an election of a new primary:

   ```
   rs.stepDown()
   ```

3. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, rotate the master key for the stepped down member:

   (a) Restart the stepped-down member, including the *--kmipRotateMasterKey* parameter. Include any other options specific to your configuration. If the member already includes the *--kmipKeyIdentifier* option, either update the *--kmipKeyIdentifier* option with the new key to use or omit.

   ```
   mongod --enableEncryption --kmipRotateMasterKey \
     --kmipServerName <KMIP Server HostName> \
     --kmipServerCAFile ca.pem --kmipClientCertificateFile client.pem
   ```

   If using a configuration file, include the `security.kmip.rotateMasterKey`.

   (b) Upon successful completion of the master key rotation and re-encryption of the database keystore, the `mongod` will exit.

   (c) Restart the stepped-down member without the *--kmipRotateMasterKey* parameter. Include any other options specific to your configuration.

   ```
   mongod --enableEncryption --kmipServerName <KMIP Server HostName> \
     --kmipServerCAFile ca.pem --kmipClientCertificateFile client.pem
   ```

   If using a configuration file, remove the `security.kmip.rotateMasterKey` setting.

9.5 Auditing

On this page

- Enable and Configure Audit Output (page 466)
- Audit Events and Filter (page 467)
- Audit Guarantee (page 467)

New in version 2.6.

MongoDB Enterprise includes an auditing capability for `mongod` and `mongos` instances. The auditing facility allows administrators and users to track system activity for deployments with multiple users and applications.

9.5.1 Enable and Configure Audit Output

The auditing facility can write audit events to the console, the *syslog*, a JSON file, or a BSON file. To enable auditing for MongoDB Enterprise, see *Configure Auditing* (page 467).

For information on the audit log messages, see *System Event Audit Messages* (page 506).

9.5.2 Audit Events and Filter

Once enabled, the auditing system can record the following operations:

- schema (DDL),
- replica set and sharded cluster,
- authentication and authorization, and
- CRUD operations (requires `auditAuthorizationSuccess` set to `true`).

For details on audited actions, see *Audit Event Actions, Details, and Results* (page 506).

With the auditing system, you can *set up filters* (page 469) to restrict the events captured. To set up filters, see *Configure Audit Filters* (page 469).

9.5.3 Audit Guarantee

The auditing system writes every audit event [66] to an in-memory buffer of audit events. MongoDB writes this buffer to disk periodically. For events collected from any single connection, the events have a total order: if MongoDB writes one event to disk, the system guarantees that it has written all prior events for that connection to disk.

If an audit event entry corresponds to an operation that affects the durable state of the database, such as a modification to data, MongoDB will always write the audit event to disk *before* writing to the *journal* for that entry.

That is, before adding an operation to the journal, MongoDB writes all audit events on the connection that triggered the operation, up to and including the entry for the operation.

These auditing guarantees require that MongoDB run with `journaling` enabled.

Warning: MongoDB may lose events **if** the server terminates before it commits the events to the audit log. The client may receive confirmation of the event before MongoDB commits to the audit log. For example, while auditing an aggregation operation, the server might crash after returning the result but before the audit log flushes.

Configure Auditing

On this page

- Enable and Configure Audit Output (page 468)

New in version 2.6.

MongoDB Enterprise[67] supports *auditing* (page 466) of various operations. A complete auditing solution must involve **all** `mongod` server and `mongos` router processes.

The audit facility can write audit events to the console, the *syslog* (option is unavailable on Windows), a JSON file, or a BSON file. For details on the audited operations and the audit log messages, see *System Event Audit Messages* (page 506).

[66] Audit configuration can include a *filter* (page 469) to limit events to audit.
[67] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

Enable and Configure Audit Output

Use the *--auditDestination* option to enable auditing and specify where to output the audit events.

> **Warning:** For sharded clusters, if you enable auditing on `mongos` instances, you must enable auditing on all `mongod` instances in the cluster, i.e. shards and config servers.

Output to Syslog To enable auditing and print audit events to the syslog (option is unavailable on Windows) in JSON format, specify `syslog` for the *--auditDestination* setting. For example:

```
mongod --dbpath data/db --auditDestination syslog
```

> **Warning:** The syslog message limit can result in the truncation of the audit messages. The auditing system will neither detect the truncation nor error upon its occurrence.

You may also specify these options in the `configuration file`:

```
storage:
   dbPath: data/db
auditLog:
   destination: syslog
```

Output to Console To enable auditing and print the audit events to standard output (i.e. `stdout`), specify `console` for the *--auditDestination* setting. For example:

```
mongod --dbpath data/db --auditDestination console
```

You may also specify these options in the `configuration file`:

```
storage:
   dbPath: data/db
auditLog:
   destination: console
```

Output to JSON File To enable auditing and print audit events to a file in JSON format, specify `file` for the *--auditDestination* setting, `JSON` for the *--auditFormat* setting, and the output filename for the *--auditPath*. The *--auditPath* option accepts either full path name or relative path name. For example, the following enables auditing and records audit events to a file with the relative path name of `data/db/auditLog.json`:

```
mongod --dbpath data/db --auditDestination file --auditFormat JSON --auditPath data/db/auditLog.json
```

The audit file rotates at the same time as the server log file.

You may also specify these options in the `configuration file`:

```
storage:
   dbPath: data/db
auditLog:
   destination: file
   format: JSON
   path: data/db/auditLog.json
```

Note: Printing audit events to a file in JSON format degrades server performance more than printing to a file in BSON format.

Output to BSON File To enable auditing and print audit events to a file in BSON binary format, specify `file` for the *`--auditDestination`* setting, `BSON` for the *`--auditFormat`* setting, and the output filename for the *`--auditPath`*. The *`--auditPath`* option accepts either full path name or relative path name. For example, the following enables auditing and records audit events to a BSON file with the relative path name of `data/db/auditLog.bson`:

```
mongod --dbpath data/db --auditDestination file --auditFormat BSON --auditPath data/db/auditLog.bson
```

The audit file rotates at the same time as the server log file.

You may also specify these options in the `configuration file`:

```
storage:
   dbPath: data/db
auditLog:
   destination: file
   format: BSON
   path: data/db/auditLog.bson
```

To view the contents of the file, pass the file to the MongoDB utility `bsondump`. For example, the following converts the audit log into a human-readable form and output to the terminal:

```
bsondump data/db/auditLog.bson
```

See also:

Configure Audit Filters (page 469), *Auditing* (page 466), *System Event Audit Messages* (page 506)

Configure Audit Filters

On this page

- `--auditFilter` Option (page 469)
- Examples (page 470)

MongoDB Enterprise[68] supports *auditing* (page 466) of various operations. When *enabled* (page 467), the audit facility, by default, records all auditable operations as detailed in *Audit Event Actions, Details, and Results* (page 506). To specify which events to record, the audit feature includes the *`--auditFilter`* option.

`--auditFilter` Option

The *`--auditFilter`* option takes a string representation of a query document of the form:

```
{ <field1>: <expression1>, ... }
```

- The `<field>` can be *any field in the audit message* (page 506), including fields returned in the *param* (page 506) document.
- The `<expression>` is a *query condition expression*.

[68] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

To specify an audit filter, enclose the filter document in single quotes to pass the document as a string.

To specify the audit filter in a `configuration file`, you must use the YAML format of the configuration file.

Examples

Filter for Multiple Operation Types The following example audits only the `createCollection` (page 501) and `dropCollection` (page 501) actions by using the filter:

```
{ atype: { $in: [ "createCollection", "dropCollection" ] } }
```

To specify an audit filter, enclose the filter document in single quotes to pass the document as a string.

```
mongod --dbpath data/db --auditDestination file --auditFilter '{ atype: { $in: [ "createCollection",
```

To specify the audit filter in a `configuration file`, you must use the YAML format of the configuration file.

```
storage:
   dbPath: data/db
auditLog:
   destination: file
   format: BSON
   path: data/db/auditLog.bson
   filter: '{ atype: { $in: [ "createCollection", "dropCollection" ] } }'
```

Filter on Authentication Operations on a Single Database The `<field>` can include *any field in the audit message* (page 506). For authentication operations (i.e. `atype: "authenticate"`), the audit messages include a `db` field in the `param` document.

The following example audits only the `authenticate` operations that occur against the `test` database by using the filter:

```
{ atype: "authenticate", "param.db": "test" }
```

To specify an audit filter, enclose the filter document in single quotes to pass the document as a string.

```
mongod --dbpath data/db --auth --auditDestination file --auditFilter '{ atype: "authenticate", "param
```

To specify the audit filter in a `configuration file`, you must use the YAML format of the configuration file.

```
storage:
   dbPath: data/db
security:
   authorization: enabled
auditLog:
   destination: file
   format: BSON
   path: data/db/auditLog.bson
   filter: '{ atype: "authenticate", "param.db": "test" }'
```

To filter on all `authenticate` operations across databases, use the filter `{ atype: "authenticate" }`.

Filter on Collection Creation and Drop Operations for a Single Database The `<field>` can include *any field in the audit message* (page 506). For collection creation and drop operations (i.e. `atype: "createCollection"` and `atype: "dropCollection"`), the audit messages include a namespace `ns` field in the `param` document.

The following example audits only the `createCollection` and `dropCollection` operations that occur against the `test` database by using the filter:

Note: The regular expression requires two backslashes (\\) to escape the dot (.).

```
{ atype: { $in: [ "createCollection", "dropCollection" ] }, "param.ns": /^test\\./ } }
```

To specify an audit filter, enclose the filter document in single quotes to pass the document as a string.

```
mongod --dbpath data/db --auth --auditDestination file --auditFilter '{ atype: { $in: [ "createColle
```

To specify the audit filter in a `configuration file`, you must use the YAML format of the configuration file.

```
storage:
   dbPath: data/db
security:
   authorization: enabled
auditLog:
   destination: file
   format: BSON
   path: data/db/auditLog.bson
   filter: '{ atype: { $in: [ "createCollection", "dropCollection" ] }, "param.ns": /^test\\./ } }'
```

Filter by Authorization Role The following example audits operations by users with `readWrite` (page 486) role on the `test` database, including users with roles that inherit from `readWrite` (page 486), by using the filter:

```
{ roles: { role: "readWrite", db: "test" } }
```

To specify an audit filter, enclose the filter document in single quotes to pass the document as a string.

```
mongod --dbpath data/db --auth --auditDestination file --auditFilter '{ roles: { role: "readWrite",
```

To specify the audit filter in a `configuration file`, you must use the YAML format of the configuration file.

```
storage:
   dbPath: data/db
security:
   authorization: enabled
auditLog:
   destination: file
   format: BSON
   path: data/db/auditLog.bson
   filter: '{ roles: { role: "readWrite", db: "test" } }'
```

Filter on Read and Write Operations To capture `read` and `write` operations in the audit, you must also enable the audit system to log authorization successes using the `auditAuthorizationSuccess` parameter. [69]

Note: Enabling `auditAuthorizationSuccess` degrades performance more than logging only the authorization failures.

The following example audits the `find()`, `insert()`, `remove()`, `update()`, `save()`, and `findAndModify()` operations by using the filter:

```
{ atype: "authCheck", "param.command": { $in: [ "find", "insert", "delete", "update", "findandmodify"
```

To specify an audit filter, enclose the filter document in single quotes to pass the document as a string.

[69] You can enable `auditAuthorizationSuccess` parameter without enabling `--auth`; however, all operations will return success for authorization checks.

```
mongod --dbpath data/db --auth --setParameter auditAuthorizationSuccess=true --auditDestination file
```

To specify the audit filter in a `configuration file`, you must use the YAML format of the configuration file.

```
storage:
   dbPath: data/db
security:
   authorization: enabled
auditLog:
   destination: file
   format: BSON
   path: data/db/auditLog.bson
   filter: '{ atype: "authCheck", "param.command": { $in: [ "find", "insert", "delete", "update", "f
setParameter: { auditAuthorizationSuccess: true }
```

Filter on Read and Write Operations for a Collection To capture `read` and `write` operations in the audit, you must also enable the audit system to log authorization successes using the `auditAuthorizationSuccess` parameter. [1]

Note: Enabling `auditAuthorizationSuccess` degrades performance more than logging only the authorization failures.

The following example audits the `find()`, `insert()`, `remove()`, `update()`, `save()`, and `findAndModify()` operations for the collection `orders` in the database `test` by using the filter:

```
{ atype: "authCheck", "param.ns": "test.orders", "param.command": { $in: [ "find", "insert", "delete"
```

To specify an audit filter, enclose the filter document in single quotes to pass the document as a string.

```
mongod --dbpath data/db --auth --setParameter auditAuthorizationSuccess=true --auditDestination file
```

To specify the audit filter in a `configuration file`, you must use the YAML format of the configuration file.

```
storage:
   dbPath: data/db
security:
   authorization: enabled
auditLog:
   destination: file
   format: BSON
   path: data/db/auditLog.bson
   filter: '{ atype: "authCheck", "param.ns": "test.orders", "param.command": { $in: [ "find", "inse
setParameter: { auditAuthorizationSuccess: true }
```

See also:

Configure Auditing (page 467), *Auditing* (page 466), *System Event Audit Messages* (page 506)

9.6 Security Hardening

On this page

- MongoDB Configuration Hardening (page 473)
- Network Hardening (page 473)

To reduce the risk exposure of the entire MongoDB system, ensure that only trusted hosts have access to MongoDB.

9.6.1 MongoDB Configuration Hardening

For MongoDB, ensure that HTTP status interface and the REST API are disabled in production to prevent potential data exposure to attackers.

Deprecated since version 3.2: HTTP interface for MongoDB

For more information, see *MongoDB Configuration Hardening* (page 473).

9.6.2 Network Hardening

To restrict exposure to MongoDB, configure firewalls to control access to MongoDB systems. Use of VPNs can also provide a secure tunnel.

For more information, see *Hardening Network Infrastructure* (page 474).

MongoDB Configuration Hardening

On this page

- HTTP Status Interface (page 473)
- REST API (page 474)
- `bind_ip` (page 474)

HTTP Status Interface

Warning: Ensure that the HTTP status interface, the REST API, and the JSON API are all disabled in production environments to prevent potential data exposure and vulnerability to attackers.

Deprecated since version 3.2: HTTP interface for MongoDB

Changed in version 2.6: The `mongod` and `mongos` instances run with the HTTP interface *disabled* by default. See `net.http.enabled` setting.

The HTTP status interface provides a web-based interface that includes a variety of operational data, logs, and status reports regarding the `mongod` or `mongos` instance. The HTTP status interface is *disabled* by default and is not recommended for production use.

The `net.http.enabled` setting enables HTTP status interface. When enabled without the `net.http.RESTInterfaceEnabled` setting, the HTTP interface is entirely read-only and limited in scope.

The HTTP interface uses the port that is `1000` greater than the primary `mongod` port. By default, the HTTP interface port is `28017`, but is indirectly set using the `net.port` option which allows you to configure the primary `mongod` port.

The HTTP status interface does not include support for authentication other than `MONGODB-CR`.

While MongoDB Enterprise does support Kerberos authentication, Kerberos is not supported in HTTP status interface in any version of MongoDB.

Changed in version 3.0: Neither the HTTP status interface nor the REST API support the *SCRAM-SHA-1* (page 399) challenge-response user authentication mechanism introduced in version 3.0.

Warning: If you enable the interface, you should only allow trusted clients to access this port. See *Firewalls* (page 475).

REST API

Warning: Ensure that the HTTP status interface, the REST API, and the JSON API are all disabled in production environments to prevent potential data exposure and vulnerability to attackers.

The REST API to MongoDB provides additional information and write access on top of the HTTP status interface. While the REST API does not provide any support for insert, update, or remove operations, it does provide administrative access, and its accessibility represents a vulnerability in a secure environment.

Deprecated since version 3.2: HTTP interface for MongoDB

The REST interface is *disabled* by default and is not recommended for production use.

The `net.http.RESTInterfaceEnabled` setting for `mongod` enables a fully interactive administrative *REST* interface, which is *disabled* by default. Enabling the REST API enables the HTTP interface, even if the HTTP interface option is disabled, and makes the HTTP interface fully interactive.

The REST API does not include support for authentication other than `MONGODB-CR`.

Warning: If you enable the interface, you should only allow trusted clients to access this port. See *Firewalls* (page 475).

Changed in version 3.0: Neither the HTTP status interface nor the REST API support the *SCRAM-SHA-1* (page 399) challenge-response user authentication mechanism introduced in version 3.0.

bind_ip

The `net.bindIp` setting (or the `--bind_ip` command line option) for `mongod` and `mongos` instances limits the network interfaces on which MongoDB programs will listen for incoming connections.

Warning: Make sure that your `mongod` and `mongos` instances are only accessible on trusted networks. If your system has more than one network interface, bind MongoDB programs to the private or internal network interface.

See also:

Firewalls (page 475), *Security Considerations* (page 292)

Hardening Network Infrastructure

On this page

- Firewalls (page 475)
- Virtual Private Networks (page 475)

Firewalls

Firewalls allow administrators to filter and control access to a system by providing granular control over network communications. For administrators of MongoDB, the following capabilities are important: limiting incoming traffic on a specific port to specific systems and limiting incoming traffic from untrusted hosts.

On Linux systems, the `iptables` interface provides access to the underlying `netfilter` firewall. On Windows systems, `netsh` command line interface provides access to the underlying Windows Firewall. For additional information about firewall configuration, see:

- *Configure Linux iptables Firewall for MongoDB* (page 475) and
- *Configure Windows netsh Firewall for MongoDB* (page 479).

For best results and to minimize overall exposure, ensure that *only* traffic from trusted sources can reach `mongod` and `mongos` instances and that the `mongod` and `mongos` instances can only connect to trusted outputs.

See also:

For MongoDB deployments on Amazon's web services, see the Amazon EC2[70] page, which addresses Amazon's Security Groups and other EC2-specific security features.

Virtual Private Networks

Virtual private networks, or VPNs, make it possible to link two networks over an encrypted and limited-access trusted network. Typically, MongoDB users who use VPNs use TLS/SSL rather than IPSEC VPNs for performance issues.

Depending on configuration and implementation, VPNs provide for certificate validation and a choice of encryption protocols, which requires a rigorous level of authentication and identification of all clients. Furthermore, because VPNs provide a secure tunnel, by using a VPN connection to control access to your MongoDB instance, you can prevent tampering and "man-in-the-middle" attacks.

Configure Linux `iptables` Firewall for MongoDB

On this page

- Overview (page 476)
- Patterns (page 476)
- Change Default Policy to `DROP` (page 478)
- Manage and Maintain `iptables` Configuration (page 478)

On contemporary Linux systems, the `iptables` program provides methods for managing the Linux Kernel's `netfilter` or network packet filtering capabilities. These firewall rules make it possible for administrators to control what hosts can connect to the system, and limit risk exposure by limiting the hosts that can connect to a system.

This document outlines basic firewall configurations for `iptables` firewalls on Linux. Use these approaches as a starting point for your larger networking organization. For a detailed overview of security practices and risk management for MongoDB, see *Security* (page 391).

See also:

For MongoDB deployments on Amazon's web services, see the Amazon EC2[71] page, which addresses Amazon's Security Groups and other EC2-specific security features.

[70] https://docs.mongodb.org/ecosystem/platforms/amazon-ec2
[71] https://docs.mongodb.org/ecosystem/platforms/amazon-ec2

Overview Rules in `iptables` configurations fall into chains, which describe the process for filtering and processing specific streams of traffic. Chains have an order, and packets must pass through earlier rules in a chain to reach later rules. This document addresses only the following two chains:

`INPUT` Controls all incoming traffic.

`OUTPUT` Controls all outgoing traffic.

Given the `default ports` of all MongoDB processes, you must configure networking rules that permit *only* required communication between your application and the appropriate `mongod` and `mongos` instances.

Be aware that, by default, the default policy of `iptables` is to allow all connections and traffic unless explicitly disabled. The configuration changes outlined in this document will create rules that explicitly allow traffic from specific addresses and on specific ports, using a default policy that drops all traffic that is not explicitly allowed. When you have properly configured your `iptables` rules to allow only the traffic that you want to permit, you can *Change Default Policy to DROP* (page 478).

Patterns This section contains a number of patterns and examples for configuring `iptables` for use with MongoDB deployments. If you have configured different ports using the `port` configuration setting, you will need to modify the rules accordingly.

Traffic to and from `mongod` Instances This pattern is applicable to all `mongod` instances running as standalone instances or as part of a *replica set*.

The goal of this pattern is to explicitly allow traffic to the `mongod` instance from the application server. In the following examples, replace `<ip-address>` with the IP address of the application server:

```
iptables -A INPUT -s <ip-address> -p tcp --destination-port 27017 -m state --state NEW,ESTABLISHED -
iptables -A OUTPUT -d <ip-address> -p tcp --source-port 27017 -m state --state ESTABLISHED -j ACCEPT
```

The first rule allows all incoming traffic from `<ip-address>` on port `27017`, which allows the application server to connect to the `mongod` instance. The second rule, allows outgoing traffic from the `mongod` to reach the application server.

Optional

If you have only one application server, you can replace `<ip-address>` with either the IP address itself, such as: `198.51.100.55`. You can also express this using CIDR notation as `198.51.100.55/32`. If you want to permit a larger block of possible IP addresses you can allow traffic from a `/24` using one of the following specifications for the `<ip-address>`, as follows:

```
10.10.10.10/24
10.10.10.10/255.255.255.0
```

Traffic to and from `mongos` Instances `mongos` instances provide query routing for *sharded clusters*. Clients connect to `mongos` instances, which behave from the client's perspective as `mongod` instances. In turn, the `mongos` connects to all `mongod` instances that are components of the sharded cluster.

Use the same `iptables` command to allow traffic to and from these instances as you would from the `mongod` instances that are members of the replica set. Take the configuration outlined in the *Traffic to and from mongod Instances* (page 476) section as an example.

Traffic to and from a MongoDB Config Server Config servers, host the *config database* that stores metadata for sharded clusters. Each production cluster has three config servers, initiated using the *`mongod --configsvr`*

option. [72] Config servers listen for connections on port `27019`. As a result, add the following `iptables` rules to the config server to allow incoming and outgoing connection on port `27019`, for connection to the other config servers.

```
iptables -A INPUT -s <ip-address> -p tcp --destination-port 27019 -m state --state NEW,ESTABLISHED -
iptables -A OUTPUT -d <ip-address> -p tcp --source-port 27019 -m state --state ESTABLISHED -j ACCEPT
```

Replace `<ip-address>` with the address or address space of *all* the `mongod` that provide config servers.

Additionally, config servers need to allow incoming connections from all of the `mongos` instances in the cluster *and* all `mongod` instances in the cluster. Add rules that resemble the following:

```
iptables -A INPUT -s <ip-address> -p tcp --destination-port 27019 -m state --state NEW,ESTABLISHED -
```

Replace `<ip-address>` with the address of the `mongos` instances and the shard `mongod` instances.

Traffic to and from a MongoDB Shard Server For shard servers, running as *`mongod --shardsvr`* [73] Because the default port number is `27018` when running with the `shardsvr` value for the `clusterRole` setting, you must configure the following `iptables` rules to allow traffic to and from each shard:

```
iptables -A INPUT -s <ip-address> -p tcp --destination-port 27018 -m state --state NEW,ESTABLISHED -
iptables -A OUTPUT -d <ip-address> -p tcp --source-port 27018 -m state --state ESTABLISHED -j ACCEPT
```

Replace the `<ip-address>` specification with the IP address of all `mongod`. This allows you to permit incoming and outgoing traffic between all shards including constituent replica set members, to:

- all `mongod` instances in the shard's replica sets.
- all `mongod` instances in other shards. [74]

Furthermore, shards need to be able make outgoing connections to:

- all `mongod` instances in the config servers.

Create a rule that resembles the following, and replace the `<ip-address>` with the address of the config servers and the `mongos` instances:

```
iptables -A OUTPUT -d <ip-address> -p tcp --source-port 27018 -m state --state ESTABLISHED -j ACCEPT
```

Provide Access For Monitoring Systems

1. The `mongostat` diagnostic tool, when running with the *`--discover`* needs to be able to reach all components of a cluster, including the config servers, the shard servers, and the `mongos` instances.
2. If your monitoring system needs access the HTTP interface, insert the following rule to the chain:

   ```
   iptables -A INPUT -s <ip-address> -p tcp --destination-port 28017 -m state --state NEW,ESTABLISH
   ```

   Replace `<ip-address>` with the address of the instance that needs access to the HTTP or REST interface. For *all* deployments, you should restrict access to this port to *only* the monitoring instance.

   Optional

   For config server `mongod` instances running with the `shardsvr` value for the `clusterRole` setting, the rule would resemble the following:

[72] You also can run a config server by using the `configsvr` value for the `clusterRole` setting in a configuration file.

[73] You can also specify the shard server option with the `shardsvr` value for the `clusterRole` setting in the configuration file. Shard members are also often conventional replica sets using the default port.

[74] All shards in a cluster need to be able to communicate with all other shards to facilitate *chunk* and balancing operations.

```
iptables -A INPUT -s <ip-address> -p tcp --destination-port 28018 -m state --state NEW,ESTABLISH
```

For config server `mongod` instances running with the `configsvr` value for the `clusterRole` setting, the rule would resemble the following:

```
iptables -A INPUT -s <ip-address> -p tcp --destination-port 28019 -m state --state NEW,ESTABLISH
```

Change Default Policy to `DROP` The default policy for `iptables` chains is to allow all traffic. After completing all `iptables` configuration changes, you *must* change the default policy to `DROP` so that all traffic that isn't explicitly allowed as above will not be able to reach components of the MongoDB deployment. Issue the following commands to change this policy:

```
iptables -P INPUT DROP

iptables -P OUTPUT DROP
```

Manage and Maintain `iptables` Configuration This section contains a number of basic operations for managing and using `iptables`. There are various front end tools that automate some aspects of `iptables` configuration, but at the core all `iptables` front ends provide the same basic functionality:

Make all `iptables` Rules Persistent By default all `iptables` rules are only stored in memory. When your system restarts, your firewall rules will revert to their defaults. When you have tested a rule set and have guaranteed that it effectively controls traffic you can use the following operations to you should make the rule set persistent.

On Red Hat Enterprise Linux, Fedora Linux, and related distributions you can issue the following command:

```
service iptables save
```

On Debian, Ubuntu, and related distributions, you can use the following command to dump the `iptables` rules to the `/etc/iptables.conf` file:

```
iptables-save > /etc/iptables.conf
```

Run the following operation to restore the network rules:

```
iptables-restore < /etc/iptables.conf
```

Place this command in your `rc.local` file, or in the `/etc/network/if-up.d/iptables` file with other similar operations.

List all `iptables` Rules To list all of currently applied `iptables` rules, use the following operation at the system shell.

```
iptables -L
```

Flush all `iptables` Rules If you make a configuration mistake when entering `iptables` rules or simply need to revert to the default rule set, you can use the following operation at the system shell to flush all rules:

```
iptables -F
```

If you've already made your `iptables` rules persistent, you will need to repeat the appropriate procedure in the *Make all iptables Rules Persistent* (page 478) section.

Configure Windows `netsh` Firewall for MongoDB

On this page

- Overview (page 479)
- Patterns (page 479)
- Manage and Maintain *Windows Firewall* Configurations (page 481)

On Windows Server systems, the `netsh` program provides methods for managing the *Windows Firewall*. These firewall rules make it possible for administrators to control what hosts can connect to the system, and limit risk exposure by limiting the hosts that can connect to a system.

This document outlines basic *Windows Firewall* configurations. Use these approaches as a starting point for your larger networking organization. For a detailed over view of security practices and risk management for MongoDB, see *Security* (page 391).

See also:

Windows Firewall[75] documentation from Microsoft.

Overview *Windows Firewall* processes rules in an ordered determined by rule type, and parsed in the following order:

1. `Windows Service Hardening`
2. `Connection security rules`
3. `Authenticated Bypass Rules`
4. `Block Rules`
5. `Allow Rules`
6. `Default Rules`

By default, the policy in *Windows Firewall* allows all outbound connections and blocks all incoming connections.

Given the `default ports` of all MongoDB processes, you must configure networking rules that permit *only* required communication between your application and the appropriate `mongod.exe` and `mongos.exe` instances.

The configuration changes outlined in this document will create rules which explicitly allow traffic from specific addresses and on specific ports, using a default policy that drops all traffic that is not explicitly allowed.

You can configure the *Windows Firewall* with using the `netsh` command line tool or through a windows application. On Windows Server 2008 this application is *Windows Firewall With Advanced Security* in *Administrative Tools*. On previous versions of Windows Server, access the *Windows Firewall* application in the *System and Security* control panel.

The procedures in this document use the `netsh` command line tool.

Patterns This section contains a number of patterns and examples for configuring *Windows Firewall* for use with MongoDB deployments. If you have configured different ports using the `port` configuration setting, you will need to modify the rules accordingly.

Traffic to and from `mongod.exe` Instances This pattern is applicable to all `mongod.exe` instances running as standalone instances or as part of a *replica set*. The goal of this pattern is to explicitly allow traffic to the `mongod.exe` instance from the application server.

[75] http://technet.microsoft.com/en-us/network/bb545423.aspx

```
netsh advfirewall firewall add rule name="Open mongod port 27017" dir=in action=allow protocol=TCP lo
```

This rule allows all incoming traffic to port `27017`, which allows the application server to connect to the `mongod.exe` instance.

Windows Firewall also allows enabling network access for an entire application rather than to a specific port, as in the following example:

```
netsh advfirewall firewall add rule name="Allowing mongod" dir=in action=allow program=" C:\mongodb\b
```

You can allow all access for a `mongos.exe` server, with the following invocation:

```
netsh advfirewall firewall add rule name="Allowing mongos" dir=in action=allow program=" C:\mongodb\b
```

Traffic to and from `mongos.exe` Instances `mongos.exe` instances provide query routing for *sharded clusters*. Clients connect to `mongos.exe` instances, which behave from the client's perspective as `mongod.exe` instances. In turn, the `mongos.exe` connects to all `mongod.exe` instances that are components of the sharded cluster.

Use the same *Windows Firewall* command to allow traffic to and from these instances as you would from the `mongod.exe` instances that are members of the replica set.

```
netsh advfirewall firewall add rule name="Open mongod shard port 27018" dir=in action=allow protocol=
```

Traffic to and from a MongoDB Config Server Configuration servers, host the *config database* that stores metadata for sharded clusters. Each production cluster has three configuration servers, initiated using the *`mongod --configsvr`* option. [76] Configuration servers listen for connections on port `27019`. As a result, add the following *Windows Firewall* rules to the config server to allow incoming and outgoing connection on port `27019`, for connection to the other config servers.

```
netsh advfirewall firewall add rule name="Open mongod config svr port 27019" dir=in action=allow prot
```

Additionally, config servers need to allow incoming connections from all of the `mongos.exe` instances in the cluster *and* all `mongod.exe` instances in the cluster. Add rules that resemble the following:

```
netsh advfirewall firewall add rule name="Open mongod config svr inbound" dir=in action=allow protoco
```

Replace `<ip-address>` with the addresses of the `mongos.exe` instances and the shard `mongod.exe` instances.

Traffic to and from a MongoDB Shard Server For shard servers, running as *`mongod --shardsvr`* [77] Because the default port number is `27018` when running with the `shardsvr` value for the `clusterRole` setting, you must configure the following *Windows Firewall* rules to allow traffic to and from each shard:

```
netsh advfirewall firewall add rule name="Open mongod shardsvr inbound" dir=in action=allow protocol=
netsh advfirewall firewall add rule name="Open mongod shardsvr outbound" dir=out action=allow protoco
```

Replace the `<ip-address>` specification with the IP address of all `mongod.exe` instances. This allows you to permit incoming and outgoing traffic between all shards including constituent replica set members to:

- all `mongod.exe` instances in the shard's replica sets.
- all `mongod.exe` instances in other shards. [78]

Furthermore, shards need to be able make outgoing connections to:

[76] You also can run a config server by using the `configsrv` value for the `clusterRole` setting in a configuration file.

[77] You can also specify the shard server option with the `shardsvr` value for the `clusterRole` setting in the configuration file. Shard members are also often conventional replica sets using the default port.

[78] All shards in a cluster need to be able to communicate with all other shards to facilitate *chunk* and balancing operations.

- all `mongos.exe` instances.
- all `mongod.exe` instances in the config servers.

Create a rule that resembles the following, and replace the `<ip-address>` with the address of the config servers and the `mongos.exe` instances:

```
netsh advfirewall firewall add rule name="Open mongod config svr outbound" dir=out action=allow proto
```

Provide Access For Monitoring Systems

1. The `mongostat` diagnostic tool, when running with the `--discover` needs to be able to reach all components of a cluster, including the config servers, the shard servers, and the `mongos.exe` instances.
2. If your monitoring system needs access the HTTP interface, insert the following rule to the chain:

   ```
   netsh advfirewall firewall add rule name="Open mongod HTTP monitoring inbound" dir=in action=all
   ```

   Replace `<ip-address>` with the address of the instance that needs access to the HTTP or REST interface. For *all* deployments, you should restrict access to this port to *only* the monitoring instance.

   Optional

   For config server `mongod` instances running with the `shardsvr` value for the `clusterRole` setting, the rule would resemble the following:

   ```
   netsh advfirewall firewall add rule name="Open mongos HTTP monitoring inbound" dir=in action=all
   ```

   For config server `mongod` instances running with the `configsvr` value for the `clusterRole` setting, the rule would resemble the following:

   ```
   netsh advfirewall firewall add rule name="Open mongod configsvr HTTP monitoring inbound" dir=in
   ```

Manage and Maintain *Windows Firewall* Configurations This section contains a number of basic operations for managing and using `netsh`. While you can use the GUI front ends to manage the *Windows Firewall*, all core functionality is accessible is accessible from `netsh`.

Delete all *Windows Firewall* Rules To delete the firewall rule allowing `mongod.exe` traffic:

```
netsh advfirewall firewall delete rule name="Open mongod port 27017" protocol=tcp localport=27017
```

```
netsh advfirewall firewall delete rule name="Open mongod shard port 27018" protocol=tcp localport=270
```

List All *Windows Firewall* Rules To return a list of all *Windows Firewall* rules:

```
netsh advfirewall firewall show rule name=all
```

Reset *Windows Firewall* To reset the *Windows Firewall* rules:

```
netsh advfirewall reset
```

Backup and Restore *Windows Firewall* Rules To simplify administration of larger collection of systems, you can export or import firewall systems from different servers) rules very easily on Windows:

Export all firewall rules with the following command:

```
netsh advfirewall export "C:\temp\MongoDBfw.wfw"
```

Replace `"C:\temp\MongoDBfw.wfw"` with a path of your choosing. You can use a command in the following form to import a file created using this operation:

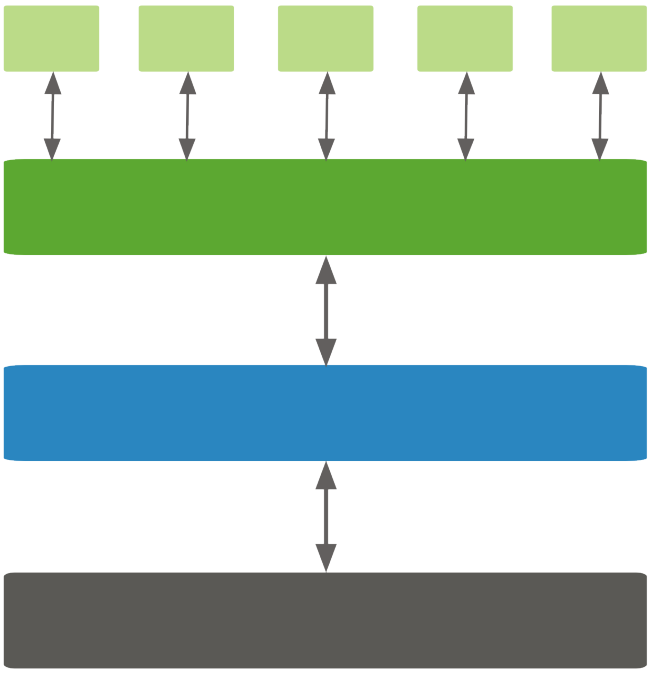
```
netsh advfirewall import "C:\temp\MongoDBfw.wfw"
```

9.7 Implement Field Level Redaction

On this page

- Procedure (page 483)

The `$redact` pipeline operator restricts the contents of the documents based on information stored in the documents themselves.



To store the access criteria data, add a field to the documents and embedded documents. To allow for multiple combinations of access levels for the same data, consider setting the access field to an array of arrays. Each array element contains a required set that allows a user with that set to access the data.

Then, include the `$redact` stage in the `db.collection.aggregate()` operation to restrict contents of the result set based on the access required to view the data.

For more information on the `$redact` pipeline operator, including its syntax and associated system variables as well as additional examples, see `$redact`.

9.7.1 Procedure

For example, a `forecasts` collection contains documents of the following form where the `tags` field determines the access levels required to view the data:

```
{
   _id: 1,
   title: "123 Department Report",
   tags: [ [ "G" ], [ "FDW" ] ],
   year: 2014,
   subsections: [
       {
           subtitle: "Section 1: Overview",
           tags: [ [ "SI", "G" ], [ "FDW" ] ],
           content:  "Section 1: This is the content of section 1."
       },
       {
           subtitle: "Section 2: Analysis",
           tags: [ [ "STLW" ] ],
           content: "Section 2: This is the content of section 2."
       },
       {
           subtitle: "Section 3: Budgeting",
           tags: [ [ "TK" ], [ "FDW", "TGE" ] ],
           content: {
               text: "Section 3: This is the content of section3.",
               tags: [ [ "HCS"], [ "FDW", "TGE", "BX" ] ]
           }
       }
   ]
}
```

For each document, the `tags` field contains various access groupings necessary to view the data. For example, the value `[ [ "G" ], [ "FDW", "TGE" ] ]` can specify that a user requires either access level `["G"]` or both `[ "FDW", "TGE" ]` to view the data.

Consider a user who only has access to view information tagged with either `"FDW"` or `"TGE"`. To run a query on all documents with year `2014` for this user, include a `$redact` stage as in the following:

```
var userAccess = [ "FDW", "TGE" ];
db.forecasts.aggregate(
   [
     { $match: { year: 2014 } },
     { $redact:
        {
          $cond: {
                   if: { $anyElementTrue:
                          {
                            $map: {
                                    input: "$tags" ,
                                    as: "fieldTag",
                                    in: { $setIsSubset: [ "$$fieldTag", userAccess ] }
                                  }
                          }
                       },
```

```
                    then: "$$DESCEND",
                    else: "$$PRUNE"
                }
            }
        }
    ]
)
```

The aggregation operation returns the following "redacted" document for the user:

```
{ "_id" : 1,
  "title" : "123 Department Report",
  "tags" : [ [ "G" ], [ "FDW" ] ],
  "year" : 2014,
  "subsections" :
     [
        {
          "subtitle" : "Section 1: Overview",
          "tags" : [ [ "SI", "G" ], [ "FDW" ] ],
          "content" : "Section 1: This is the content of section 1."
        },
       {
         "subtitle" : "Section 3: Budgeting",
         "tags" : [ [ "TK" ], [ "FDW", "TGE" ] ]
       }
     ]
}
```

See also:

`$map`, `$setIsSubset`, `$anyElementTrue`

9.8 Security Reference

On this page

- Security Methods in the `mongo` Shell (page 485)
- Security Reference Documentation (page 485)

The following lists the security related methods available in the `mongo` shell as well as additional *security reference material* (page 485).

9.8.1 Security Methods in the mongo Shell

User Management and Authentication Methods

Name	Description
db.auth()	Authenticates a user to a database.
db.createUser()	Creates a new user.
db.updateUser()	Updates user data.
db.changeUserPassword()	Changes an existing user's password.
db.removeUser()	Deprecated. Removes a user from a database.
db.dropAllUsers()	Deletes all users associated with a database.
db.dropUser()	Removes a single user.
db.grantRolesToUser()	Grants a role and its privileges to a user.
db.revokeRolesFromUser()	Removes a role from a user.
db.getUser()	Returns information about the specified user.
db.getUsers()	Returns information about all users associated with a database.

Role Management Methods

Name	Description
db.createRole()	Creates a role and specifies its privileges.
db.updateRole()	Updates a user-defined role.
db.dropRole()	Deletes a user-defined role.
db.dropAllRoles()	Deletes all user-defined roles associated with a database.
db.grantPrivilegesToRole()	Assigns privileges to a user-defined role.
db.revokePrivilegesFromRole()	Removes the specified privileges from a user-defined role.
db.grantRolesToRole()	Specifies roles from which a user-defined role inherits privileges.
db.revokeRolesFromRole()	Removes inherited roles from a role.
db.getRole()	Returns information for the specified role.
db.getRoles()	Returns information for all the user-defined roles in a database.

9.8.2 Security Reference Documentation

Built-In Roles **(page 485)** Reference on MongoDB provided roles and corresponding access.

system.roles Collection **(page 494)** Describes the content of the collection that stores user-defined roles.

system.users Collection **(page 497)** Describes the content of the collection that stores users' credentials and role assignments.

Resource Document **(page 498)** Describes the resource document for roles.

Privilege Actions **(page 500)** List of the actions available for privileges.

System Event Audit Messages **(page 506)** Reference on system event audit messages.

Built-In Roles

On this page

- Database User Roles (page 486)
- Database Administration Roles (page 487)
- Cluster Administration Roles (page 488)
- Backup and Restoration Roles (page 491)
- All-Database Roles (page 493)
- Superuser Roles (page 493)
- Internal Role (page 494)

MongoDB grants access to data and commands through *role-based authorization* (page 434) and provides built-in roles that provide the different levels of access commonly needed in a database system. You can additionally create *user-defined roles* (page 440).

A role grants privileges to perform sets of *actions* (page 500) on defined *resources* (page 498). A given role applies to the database on which it is defined and can grant access down to a collection level of granularity.

Each of MongoDB's built-in roles defines access at the database level for all *non*-system collections in the role's database and at the collection level for all *system collections* (page 376).

MongoDB provides the built-in *database user* (page 486) and *database administration* (page 487) roles on *every* database. MongoDB provides all other built-in roles only on the `admin` database.

This section describes the privileges for each built-in role. You can also view the privileges for a built-in role at any time by issuing the `rolesInfo` command with the `showPrivileges` and `showBuiltinRoles` fields both set to `true`.

Database User Roles

Every database includes the following client roles:

`read`
: Provides the ability to read data on all *non*-system collections and on the following system collections: `system.indexes` (page 377), `system.js` (page 377), and `system.namespaces` (page 377) collections. The role provides read access by granting the following *actions* (page 500):

  - `collStats` (page 505)
  - `dbHash` (page 505)
  - `dbStats` (page 505)
  - `find` (page 501)
  - `killCursors` (page 502)
  - `listIndexes` (page 505)
  - `listCollections` (page 505)

`readWrite`
: Provides all the privileges of the `read` (page 486) role plus ability to modify data on all *non*-system collections and the `system.js` (page 377) collection. The role provides the following actions on those collections:

  - `collStats` (page 505)
  - `convertToCapped` (page 504)
  - `createCollection` (page 501)

- dbHash (page 505)
- dbStats (page 505)
- dropCollection (page 501)
- createIndex (page 501)
- dropIndex (page 504)
- emptycapped (page 502)
- find (page 501)
- insert (page 501)
- killCursors (page 502)
- listIndexes (page 505)
- listCollections (page 505)
- remove (page 501)
- renameCollectionSameDB (page 504)
- update (page 501)

Database Administration Roles

Every database includes the following database administration roles:

dbAdmin

Provides the following *actions* (page 500) on the database's system.indexes (page 377), system.namespaces (page 377), and system.profile (page 377) collections:

- collStats (page 505)
- dbHash (page 505)
- dbStats (page 505)
- find (page 501)
- killCursors (page 502)
- listIndexes (page 505)
- listCollections (page 505)
- dropCollection (page 501) and createCollection (page 501) on system.profile (page 377) *only*

Changed in version 2.6.4: dbAdmin (page 487) added the createCollection (page 501) for the system.profile (page 377) collection. Previous versions only had the dropCollection (page 501) on the system.profile (page 377) collection.

Provides the following actions on all *non*-system collections. This role *does not* include full read access on non-system collections:

- bypassDocumentValidation (page 501)
- collMod (page 504)
- collStats (page 505)
- compact (page 504)

- convertToCapped (page 504)
- createCollection (page 501)
- createIndex (page 501)
- dbStats (page 505)
- dropCollection (page 501)
- dropDatabase (page 504)
- dropIndex (page 504)
- enableProfiler (page 502)
- indexStats (page 505)
- reIndex (page 504)
- renameCollectionSameDB (page 504)
- repairDatabase (page 504)
- storageDetails (page 502)
- validate (page 505)

dbOwner

The database owner can perform any administrative action on the database. This role combines the privileges granted by the readWrite (page 486), dbAdmin (page 487) and userAdmin (page 488) roles.

userAdmin

Provides the ability to create and modify roles and users on the current database. This role also indirectly provides *superuser* (page 493) access to either the database or, if scoped to the admin database, the cluster. The userAdmin (page 488) role allows users to grant any user any privilege, including themselves.

The userAdmin (page 488) role explicitly provides the following actions:

- changeCustomData (page 501)
- changePassword (page 501)
- createRole (page 501)
- createUser (page 501)
- dropRole (page 501)
- dropUser (page 502)
- grantRole (page 502)
- revokeRole (page 502)
- viewRole (page 502)
- viewUser (page 502)

Cluster Administration Roles

The admin database includes the following roles for administering the whole system rather than just a single database. These roles include but are not limited to *replica set* and *sharded cluster* administrative functions.

clusterAdmin

Provides the greatest cluster-management access. This role combines the privileges granted by the clusterManager (page 489), clusterMonitor (page 490), and hostManager (page 490) roles. Additionally, the role provides the dropDatabase (page 504) action.

clusterManager

Provides management and monitoring actions on the cluster. A user with this role can access the config and local databases, which are used in sharding and replication, respectively.

Provides the following actions on the cluster as a whole:

- addShard (page 503)
- applicationMessage (page 504)
- cleanupOrphaned (page 502)
- flushRouterConfig (page 503)
- listShards (page 503)
- removeShard (page 503)
- replSetConfigure (page 503)
- replSetGetStatus (page 503)
- replSetStateChange (page 503)
- resync (page 503)

Provides the following actions on *all* databases in the cluster:

- enableSharding (page 503)
- moveChunk (page 503)
- splitChunk (page 503)
- splitVector (page 503)

On the config database, provides the following actions on the settings (page 828) collection:

- insert (page 501)
- remove (page 501)
- update (page 501)

On the config database, provides the following actions on all configuration collections and on the system.indexes (page 377), system.js (page 377), and system.namespaces (page 377) collections:

- collStats (page 505)
- dbHash (page 505)
- dbStats (page 505)
- find (page 501)
- killCursors (page 502)

On the local database, provides the following actions on the replset (page 724) collection:

- collStats (page 505)
- dbHash (page 505)

- dbStats (page 505)
- find (page 501)
- killCursors (page 502)

clusterMonitor

Provides read-only access to monitoring tools, such as the MongoDB Cloud Manager[79] and Ops Manager[80] monitoring agent.

Provides the following actions on the cluster as a whole:

- connPoolStats (page 505)
- cursorInfo (page 505)
- getCmdLineOpts (page 505)
- getLog (page 505)
- getParameter (page 504)
- getShardMap (page 503)
- hostInfo (page 504)
- inprog (page 502)
- listDatabases (page 505)
- listShards (page 503)
- netstat (page 505)
- replSetGetStatus (page 503)
- serverStatus (page 505)
- shardingState (page 503)
- top (page 505)

Provides the following actions on *all* databases in the cluster:

- collStats (page 505)
- dbStats (page 505)
- getShardVersion (page 503)

Provides the find (page 501) action on all system.profile (page 377) collections in the cluster.

Provides the following actions on the config database's configuration collections and system.indexes (page 377), system.js (page 377), and system.namespaces (page 377) collections:

- collStats (page 505)
- dbHash (page 505)
- dbStats (page 505)
- find (page 501)
- killCursors (page 502)

[79] https://cloud.mongodb.com/?jmp=docs
[80] https://docs.opsmanager.mongodb.com/current/

hostManager
Provides the ability to monitor and manage servers.

Provides the following actions on the cluster as a whole:

- applicationMessage (page 504)
- closeAllDatabases (page 504)
- connPoolSync (page 504)
- cpuProfiler (page 502)
- diagLogging (page 505)
- flushRouterConfig (page 503)
- fsync (page 504)
- invalidateUserCache (page 502)
- killop (page 502)
- logRotate (page 504)
- resync (page 503)
- setParameter (page 504)
- shutdown (page 504)
- touch (page 504)
- unlock (page 502)

Provides the following actions on *all* databases in the cluster:

- killCursors (page 502)
- repairDatabase (page 504)

Backup and Restoration Roles

The admin database includes the following roles for backing up and restoring data:

backup
Provides minimal privileges needed for backing up data. This role provides sufficient privileges to use the MongoDB Cloud Manager[81] backup agent, Ops Manager[82] backup agent, or to use mongodump to back up an entire mongod instance.

Provides the following *actions* (page 500) on the mms.backup collection in the admin database:

- insert (page 501)
- update (page 501)

Provides the listDatabases (page 505) action on the cluster as a whole.

Provides the listCollections (page 505) action on all databases.

Provides the listIndexes (page 505) action for all collections.

Provides the bypassDocumentValidation (page 501) action for collections that have *document validation* (page 250).

[81] https://cloud.mongodb.com/?jmp=docs
[82] https://docs.opsmanager.mongodb.com/current/

Provides the find (page 501) action on the following:

- all *non*-system collections in the cluster
- all the following system collections in the cluster: system.indexes (page 377), system.namespaces (page 377), and system.js (page 377)
- the admin.system.users (page 377) and admin.system.roles (page 377) collections
- legacy system.users collections from versions of MongoDB prior to 2.6

Changed in version 3.2.1: The backup (page 491) role provides additional privileges to back up the system.profile (page 377) collections that exist when running with *database profiling* (page 312). Previously, users required an additional read access on this collection.

restore

Provides privileges needed to restore data from backups that do not include system.profile (page 377) collection data. This role is sufficient when restoring data with mongorestore without the *--oplogReplay* option.

- If the backup data includes system.profile (page 377) collection data and the target database does not contain the system.profile (page 377) collection, mongorestore attempts to create the collection even though the program does not actually restore system.profile documents. As such, the user requires additional privileges to perform createCollection (page 501) and convertToCapped (page 504) actions on the system.profile (page 377) collection for a database.

  The built-in roles dbAdmin (page 487) and dbAdminAnyDatabase (page 493) provide the additional privileges.

- If running mongorestore with *--oplogReplay*, the restore (page 492) role is insufficient to replay the oplog. To replay the oplog, create a *user-defined role* (page 442) that has anyAction (page 505) on *anyResource* (page 500) and grant only to users who must run mongorestore with *--oplogReplay*.

Provides the following actions on all *non*-system collections and system.js (page 377) collections in the cluster; on the admin.system.users (page 377) and admin.system.roles (page 377) collections in the admin database; and on legacy system.users collections from versions of MongoDB prior to 2.6:

- collMod (page 504)
- createCollection (page 501)
- createIndex (page 501)
- dropCollection (page 501)
- insert (page 501)

Provides the listCollections (page 505) action on all databases.

Provides the following *additional* actions on admin.system.users (page 377) and legacy system.users collections:

- find (page 501)
- remove (page 501)
- update (page 501)

Provides the find (page 501) action on all the system.namespaces (page 377) collections in the cluster.

Although, restore (page 492) includes the ability to modify the documents in the admin.system.users (page 377) collection using normal modification operations, *only* modify these data using the *user management methods*.

All-Database Roles

The admin database provides the following roles that apply to all databases in a mongod instance and are roughly equivalent to their single-database equivalents:

readAnyDatabase
Provides the same read-only permissions as read (page 486), except it applies to *all* databases in the cluster. The role also provides the listDatabases (page 505) action on the cluster as a whole.

readWriteAnyDatabase
Provides the same read and write permissions as readWrite (page 486), except it applies to *all* databases in the cluster. The role also provides the listDatabases (page 505) action on the cluster as a whole.

userAdminAnyDatabase
Provides the same access to user administration operations as userAdmin (page 488), except it applies to *all* databases in the cluster. The role also provides the following actions on the cluster as a whole:

- authSchemaUpgrade (page 502)
- invalidateUserCache (page 502)
- listDatabases (page 505)

The role also provides the following actions on the admin.system.users (page 377) and admin.system.roles (page 377) collections on the admin database, and on legacy system.users collections from versions of MongoDB prior to 2.6:

- collStats (page 505)
- dbHash (page 505)
- dbStats (page 505)
- find (page 501)
- killCursors (page 502)
- planCacheRead (page 502)

Changed in version 2.6.4: userAdminAnyDatabase (page 493) added the following permissions on the admin.system.users (page 377) and admin.system.roles (page 377) collections:

- createIndex (page 501)
- dropIndex (page 504)

The userAdminAnyDatabase (page 493) role does not restrict the permissions that a user can grant. As a result, userAdminAnyDatabase (page 493) users can grant themselves privileges in excess of their current privileges and even can grant themselves *all privileges*, even though the role does not explicitly authorize privileges beyond user administration. This role is effectively a MongoDB system *superuser* (page 493).

dbAdminAnyDatabase
Provides the same access to database administration operations as dbAdmin (page 487), except it applies to *all* databases in the cluster. The role also provides the listDatabases (page 505) action on the cluster as a whole.

Superuser Roles

Several roles provide either indirect or direct system-wide superuser access.

The following roles provide the ability to assign any user any privilege on any database, which means that users with one of these roles can assign *themselves* any privilege on any database:

- dbOwner (page 488) role, when scoped to the admin database
- userAdmin (page 488) role, when scoped to the admin database
- userAdminAnyDatabase (page 493) role

The following role provides full privileges on all resources:

root
Provides access to the operations and all the resources of the readWriteAnyDatabase (page 493), dbAdminAnyDatabase (page 493), userAdminAnyDatabase (page 493), clusterAdmin (page 488) roles, restore (page 492) *combined*.

Changed in version 3.0.7: The root (page 494) has validate (page 505) action on system. collections. Previously, root (page 494) does **not** include any access to collections that begin with the system. prefix.

The root (page 494) role includes privileges from the restore (page 492) role.

Internal Role

__system
MongoDB assigns this role to user objects that represent cluster members, such as replica set members and mongos instances. The role entitles its holder to take any action against any object in the database.

Do not assign this role to user objects representing applications or human administrators, other than in exceptional circumstances.

If you need access to all actions on all resources, for example to run applyOps commands, do not assign this role. Instead, *create a user-defined role* (page 442) that grants anyAction (page 505) on *anyResource* (page 500) and ensure that only the users who need access to these operations have this access.

system.roles Collection

New in version 2.6.

On this page

- system.roles Schema (page 494)
- Examples (page 496)

The system.roles collection in the admin database stores the user-defined roles. To create and manage these user-defined roles, MongoDB provides *role management commands*.

system.roles Schema

The documents in the system.roles collection have the following schema:

```
{
  _id: <system-defined id>,
  role: "<role name>",
  db: "<database>",
  privileges:
      [
          {
              resource: { <resource> },
              actions: [ "<action>", ... ]
```

```
            },
            ...
        ],
    roles:
        [
            { role: "<role name>", db: "<database>" },
            ...
        ]
}
```

A `system.roles` document has the following fields:

`admin.system.roles.`**`role`**
The `role` (page 495) field is a string that specifies the name of the role.

`admin.system.roles.`**`db`**
The `db` (page 495) field is a string that specifies the database to which the role belongs. MongoDB uniquely identifies each role by the pairing of its name (i.e. `role` (page 495)) and its database.

`admin.system.roles.`**`privileges`**
The `privileges` (page 495) array contains the privilege documents that define the *privileges* (page 434) for the role.

A privilege document has the following syntax:

```
{
  resource: { <resource> },
  actions: [ "<action>", ... ]
}
```

Each privilege document has the following fields:

`admin.system.roles.privileges[n].`**`resource`**
A document that specifies the resources upon which the privilege `actions` (page 495) apply. The document has one of the following form:

```
{ db: <database>, collection: <collection> }
```

or

```
{ cluster : true }
```

See *Resource Document* (page 498) for more details.

`admin.system.roles.privileges[n].`**`actions`**
An array of actions permitted on the resource. For a list of actions, see *Privilege Actions* (page 500).

`admin.system.roles.`**`roles`**
The `roles` (page 495) array contains role documents that specify the roles from which this role *inherits* (page 434) privileges.

A role document has the following syntax:

```
{ role: "<role name>", db: "<database>" }
```

A role document has the following fields:

`admin.system.roles.roles[n].`**`role`**
The name of the role. A role can be a *built-in role* (page 485) provided by MongoDB or a *user-defined role* (page 440).

admin.system.roles.roles[n].**db**
The name of the database where the role is defined.

Examples

Consider the following sample documents found in `system.roles` collection of the `admin` database.

A User-Defined Role Specifies Privileges The following is a sample document for a user-defined role `appUser` defined for the `myApp` database:

```
{
  _id: "myApp.appUser",
  role: "appUser",
  db: "myApp",
  privileges: [
       { resource: { db: "myApp" , collection: "" },
         actions: [ "find", "createCollection", "dbStats", "collStats" ] },
       { resource: { db: "myApp", collection: "logs" },
         actions: [ "insert" ] },
       { resource: { db: "myApp", collection: "data" },
         actions: [ "insert", "update", "remove", "compact" ] },
       { resource: { db: "myApp", collection: "system.js" },
         actions: [ "find" ] },
  ],
  roles: []
}
```

The `privileges` array lists the five privileges that the `appUser` role specifies:

- The first privilege permits its actions (`"find"`, `"createCollection"`, `"dbStats"`, `"collStats"`) on all the collections in the `myApp` database *excluding* its system collections. See *Specify a Database as Resource* (page 499).
- The next two privileges permits *additional* actions on specific collections, `logs` and `data`, in the `myApp` database. See *Specify a Collection of a Database as Resource* (page 499).
- The last privilege permits actions on one *system collections* (page 376) in the `myApp` database. While the first privilege gives database-wide permission for the `find` action, the action does not apply to `myApp`'s system collections. To give access to a system collection, a privilege must explicitly specify the collection. See *Resource Document* (page 498).

As indicated by the empty `roles` array, `appUser` inherits no additional privileges from other roles.

User-Defined Role Inherits from Other Roles The following is a sample document for a user-defined role `appAdmin` defined for the `myApp` database: The document shows that the `appAdmin` role specifies privileges as well as inherits privileges from other roles:

```
{
  _id: "myApp.appAdmin",
  role: "appAdmin",
  db: "myApp",
  privileges: [
                {
                  resource: { db: "myApp", collection: "" },
                  actions: [ "insert", "dbStats", "collStats", "compact", "repairDatabase" ]
                }
              ],
```

```
  roles: [
           { role: "appUser", db: "myApp" }
         ]
}
```

The `privileges` array lists the privileges that the `appAdmin` role specifies. This role has a single privilege that permits its actions (`"insert"`, `"dbStats"`, `"collStats"`, `"compact"`, `"repairDatabase"`) on all the collections in the `myApp` database *excluding* its system collections. See *Specify a Database as Resource* (page 499).

The `roles` array lists the roles, identified by the role names and databases, from which the role `appAdmin` inherits privileges.

`system.users` Collection

Changed in version 2.6.

On this page

- `system.users` Schema (page 497)
- Example (page 498)

The `system.users` collection in the `admin` database stores user *authentication* (page 393) and *authorization* (page 433) information. To manage data in this collection, MongoDB provides *user management commands*.

`system.users` Schema

The documents in the `system.users` collection have the following schema:

```
{
  _id: <system defined id>,
  user: "<name>",
  db: "<database>",
  credentials: { <authentication credentials> },
  roles: [
           { role: "<role name>", db: "<database>" },
           ...
         ],
  customData: <custom information>
 }
```

Each `system.users` document has the following fields:

`admin.system.users.`**`user`**
: The `user` (page 497) field is a string that identifies the user. A user exists in the context of a single logical database but can have access to other databases through roles specified in the `roles` (page 497) array.

`admin.system.users.`**`db`**
: The `db` (page 497) field specifies the database associated with the user. The user's privileges are not necessarily limited to this database. The user can have privileges in additional databases through the `roles` (page 497) array.

`admin.system.users.`**`credentials`**
: The `credentials` (page 497) field contains the user's authentication information. For users with externally stored authentication credentials, such as users that use *Kerberos* (page 409) or x.509 certificates for authentication, the `system.users` document for that user does not contain the `credentials` (page 497) field.

admin.system.users.**roles**
: The roles (page 497) array contains role documents that specify the roles granted to the user. The array contains both *built-in roles* (page 485) and *user-defined role* (page 440).

  A role document has the following syntax:

  ```
  { role: "<role name>", db: "<database>" }
  ```

  A role document has the following fields:

  admin.system.users.roles[n].**role**
  : The name of a role. A role can be a *built-in role* (page 485) provided by MongoDB or a *custom user-defined role* (page 440).

  admin.system.users.roles[n].**db**
  : The name of the database where role is defined.

  When specifying a role using the *role management* or *user management* commands, you can specify the role name alone (e.g. "readWrite") if the role that exists on the database on which the command is run.

admin.system.users.**customData**
: The customData (page 498) field contains optional custom information about the user.

Example

Changed in version 3.0.0.

Consider the following document in the system.users collection:

```
{
  _id : "home.Kari",
  user : "Kari",
  db : "home",
  credentials : {
         "SCRAM-SHA-1" : {
                 "iterationCount" : 10000,
                 "salt" : nkHYXEZTTYmn+hrY994y1Q==",
                 "storedKey" : "wxWGN3ElQ25WbPjACeXdUmN4nNo=",
                 "serverKey" : "h7vBq5tACT/BtrIElY2QTm+pQzM="
         }
  },
  roles : [
             { role: "read", db: "home" },
             { role: "readWrite", db: "test" },
             { role: "appUser", db: "myApp" }
           ],
  customData : { zipCode: "64157" }
}
```

The document shows that a user Kari is associated with the home database. Kari has the read (page 486) role in the home database, the readWrite (page 486) role in the test database, and the appUser role in the myApp database.

Resource Document

On this page

- Database and/or Collection Resource (page 499)
- Cluster Resource (page 500)
- anyResource (page 500)

The resource document specifies the resources upon which a privilege permits `actions`.

Database and/or Collection Resource

To specify databases and/or collections, use the following syntax:

```
{ db: <database>, collection: <collection> }
```

Specify a Collection of a Database as Resource If the resource document species both the `db` and `collection` fields as non-empty strings, the resource is the specified collection in the specified database. For example, the following document specifies a resource of the `inventory` collection in the `products` database:

```
{ db: "products", collection: "inventory" }
```

For a user-defined role scoped for a non-`admin` database, the resource specification for its privileges must specify the same database as the role. User-defined roles scoped for the `admin` database can specify other databases.

Specify a Database as Resource If only the `collection` field is an empty string (`""`), the resource is the specified database, excluding the *system collections* (page 376). For example, the following resource document specifies the resource of the `test` database, excluding the system collections:

```
{ db: "test", collection: "" }
```

For a user-defined role scoped for a non-`admin` database, the resource specification for its privileges must specify the same database as the role. User-defined roles scoped for the `admin` database can specify other databases.

Note: When you specify a database as the resource, system collections are excluded, unless you name them explicitly, as in the following:

```
{ db: "test", collection: "system.js" }
```

System collections include but are not limited to the following:

- `<database>.system.profile` (page 377)
- `<database>.system.js` (page 377)
- *system.users Collection* (page 497) in the `admin` database
- *system.roles Collection* (page 494) in the `admin` database

Specify Collections Across Databases as Resource If only the `db` field is an empty string (`""`), the resource is all collections with the specified name across all databases. For example, the following document specifies the resource of all the `accounts` collections across all the databases:

```
{ db: "", collection: "accounts" }
```

For user-defined roles, only roles scoped for the `admin` database can have this resource specification for their privileges.

Specify All Non-System Collections in All Databases If both the `db` and `collection` fields are empty strings (`""`), the resource is all collections, excluding the *system collections* (page 376), in all the databases:

```
{ db: "", collection: "" }
```

For user-defined roles, only roles scoped for the `admin` database can have this resource specification for their privileges.

Cluster Resource

To specify the cluster as the resource, use the following syntax:

```
{ cluster : true }
```

Use the `cluster` resource for actions that affect the state of the system rather than act on specific set of databases or collections. Examples of such actions are `shutdown`, `replSetReconfig`, and `addShard`. For example, the following document grants the action `shutdown` on the `cluster`.

```
{ resource: { cluster : true }, actions: [ "shutdown" ] }
```

For user-defined roles, only roles scoped for the `admin` database can have this resource specification for their privileges.

anyResource

The internal resource `anyResource` gives access to every resource in the system and is intended for internal use. **Do not** use this resource, other than in exceptional circumstances. The syntax for this resource is `{ anyResource: true }`.

Privilege Actions

New in version 2.6.

On this page

- Query and Write Actions (page 501)
- Database Management Actions (page 501)
- Deployment Management Actions (page 502)
- Replication Actions (page 503)
- Sharding Actions (page 503)
- Server Administration Actions (page 504)
- Diagnostic Actions (page 505)
- Internal Actions (page 505)

Privilege actions define the operations a user can perform on a *resource* (page 498). A MongoDB *privilege* (page 434) comprises a *resource* (page 498) and the permitted actions. This page lists available actions grouped by common purpose.

MongoDB provides built-in roles with pre-defined pairings of resources and permitted actions. For lists of the actions granted, see *Built-In Roles* (page 485). To define custom roles, see *Create a User-Defined Role* (page 442).

Query and Write Actions

find
User can perform the `db.collection.find()` method. Apply this action to database or collection resources.

insert
User can perform the `insert` command. Apply this action to database or collection resources.

remove
User can perform the `db.collection.remove()` method. Apply this action to database or collection resources.

update
User can perform the `update` command. Apply this action to database or collection resources.

bypassDocumentValidation
New in version 3.2.

User can bypass document validation on commands that support the `bypassDocumentValidation` option. For a list of commands that support the `bypassDocumentValidation` option, see *Document Validation* (page 890). Apply this action to database or collection resources.

Database Management Actions

changeCustomData
User can change the custom information of any user in the given database. Apply this action to database resources.

changeOwnCustomData
Users can change their own custom information. Apply this action to database resources. See also *Change Your Password and Custom Data* (page 447).

changeOwnPassword
Users can change their own passwords. Apply this action to database resources. See also *Change Your Password and Custom Data* (page 447).

changePassword
User can change the password of any user in the given database. Apply this action to database resources.

createCollection
User can perform the `db.createCollection()` method. Apply this action to database or collection resources.

createIndex
Provides access to the `db.collection.createIndex()` method and the `createIndexes` command. Apply this action to database or collection resources.

createRole
User can create new roles in the given database. Apply this action to database resources.

createUser
User can create new users in the given database. Apply this action to database resources.

dropCollection
User can perform the `db.collection.drop()` method. Apply this action to database or collection resources.

dropRole
User can delete any role from the given database. Apply this action to database resources.

dropUser
User can remove any user from the given database. Apply this action to database resources.

emptycapped
User can perform the `emptycapped` command. Apply this action to database or collection resources.

enableProfiler
User can perform the `db.setProfilingLevel()` method. Apply this action to database resources.

grantRole
User can grant any role in the database to any user from any database in the system. Apply this action to database resources.

killCursors
User can kill cursors on the target collection.

revokeRole
User can remove any role from any user from any database in the system. Apply this action to database resources.

unlock
User can perform the `db.fsyncUnlock()` method. Apply this action to the `cluster` resource.

viewRole
User can view information about any role in the given database. Apply this action to database resources.

viewUser
User can view the information of any user in the given database. Apply this action to database resources.

Deployment Management Actions

authSchemaUpgrade
User can perform the `authSchemaUpgrade` command. Apply this action to the `cluster` resource.

cleanupOrphaned
User can perform the `cleanupOrphaned` command. Apply this action to the `cluster` resource.

cpuProfiler
User can enable and use the CPU profiler. Apply this action to the `cluster` resource.

inprog
User can use the `db.currentOp()` method to return pending and active operations. Apply this action to the `cluster` resource.

invalidateUserCache
Provides access to the `invalidateUserCache` command. Apply this action to the `cluster` resource.

killop
User can perform the `db.killOp()` method. Apply this action to the `cluster` resource.

planCacheRead
User can perform the `planCacheListPlans` and `planCacheListQueryShapes` commands and the `PlanCache.getPlansByQuery()` and `PlanCache.listQueryShapes()` methods. Apply this action to database or collection resources.

planCacheWrite
User can perform the `planCacheClear` command and the `PlanCache.clear()` and `PlanCache.clearPlansByQuery()` methods. Apply this action to database or collection resources.

storageDetails
User can perform the `storageDetails` command. Apply this action to database or collection resources.

Replication Actions

appendOplogNote
User can append notes to the oplog. Apply this action to the `cluster` resource.

replSetConfigure
User can configure a replica set. Apply this action to the `cluster` resource.

replSetGetStatus
User can perform the `replSetGetStatus` command. Apply this action to the `cluster` resource.

replSetHeartbeat
User can perform the `replSetHeartbeat` command. Apply this action to the `cluster` resource.

replSetStateChange
User can change the state of a replica set through the `replSetFreeze`, `replSetMaintenance`, `replSetStepDown`, and `replSetSyncFrom` commands. Apply this action to the `cluster` resource.

resync
User can perform the `resync` command. Apply this action to the `cluster` resource.

Sharding Actions

addShard
User can perform the `addShard` command. Apply this action to the `cluster` resource.

enableSharding
User can enable sharding on a database using the `enableSharding` command and can shard a collection using the `shardCollection` command. Apply this action to database or collection resources.

flushRouterConfig
User can perform the `flushRouterConfig` command. Apply this action to the `cluster` resource.

getShardMap
User can perform the `getShardMap` command. Apply this action to the `cluster` resource.

getShardVersion
User can perform the `getShardVersion` command. Apply this action to database resources.

listShards
User can perform the `listShards` command. Apply this action to the `cluster` resource.

moveChunk
User can perform the `moveChunk` command. In addition, user can perform the `movePrimary` command provided that the privilege is applied to an appropriate database resource. Apply this action to database or collection resources.

removeShard
User can perform the `removeShard` command. Apply this action to the `cluster` resource.

shardingState
User can perform the `shardingState` command. Apply this action to the `cluster` resource.

splitChunk
User can perform the `splitChunk` command. Apply this action to database or collection resources.

splitVector
User can perform the `splitVector` command. Apply this action to database or collection resources.

Server Administration Actions

applicationMessage
User can perform the `logApplicationMessage` command. Apply this action to the `cluster` resource.

closeAllDatabases
User can perform the `closeAllDatabases` command. Apply this action to the `cluster` resource.

collMod
User can perform the `collMod` command. Apply this action to database or collection resources.

compact
User can perform the `compact` command. Apply this action to database or collection resources.

connPoolSync
User can perform the `connPoolSync` command. Apply this action to the `cluster` resource.

convertToCapped
User can perform the `convertToCapped` command. Apply this action to database or collection resources.

dropDatabase
User can perform the `dropDatabase` command. Apply this action to database resources.

dropIndex
User can perform the `dropIndexes` command. Apply this action to database or collection resources.

fsync
User can perform the `fsync` command. Apply this action to the `cluster` resource.

getParameter
User can perform the `getParameter` command. Apply this action to the `cluster` resource.

hostInfo
Provides information about the server the MongoDB instance runs on. Apply this action to the `cluster` resource.

logRotate
User can perform the `logRotate` command. Apply this action to the `cluster` resource.

reIndex
User can perform the `reIndex` command. Apply this action to database or collection resources.

renameCollectionSameDB
Allows the user to rename collections on the current database using the `renameCollection` command. Apply this action to database resources.

Additionally, the user must either *have* `find` (page 501) on the source collection or *not have* `find` (page 501) on the destination collection.

If a collection with the new name already exists, the user must also have the `dropCollection` (page 501) action on the destination collection.

repairDatabase
User can perform the `repairDatabase` command. Apply this action to database resources.

setParameter
User can perform the `setParameter` command. Apply this action to the `cluster` resource.

shutdown
User can perform the `shutdown` command. Apply this action to the `cluster` resource.

touch
User can perform the `touch` command. Apply this action to the `cluster` resource.

Diagnostic Actions

collStats
User can perform the `collStats` command. Apply this action to database or collection resources.

connPoolStats
User can perform the `connPoolStats` and `shardConnPoolStats` commands. Apply this action to the `cluster` resource.

cursorInfo
User can perform the `cursorInfo` command. Apply this action to the `cluster` resource.

dbHash
User can perform the `dbHash` command. Apply this action to database or collection resources.

dbStats
User can perform the `dbStats` command. Apply this action to database resources.

diagLogging
User can perform the `diagLogging` command. Apply this action to the `cluster` resource.

getCmdLineOpts
User can perform the `getCmdLineOpts` command. Apply this action to the `cluster` resource.

getLog
User can perform the `getLog` command. Apply this action to the `cluster` resource.

indexStats
User can perform the `indexStats` command. Apply this action to database or collection resources.

Changed in version 3.0: MongoDB 3.0 removes the `indexStats` command.

listDatabases
User can perform the `listDatabases` command. Apply this action to the `cluster` resource.

listCollections
User can perform the `listCollections` command. Apply this action to database resources.

listIndexes
User can perform the `ListIndexes` command. Apply this action to database or collection resources.

netstat
User can perform the `netstat` command. Apply this action to the `cluster` resource.

serverStatus
User can perform the `serverStatus` command. Apply this action to the `cluster` resource.

validate
User can perform the `validate` command. Apply this action to database or collection resources.

top
User can perform the `top` command. Apply this action to the `cluster` resource.

Internal Actions

anyAction
Allows any action on a resource. **Do not** assign this action except for exceptional circumstances.

internal
Allows internal actions. **Do not** assign this action except for exceptional circumstances.

System Event Audit Messages

On this page

- Audit Message (page 506)
- Audit Event Actions, Details, and Results (page 506)

Note: Available only in MongoDB Enterprise[83].

Audit Message

The *event auditing feature* (page 466) can record events in JSON format. To configure auditing output, see *Configure Auditing* (page 467)

The recorded JSON messages have the following syntax:

```
{
  atype: <String>,
  ts : { "$date": <timestamp> },
  local: { ip: <String>, port: <int> },
  remote: { ip: <String>, port: <int> },
  users : [ { user: <String>, db: <String> }, ... ],
  roles: [ { role: <String>, db: <String> }, ... ],
  param: <document>,
  result: <int>
}
```

field string atype Action type. See *Audit Event Actions, Details, and Results* (page 506).

field document ts Document that contains the date and UTC time of the event, in ISO 8601 format.

field document local Document that contains the local `ip` address and the `port` number of the running instance.

field document remote Document that contains the remote `ip` address and the `port` number of the incoming connection associated with the event.

field array users Array of user identification documents. Because MongoDB allows a session to log in with different user per database, this array can have more than one user. Each document contains a `user` field for the username and a `db` field for the authentication database for that user.

field array roles Array of documents that specify the *roles* (page 433) granted to the user. Each document contains a `role` field for the name of the role and a `db` field for the database associated with the role.

field document param Specific details for the event. See *Audit Event Actions, Details, and Results* (page 506).

field integer result Error code. See *Audit Event Actions, Details, and Results* (page 506).

Audit Event Actions, Details, and Results

The following table lists for each `atype` or action type, the associated `param` details and the `result` values, if any.

[83] http://www.mongodb.com/products/mongodb-enterprise?jmp=docs

atype	param	result
authenticate	{ user: <user name>, db: <database>, mechanism: <mechanism> }	0 - Success 18 - Authentication Failed
authCheck	{ command: <name>, ns: <database>.<collection>, args: <command object> } ns field is optional. args field may be redacted.	0 - Success 13 - Unauthorized to perform the operation. By default, the auditing system logs only the authorization failures. To enable the system to log authorization successes, use the auditAuthorizationSuccess parameter. [84]
createCollection (page 501)	{ ns: <database>.<collection> }	0 - Success
createDatabase	{ ns: <database> }	0 - Success
createIndex (page 501)	{ ns: <database>.<collection>, indexName: <index name>, indexSpec: <index specification> }	0 - Success
renameCollection	{ old: <database>.<collection>, new: <database>.<collection> }	0 - Success
dropCollection (page 501)	{ ns: <database>.<collection> }	0 - Success
dropDatabase (page 504)	{ ns: <database> }	0 - Success
dropIndex (page 504)	{ ns: <database>.<collection>, indexName: <index name> }	0 - Success

Continued on next page

[84] Enabling auditAuthorizationSuccess degrades performance more than logging only the authorization failures.

Table 9.1 – continued from previous page

atype	param	result
createUser (page 501)	{ user: <user name>, db: <database>, customData: <document>, roles: [{ role: <role name>, db: <database> }, ...] } The customData field is optional.	0 - Success
dropUser (page 502)	{ user: <user name>, db: <database> }	0 - Success
dropAllUsersFromDatabase	{ db: <database> }	0 - Success
updateUser	{ user: <user name>, db: <database>, passwordChanged: <boolean>, customData: <document>, roles: [{ role: <role name>, db: <database> }, ...] } The customData field is optional.	0 - Success
grantRolesToUser	{ user: <user name>, db: <database>, roles: [{ role: <role name>, db: <database> }, ...] }	0 - Success

Continued on next page

Table 9.1 – continued from previous page

atype	param	result
revokeRolesFromUser	{ user: <user name>, db: <database>, roles: [{ role: <role name>, db: <database> }, ...] }	0 - Success
createRole (page 501)	{ role: <role name>, db: <database>, roles: [{ role: <role name>, db: <database> }, ...], privileges: [{ resource: <resource document>, actions: [<action>, ...] }, ...] } The roles and the privileges fields are optional. For details on the resource document, see *Resource Document* (page 498). For a list of actions, see *Privilege Actions* (page 500).	0 - Success

Continued on next page

Table 9.1 – continued from previous page

atype	param	result
updateRole	{ role: <role name>, db: <database>, roles: [{ role: <role name>, db: <database> }, ...], privileges: [{ resource: <resource document>, actions: [<action>, ...] }, ...] } The roles and the privileges fields are optional. For details on the resource document, see *Resource Document* (page 498). For a list of actions, see *Privilege Actions* (page 500).	0 - Success
dropRole (page 501)	{ role: <role name>, db: <database> }	0 - Success
dropAllRolesFromDatabase	{ db: <database> }	0 - Success
grantRolesToRole	{ role: <role name>, db: <database>, roles: [{ role: <role name>, db: <database> }, ...] }	0 - Success

Continued on next page

Table 9.1 – continued from previous page

atype	param	result
revokeRolesFromRole	{ role: <role name>, db: <database>, roles: [{ role: <role name>, db: <database> }, ...] }	0 - Success
grantPrivilegesToRole	{ role: <role name>, db: <database>, privileges: [{ resource: <resource document>, actions: [<action>, ...] }, ...] } For details on the resource document, see *Resource Document* (page 498). For a list of actions, see *Privilege Actions* (page 500).	0 - Success
revokePrivilegesFromRole	{ role: <role name>, db: <database name>, privileges: [{ resource: <resource document>, actions: [<action>, ...] }, ...] } For details on the resource document, see *Resource Document* (page 498). For a list of actions, see *Privilege Actions* (page 500).	0 - Success

Continued on next page

Table 9.1 – continued from previous page

atype	param	result
replSetReconfig	{ old: <configuration>, new: <configuration> } Indicates membership change in the replica set. The old field is optional.	0 - Success
enableSharding (page 503)	{ ns: <database> }	0 - Success
shardCollection	{ ns: <database>.<collection>, key: <shard key pattern>, options: { unique: <boolean> } }	0 - Success
addShard (page 503)	{ shard: <shard name>, connectionString: <hostname>:<port>, maxSize: <maxSize> } When a shard is a replica set, the connectionString includes the replica set name and can include other members of the replica set.	0 - Success
removeShard (page 503)	{ shard: <shard name> }	0 - Success
shutdown (page 504)	{ } Indicates commencement of database shutdown.	0 - Success
applicationMessage (page 504)	{ msg: <custom message string> } See logApplicationMessage.	0 - Success

9.9 Create a Vulnerability Report

On this page

- Create the Report in JIRA (page 513)
- Information to Provide (page 513)
- Send the Report via Email (page 513)
- Evaluation of a Vulnerability Report (page 513)
- Disclosure (page 513)

If you believe you have discovered a vulnerability in MongoDB or have experienced a security incident related to MongoDB, please report the issue to aid in its resolution.

To report an issue, we strongly suggest filing a ticket in the SECURITY[85] project in JIRA. MongoDB, Inc responds to vulnerability notifications within 48 hours.

9.9.1 Create the Report in JIRA

Submit a Ticket[86] in the Security[87] project on our JIRA. The ticket number will become the reference identification for the issue for its lifetime. You can use this identifier for tracking purposes.

9.9.2 Information to Provide

All vulnerability reports should contain as much information as possible so MongoDB's developers can move quickly to resolve the issue. In particular, please include the following:

- The name of the product.
- *Common Vulnerability* information, if applicable, including:
- CVSS (Common Vulnerability Scoring System) Score.
- CVE (Common Vulnerability and Exposures) Identifier.
- Contact information, including an email address and/or phone number, if applicable.

9.9.3 Send the Report via Email

While JIRA is the preferred reporting method, you may also report vulnerabilities via email to security@mongodb.com[88].

You may encrypt email using MongoDB's public key at https://docs.mongodb.org/10gen-security-gpg-key.asc.

MongoDB, Inc. responds to vulnerability reports sent via email with a response email that contains a reference number for a JIRA ticket posted to the SECURITY[89] project.

9.9.4 Evaluation of a Vulnerability Report

MongoDB, Inc. validates all submitted vulnerabilities and uses Jira to track all communications regarding a vulnerability, including requests for clarification or additional information. If needed, MongoDB representatives set up a conference call to exchange information regarding the vulnerability.

9.9.5 Disclosure

MongoDB, Inc. requests that you do *not* publicly disclose any information regarding the vulnerability or exploit the issue until it has had the opportunity to analyze the vulnerability, to respond to the notification, and to notify key users, customers, and partners.

[85] https://jira.mongodb.org/browse/SECURITY
[86] https://jira.mongodb.org/secure/CreateIssue!default.jspa?project-field=%22Security%22
[87] https://jira.mongodb.org/browse/SECURITY
[88] security@mongodb.com
[89] https://jira.mongodb.org/browse/SECURITY

The amount of time required to validate a reported vulnerability depends on the complexity and severity of the issue. MongoDB, Inc. takes all required vulnerabilities very seriously and will always ensure that there is a clear and open channel of communication with the reporter.

After validating an issue, MongoDB, Inc. coordinates public disclosure of the issue with the reporter in a mutually agreed timeframe and format. If required or requested, the reporter of a vulnerability will receive credit in the published security bulletin.

9.10 Additional Resources

- Making HIPAA Compliant MongoDB Applications[90]
- Security Architecture White Paper[91]
- Webinar: Securing Your MongoDB Deployment[92]

[90] https://www.mongodb.com/blog/post/making-hipaa-compliant-applications-mongodb?jmp=docs
[91] https://www.mongodb.com/lp/white-paper/mongodb-security-architecture?jmp=docs
[92] http://www.mongodb.com/presentations/webinar-securing-your-mongodb-deployment?jmp=docs

CHAPTER 10

Indexes

On this page

- Default _id Index (page 515)
- Create an Index (page 516)
- Index Types (page 516)
- Index Properties (page 518)
- Index Use (page 519)
- Covered Queries (page 519)
- Index Intersection (page 520)
- Restrictions (page 520)
- Additional Considerations (page 520)
- Additional Resources (page 593)

Indexes support the efficient execution of queries in MongoDB. Without indexes, MongoDB must perform a *collection scan*, i.e. scan every document in a collection, to select those documents that match the query statement. If an appropriate index exists for a query, MongoDB can use the index to limit the number of documents it must inspect.

Indexes are special data structures [1] that store a small portion of the collection's data set in an easy to traverse form. The index stores the value of a specific field or set of fields, ordered by the value of the field. The ordering of the index entries supports efficient equality matches and range-based query operations. In addition, MongoDB can return sorted results by using the ordering in the index.

The following diagram illustrates a query that selects and orders the matching documents using an index:

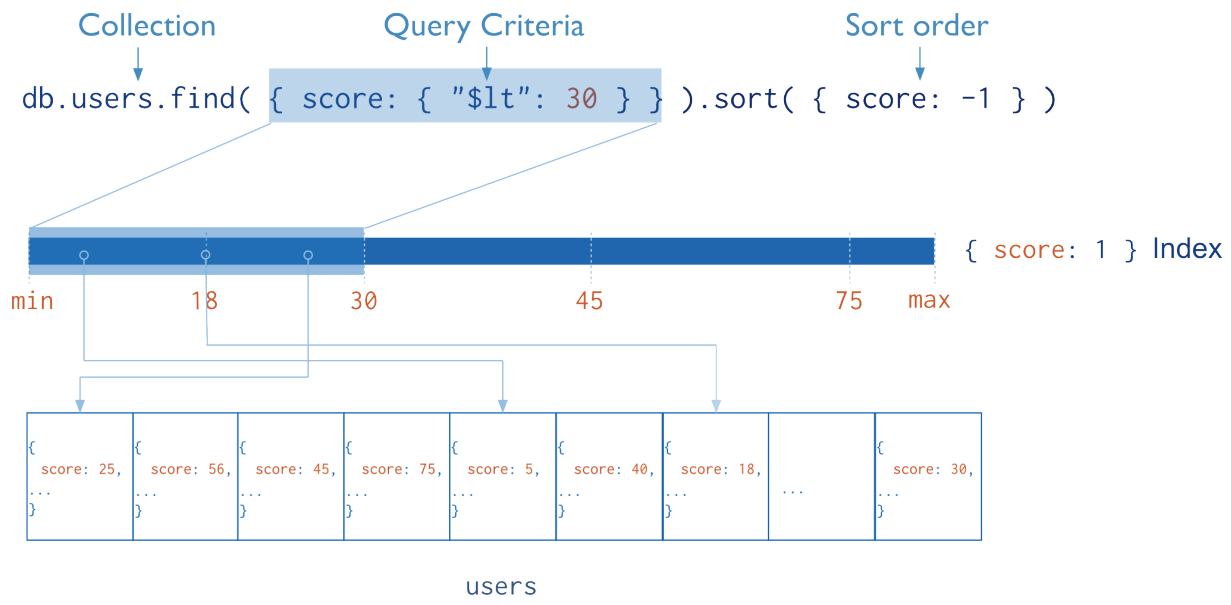
Fundamentally, indexes in MongoDB are similar to indexes in other database systems. MongoDB defines indexes at the *collection* level and supports indexes on any field or sub-field of the documents in a MongoDB collection.

10.1 Default _id Index

MongoDB creates a *unique index* (page 568) on the *_id* (page 11) field during the creation of a collection. The _id index prevents clients from inserting two documents with the same value for the _id field. You cannot drop this index on the _id field.

Note: In *sharded clusters*, if you do *not* use the _id field as the *shard key*, then your application **must** ensure the uniqueness of the values in the _id field to prevent errors. This is most-often done by using a standard auto-generated *ObjectId*.

[1] MongoDB indexes use a B-tree data structure.



10.2 Create an Index

To create an index, use `db.collection.createIndex()` or a similar method from your driver[2].

```
db.collection.createIndex( <key and index type specification>, <options> )
```

The `db.collection.createIndex()` method only creates an index if an index of the same specification does not already exist.

10.3 Index Types

MongoDB provides a number of different index types to support specific types of data and queries.

10.3.1 Single Field

In addition to the MongoDB-defined `_id` index, MongoDB supports the creation of user-defined ascending/descending indexes on a *single field of a document* (page 520).

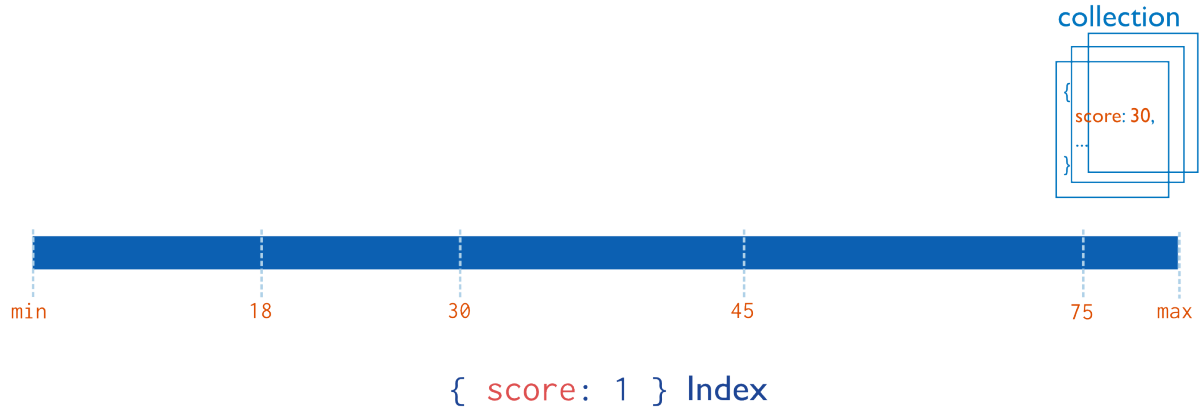
For a single-field index and sort operations, the sort order (i.e. ascending or descending) of the index key does not matter because MongoDB can traverse the index in either direction.

See *Single Field Indexes* (page 520) and *Sort with a Single Field Index* (page 588) for more information on single-field indexes.

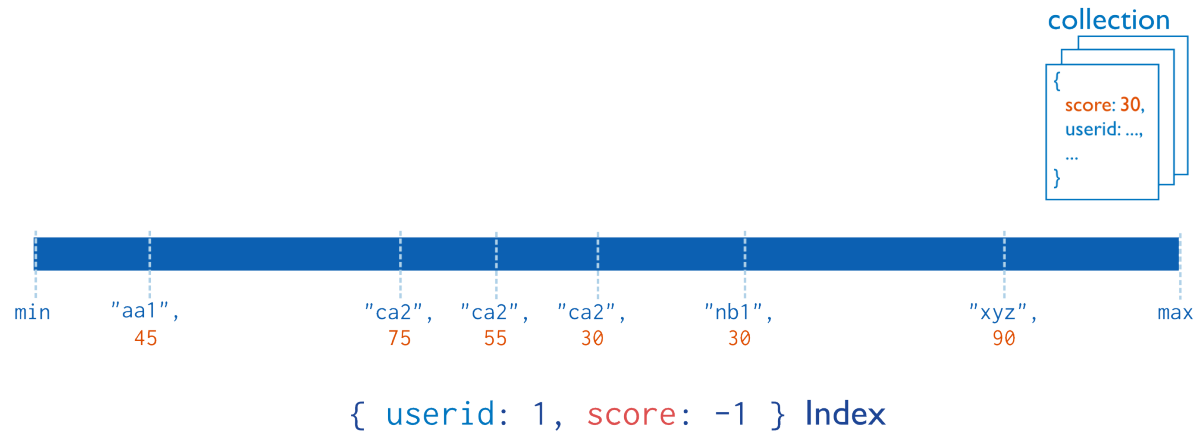
10.3.2 Compound Index

MongoDB also supports user-defined indexes on multiple fields, i.e. *compound indexes* (page 522).

[2] https://api.mongodb.org/



The order of fields listed in a compound index has significance. For instance, if a compound index consists of `{ userid: 1, score: -1 }`, the index sorts first by `userid` and then, within each `userid` value, sorts by `score`.



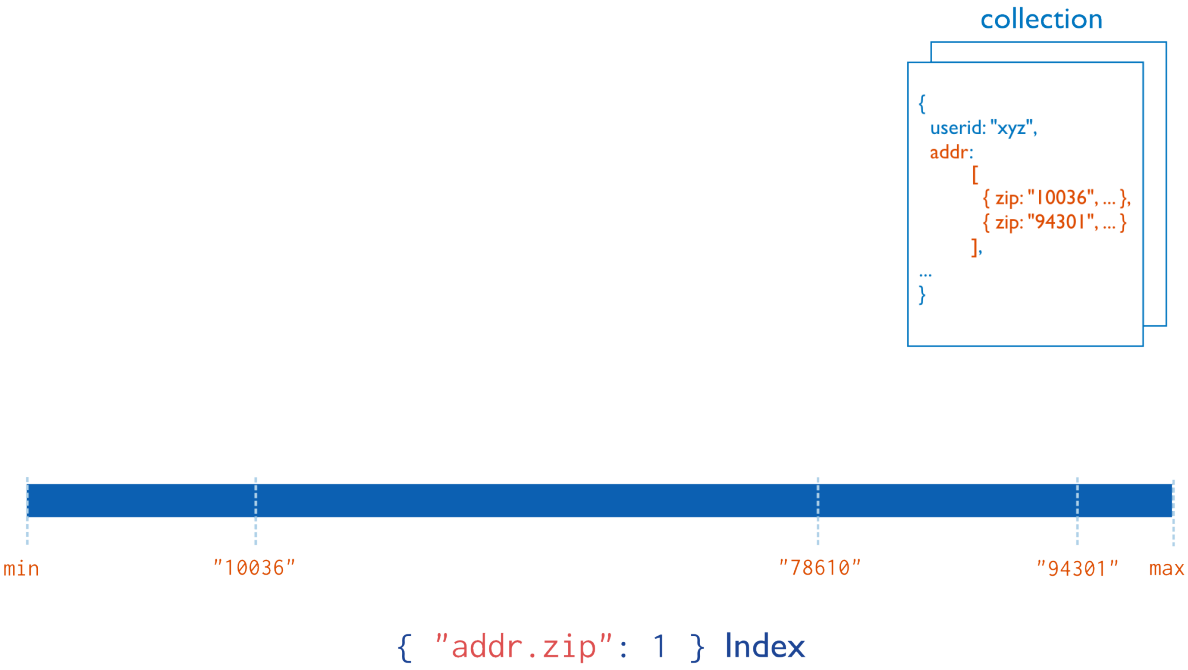
For compound indexes and sort operations, the sort order (i.e. ascending or descending) of the index keys can determine whether the index can support a sort operation. See *Sort Order* (page 524) for more information on the impact of index order on results in compound indexes.

See *Compound Indexes* (page 522) and *Sort on Multiple Fields* (page 588) for more information on compound indexes.

10.3.3 Multikey Index

MongoDB uses *multikey indexes* (page 525) to index the content stored in arrays. If you index a field that holds an array value, MongoDB creates separate index entries for *every* element of the array. These *multikey indexes* (page 525) allow queries to select documents that contain arrays by matching on element or elements of the arrays. MongoDB automatically determines whether to create a multikey index if the indexed field contains an array value; you do not need to explicitly specify the multikey type.

See *Multikey Indexes* (page 525) and *Multikey Index Bounds* (page 529) for more information on multikey indexes.



10.3.4 Geospatial Index

To support efficient queries of geospatial coordinate data, MongoDB provides two special indexes: *2d indexes* (page 557) that uses planar geometry when returning results and *2sphere indexes* (page 543) that use spherical geometry to return results.

See *2d Index Internals* (page 561) for a high level introduction to geospatial indexes.

10.3.5 Text Indexes

MongoDB provides a `text` index type that supports searching for string content in a collection. These text indexes do not store language-specific *stop* words (e.g. "the", "a", "or") and *stem* the words in a collection to only store root words.

See *Text Indexes* (page 533) for more information on text indexes and search.

10.3.6 Hashed Indexes

To support *hash based sharding* (page 748), MongoDB provides a *hashed index* (page 564) type, which indexes the hash of the value of a field. These indexes have a more random distribution of values along their range, but *only* support equality matches and cannot support range-based queries.

10.4 Index Properties

10.4.1 Unique Indexes

The *unique* (page 568) property for an index causes MongoDB to reject duplicate values for the indexed field. Other than the unique constraint, unique indexes are functionally interchangeable with other MongoDB indexes.

10.4.2 Partial Indexes

New in version 3.2.

Partial indexes (page 570) only index the documents in a collection that meet a specified filter expression. By indexing a subset of the documents in a collection, partial indexes have lower storage requirements and reduced performance costs for index creation and maintenance.

Partial indexes offer a superset of the functionality of sparse indexes and should be preferred over sparse indexes.

10.4.3 Sparse Indexes

The *sparse* (page 574) property of an index ensures that the index only contain entries for documents that have the indexed field. The index skips documents that *do not* have the indexed field.

You can combine the sparse index option with the unique index option to reject documents that have duplicate values for a field but ignore documents that do not have the indexed key.

10.4.4 TTL Indexes

TTL indexes (page 566) are special indexes that MongoDB can use to automatically remove documents from a collection after a certain amount of time. This is ideal for certain types of information like machine generated event data, logs, and session information that only need to persist in a database for a finite amount of time.

See: *Expire Data from Collections by Setting TTL* (page 567) for implementation instructions.

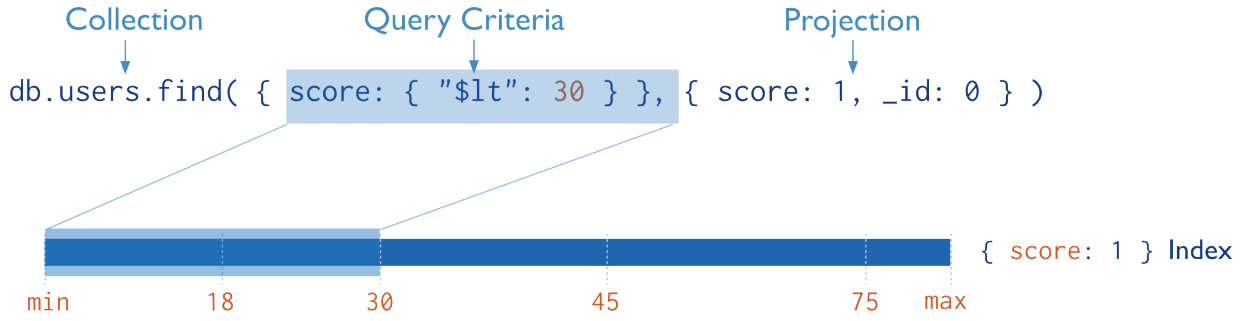
10.5 Index Use

Indexes can improve the efficiency of read operations. The *Analyze Query Performance* (page 159) tutorial provides an example of the execution statistics of a query with and without an index.

For information on how MongoDB chooses an index to use, see *query optimizer* (page 108).

10.6 Covered Queries

When the query criteria and the *projection* of a query include *only* the indexed fields, MongoDB will return results directly from the index *without* scanning any documents or bringing documents into memory. These covered queries can be *very* efficient.



For more information on covered queries, see *Covered Query* (page 106).

10.7 Index Intersection

New in version 2.6.

MongoDB can use the *intersection of indexes* (page 581) to fulfill queries. For queries that specify compound query conditions, if one index can fulfill a part of a query condition, and another index can fulfill another part of the query condition, then MongoDB can use the intersection of the two indexes to fulfill the query. Whether the use of a compound index or the use of an index intersection is more efficient depends on the particular query and the system.

For details on index intersection, see *Index Intersection* (page 581).

10.8 Restrictions

Certain restrictions apply to indexes, such as the length of the index keys or the number of indexes per collection. See *Index Limitations* for details.

10.9 Additional Considerations

Although indexes can improve query performances, indexes also present some operational considerations. See *Operational Considerations for Indexes* (page 256) for more information.

If your collection holds a large amount of data, and your application needs to be able to access the data while building the index, consider building the index in the background, as described in *Background Construction* (page 577).

To build or rebuild indexes for a *replica set*, see *Build Indexes on Replica Sets* (page 579).

Some drivers may specify indexes, using `NumberLong(1)` rather than `1` as the specification. This does not have any affect on the resulting index.

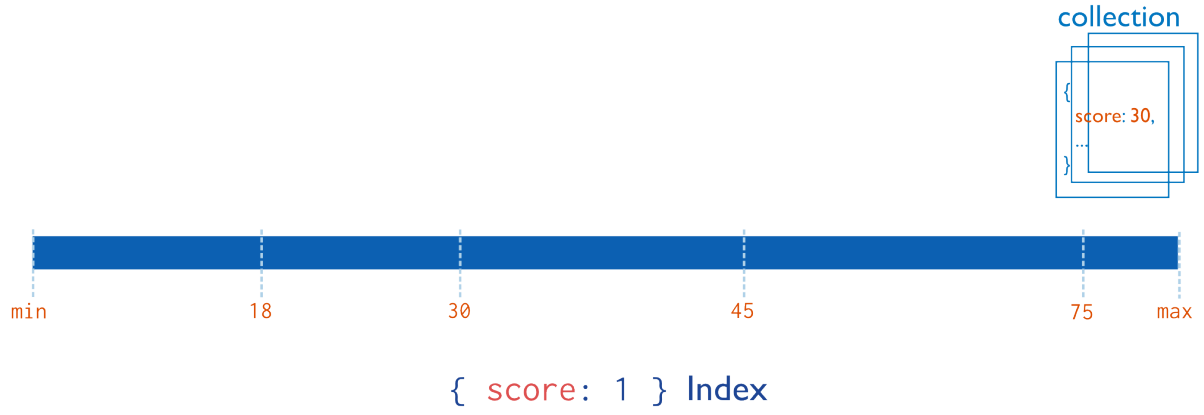
10.9.1 Single Field Indexes

On this page

- Create an Ascending Index on a Single Field (page 521)
- Create an Index on an Embedded Field (page 521)
- Create an Index on Embedded Document (page 522)
- Additional Considerations (page 522)

MongoDB provides complete support for indexes on any field in a *collection* of *documents*. By default, all collections have an index on the *_id field* (page 515), and applications and users may add additional indexes to support important queries and operations.

This document describes ascending/descending indexes on a single field.



Create an Ascending Index on a Single Field

Consider a collection named `records` that holds documents that resemble the following sample document:

```
{
  "_id": ObjectId("570c04a4ad233577f97dc459"),
  "score": 1034,
  "location": { state: "NY", city: "New York" }
}
```

The following operation creates an ascending index on the `score` field of the `records` collection:

```
db.records.createIndex( { score: 1 } )
```

The value of the field in the index specification describes the kind of index for that field. For example, a value of `1` specifies an index that orders items in ascending order. A value of `-1` specifies an index that orders items in descending order. For additional index types, see *index types* (page 516).

The created index will support queries that select on the field `score`, such as the following:

```
db.records.find( { score: 2 } )
db.records.find( { score: { $gt: 10 } } )
```

Create an Index on an Embedded Field

You can create indexes on fields within embedded documents, just as you can index top-level fields in documents. Indexes on embedded fields differ from *indexes on embedded documents* (page 521), which include the full content up to the maximum `index size` of the embedded document in the index. Instead, indexes on embedded fields allow you to use a "dot notation," to introspect into embedded documents.

Consider a collection named `records` that holds documents that resemble the following sample document:

```
{
  "_id": ObjectId("570c04a4ad233577f97dc459"),
  "score": 1034,
  "location": { state: "NY", city: "New York" }
}
```

The following operation creates an index on the `location.state` field:

```
db.records.createIndex( { "location.state": 1 } )
```

The created index will support queries that select on the field `location.state`, such as the following:

```
db.records.find( { "location.state": "CA" } )
db.records.find( { "location.city": "Albany", "location.state": "NY" } )
```

Create an Index on Embedded Document

You can also create indexes on embedded document as a whole.

Consider a collection named `records` that holds documents that resemble the following sample document:

```
{
  "_id": ObjectId("570c04a4ad233577f97dc459"),
  "score": 1034,
  "location": { state: "NY", city: "New York" }
}
```

The `location` field is an embedded document, containing the embedded fields `city` and `state`. The following command creates an index on the `location` field as a whole:

```
db.records.createIndex( { location: 1 } )
```

The following query can use the index on the `location` field:

```
db.records.find( { location: { city: "New York", state: "NY" } } )
```

Note: Although the query can use the index, the result set does not include the sample document above. When performing equality matches on embedded documents, field order matters and the embedded documents must match exactly. See *query-embedded-documents* for more information regarding querying on embedded documents.

Additional Considerations

If your collection holds a large amount of data, and your application needs to be able to access the data while building the index, consider building the index in the background, as described in *Background Construction* (page 577).

To build or rebuild indexes for a *replica set*, see *Build Indexes on Replica Sets* (page 579).

Some drivers may specify indexes, using `NumberLong(1)` rather than `1` as the specification. This does not have any affect on the resulting index.

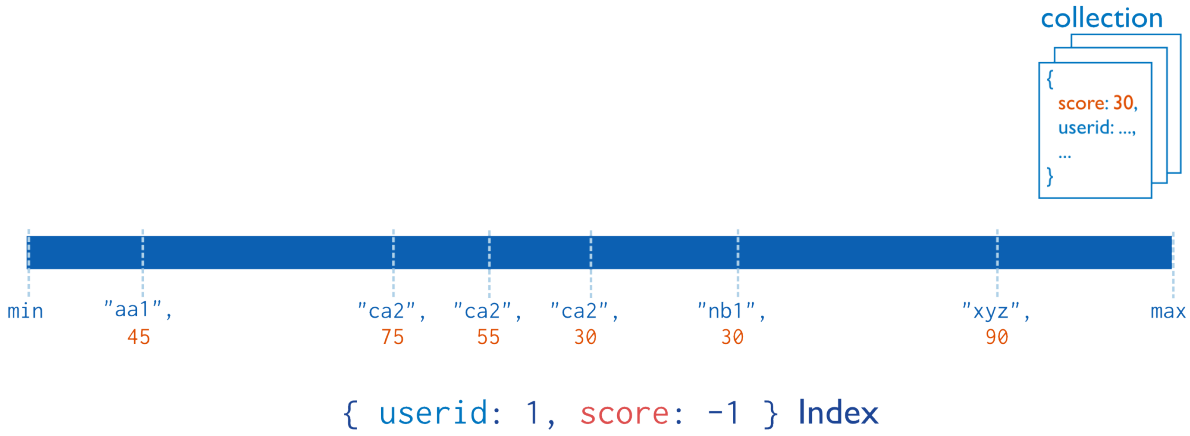
10.9.2 Compound Indexes

On this page

- Create a Compound Index (page 523)
- Sort Order (page 524)
- Prefixes (page 524)
- Index Intersection (page 525)
- Additional Considerations (page 525)

MongoDB supports *compound indexes*, where a single index structure holds references to multiple fields [3] within a collection's documents. The following diagram illustrates an example of a compound index on two fields:

[3] MongoDB imposes a `limit of 31 fields for any compound index`.



Compound indexes can support queries that match on multiple fields.

Create a Compound Index

To create a *compound index* (page 522) use an operation that resembles the following prototype:

```
db.collection.createIndex( { <field1>: <type>, <field2>: <type2>, ... } )
```

The value of the field in the index specification describes the kind of index for that field. For example, a value of `1` specifies an index that orders items in ascending order. A value of `-1` specifies an index that orders items in descending order. For additional index types, see *index types* (page 516).

Important: You may not create compound indexes that have `hashed` index type. You will receive an error if you attempt to create a compound index that includes *a hashed index field* (page 564).

Consider a collection named `products` that holds documents that resemble the following document:

```
{
 "_id": ObjectId(...),
 "item": "Banana",
 "category": ["food", "produce", "grocery"],
 "location": "4th Street Store",
 "stock": 4,
 "type": "cases"
}
```

The following operation creates an ascending index on the `item` and `stock` fields:

```
db.products.createIndex( { "item": 1, "stock": 1 } )
```

The order of the fields listed in a compound index is important. The index will contain references to documents sorted first by the values of the `item` field and, within each value of the `item` field, sorted by values of the stock field. See *Sort Order* (page 524) for more information.

In addition to supporting queries that match on all the index fields, compound indexes can support queries that match on the prefix of the index fields. That is, the index supports queries on the `item` field as well as both `item` and `stock` fields:

```
db.products.find( { item: "Banana" } )
db.products.find( { item: "Banana", stock: { gt: 5 } } )
```

For details, see *Prefixes* (page 524).

Sort Order

Indexes store references to fields in either ascending (`1`) or descending (`-1`) sort order. For single-field indexes, the sort order of keys doesn't matter because MongoDB can traverse the index in either direction. However, for *compound indexes* (page 522), sort order can matter in determining whether the index can support a sort operation.

Consider a collection `events` that contains documents with the fields `username` and `date`. Applications can issue queries that return results sorted first by ascending `username` values and then by descending (i.e. more recent to last) `date` values, such as:

```
db.events.find().sort( { username: 1, date: -1 } )
```

or queries that return results sorted first by descending `username` values and then by ascending `date` values, such as:

```
db.events.find().sort( { username: -1, date: 1 } )
```

The following index can support both these sort operations:

```
db.events.createIndex( { "username" : 1, "date" : -1 } )
```

However, the above index **cannot** support sorting by ascending `username` values and then by ascending `date` values, such as the following:

```
db.events.find().sort( { username: 1, date: 1 } )
```

For more information on sort order and compound indexes, see *Use Indexes to Sort Query Results* (page 587).

Prefixes

Index prefixes are the *beginning* subsets of indexed fields. For example, consider the following compound index:

```
{ "item": 1, "location": 1, "stock": 1 }
```

The index has the following index prefixes:

- `{ item: 1 }`
- `{ item: 1, location: 1 }`

For a compound index, MongoDB can use the index to support queries on the index prefixes. As such, MongoDB can use the index for queries on the following fields:

- the `item` field,
- the `item` field *and* the `location` field,
- the `item` field *and* the `location` field *and* the `stock` field.

MongoDB can also use the index to support a query on `item` and `stock` fields since `item` field corresponds to a prefix. However, the index would not be as efficient in supporting the query as would be an index on only `item` and `stock`.

However, MongoDB cannot use the index to support queries that include the following fields since without the `item` field, none of the listed fields correspond to a prefix index:

- the `location` field,
- the `stock` field, or

- the `location` and `stock` fields.

If you have a collection that has both a compound index and an index on its prefix (e.g. `{ a: 1, b: 1 }` and `{ a: 1 }`), if neither index has a sparse or unique constraint, then you can remove the index on the prefix (e.g. `{ a: 1 }`). MongoDB will use the compound index in all of the situations that it would have used the prefix index.

Index Intersection

Starting in version 2.6, MongoDB can use *index intersection* (page 581) to fulfill queries. The choice between creating compound indexes that support your queries or relying on index intersection depends on the specifics of your system. See *Index Intersection and Compound Indexes* (page 581) for more details.

Additional Considerations

If your collection holds a large amount of data, and your application needs to be able to access the data while building the index, consider building the index in the background, as described in *Background Construction* (page 577).

To build or rebuild indexes for a *replica set*, see *Build Indexes on Replica Sets* (page 579).

Some drivers may specify indexes, using `NumberLong(1)` rather than `1` as the specification. This does not have any affect on the resulting index.

10.9.3 Multikey Indexes

On this page

- Create Multikey Index (page 525)
- Index Bounds (page 525)
- Limitations (page 526)
- Examples (page 527)

To index a field that holds an array value, MongoDB creates an index key for each element in the array. These *multikey* indexes support efficient queries against array fields. Multikey indexes can be constructed over arrays that hold both scalar values (e.g. strings, numbers) *and* nested documents.

Create Multikey Index

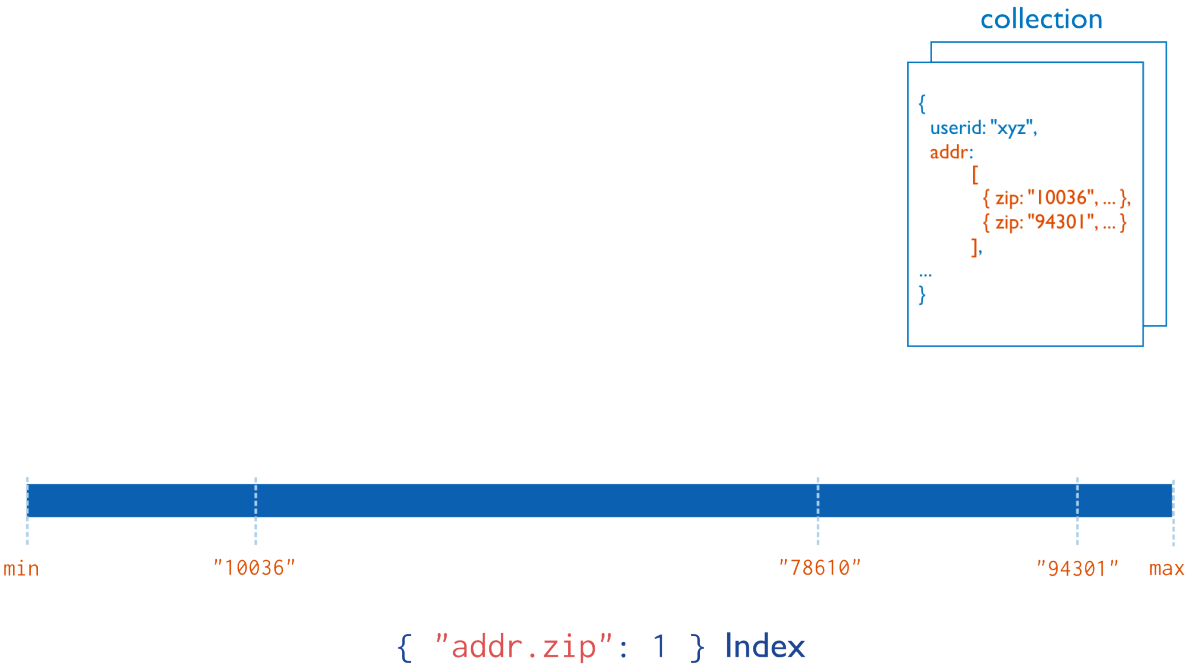
To create a multikey index, use the `db.collection.createIndex()` method:

```
db.coll.createIndex( { <field>: < 1 or -1 > } )
```

MongoDB automatically creates a multikey index if any indexed field is an array; you do not need to explicitly specify the multikey type.

Index Bounds

If an index is multikey, then computation of the index bounds follows special rules. For details on multikey index bounds, see *Multikey Index Bounds* (page 529).



Limitations

Compound Multikey Indexes

For a *compound* (page 522) multikey index, each indexed document can have *at most* one indexed field whose value is an array. As such, you cannot create a compound multikey index if more than one to-be-indexed field of a document is an array. Or, if a compound multikey index already exists, you cannot insert a document that would violate this restriction.

For example, consider a collection that contains the following document:

```
{ _id: 1, a: [ 1, 2 ], b: [ 1, 2 ], category: "AB - both arrays" }
```

You cannot create a compound multikey index `{ a: 1, b: 1 }` on the collection since both the `a` and `b` fields are arrays.

But consider a collection that contains the following documents:

```
{ _id: 1, a: [1, 2], b: 1, category: "A array" }
{ _id: 2, a: 1, b: [1, 2], category: "B array" }
```

A compound multikey index `{ a: 1, b: 1 }` is permissible since for each document, only one field indexed by the compound multikey index is an array; i.e. no document contains array values for both `a` and `b` fields. After creating the compound multikey index, if you attempt to insert a document where both `a` and `b` fields are arrays, MongoDB will fail the insert.

Shard Keys

You **cannot** specify a multikey index as the shard key index.

Changed in version 2.6: However, if the shard key index is a *prefix* (page 524) of a compound index, the compound index is allowed to become a compound *multikey* index if one of the other keys (i.e. keys that are not part of the shard

key) indexes an array. Compound multikey indexes can have an impact on performance.

Hashed Indexes

Hashed (page 564) indexes **cannot** be multikey.

Covered Queries

A *multikey index* (page 525) cannot support a *covered query* (page 106).

Query on the Array Field as a Whole

When a query filter specifies an *exact match for an array as a whole* (page 143), MongoDB can use the multikey index to look up the first element of the query array but cannot use the multikey index scan to find the whole array. Instead, after using the multikey index to look up the first element of the query array, MongoDB retrieves the associated documents and filters for documents whose array matches the array in the query.

For example, consider an `inventory` collection that contains the following documents:

```
{ _id: 5, type: "food", item: "aaa", ratings: [ 5, 8, 9 ] }
{ _id: 6, type: "food", item: "bbb", ratings: [ 5, 9 ] }
{ _id: 7, type: "food", item: "ccc", ratings: [ 9, 5, 8 ] }
{ _id: 8, type: "food", item: "ddd", ratings: [ 9, 5 ] }
{ _id: 9, type: "food", item: "eee", ratings: [ 5, 9, 5 ] }
```

The collection has a multikey index on the `ratings` field:

```
db.inventory.createIndex( { ratings: 1 } )
```

The following query looks for documents where the `ratings` field is the array `[ 5, 9 ]`:

```
db.inventory.find( { ratings: [ 5, 9 ] } )
```

MongoDB can use the multikey index to find documents that have `5` at any position in the `ratings` array. Then, MongoDB retrieves these documents and filters for documents whose `ratings` array equals the query array `[ 5, 9 ]`.

Examples

Index Basic Arrays

Consider a `survey` collection with the following document:

```
{ _id: 1, item: "ABC", ratings: [ 2, 5, 9 ] }
```

Create an index on the field `ratings`:

```
db.survey.createIndex( { ratings: 1 } )
```

Since the `ratings` field contains an array, the index on `ratings` is multikey. The multikey index contains the following three index keys, each pointing to the same document:

- `2`,
- `5`, and

- 9.

Index Arrays with Embedded Documents

You can create multikey indexes on array fields that contain nested objects.

Consider an `inventory` collection with documents of the following form:

```
{
  _id: 1,
  item: "abc",
  stock: [
    { size: "S", color: "red", quantity: 25 },
    { size: "S", color: "blue", quantity: 10 },
    { size: "M", color: "blue", quantity: 50 }
  ]
}
{
  _id: 2,
  item: "def",
  stock: [
    { size: "S", color: "blue", quantity: 20 },
    { size: "M", color: "blue", quantity: 5 },
    { size: "M", color: "black", quantity: 10 },
    { size: "L", color: "red", quantity: 2 }
  ]
}
{
  _id: 3,
  item: "ijk",
  stock: [
    { size: "M", color: "blue", quantity: 15 },
    { size: "L", color: "blue", quantity: 100 },
    { size: "L", color: "red", quantity: 25 }
  ]
}

...
```

The following operation creates a multikey index on the `stock.size` and `stock.quantity` fields:

```
db.inventory.createIndex( { "stock.size": 1, "stock.quantity": 1 } )
```

The compound multikey index can support queries with predicates that include both indexed fields as well as predicates that include only the index prefix `"stock.size"`, as in the following examples:

```
db.inventory.find( { "stock.size": "M" } )
db.inventory.find( { "stock.size": "S", "stock.quantity": { $gt: 20 } } )
```

For details on how MongoDB can combine multikey index bounds, see *Multikey Index Bounds* (page 529). For more information on behavior of compound indexes and prefixes, see *compound indexes and prefixes* (page 524).

The compound multikey index can also support sort operations, such as the following examples:

```
db.inventory.find( ).sort( { "stock.size": 1, "stock.quantity": 1 } )
db.inventory.find( { "stock.size": "M" } ).sort( { "stock.quantity": 1 } )
```

For more information on behavior of compound indexes and sort operations, see *Use Indexes to Sort Query Results* (page 587).

Multikey Index Bounds

On this page

- Intersect Bounds for Multikey Index (page 529)
- Compound Bounds for Multikey Index (page 529)

The bounds of an index scan define the portions of an index to search during a query. When multiple predicates over an index exist, MongoDB will attempt to combine the bounds for these predicates by either *intersection* or *compounding* in order to produce a scan with smaller bounds.

Intersect Bounds for Multikey Index Bounds intersection refers to a logical conjunction (i.e. `AND`) of multiple bounds. For instance, given two bounds `[ [ 3, Infinity ] ]` and `[ [ -Infinity, 6 ] ]`, the intersection of the bounds results in `[ [ 3, 6 ] ]`.

Given an *indexed* (page 525) array field, consider a query that specifies multiple predicates on the array and can use a *multikey index* (page 525). MongoDB can intersect *multikey index* (page 525) bounds if an `$elemMatch` joins the predicates.

For example, a collection `survey` contains documents with a field `item` and an array field `ratings`:

```
{ _id: 1, item: "ABC", ratings: [ 2, 9 ] }
{ _id: 2, item: "XYZ", ratings: [ 4, 3 ] }
```

Create a *multikey index* (page 525) on the `ratings` array:

```
db.survey.createIndex( { ratings: 1 } )
```

The following query uses `$elemMatch` to require that the array contains at least one *single* element that matches both conditions:

```
db.survey.find( { ratings : { $elemMatch: { $gte: 3, $lte: 6 } } } )
```

Taking the predicates separately:

- the bounds for the greater than or equal to 3 predicate (i.e. `$gte: 3`) are `[ [ 3, Infinity ] ]`;
- the bounds for the less than or equal to 6 predicate (i.e. `$lte: 6`) are `[ [ -Infinity, 6 ] ]`.

Because the query uses `$elemMatch` to join these predicates, MongoDB can intersect the bounds to:

```
ratings: [ [ 3, 6 ] ]
```

If the query does *not* join the conditions on the array field with `$elemMatch`, MongoDB cannot intersect the multikey index bounds. Consider the following query:

```
db.survey.find( { ratings : { $gte: 3, $lte: 6 } } )
```

The query searches the `ratings` array for at least one element greater than or equal to 3 and at least one element less than or equal to 6. Because a single element does not need to meet both criteria, MongoDB does *not* intersect the bounds and uses either `[ [ 3, Infinity ] ]` or `[ [ -Infinity, 6 ] ]`. MongoDB makes no guarantee as to which of these two bounds it chooses.

Compound Bounds for Multikey Index Compounding bounds refers to using bounds for multiple keys of *compound index* (page 522). For instance, given a compound index `{ a: 1, b: 1 }` with bounds on field `a` of `[ [ 3, Infinity ] ]` and bounds on field `b` of `[ [ -Infinity, 6 ] ]`, compounding the bounds results in the use of both bounds:

```
{ a: [ [ 3, Infinity ] ], b: [ [ -Infinity, 6 ] ] }
```

If MongoDB cannot compound the two bounds, MongoDB always constrains the index scan by the bound on its leading field, in this case, `a: [ [ 3, Infinity ] ]`.

Compound Index on an Array Field Consider a compound multikey index; i.e. a *compound index* (page 522) where one of the indexed fields is an array. For example, a collection `survey` contains documents with a field `item` and an array field `ratings`:

```
{ _id: 1, item: "ABC", ratings: [ 2, 9 ] }
{ _id: 2, item: "XYZ", ratings: [ 4, 3 ] }
```

Create a *compound index* (page 522) on the `item` field and the `ratings` field:

```
db.survey.createIndex( { item: 1, ratings: 1 } )
```

The following query specifies a condition on both keys of the index:

```
db.survey.find( { item: "XYZ", ratings: { $gte: 3 } } )
```

Taking the predicates separately:

- the bounds for the `item: "XYZ"` predicate are `[ [ "XYZ", "XYZ" ] ]`;
- the bounds for the `ratings: { $gte: 3 }` predicate are `[ [ 3, Infinity ] ]`.

MongoDB can compound the two bounds to use the combined bounds of:

```
{ item: [ [ "XYZ", "XYZ" ] ], ratings: [ [ 3, Infinity ] ] }
```

Compound Index on Fields from an Array of Embedded Documents If an array contains embedded documents, to index on fields contained in the embedded documents, use the *dotted field name* (page 9) in the index specification. For instance, given the following array of embedded documents:

```
ratings: [ { score: 2, by: "mn" }, { score: 9, by: "anon" } ]
```

The dotted field name for the `score` field is `"ratings.score"`.

Compound Bounds of Non-array Field and Field from an Array Consider a collection `survey2` contains documents with a field `item` and an array field `ratings`:

```
{
  _id: 1,
  item: "ABC",
  ratings: [ { score: 2, by: "mn" }, { score: 9, by: "anon" } ]
}
{
  _id: 2,
  item: "XYZ",
  ratings: [ { score: 5, by: "anon" }, { score: 7, by: "wv" } ]
}
```

Create a *compound index* (page 522) on the non-array field `item` as well as two fields from an array `ratings.score` and `ratings.by`:

```
db.survey2.createIndex( { "item": 1, "ratings.score": 1, "ratings.by": 1 } )
```

The following query specifies a condition on all three fields:

```
db.survey2.find( { item: "XYZ",  "ratings.score": { $lte: 5 }, "ratings.by": "anon" } )
```

Taking the predicates separately:

- the bounds for the `item: "XYZ"` predicate are `[ [ "XYZ", "XYZ" ] ]`;
- the bounds for the `score: { $lte: 5 }` predicate are `[ [ -Infinity, 5 ] ]`;
- the bounds for the `by: "anon"` predicate are `[ "anon", "anon" ]`.

MongoDB can compound the bounds for the `item` key with *either* the bounds for `"ratings.score"` or the bounds for `"ratings.by"`, depending upon the query predicates and the index key values. MongoDB makes no guarantee as to which bounds it compounds with the `item` field. For instance, MongoDB will either choose to compound the `item` bounds with the `"ratings.score"` bounds:

```
{
  "item" : [ [ "XYZ", "XYZ" ] ],
  "ratings.score" : [ [ -Infinity, 5 ] ],
  "ratings.by" : [ [ MinKey, MaxKey ] ]
}
```

Or, MongoDB may choose to compound the `item` bounds with `"ratings.by"` bounds:

```
{
  "item" : [ [ "XYZ", "XYZ" ] ],
  "ratings.score" : [ [ MinKey, MaxKey ] ],
  "ratings.by" : [ [ "anon", "anon" ] ]
}
```

However, to compound the bounds for `"ratings.score"` with the bounds for `"ratings.by"`, the query must use `$elemMatch`. See *Compound Bounds of Index Fields from an Array* (page 531) for more information.

Compound Bounds of Index Fields from an Array To compound together the bounds for index keys from the same array:

- the index keys must share the same field path up to but excluding the field names, and
- the query must specify predicates on the fields using `$elemMatch` on that path.

For a field in an embedded document, the *dotted field name* (page 9), such as `"a.b.c.d"`, is the field path for `d`. To compound the bounds for index keys from the same array, the `$elemMatch` must be on the path up to *but excluding* the field name itself; i.e. `"a.b.c"`.

For instance, create a *compound index* (page 522) on the `ratings.score` and the `ratings.by` fields:

```
db.survey2.createIndex( { "ratings.score": 1, "ratings.by": 1 } )
```

The fields `"ratings.score"` and `"ratings.by"` share the field path `ratings`. The following query uses `$elemMatch` on the field `ratings` to require that the array contains at least one *single* element that matches both conditions:

```
db.survey2.find( { ratings: { $elemMatch: { score: { $lte: 5 }, by: "anon" } } } )
```

Taking the predicates separately:

- the bounds for the `score: { $lte: 5 }` predicate is `[ -Infinity, 5 ]`;
- the bounds for the `by: "anon"` predicate is `[ "anon", "anon" ]`.

MongoDB can compound the two bounds to use the combined bounds of:

```
{ "ratings.score" : [ [ -Infinity, 5 ] ], "ratings.by" : [ [ "anon", "anon" ] ] }
```

Query Without `$elemMatch` If the query does *not* join the conditions on the indexed array fields with `$elemMatch`, MongoDB *cannot* compound their bounds. Consider the following query:

```
db.survey2.find( { "ratings.score": { $lte: 5 }, "ratings.by": "anon" } )
```

Because a single embedded document in the array does not need to meet both criteria, MongoDB does *not* compound the bounds. When using a compound index, if MongoDB cannot constrain all the fields of the index, MongoDB always constrains the leading field of the index, in this case `"ratings.score"`:

```
{
  "ratings.score": [ [ -Infinity, 5 ] ],
  "ratings.by": [ [ MinKey, MaxKey ] ]
}
```

`$elemMatch` on Incomplete Path If the query does not specify `$elemMatch` on the path of the embedded fields, up to but excluding the field names, MongoDB **cannot** compound the bounds of index keys from the same array.

For example, a collection `survey3` contains documents with a field `item` and an array field `ratings`:

```
{
  _id: 1,
  item: "ABC",
  ratings: [ { score: { q1: 2, q2: 5 } }, { score: { q1: 8, q2: 4 } } ]
}
{
  _id: 2,
  item: "XYZ",
  ratings: [ { score: { q1: 7, q2: 8 } }, { score: { q1: 9, q2: 5 } } ]
}
```

Create a *compound index* (page 522) on the `ratings.score.q1` and the `ratings.score.q2` fields:

```
db.survey3.createIndex( { "ratings.score.q1": 1, "ratings.score.q2": 1 } )
```

The fields `"ratings.score.q1"` and `"ratings.score.q2"` share the field path `"ratings.score"` and the `$elemMatch` must be on that path.

The following query, however, uses an `$elemMatch` but not on the required path:

```
db.survey3.find( { ratings: { $elemMatch: { 'score.q1': 2, 'score.q2': 8 } } } )
```

As such, MongoDB **cannot** compound the bounds, and the `"ratings.score.q2"` field will be unconstrained during the index scan. To compound the bounds, the query must use `$elemMatch` on the path `"ratings.score"`:

```
db.survey3.find( { 'ratings.score': { $elemMatch: { 'q1': 2, 'q2': 8 } } } )
```

Compound `$elemMatch` Clauses Consider a query that contains multiple `$elemMatch` clauses on different field paths, for instance, `"a.b": { $elemMatch: ... }, "a.c": { $elemMatch: ... }`. MongoDB cannot combine the bounds of the `"a.b"` with the bounds of `"a.c"` since `"a.b"` and `"a.c"` also require `$elemMatch` on the path `a`.

For example, a collection `survey4` contains documents with a field `item` and an array field `ratings`:

```
{
  _id: 1,
  item: "ABC",
  ratings: [
    { score: { q1: 2, q2: 5 }, certainty: { q1: 2, q2: 3 } },
    { score: { q1: 8, q2: 4 }, certainty: { q1: 10, q2: 10 } }
  ]
}
{
  _id: 2,
  item: "XYZ",
  ratings: [
    { score: { q1: 7, q2: 8 }, certainty: { q1: 5, q2: 5 } },
    { score: { q1: 9, q2: 5 }, certainty: { q1: 7, q2: 7 } }
  ]
}
```

Create a *compound index* (page 522) on the `ratings.score.q1` and the `ratings.score.q2` fields:

```
db.survey4.createIndex( {
   "ratings.score.q1": 1,
   "ratings.score.q2": 1,
   "ratings.certainty.q1": 1,
   "ratings.certainty.q2": 1
} )
```

Consider the following query with two `$elemMatch` clauses:

```
db.survey4.find(
  {
    "ratings.score": { $elemMatch: { q1: 5, q2: 5 } },
    "ratings.certainty": { $elemMatch: { q1: 7, q2: 7 } },
  }
)
```

Taking the predicates separately:

- the bounds for the `"ratings.score"` predicate are the compound bounds:

  ```
  { "ratings.score.q1" : [ [ 5, 5 ] ], "ratings.score.q2" : [ [ 5, 5 ] ] }
  ```

- the bounds for the `"ratings.certainty"` predicate are the compound bounds:

  ```
  { "ratings.certainty.q1" : [ [ 7, 7 ] ], "ratings.certainty.q2" : [ [ 7, 7 ] ] }
  ```

However, MongoDB cannot compound the bounds for `"ratings.score"` and `"ratings.certainty"` since `$elemMatch` does not join the two. Instead, MongoDB constrains the leading field of the index `"ratings.score.q1"` which can be compounded with the bounds for `"ratings.score.q2"`:

```
{
  "ratings.score.q1" : [ [ 5, 5 ] ],
  "ratings.score.q2" : [ [ 5, 5 ] ],
  "ratings.certainty.q1" : [ [ MinKey, MaxKey ] ],
  "ratings.certainty.q2" : [ [ MinKey, MaxKey ] ]
}
```

10.9.4 Text Indexes

On this page

- Overview (page 534)
- Create Text Index (page 534)
- Case Insensitivity (page 535)
- Diacritic Insensitivity (page 536)
- Tokenization Delimiters (page 536)
- Index Entries (page 536)
- Supported Languages and Stop Words (page 536)
- `sparse` Property (page 536)
- Restrictions (page 537)
- Storage Requirements and Performance Costs (page 537)
- Text Search Support (page 538)

Changed in version 3.2.

Starting in MongoDB 3.2, MongoDB introduces a version 3 of the `text` index. Key features of the new version of the index are:

- Improved *case insensitivity* (page 535)
- *Diacritic insensitivity* (page 536)
- Additional *delimiters for tokenization* (page 536)

Starting in MongoDB 3.2, version 3 is the default version for new `text` indexes.

Overview

MongoDB provides *text indexes* (page 533) to support text search queries on string content. `text` indexes can include any field whose value is a string or an array of string elements.

Create Text Index

Important: A collection can have at most **one** `text` index.

To create a `text` index, use the `db.collection.createIndex()` method. To index a field that contains a string or an array of string elements, include the field and specify the string literal `"text"` in the index document, as in the following example:

```
db.reviews.createIndex( { comments: "text" } )
```

You can index multiple fields for the `text` index. The following example creates a `text` index on the fields `subject` and `comments`:

```
db.reviews.createIndex(
   {
     subject: "text",
     comments: "text"
   }
 )
```

A *compound index* (page 522) can include `text` index keys in combination with ascending/descending index keys. For more information, see *Compound Index* (page 537).

In order to drop a `text` index, use the index name. See *Use the Index Name to Drop a text Index* (page 541) for more information.

Specify Weights

For a `text` index, the *weight* of an indexed field denotes the significance of the field relative to the other indexed fields in terms of the text search score.

For each indexed field in the document, MongoDB multiplies the number of matches by the weight and sums the results. Using this sum, MongoDB then calculates the score for the document. See `$meta` operator for details on returning and sorting by text scores.

The default weight is 1 for the indexed fields. To adjust the weights for the indexed fields, include the `weights` option in the `db.collection.createIndex()` method.

For more information using weights to control the results of a text search, see *Control Search Results with Weights* (page 541).

Wildcard Text Indexes

When creating a `text` index on multiple fields, you can also use the wildcard specifier (`$**`). With a wildcard text index, MongoDB indexes every field that contains string data for each document in the collection. The following example creates a text index using the wildcard specifier:

```
db.collection.createIndex( { "$**": "text" } )
```

This index allows for text search on all fields with string content. Such an index can be useful with highly unstructured data if it is unclear which fields to include in the text index or for ad-hoc querying.

Wildcard text indexes are `text` indexes on multiple fields. As such, you can assign weights to specific fields during index creation to control the ranking of the results. For more information using weights to control the results of a text search, see *Control Search Results with Weights* (page 541).

Wildcard text indexes, as with all text indexes, can be part of a compound indexes. For example, the following creates a compound index on the field `a` as well as the wildcard specifier:

```
db.collection.createIndex( { a: 1, "$**": "text" } )
```

As with all *compound text indexes* (page 537), since the `a` precedes the text index key, in order to perform a `$text` search with this index, the query predicate must include an equality match conditions `a`. For information on compound text indexes, see *Compound Text Indexes* (page 537).

Case Insensitivity

Changed in version 3.2.

The version 3 `text` index supports the common `C`, simple `S`, and for Turkish languages, the special `T` case foldings as specified in Unicode 8.0 Character Database Case Folding[4].

The case foldings expands the case insensitivity of the `text` index to include characters with diacritics, such as `é` and `É`, and characters from non-Latin alphabets, such as characters from Cyrillic alphabet.

Version 3 of the `text` index is also *diacritic insensitive* (page 536). As such, the index also does not distinguish between `é`, `É`, `e`, and `E`.

[4] http://www.unicode.org/Public/8.0.0/ucd/CaseFolding.txt

Previous versions of the `text` index are case insensitive for `[A-z]` only; i.e. case insensitive for non-diacritics Latin characters only . For all other characters, earlier versions of the text index treat them as distinct.

Diacritic Insensitivity

Changed in version 3.2.

With version 3, `text` index is diacritic insensitive. That is, the index does not distinguish between characters that contain diacritical marks and their non-marked counterpart, such as `é`, `ê`, and `e`. More specifically, the `text` index strips the characters categorized as diacritics in Unicode 8.0 Character Database Prop List[5].

Version 3 of the `text` index is also *case insensitive* (page 535) to characters with diacritics. As such, the index also does not distinguish between `é`, `É`, `e`, and `E`.

Previous versions of the `text` index treat characters with diacritics as distinct.

Tokenization Delimiters

Changed in version 3.2.

For tokenization, version 3 `text` index uses the delimiters categorized under `Dash`, `Hyphen`, `Pattern_Syntax`, `Quotation_Mark`, `Terminal_Punctuation`, and `White_Space` in Unicode 8.0 Character Database Prop List[6].

For example, if given a string `"Il a dit qu'il «était le meilleur joueur du monde»"`, the `text` index treats «, », and spaces as delimiters.

Previous versions of the index treat « as part of the term `"«était"` and » as part of the term `"monde»"`.

Index Entries

`text` index tokenizes and stems the terms in the indexed fields for the index entries. `text` index stores one index entry for each unique stemmed term in each indexed field for each document in the collection. The index uses simple *language-specific* (page 536) suffix stemming.

Supported Languages and Stop Words

MongoDB supports text search for various languages. `text` indexes drop language-specific stop words (e.g. in English, `the`, `an`, `a`, `and`, etc.) and use simple language-specific suffix stemming. For a list of the supported languages, see *Text Search Languages* (page 245).

If you specify a language value of `"none"`, then the `text` index uses simple tokenization with no list of stop words and no stemming.

To specify a language for the `text` index, see *Specify a Language for Text Index* (page 538).

`sparse` Property

`text` indexes are *sparse* (page 574) by default and ignore the *sparse: true* (page 574) option. If a document lacks a `text` index field (or the field is `null` or an empty array), MongoDB does not add an entry for the document to the `text` index. For inserts, MongoDB inserts the document but does not add to the `text` index.

[5] http://www.unicode.org/Public/8.0.0/ucd/PropList.txt
[6] http://www.unicode.org/Public/8.0.0/ucd/PropList.txt

For a compound index that includes a `text` index key along with keys of other types, only the `text` index field determines whether the index references a document. The other keys do not determine whether the index references the documents or not.

Restrictions

One Text Index Per Collection

A collection can have at most **one** `text` index.

Text Search and Hints

You cannot use `hint()` if the query includes a `$text` query expression.

Text Index and Sort

Sort operations cannot obtain sort order from a `text` index, even from a *compound text index* (page 537); i.e. sort operations cannot use the ordering in the text index.

Compound Index

A *compound index* (page 522) can include a `text` index key in combination with ascending/descending index keys. However, these compound indexes have the following restrictions:

- A compound `text` index cannot include any other special index types, such as *multi-key* (page 525) or *geospatial* index fields.
- If the compound `text` index includes keys **preceding** the `text` index key, to perform a `$text` search, the query predicate must include **equality match conditions** on the preceding keys.

See also *Text Index and Sort* (page 537) for additional limitations.

For an example of a compound text index, see *Limit the Number of Entries Scanned* (page 542).

Drop a Text Index

To drop a `text` index, pass the *name* of the index to the `db.collection.dropIndex()` method. To get the name of the index, run the `db.collection.getIndexes()` method.

For information on the default naming scheme for `text` indexes as well as overriding the default name, see *Specify Name for text Index* (page 540).

Storage Requirements and Performance Costs

`text` indexes have the following storage requirements and performance costs:

- `text` indexes can be large. They contain one index entry for each unique post-stemmed word in each indexed field for each document inserted.
- Building a `text` index is very similar to building a large multi-key index and will take longer than building a simple ordered (scalar) index on the same data.

- When building a large `text` index on an existing collection, ensure that you have a sufficiently high limit on open file descriptors. See the *recommended settings* (page 372).
- `text` indexes will impact insertion throughput because MongoDB must add an index entry for each unique post-stemmed word in each indexed field of each new source document.
- Additionally, `text` indexes do not store phrases or information about the proximity of words in the documents. As a result, phrase queries will run much more effectively when the entire collection fits in RAM.

Text Search Support

The `text` index supports `$text` query operations. For examples of text search, see the `$text reference page`. For examples of `$text` operations in aggregation pipelines, see *Text Search in the Aggregation Pipeline* (page 242).

Specify a Language for Text Index

On this page

- Specify the Default Language for a `text` Index (page 538)
- Create a `text` Index for a Collection in Multiple Languages (page 538)

This tutorial describes how to *specify the default language associated with the text index* (page 538) and also how to *create text indexes for collections that contain documents in different languages* (page 538).

Specify the Default Language for a `text` Index The default language associated with the indexed data determines the rules to parse word roots (i.e. stemming) and ignore stop words. The default language for the indexed data is `english`.

To specify a different language, use the `default_language` option when creating the `text` index. See *Text Search Languages* (page 245) for the languages available for `default_language`.

The following example creates for the `quotes` collection a `text` index on the `content` field and sets the `default_language` to `spanish`:

```
db.quotes.createIndex(
   { content : "text" },
   { default_language: "spanish" }
)
```

Create a `text` Index for a Collection in Multiple Languages Changed in version 2.6: Added support for language overrides within embedded documents.

Specify the Index Language within the Document If a collection contains documents or embedded documents that are in different languages, include a field named `language` in the documents or embedded documents and specify as its value the language for that document or embedded document.

MongoDB will use the specified language for that document or embedded document when building the `text` index:

- The specified language in the document overrides the default language for the `text` index.
- The specified language in an embedded document override the language specified in an enclosing document or the default language for the index.

See *Text Search Languages* (page 245) for a list of supported languages.

For example, a collection `quotes` contains multi-language documents that include the `language` field in the document and/or the embedded document as needed:

```
{
   _id: 1,
   language: "portuguese",
   original: "A sorte protege os audazes.",
   translation:
      [
         {
            language: "english",
            quote: "Fortune favors the bold."
         },
         {
            language: "spanish",
            quote: "La suerte protege a los audaces."
         }
    ]
}
{
   _id: 2,
   language: "spanish",
   original: "Nada hay más surrealista que la realidad.",
   translation:
      [
         {
           language: "english",
           quote: "There is nothing more surreal than reality."
         },
         {
           language: "french",
           quote: "Il n'y a rien de plus surréaliste que la réalité."
         }
      ]
}
{
   _id: 3,
   original: "is this a dagger which I see before me.",
   translation:
   {
      language: "spanish",
      quote: "Es este un puñal que veo delante de mí."
   }
}
```

If you create a `text` index on the `quote` field with the default language of English.

```
db.quotes.createIndex( { original: "text", "translation.quote": "text" } )
```

Then, for the documents and embedded documents that contain the `language` field, the `text` index uses that language to parse word stems and other linguistic characteristics.

For embedded documents that do not contain the `language` field,

- If the enclosing document contains the `language` field, then the index uses the document's language for the embedded document.
- Otherwise, the index uses the default language for the embedded documents.

For documents that do not contain the `language` field, the index uses the default language, which is English.

Use any Field to Specify the Language for a Document To use a field with a name other than `language`, include the `language_override` option when creating the index.

For example, give the following command to use `idioma` as the field name instead of `language`:

```
db.quotes.createIndex( { quote : "text" },
                       { language_override: "idioma" } )
```

The documents of the `quotes` collection may specify a language with the `idioma` field:

```
{ _id: 1, idioma: "portuguese", quote: "A sorte protege os audazes" }
{ _id: 2, idioma: "spanish", quote: "Nada hay más surrealista que la realidad." }
{ _id: 3, idioma: "english", quote: "is this a dagger which I see before me" }
```

Specify Name for `text` Index

On this page

- Specify a Name for `text` Index (page 540)
- Use the Index Name to Drop a `text` Index (page 541)

The default name for the index consists of each indexed field name concatenated with `_text`. For example, the following command creates a `text` index on the fields `content`, `users.comments`, and `users.profiles`:

```
db.collection.createIndex(
   {
     content: "text",
     "users.comments": "text",
     "users.profiles": "text"
   }
)
```

The default name for the index is:

```
"content_text_users.comments_text_users.profiles_text"
```

The `text` index, like other indexes, must fall within the `index name length limit`.

Specify a Name for `text` Index To avoid creating an index with a name that exceeds the `index name length limit`, you can pass the `name` option to the `db.collection.createIndex()` method:

```
db.collection.createIndex(
   {
     content: "text",
     "users.comments": "text",
     "users.profiles": "text"
   },
   {
     name: "MyTextIndex"
   }
)
```

Use the Index Name to Drop a `text` Index Whether the *text* (page 533) index has the default name or you specified a name for the *text* (page 533) index, to drop the *text* (page 533) index, pass the index name to the `db.collection.dropIndex()` method.

For example, consider the index created by the following operation:

```
db.collection.createIndex(
   {
     content: "text",
     "users.comments": "text",
     "users.profiles": "text"
   },
   {
     name: "MyTextIndex"
   }
)
```

Then, to remove this text index, pass the name `"MyTextIndex"` to the `db.collection.dropIndex()` method, as in the following:

```
db.collection.dropIndex("MyTextIndex")
```

To get the names of the indexes, use the `db.collection.getIndexes()` method.

Control Search Results with Weights

Text search assigns a score to each document that contains the search term in the indexed fields. The score determines the relevance of a document to a given search query.

For a `text` index, the *weight* of an indexed field denotes the significance of the field relative to the other indexed fields in terms of the text search score.

For each indexed field in the document, MongoDB multiplies the number of matches by the weight and sums the results. Using this sum, MongoDB then calculates the score for the document. See `$meta` operator for details on returning and sorting by text scores.

The default weight is 1 for the indexed fields. To adjust the weights for the indexed fields, include the `weights` option in the `db.collection.createIndex()` method.

> **Warning:** Choose the weights carefully in order to prevent the need to reindex.

A collection `blog` has the following documents:

```
{
  _id: 1,
  content: "This morning I had a cup of coffee.",
  about: "beverage",
  keywords: [ "coffee" ]
}

{
  _id: 2,
  content: "Who doesn't like cake?",
  about: "food",
  keywords: [ "cake", "food", "dessert" ]
}
```

To create a `text` index with different field weights for the `content` field and the `keywords` field, include the `weights` option to the `createIndex()` method. For example, the following command creates an index on three fields and assigns weights to two of the fields:

```
db.blog.createIndex(
   {
     content: "text",
     keywords: "text",
     about: "text"
   },
   {
     weights: {
       content: 10,
       keywords: 5
     },
     name: "TextIndex"
   }
 )
```

The `text` index has the following fields and weights:

- `content` has a weight of 10,
- `keywords` has a weight of 5, and
- `about` has the default weight of 1.

These weights denote the relative significance of the indexed fields to each other. For instance, a term match in the `content` field has:

- `2` times (i.e. `10:5`) the impact as a term match in the `keywords` field and
- `10` times (i.e. `10:1`) the impact as a term match in the `about` field.

Limit the Number of Entries Scanned

This tutorial describes how to create indexes to limit the number of index entries scanned for queries that includes a `$text` expression and equality conditions.

A collection `inventory` contains the following documents:

```
{ _id: 1, dept: "tech", description: "lime green computer" }
{ _id: 2, dept: "tech", description: "wireless red mouse" }
{ _id: 3, dept: "kitchen", description: "green placemat" }
{ _id: 4, dept: "kitchen", description: "red peeler" }
{ _id: 5, dept: "food", description: "green apple" }
{ _id: 6, dept: "food", description: "red potato" }
```

Consider the common use case that performs text searches by *individual* departments, such as:

```
db.inventory.find( { dept: "kitchen", $text: { $search: "green" } } )
```

To limit the text search to scan only those documents within a specific `dept`, create a compound index that *first* specifies an ascending/descending index key on the field `dept` and then a `text` index key on the field `description`:

```
db.inventory.createIndex(
   {
     dept: 1,
     description: "text"
   }
)
```

Then, the text search within a particular department will limit the scan of indexed documents. For example, the following query scans only those documents with `dept` equal to `kitchen`:

```
db.inventory.find( { dept: "kitchen", $text: { $search: "green" } } )
```

Note:

- A compound `text` index cannot include any other special index types, such as *multi-key* (page 525) or *geospatial* index fields.
- If the compound `text` index includes keys **preceding** the `text` index key, to perform a `$text` search, the query predicate must include **equality match conditions** on the preceding keys.

See also:

Text Indexes (page 533)

10.9.5 `2dsphere` Indexes

On this page

- Overview (page 543)
- `2dsphere (Version 2)` (page 543)
- Considerations (page 544)
- Create a `2dsphere` Index (page 544)

New in version 2.4.

Overview

A `2dsphere` index supports queries that calculate geometries on an earth-like sphere. `2dsphere` index supports all MongoDB geospatial queries: queries for inclusion, intersection and proximity. See the `https://docs.mongodb.org/manual/reference/operator/query-geospatial` for the query operators that support geospatial queries.

The `2dsphere` index supports data stored as *GeoJSON* (page 554) objects and as legacy coordinate pairs (See also *2dsphere Indexed Field Restrictions* (page 544)). For legacy coordinate pairs, the index converts the data to GeoJSON *Point* (page 554). For details on the supported GeoJSON objects, see *GeoJSON Objects* (page 554).

The default datum for an earth-like sphere is *WGS84*. Coordinate-axis order is **longitude, latitude**.

`2dsphere (Version 2)`

Changed in version 2.6.

MongoDB 2.6 introduces a version 2 of `2dsphere` indexes. Version 2 is the default version of `2dsphere` indexes created in MongoDB 2.6 and later series. To override the default version 2 and create a version 1 index, include the option `{ "2dsphereIndexVersion": 1 }` when creating the index.

`sparse` Property

Changed in version 2.6.

`2dsphere (Version 2)` indexes are *sparse* (page 574) by default and ignores the *sparse: true* (page 574) option. If a document lacks a `2dsphere` index field (or the field is `null` or an empty array), MongoDB does not add an entry for the document to the index. For inserts, MongoDB inserts the document but does not add to the `2dsphere` index.

For a compound index that includes a `2dsphere` index key along with keys of other types, only the `2dsphere` index field determines whether the index references a document.

Earlier versions of MongoDB only support `2dsphere (Version 1)` indexes. `2dsphere (Version 1)` indexes are *not* sparse by default and will reject documents with `null` location fields.

Additional GeoJSON Objects

`2dsphere (Version 2)` includes support for additional GeoJSON object: *MultiPoint* (page 555), *MultiLineString* (page 556), *MultiPolygon* (page 556), and *GeometryCollection* (page 556). For details on all supported GeoJSON objects, see *GeoJSON Objects* (page 554).

Considerations

`geoNear` and `$geoNear` Restrictions

The `geoNear` command and the `$geoNear` pipeline stage require that a collection have *at most* only one `2dsphere` index and/or only one *2d* (page 557) index whereas *geospatial query operators* (e.g. `$near` and `$geoWithin`) permit collections to have multiple geospatial indexes.

The geospatial index restriction for the `geoNear` command and the `$geoNear` pipeline stage exists because neither the `geoNear` command nor the `$geoNear` pipeline stage syntax includes the location field. As such, index selection among multiple `2d` indexes or `2dsphere` indexes is ambiguous.

No such restriction applies for *geospatial query operators* since these operators take a location field, eliminating the ambiguity.

Shard Key Restrictions

You cannot use a `2dsphere` index as a shard key when sharding a collection. However, you can create and maintain a geospatial index on a sharded collection by using a different field as the shard key.

`2dsphere` Indexed Field Restrictions

Fields with *2dsphere* (page 543) indexes must hold geometry data in the form of *coordinate pairs* or *GeoJSON* data. If you attempt to insert a document with non-geometry data in a `2dsphere` indexed field, or build a `2dsphere` index on a collection where the indexed field has non-geometry data, the operation will fail.

Create a `2dsphere` Index

To create a `2dsphere` index, use the `db.collection.createIndex()` method, specifying the location field as the key and the string literal `"2dsphere"` as the index type:

```
db.collection.createIndex( { <location field> : "2dsphere" } )
```

Unlike a compound *2d* (page 557) index which can reference one location field and one other field, a *compound* (page 522) `2dsphere` index can reference multiple location and non-location fields.

For the following examples, consider a collection `places` with documents that store location data as *GeoJSON Point* (page 554) in a field named `loc`:

```
db.places.insert(
   {
      loc : { type: "Point", coordinates: [ -73.97, 40.77 ] },
      name: "Central Park",
      category : "Parks"
   }
)

db.places.insert(
   {
      loc : { type: "Point", coordinates: [ -73.88, 40.78 ] },
      name: "La Guardia Airport",
      category : "Airport"
   }
)
```

Create a `2dsphere` Index

The following operation creates a *2dsphere* (page 543) index on the location field `loc`:

```
db.places.createIndex( { loc : "2dsphere" } )
```

Create a Compound Index with `2dsphere` Index Key

A *compound index* (page 522) can include a `2dsphere` index key in combination with non-geospatial index keys. For example, the following operation creates a compound index where the first key `loc` is a `2dsphere` index key, and the remaining keys `category` and `names` are non-geospatial index keys, specifically descending (`-1`) and ascending (`1`) keys respectively.

```
db.places.createIndex( { loc : "2dsphere" , category : -1, name: 1 } )
```

Unlike the *2d* (page 557) index, a compound `2dsphere` index does not require the location field to be the first field indexed. For example:

```
db.places.createIndex( { category : 1 , loc : "2dsphere" } )
```

Find Restaurants with Geospatial Queries

On this page

- Overview (page 545)
- Differences Between Flat and Spherical Geometry (page 546)
- Distortion (page 546)
- Searching for Restaurants (page 546)

Overview MongoDB's *geospatial* indexing allows you to efficiently execute spatial queries on a collection that contains geospatial shapes and points. This tutorial will briefly introduce the concepts of geospatial indexes, and then demonstrate their use with `$geoWithin`, `$geoIntersects`, and `geoNear`.

To showcase the capabilities of geospatial features and compare different approaches, this tutorial will guide you through the process of writing queries for a simple geospatial application.

Suppose you are designing a mobile application to help users find restaurants in New York City. The application must:

- Determine the user's current neighborhood using `$geoIntersects`,
- Show the number of restaurants in that neighborhood using `$geoWithin`, and
- Find restaurants within a specified distance of the user using `$nearSphere`.

This tutorial will use a `2dsphere` index to query for this data on spherical geometry.

Differences Between Flat and Spherical Geometry Geospatial queries can use either flat or spherical geometries, depending on both the query and the type of index in use. `2dsphere` indexes support only spherical geometries, while `2d` indexes support both flat and spherical geometries.

However, queries using spherical geometries will be more performant and accurate with a `2dsphere` index, so you should always use `2dsphere` indexes on geographical geospatial fields.

The following table shows what kind of geometry each geospatial operator will use:

Query Type	Geometry Type	Notes
`$near` (*GeoJSON* point, `2dsphere` index)	Spherical	
`$near` (*legacy coordinates*, `2d` index)	Flat	
`$nearSphere` (*GeoJSON* point, `2dsphere` index)	Spherical	
`$nearSphere` (*legacy coordinates*, `2d` index)	Spherical	Use *GeoJSON* points instead.
`$geoWithin : { $geometry: ... }`	Spherical	
`$geoWithin : { $box: ... }`	Flat	
`$geoWithin : { $polygon: ... }`	Flat	
`$geoWithin : { $center: ... }`	Flat	
`$geoWithin : { $centerSphere: ... }`	Spherical	
`$geoIntersects`	Spherical	

The `geoNear` command and the `$geoNear` aggregation operator both operate in radians when using *legacy coordinates*, and meters when using *GeoJSON* points.

Distortion Spherical geometry will appear distorted when visualized on a map due to the nature of projecting a three dimensional sphere, such as the earth, onto a flat plane.

For example, take the specification of the spherical square defined by the longitude latitude points `(0,0)`, `(80,0)`, `(80,80)`, and `(0,80)`. The following figure depicts the area covered by this region:

Searching for Restaurants

Prerequisites Download the example datasets from https://raw.githubusercontent.com/mongodb/docs-assets/geospatial/neighborhoods.json and https://raw.githubusercontent.com/mongodb/docs-assets/geospatial/restaurants.json. These contain the collections `restaurants` and `neighborhoods` respectively.

After downloading the datasets, import them into the database:

```
mongoimport <path to restaurants.json> -c restaurants
mongoimport <path to neighborhoods.json> -c neighborhoods
```

The `geoNear` command requires a geospatial index, and almost always improves performance of `$geoWithin` and `$geoIntersects` queries.

Because this data is geographical, create a `2dsphere` index on each collection using the `mongo` shell:


Arctic Ocean
Iceland
Finland
Sweden
Norway
Russia
United Kingdom
Poland
Germany
Ukraine
Kazakhstan
Mongolia
France
Italy
Spain
Turkey
China
Japan
South Korea
Afghanistan
Iraq
Iran
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
India
Thailand
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Kenya
DR Congo
Tanzania
Indonesia
Papua New Guinea
Angola
Namibia
Botswana
Madagascar
Indian Ocean
South Atlantic
Australia

```
db.restaurants.createIndex({ location: "2dsphere" })
db.neighborhoods.createIndex({ geometry: "2dsphere" })
```

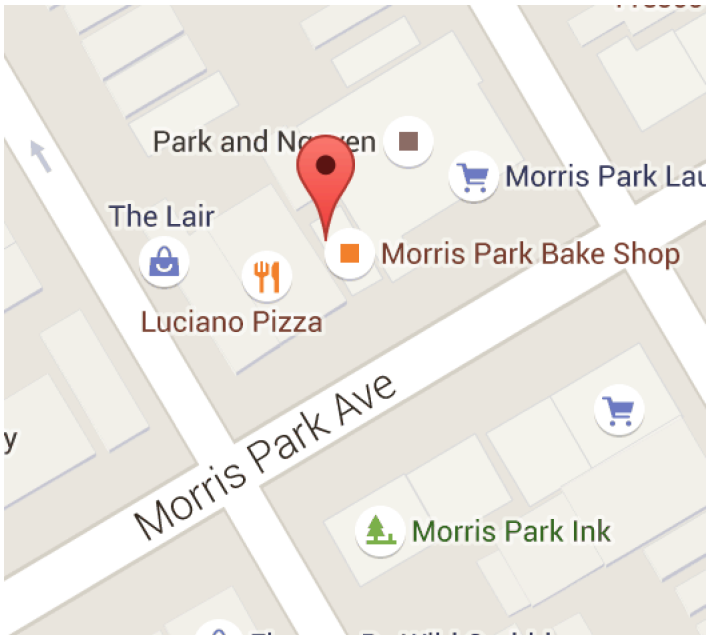
Exploring the Data Inspect an entry in the newly-created `restaurants` collection from within the `mongo` shell:

```
db.restaurants.findOne()
```

This query returns a document like the following:

```
{
   location: {
      type: "Point",
      coordinates: [-73.856077, 40.848447]
   },
   name: "Morris Park Bake Shop"
}
```

This restaurant document corresponds to the location shown in the following figure:



Because the tutorial uses a `2dsphere` index, the geometry data in the `location` field must follow the doc:*GeoJSON format </reference/geojson>*.

Now inspect an entry in the `neighborhoods` collection:

```
db.neighborhoods.findOne()
```

This query will return a document like the following:

```
{
   geometry: {
      type: "Polygon",
      coordinates: [[
         [ -73.99, 40.75 ],
         ...
         [ -73.98, 40.76 ],
```

```
            [ -73.99, 40.75 ]
        ]]
    },
    name: "Hell's Kitchen"
}
```

This geometry corresponds to the region depicted in the following figure:

Find the Current Neighborhood Assuming the user's mobile device can give a reasonably accurate location for the user, it is simple to find the user's current neighborhood with `$geoIntersects`.

Suppose the user is located at -73.93414657 longitude and 40.82302903 latitude. To find the current neighborhood, you will specify a point using the special `$geometry` field in *GeoJSON* format:

```
db.neighborhoods.findOne({ geometry: { $geoIntersects: { $geometry: { type: "Point", coordinates: [ -
```

This query will return the following result:

```
{
    "_id" : ObjectId("55cb9c666c522cafdb053a68"),
    "geometry" : {
        "type" : "Polygon",
```

```
        "coordinates" : [
            [
                [
                    -73.93383000695911,
                    40.81949109558767
                ],
                ...
            ]
        ]
    },
    "name" : "Central Harlem North-Polo Grounds"
}
```

Find all Restaurants in the Neighborhood You can also query to find all restaurants contained in a given neighborhood. Run the following in the `mongo` shell to find the neighborhood containing the user, and then count the restaurants within that neighborhood:

```
var neighborhood = db.neighborhoods.findOne( { geometry: { $geoIntersects: { $geometry: { type: "Poi
db.restaurants.find( { location: { $geoWithin: { $geometry: neighborhood.geometry } } } ).count()
```

This query will tell you that there are 127 restaurants in the requested neighborhood, visualized in the following figure:

Find Restaurants within a Distance To find restaurants within a specified distance of a point, you can use either `$geoWithin` with `$centerSphere` to return results in unsorted order, or `nearSphere` with `$maxDistance` if you need results sorted by distance.

Unsorted with `$geoWithin` To find restaurants within a circular region, use `$geoWithin` with `$centerSphere`. `$centerSphere` is a MongoDB-specific syntax to denote a circular region by specifying the center and the radius in radians.

`$geoWithin` does not return the documents in any specific order, so it may show the user the furthest documents first.

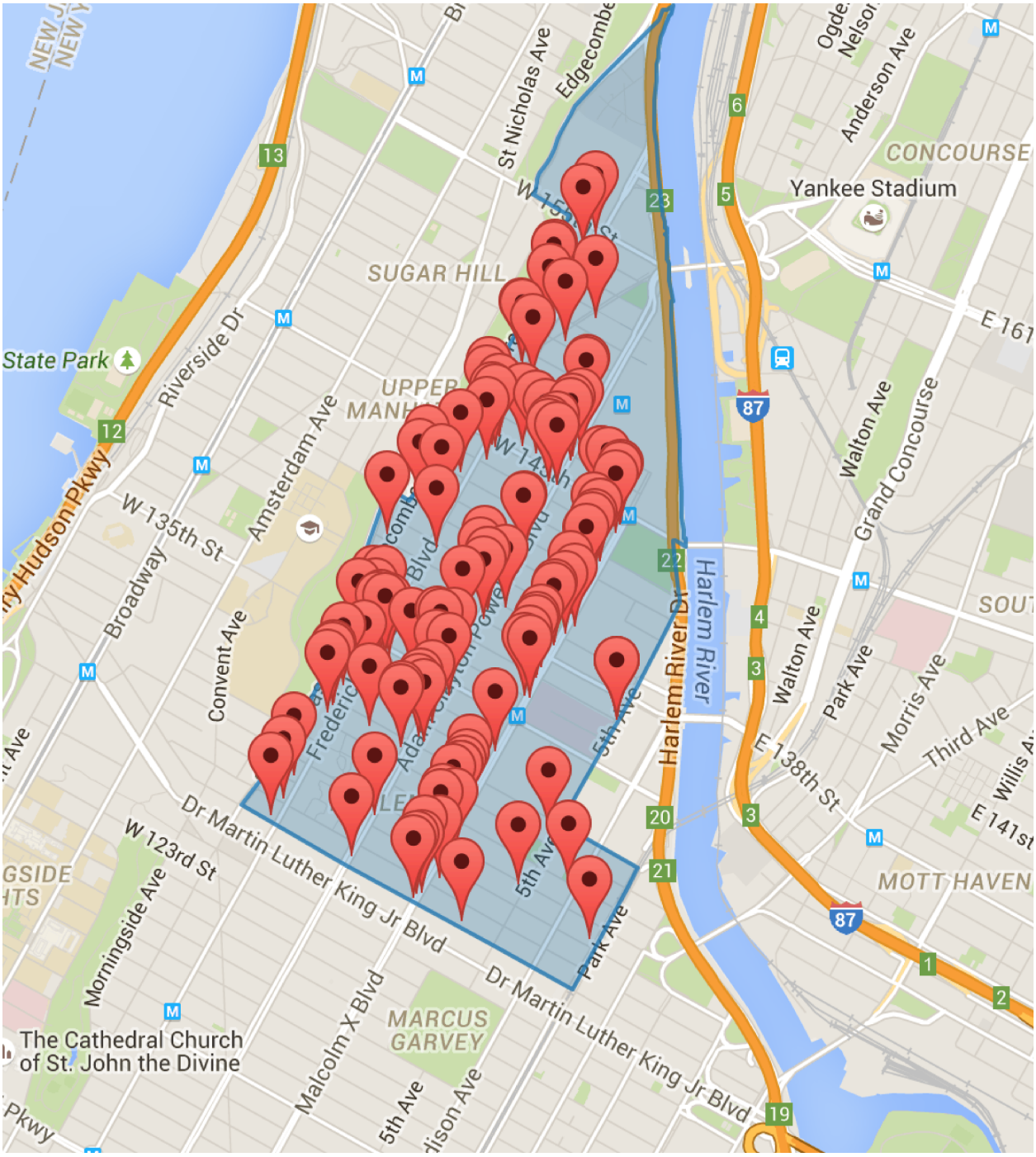
The following will find all restaurants within five miles of the user:

```
db.restaurants.find({ location:
   { $geoWithin:
      { $centerSphere: [ [ -73.93414657, 40.82302903 ], 5 / 3963.2 ] } } })
```

`$centerSphere`'s second argument accepts the radius in radians, so you must divide it by the radius of the earth in miles. See *Calculate Distance Using Spherical Geometry* (page 562) for more information on converting between distance units.

Sorted with `$nearSphere` You may also use `$nearSphere` and specify a `$maxDistance` term in meters. This will return all restaurants within five miles of the user in sorted order from nearest to farthest:

```
var METERS_PER_MILE = 1609.34
db.restaurants.find({ location: { $nearSphere: { $geometry: { type: "Point", coordinates: [ -73.9341
```


SUGAR HILL
Yankee Stadium
CONCOURSE
State Park
Riverside Dr
Amsterdam Ave
Broadway
W 135th St
Convent Ave
St Nicholas Ave
Edgecombe
Anderson Ave
Walton Ave
Grand Concourse
Harlem River
Harlem River Dr
Park Ave
Morris Ave
Third Ave
E 138th St
E 141st
MOTT HAVEN
Dr Martin Luther King Jr Blvd
W 123rd St
Morningside Ave
Malcolm X Blvd
MARCUS GARVEY
5th Ave
The Cathedral Church of St. John the Divine

Query a `2dsphere` Index

On this page

- GeoJSON Objects Bounded by a Polygon (page 552)
- Intersections of GeoJSON Objects (page 552)
- Proximity to a GeoJSON Point (page 553)
- Points within a Circle Defined on a Sphere (page 553)

The following sections describe queries supported by the `2dsphere` index.

GeoJSON Objects Bounded by a Polygon The `$geoWithin` operator queries for location data found within a GeoJSON polygon. Your location data must be stored in GeoJSON format. Use the following syntax:

```
db.<collection>.find( { <location field> :
                         { $geoWithin :
                           { $geometry :
                             { type : "Polygon" ,
                               coordinates : [ <coordinates> ]
                     } } } } )
```

The following example selects all points and shapes that exist entirely within a GeoJSON polygon:

```
db.places.find( { loc :
                  { $geoWithin :
                    { $geometry :
                      { type : "Polygon" ,
                        coordinates : [ [
                                          [ 0 , 0 ] ,
                                          [ 3 , 6 ] ,
                                          [ 6 , 1 ] ,
                                          [ 0 , 0 ]
                                        ] ]
                } } } } )
```

Intersections of GeoJSON Objects New in version 2.4.

The `$geoIntersects` operator queries for locations that intersect a specified GeoJSON object. A location intersects the object if the intersection is non-empty. This includes documents that have a shared edge.

The `$geoIntersects` operator uses the following syntax:

```
db.<collection>.find( { <location field> :
                         { $geoIntersects :
                           { $geometry :
                             { type : "<GeoJSON object type>" ,
                               coordinates : [ <coordinates> ]
                     } } } } )
```

The following example uses `$geoIntersects` to select all indexed points and shapes that intersect with the polygon defined by the `coordinates` array.

```
db.places.find( { loc :
                  { $geoIntersects :
                    { $geometry :
                      { type : "Polygon" ,
                        coordinates: [ [
                                         [ 0 , 0 ] ,
```

```
                                        [ 3 , 6 ] ,
                                        [ 6 , 1 ] ,
                                        [ 0 , 0 ]
                                      ] ]
                } } } } )
```

Proximity to a GeoJSON Point Proximity queries return the points closest to the defined point and sorts the results by distance. A proximity query on GeoJSON data requires a `2dsphere` index.

To query for proximity to a GeoJSON point, use either the `$near` operator or `geoNear` command. Distance is in meters.

The `$near` uses the following syntax:

```
db.<collection>.find( { <location field> :
                        { $near :
                          { $geometry :
                             { type : "Point" ,
                               coordinates : [ <longitude> , <latitude> ] } ,
                            $maxDistance : <distance in meters>
                      } } } )
```

For examples, see `$near`.

The `geoNear` command uses the following syntax:

```
db.runCommand( { geoNear : <collection> ,
                 near : { type : "Point" ,
                          coordinates: [ <longitude>, <latitude> ] } ,
                 spherical : true } )
```

The `geoNear` command offers more options and returns more information than does the `$near` operator. To run the command, see `geoNear`.

Points within a Circle Defined on a Sphere To select all grid coordinates in a "spherical cap" on a sphere, use `$geoWithin` with the `$centerSphere` operator. Specify an array that contains:

- The grid coordinates of the circle's center point
- The circle's radius measured in radians. To calculate radians, see *Calculate Distance Using Spherical Geometry* (page 562).

Use the following syntax:

```
db.<collection>.find( { <location field> :
                        { $geoWithin :
                          { $centerSphere :
                             [ [ <x>, <y> ] , <radius> ] }
                      } } )
```

The following example queries grid coordinates and returns all documents within a 10 mile radius of longitude `88 W` and latitude `30 N`. The example converts the distance, 10 miles, to radians by dividing by the approximate equatorial radius of the earth, 3963.2 miles:

```
db.places.find( { loc :
                  { $geoWithin :
                    { $centerSphere :
                       [ [ -88 , 30 ] , 10 / 3963.2 ]
                } } } )
```

GeoJSON Objects

On this page

- Overview (page 554)
- `Point` (page 554)
- `LineString` (page 554)
- `Polygon` (page 554)
- `MultiPoint` (page 555)
- `MultiLineString` (page 556)
- `MultiPolygon` (page 556)
- `GeometryCollection` (page 556)

Overview MongoDB supports the GeoJSON object types listed on this page.

To specify GeoJSON data, use a document with a `type` field specifying the GeoJSON object type and a `coordinates` field specifying the object's coordinates:

```
{ type: "<GeoJSON type>" , coordinates: <coordinates> }
```

Important: Always list coordinates in `longitude, latitude` order.

The default coordinate reference system for GeoJSON uses the *WGS84* datum.

`Point` New in version 2.4.

The following example specifies a GeoJSON Point[7]:

```
{ type: "Point", coordinates: [ 40, 5 ] }
```

`LineString` New in version 2.4.

The following example specifies a GeoJSON LineString[8]:

```
{ type: "LineString", coordinates: [ [ 40, 5 ], [ 41, 6 ] ] }
```

`Polygon` New in version 2.4.

Polygons[9] consist of an array of GeoJSON `LinearRing` coordinate arrays. These `LinearRings` are closed `LineStrings`. Closed `LineStrings` have at least four coordinate pairs and specify the same position as the first and last coordinates.

The line that joins two points on a curved surface may or may not contain the same set of co-ordinates that joins those two points on a flat surface. The line that joins two points on a curved surface will be a geodesic. Carefully check points to avoid errors with shared edges, as well as overlaps and other types of intersections.

Polygons with a Single Ring The following example specifies a GeoJSON `Polygon` with an exterior ring and no interior rings (or holes). The first and last coordinates must match in order to close the polygon:

[7] http://geojson.org/geojson-spec.html#point
[8] http://geojson.org/geojson-spec.html#linestring
[9] http://geojson.org/geojson-spec.html#polygon

```
{
  type: "Polygon",
  coordinates: [ [ [ 0 , 0 ] , [ 3 , 6 ] , [ 6 , 1 ] , [ 0 , 0  ] ] ]
}
```

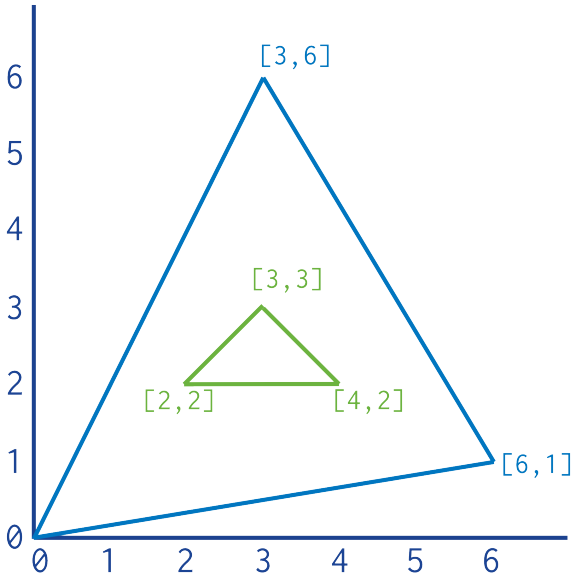
For Polygons with a single ring, the ring cannot self-intersect.

Polygons with Multiple Rings For Polygons with multiple rings:

- The first described ring must be the exterior ring.
- The exterior ring cannot self-intersect.
- Any interior ring must be entirely contained by the outer ring.
- Interior rings cannot intersect or overlap each other. Interior rings cannot share an edge.

The following example represents a GeoJSON polygon with an interior ring:

```
{
  type : "Polygon",
  coordinates : [
     [ [ 0 , 0 ] , [ 3 , 6 ] , [ 6 , 1 ] , [ 0 , 0 ] ],
     [ [ 2 , 2 ] , [ 3 , 3 ] , [ 4 , 2 ] , [ 2 , 2 ] ]
  ]
}
```



`MultiPoint` New in version 2.6: Requires *2dsphere (Version 2)* (page 543)

GeoJSON *MultiPoint <http://geojson.org/geojson-spec.html#multipoint>* embedded documents encode a list of points.

```
{
  type: "MultiPoint",
  coordinates: [
```

```
      [ -73.9580, 40.8003 ],
      [ -73.9498, 40.7968 ],
      [ -73.9737, 40.7648 ],
      [ -73.9814, 40.7681 ]
  ]
}
```

`MultiLineString` New in version 2.6: Requires *2dsphere (Version 2)* (page 543)

The following example specifies a GeoJSON MultiLineString[10]:

```
{
  type: "MultiLineString",
  coordinates: [
      [ [ -73.96943, 40.78519 ], [ -73.96082, 40.78095 ] ],
      [ [ -73.96415, 40.79229 ], [ -73.95544, 40.78854 ] ],
      [ [ -73.97162, 40.78205 ], [ -73.96374, 40.77715 ] ],
      [ [ -73.97880, 40.77247 ], [ -73.97036, 40.76811 ] ]
  ]
}
```

`MultiPolygon` New in version 2.6: Requires *2dsphere (Version 2)* (page 543)

The following example specifies a GeoJSON MultiPolygon[11]:

```
{
  type: "MultiPolygon",
  coordinates: [
      [ [ [ -73.958, 40.8003 ], [ -73.9498, 40.7968 ], [ -73.9737, 40.7648 ], [ -73.9814, 40.7681 ],
      [ [ [ -73.958, 40.8003 ], [ -73.9498, 40.7968 ], [ -73.9737, 40.7648 ], [ -73.958, 40.8003 ] ] ]
  ]
}
```

`GeometryCollection` New in version 2.6: Requires *2dsphere (Version 2)* (page 543)

The following example stores coordinates of GeoJSON type GeometryCollection[12]:

```
{
  type: "GeometryCollection",
  geometries: [
      {
        type: "MultiPoint",
        coordinates: [
           [ -73.9580, 40.8003 ],
           [ -73.9498, 40.7968 ],
           [ -73.9737, 40.7648 ],
           [ -73.9814, 40.7681 ]
        ]
      },
      {
        type: "MultiLineString",
        coordinates: [
           [ [ -73.96943, 40.78519 ], [ -73.96082, 40.78095 ] ],
```

[10] http://geojson.org/geojson-spec.html#multilinestring
[11] http://geojson.org/geojson-spec.html#multipolygon
[12] http://geojson.org/geojson-spec.html#geometrycollection

```
            [ [ -73.96415, 40.79229 ], [ -73.95544, 40.78854 ] ],
            [ [ -73.97162, 40.78205 ], [ -73.96374, 40.77715 ] ],
            [ [ -73.97880, 40.77247 ], [ -73.97036, 40.76811 ] ]
        ]
      }
  ]
}
```

10.9.6 `2d` Indexes

On this page

- Considerations (page 557)
- Behavior (page 557)
- Points on a 2D Plane (page 558)
- `sparse` Property (page 558)

Use a `2d` index for data stored as points on a two-dimensional plane. The `2d` index is intended for legacy coordinate pairs used in MongoDB 2.2 and earlier.

Use a `2d` index if:

- your database has legacy location data from MongoDB 2.2 or earlier, *and*
- you do not intend to store any location data as *GeoJSON* objects.

See the `https://docs.mongodb.org/manual/reference/operator/query-geospatial` for the query operators that support geospatial queries.

Considerations

The `geoNear` command and the `$geoNear` pipeline stage require that a collection have *at most* only one `2d` index and/or only one *2dsphere index* (page 543) whereas *geospatial query operators* (e.g. `$near` and `$geoWithin`) permit collections to have multiple geospatial indexes.

The geospatial index restriction for the `geoNear` command and the `$geoNear` pipeline stage exists because neither the `geoNear` command nor the `$geoNear` pipeline stage syntax includes the location field. As such, index selection among multiple `2d` indexes or `2dsphere` indexes is ambiguous.

No such restriction applies for *geospatial query operators* since these operators take a location field, eliminating the ambiguity.

Do not use a `2d` index if your location data includes GeoJSON objects. To index on both legacy coordinate pairs *and* GeoJSON objects, use a *2dsphere* (page 543) index.

You cannot use a `2d` index as a shard key when sharding a collection. However, you can create and maintain a geospatial index on a sharded collection by using a different field as the shard key.

Behavior

The `2d` index supports calculations on a flat, Euclidean plane. The `2d` index also supports *distance-only* calculations on a sphere, but for *geometric* calculations (e.g. `$geoWithin`) on a sphere, store data as GeoJSON objects and use the `2dsphere` index type.

A 2d index can reference two fields. The first must be the location field. A 2d compound index constructs queries that select first on the location field, and then filters those results by the additional criteria. A compound 2d index can cover queries.

Points on a 2D Plane

To store location data as legacy coordinate pairs, use an array or an embedded document. When possible, use the array format:

```
loc : [ <longitude> , <latitude> ]
```

Consider the embedded document form:

```
loc : { lng : <longitude> , lat : <latitude> }
```

Arrays are preferred as certain languages do not guarantee associative map ordering.

For all points, if you use longitude and latitude, store coordinates in **longitude, latitude** order.

`sparse` Property

2d indexes are *sparse* (page 574) by default and ignores the *sparse: true* (page 574) option. If a document lacks a 2d index field (or the field is null or an empty array), MongoDB does not add an entry for the document to the 2d index. For inserts, MongoDB inserts the document but does not add to the 2d index.

For a compound index that includes a 2d index key along with keys of other types, only the 2d index field determines whether the index references a document.

Create a `2d` Index

On this page

- Define Location Range for a 2d Index (page 558)
- Define Location Precision for a 2d Index (page 559)

To build a geospatial 2d index, use the db.collection.createIndex() method and specify 2d. Use the following syntax:

```
db.<collection>.createIndex( { <location field> : "2d" ,
                               <additional field> : <value> } ,
                             { <index-specification options> } )
```

The 2d index uses the following optional index-specification options:

```
{ min : <lower bound> , max : <upper bound> ,
  bits : <bit precision> }
```

Define Location Range for a `2d` Index By default, a 2d index assumes longitude and latitude and has boundaries of -180 **inclusive** and 180 **non-inclusive**. If documents contain coordinate data outside of the specified range, MongoDB returns an error.

Important: The default boundaries allow applications to insert documents with invalid latitudes greater than 90 or less than -90. The behavior of geospatial queries with such invalid points is not defined.

On `2d` indexes you can change the location range.

You can build a `2d` geospatial index with a location range other than the default. Use the `min` and `max` options when creating the index. Use the following syntax:

```
db.collection.createIndex( { <location field> : "2d" } ,
                           { min : <lower bound> , max : <upper bound> } )
```

Define Location Precision for a `2d` Index By default, a `2d` index on legacy coordinate pairs uses 26 bits of precision, which is roughly equivalent to 2 feet or 60 centimeters of precision using the default range of -180 to 180. Precision is measured by the size in bits of the *geohash* values used to store location data. You can configure geospatial indexes with up to 32 bits of precision.

Index precision does not affect query accuracy. The actual grid coordinates are always used in the final query processing. Advantages to lower precision are a lower processing overhead for insert operations and use of less space. An advantage to higher precision is that queries scan smaller portions of the index to return results.

To configure a location precision other than the default, use the `bits` option when creating the index. Use following syntax:

```
db.<collection>.createIndex( {<location field> : "<index type>"} ,
                             { bits : <bit precision> } )
```

For information on the internals of geohash values, see *Calculation of Geohash Values for 2d Indexes* (page 561).

Query a `2d` Index

On this page

- Points within a Shape Defined on a Flat Surface (page 559)
- Points within a Circle Defined on a Sphere (page 560)
- Proximity to a Point on a Flat Surface (page 560)
- Exact Matches on a Flat Surface (page 561)

The following sections describe queries supported by the `2d` index.

Points within a Shape Defined on a Flat Surface To select all legacy coordinate pairs found within a given shape on a flat surface, use the `$geoWithin` operator along with a shape operator. Use the following syntax:

```
db.<collection>.find( { <location field> :
                        { $geoWithin :
                           { $box|$polygon|$center : <coordinates>
                     } } } )
```

The following queries for documents within a rectangle defined by `[ 0 , 0 ]` at the bottom left corner and by `[ 100 , 100 ]` at the top right corner.

```
db.places.find( { loc :
                  { $geoWithin :
                     { $box : [ [ 0 , 0 ] ,
                                [ 100 , 100 ] ]
                  } } } )
```

The following queries for documents that are within the circle centered on `[ -74 , 40.74 ]` and with a radius of `10`:

```
db.places.find( { loc: { $geoWithin :
                         { $center : [ [-74, 40.74 ] , 10 ]
                } } } )
```

For syntax and examples for each shape, see the following:

- `$box`
- `$polygon`
- `$center` (defines a circle)

Points within a Circle Defined on a Sphere MongoDB supports rudimentary spherical queries on flat `2d` indexes for legacy reasons. In general, spherical calculations should use a `2dsphere` index, as described in *2dsphere Indexes* (page 543).

To query for legacy coordinate pairs in a "spherical cap" on a sphere, use `$geoWithin` with the `$centerSphere` operator. Specify an array that contains:

- The grid coordinates of the circle's center point
- The circle's radius measured in radians. To calculate radians, see *Calculate Distance Using Spherical Geometry* (page 562).

Use the following syntax:

```
db.<collection>.find( { <location field> :
                        { $geoWithin :
                           { $centerSphere : [ [ <x>, <y> ] , <radius> ] }
                      } } )
```

The following example query returns all documents within a 10-mile radius of longitude `88 W` and latitude `30 N`. The example converts distance to radians by dividing distance by the approximate equatorial radius of the earth, 3963.2 miles:

```
db.<collection>.find( { loc : { $geoWithin :
                                { $centerSphere :
                                   [ [ 88 , 30 ] , 10 / 3963.2 ]
                      } } } )
```

Proximity to a Point on a Flat Surface Proximity queries return the legacy coordinate pairs closest to the defined point and sort the results by distance. Use either the `$near` operator or `geoNear` command. Both require a `2d` index.

The `$near` operator uses the following syntax:

```
db.<collection>.find( { <location field> :
                        { $near : [ <x> , <y> ]
                      } } )
```

For examples, see `$near`.

The `geoNear` command uses the following syntax:

```
db.runCommand( { geoNear: <collection>, near: [ <x> , <y> ] } )
```

The `geoNear` command offers more options and returns more information than does the `$near` operator. To run the command, see `geoNear`.

Exact Matches on a Flat Surface Changed in version 2.6: Previously, `2d` indexes would support exact-match queries for coordinate pairs.

You cannot use a `2d` index to return an exact match for a coordinate pair. Use a scalar, ascending or descending, index on a field that stores coordinates to return exact matches.

In the following example, the `find()` operation will return an exact match on a location if you have a `{ 'loc': 1}` index:

```
db.<collection>.find( { loc: [ <x> , <y> ] } )
```

This query will return any documents with the value of `[ <x> , <y> ]`.

2d Index Internals

On this page

- Calculation of Geohash Values for `2d` Indexes (page 561)
- Multi-location Documents for `2d` Indexes (page 561)

This document provides a more in-depth explanation of the internals of MongoDB's `2d` geospatial indexes. This material is not necessary for normal operations or application development but may be useful for troubleshooting and for further understanding.

Calculation of Geohash Values for `2d` Indexes When you create a geospatial index on *legacy coordinate pairs*, MongoDB computes *geohash* values for the coordinate pairs within the specified *location range* (page 558) and then indexes the geohash values.

To calculate a geohash value, recursively divide a two-dimensional map into quadrants. Then assign each quadrant a two-bit value. For example, a two-bit representation of four quadrants would be:

```
01  11

00  10
```

These two-bit values (`00`, `01`, `10`, and `11`) represent each of the quadrants and all points within each quadrant. For a geohash with two bits of resolution, all points in the bottom left quadrant would have a geohash of `00`. The top left quadrant would have the geohash of `01`. The bottom right and top right would have a geohash of `10` and `11`, respectively.

To provide additional precision, continue dividing each quadrant into sub-quadrants. Each sub-quadrant would have the geohash value of the containing quadrant concatenated with the value of the sub-quadrant. The geohash for the upper-right quadrant is `11`, and the geohash for the sub-quadrants would be (clockwise from the top left): `1101`, `1111`, `1110`, and `1100`, respectively.

Multi-location Documents for `2d` Indexes

Note: *2dsphere* (page 543) indexes can cover multiple geospatial fields in a document, and can express lists of points using *MultiPoint* (page 555) embedded documents.

While `2d` geospatial indexes do not support more than one geospatial field in a document, you can use a *multi-key index* (page 525) to index multiple coordinate pairs in a single document. In the simplest example you may have a field (e.g. `locs`) that holds an array of coordinates, as in the following example:

```
db.places.save( {
  locs : [ [ 55.5 , 42.3 ] ,
           [ -74 , 44.74 ] ,
           { lng : 55.5 , lat : 42.3 } ]
} )
```

The values of the array may be either arrays, as in `[ 55.5, 42.3 ]`, or embedded documents, as in `{ lng : 55.5 , lat : 42.3 }`.

You could then create a geospatial index on the `locs` field, as in the following:

```
db.places.createIndex( { "locs": "2d" } )
```

You may also model the location data as a field inside of an embedded document. In this case, the document would contain a field (e.g. `addresses`) that holds an array of documents where each document has a field (e.g. `loc:`) that holds location coordinates. For example:

```
db.records.save( {
  name : "John Smith",
  addresses : [ {
                 context : "home" ,
                 loc : [ 55.5, 42.3 ]
                } ,
                {
                 context : "work",
                 loc : [ -74 , 44.74 ]
                }
              ]
} )
```

You could then create the geospatial index on the `addresses.loc` field as in the following example:

```
db.records.createIndex( { "addresses.loc": "2d" } )
```

To include the location field with the distance field in multi-location document queries, specify `includeLocs: true` in the `geoNear` command.

Calculate Distance Using Spherical Geometry

On this page

- Distance Multiplier (page 564)

Note: While basic queries using spherical distance are supported by the `2d` index, consider moving to a `2dsphere` index if your data is primarily longitude and latitude.

The `2d` index supports queries that calculate distances on a Euclidean plane (flat surface). The index also supports the following query operators and command that calculate distances using spherical geometry:

- `$nearSphere`
- `$centerSphere`
- `$near`
- `geoNear` command with the `{ spherical: true }` option.

Important: These three queries use radians for distance. Other query types do not.

For spherical query operators to function properly, you must convert distances to radians, and convert from radians to the distances units used by your application.

To convert:

- *distance to radians*: divide the distance by the radius of the sphere (e.g. the Earth) in the same units as the distance measurement.
- *radians to distance*: multiply the radian measure by the radius of the sphere (e.g. the Earth) in the units system that you want to convert the distance to.

The equatorial radius of the Earth is approximately `3,963.2` miles or `6,378.1` kilometers.

The following query would return documents from the `places` collection within the circle described by the center `[ -74, 40.74 ]` with a radius of `100` miles:

```
db.places.find( { loc: { $geoWithin: { $centerSphere: [ [ -74, 40.74 ] ,
                                                         100 / 3963.2 ] } } } )
```

You may also use the `distanceMultiplier` option to the `geoNear` to convert radians in the `mongod` process, rather than in your application code. See *distance multiplier* (page 564).

The following spherical query, returns all documents in the collection `places` within `100` miles from the point `[ -74, 40.74 ]`.

```
db.runCommand( { geoNear: "places",
                 near: [ -74, 40.74 ],
                 spherical: true
               }  )
```

The output of the above command would be:

```
{
   // [ ... ]
   "results" : [
      {
         "dis" : 0.01853688938212826,
         "obj" : {
            "_id" : ObjectId( ... )
            "loc" : [
               -73,
               40
            ]
         }
      }
   ],
   "stats" : {
      // [ ... ]
      "avgDistance" : 0.01853688938212826,
      "maxDistance" : 0.01853714811400047
   },
   "ok" : 1
}
```

Warning: Spherical queries that wrap around the poles or at the transition from `-180` to `180` longitude raise an error.

Note: While the default Earth-like bounds for geospatial indexes are between `-180` inclusive, and `180`, valid values for latitude are between `-90` and `90`.

Distance Multiplier The `distanceMultiplier` option of the `geoNear` command returns distances only after multiplying the results by an assigned value. This allows MongoDB to return converted values, and removes the requirement to convert units in application logic.

Using `distanceMultiplier` in spherical queries provides results from the `geoNear` command that do not need radian-to-distance conversion. The following example uses `distanceMultiplier` in the `geoNear` command with a *spherical* (page 562) example:

```
db.runCommand( { geoNear: "places",
                 near: [ -74, 40.74 ],
                 spherical: true,
                 distanceMultiplier: 3963.2
               } )
```

The output of the above operation would resemble the following:

```
{
   // [ ... ]
   "results" : [
      {
         "dis" : 73.46525170413567,
         "obj" : {
            "_id" : ObjectId( ... )
            "loc" : [
               -73,
               40
            ]
         }
      }
   ],
   "stats" : {
      // [ ... ]
      "avgDistance" : 0.01853688938212826,
      "maxDistance" : 0.01853714811400047
   },
   "ok" : 1
}
```

10.9.7 Hashed Indexes

On this page

- Hashing Function (page 565)
- Create a Hashed Index (page 565)
- Considerations (page 565)

New in version 2.4.

Hashed indexes maintain entries with hashes of the values of the indexed field.

Hashed indexes support *sharding* (page 733) using hashed shard keys. *Hashed based sharding* (page 748) uses a hashed index of a field as the shard key to partition data across your sharded cluster.

Using a hashed shard key to shard a collection results in a more random distribution of data. See *Shard a Collection Using a Hashed Shard Key* (page 773) for more details.

Hashing Function

Hashed indexes uses a hashing function to compute the hash of the value of the index field. The hashing function collapses embedded documents and computes the hash for the entire value but does not support multi-key (i.e. arrays) indexes.

Tip

MongoDB automatically computes the hashes when resolving queries using hashed indexes. Applications do **not** need to compute hashes.

Create a Hashed Index

To create a *hashed index* (page 564), specify `hashed` as the value of the index key, as in the following example:

```
db.collection.createIndex( { _id: "hashed" } )
```

Considerations

MongoDB supports `hashed` indexes of any single field. The hashing function collapses embedded documents and computes the hash for the entire value, but does not support multi-key (i.e. arrays) indexes.

You may not create compound indexes that have `hashed` index fields or specify a unique constraint on a `hashed` index; however, you can create both a `hashed` index and an ascending/descending (i.e. non-hashed) index on the same field: MongoDB will use the scalar index for range queries.

Warning: MongoDB `hashed` indexes truncate floating point numbers to 64-bit integers before hashing. For example, a `hashed` index would store the same value for a field that held a value of `2.3`, `2.2`, and `2.9`. To prevent collisions, do not use a `hashed` index for floating point numbers that cannot be reliably converted to 64-bit integers (and then back to floating point). MongoDB `hashed` indexes do not support floating point values larger than 2^{53}.

10.9.8 Index Properties

In addition to the numerous *index types* (page 516) MongoDB supports, indexes can also have various properties. The following documents detail the index properties that you can select when building an index.

TTL Indexes **(page 566)** The TTL index is used for TTL collections, which expire data after a period of time.

Unique Indexes **(page 568)** A unique index causes MongoDB to reject all documents that contain a duplicate value for the indexed field.

Partial Indexes **(page 570)** A partial index indexes only documents that meet specified filter criteria.

Sparse Indexes **(page 574)** A sparse index does not index documents that do not have the indexed field.

TTL Indexes

On this page

- Behavior (page 566)
- Restrictions (page 567)

TTL indexes are special single-field indexes that MongoDB can use to automatically remove documents from a collection after a certain amount of time. Data expiration is useful for certain types of information like machine generated event data, logs, and session information that only need to persist in a database for a finite amount of time.

To create a TTL index, use the `db.collection.createIndex()` method with the `expireAfterSeconds` option on a field whose value is either a *date* (page 15) or an array that contains *date values* (page 15).

For example, to create a TTL index on the `lastModifiedDate` field of the `eventlog` collection, use the following operation in the `mongo` shell:

```
db.eventlog.createIndex( { "lastModifiedDate": 1 }, { expireAfterSeconds: 3600 } )
```

Behavior

Expiration of Data TTL indexes expire documents after the specified number of seconds has passed since the indexed field value; i.e. the expiration threshold is the indexed field value plus the specified number of seconds.

If the field is an array, and there are multiple date values in the index, MongoDB uses *lowest* (i.e. earliest) date value in the array to calculate the expiration threshold.

If the indexed field in a document is not a *date* or an array that holds a date value(s), the document will not expire.

If a document does not contain the indexed field, the document will not expire.

Delete Operations A background thread in `mongod` reads the values in the index and removes expired *documents* from the collection.

When the TTL thread is active, you will see *delete* (page 114) operations in the output of `db.currentOp()` or in the data collected by the *database profiler* (page 326).

Timing of the Delete Operation When you build a TTL index in the *background* (page 577), the TTL thread can begin deleting documents while the index is building. If you build a TTL index in the foreground, MongoDB begins removing expired documents as soon as the index finishes building.

The TTL index does not guarantee that expired data will be deleted immediately upon expiration. There may be a delay between the time a document expires and the time that MongoDB removes the document from the database.

The background task that removes expired documents runs *every 60 seconds*. As a result, documents may remain in a collection during the period between the expiration of the document and the running of the background task.

Because the duration of the removal operation depends on the workload of your `mongod` instance, expired data may exist for some time *beyond* the 60 second period between runs of the background task.

Replica Sets On *replica sets*, the TTL background thread *only* deletes documents on the *primary*. However, the TTL background thread does run on secondaries. *Secondary* members replicate deletion operations from the primary.

Support for Queries A TTL index supports queries in the same way non-TTL indexes do.

Record Allocation A collection with a TTL index has `usePowerOf2Sizes` enabled, and you cannot modify this setting for the collection. As a result of enabling `usePowerOf2Sizes`, MongoDB must allocate more disk space relative to data size. This approach helps mitigate the possibility of storage fragmentation caused by frequent delete operations and leads to more predictable storage use patterns.

Restrictions

- TTL indexes are a single-field indexes. *Compound indexes* (page 522) do not support TTL and ignores the `expireAfterSeconds` option.
- The `_id` field does not support TTL indexes.
- You cannot create a TTL index on a *capped collection* (page 6) because MongoDB cannot remove documents from a capped collection.
- You cannot use `createIndex()` to change the value of `expireAfterSeconds` of an existing index. Instead use the `collMod` database command in conjunction with the `index` collection flag. Otherwise, to change the value of the option of an existing index, you must drop the index first and recreate.
- If a non-TTL single-field index already exists for a field, you cannot create a TTL index on the same field since you cannot create indexes that have the same key specification and differ only by the options. To change a non-TTL single-field index to a TTL index, you must drop the index first and recreate with the `expireAfterSeconds` option.

Expire Data from Collections by Setting TTL

On this page

- Procedures (page 567)

This document provides an introduction to MongoDB's "*time to live*" or *TTL* collection feature. TTL collections make it possible to store data in MongoDB and have the `mongod` automatically remove data after a specified number of seconds or at a specific clock time.

Data expiration is useful for some classes of information, including machine generated event data, logs, and session information that only need to persist for a limited period of time.

A special *TTL index property* (page 566) supports the implementation of TTL collections. The TTL feature relies on a background thread in `mongod` that reads the date-typed values in the index and removes expired *documents* from the collection.

Procedures To create a *TTL index* (page 566), use the `db.collection.createIndex()` method with the `expireAfterSeconds` option on a field whose value is either a *date* (page 15) or an array that contains *date values* (page 15).

Note: The TTL index is a single field index. Compound indexes do not support the TTL property. For more information on TTL indexes, see *TTL Indexes* (page 566).

You can modify the `expireAfterSeconds` of an existing TTL index using the `collMod` command.

Expire Documents after a Specified Number of Seconds To expire data after a specified number of seconds has passed since the indexed field, create a TTL index on a field that holds values of BSON date type or an array of BSON date-typed objects *and* specify a positive non-zero value in the `expireAfterSeconds` field. A document will

expire when the number of seconds in the `expireAfterSeconds` field has passed since the time specified in its indexed field. [13]

For example, the following operation creates an index on the `log_events` collection's `createdAt` field and specifies the `expireAfterSeconds` value of `3600` to set the expiration time to be one hour after the time specified by `createdAt`.

```
db.log_events.createIndex( { "createdAt": 1 }, { expireAfterSeconds: 3600 } )
```

When adding documents to the `log_events` collection, set the `createdAt` field to the current time:

```
db.log_events.insert( {
   "createdAt": new Date(),
   "logEvent": 2,
   "logMessage": "Success!"
} )
```

MongoDB will automatically delete documents from the `log_events` collection when the document's `createdAt` value [1] is older than the number of seconds specified in `expireAfterSeconds`.

See also:

`$currentDate` operator

Expire Documents at a Specific Clock Time To expire documents at a specific clock time, begin by creating a TTL index on a field that holds values of BSON date type or an array of BSON date-typed objects *and* specify an `expireAfterSeconds` value of `0`. For each document in the collection, set the indexed date field to a value corresponding to the time the document should expire. If the indexed date field contains a date in the past, MongoDB considers the document expired.

For example, the following operation creates an index on the `log_events` collection's `expireAt` field and specifies the `expireAfterSeconds` value of `0`:

```
db.log_events.createIndex( { "expireAt": 1 }, { expireAfterSeconds: 0 } )
```

For each document, set the value of `expireAt` to correspond to the time the document should expire. For instance, the following `insert()` operation adds a document that should expire at `July 22, 2013 14:00:00`.

```
db.log_events.insert( {
   "expireAt": new Date('July 22, 2013 14:00:00'),
   "logEvent": 2,
   "logMessage": "Success!"
} )
```

MongoDB will automatically delete documents from the `log_events` collection when the documents' `expireAt` value is older than the number of seconds specified in `expireAfterSeconds`, i.e. `0` seconds older in this case. As such, the data expires at the specified `expireAt` value.

Unique Indexes

On this page

- Create a Unique Index (page 569)
- Behavior (page 569)

[13] If the field contains an array of BSON date-typed objects, data expires if at least one of BSON date-typed object is older than the number of seconds specified in `expireAfterSeconds`.

A unique index ensures that the indexed fields do not store duplicate values; i.e. enforces uniqueness for the indexed fields. By default, MongoDB creates a unique index on the *_id* (page 11) field during the creation of a collection.

Create a Unique Index

To create a unique index, use the `db.collection.createIndex()` method with the `unique` option set to `true`.

```
db.collection.createIndex( <key and index type specification>, { unique: true } )
```

Unique Index on a Single Field For example, to create a unique index on the `user_id` field of the `members` collection, use the following operation in the `mongo` shell:

```
db.members.createIndex( { "user_id": 1 }, { unique: true } )
```

Unique Compound Index You can also enforce a unique constraint on *compound indexes* (page 522). If you use the unique constraint on a *compound index* (page 522), then MongoDB will enforce uniqueness on the *combination* of the index key values.

For example, to create a unique index on `groupNumber`, `lastname`, and `firstname` fields of the `members` collection, use the following operation in the `mongo` shell:

```
db.members.createIndex( { groupNumber: 1, lastname: 1, firstname: 1 }, { unique: true } )
```

The created index enforces uniqueness for the *combination* of `groupNumber`, `lastname`, and `firstname` values.

Behavior

Restrictions MongoDB cannot create a *unique index* (page 568) on the specified index field(s) if the collection already contains data that would violate the unique constraint for the index.

You may not specify a unique constraint on a *hashed index* (page 564).

Unique Constraint Across Separate Documents The unique constraint applies to separate documents in the collection. That is, the unique index prevents *separate* documents from having the same value for the indexed key, but the index does not prevent a document from having multiple elements or embedded documents in an indexed array from having the same value. In the case of a single document with repeating values, the repeated value is inserted into the index only once.

For example, a collection has a unique index on `a.b`:

```
db.collection.createIndex( { "a.b": 1 }, { unique: true } )
```

The unique index permits the insertion of the following document into the collection if no other document in the collection has the `a.b` value of `5`:

```
db.collection.insert( { a: [ { b: 5 }, { b: 5 } ] } )
```

Unique Index and Missing Field If a document does not have a value for the indexed field in a unique index, the index will store a null value for this document. Because of the unique constraint, MongoDB will only permit one document that lacks the indexed field. If there is more than one document without a value for the indexed field or is missing the indexed field, the index build will fail with a duplicate key error.

For example, a collection has a unique index on `x`:

```
db.collection.createIndex( { "x": 1 }, { unique: true } )
```

The unique index allows the insertion of a document without the field `x` if the collection does not already contain a document missing the field `x`:

```
db.collection.insert( { y: 1 } )
```

However, the unique index errors on the insertion of a document without the field `x` if the collection already contains a document missing the field `x`:

```
db.collection.insert( { z: 1 } )
```

The operation fails to insert the document because of the violation of the unique constraint on the value of the field `x`:

```
WriteResult({
   "nInserted" : 0,
   "writeError" : {
      "code" : 11000,
      "errmsg" : "E11000 duplicate key error index: test.collection.$a.b_1 dup key: { : null }"
   }
})
```

See also:

Unique Partial Indexes (page 570)

Unique Partial Indexes New in version 3.2.

Partial indexes only index the documents in a collection that meet a specified filter expression. If you specify both the `partialFilterExpression` and a *unique constraint* (page 568), the unique constraint only applies to the documents that meet the filter expression.

A partial index with a unique constraint does not prevent the insertion of documents that do not meet the unique constraint if the documents do not meet the filter criteria. For an example, see *Partial Index with Unique Constraint* (page 573).

Partial Indexes

On this page

- Create a Partial Index (page 571)
- Behavior (page 571)
- Restrictions (page 572)
- Examples (page 573)

New in version 3.2.

Partial indexes only index the documents in a collection that meet a specified filter expression. By indexing a subset of the documents in a collection, partial indexes have lower storage requirements and reduced performance costs for index creation and maintenance.

Create a Partial Index

To create a `partial` index, use the `db.collection.createIndex()` method with the new `partialFilterExpression` option. The `partialFilterExpression` option accepts a document that specifies the filter condition using:

- equality expressions (i.e. `field: value` or using the `$eq` operator),
- `$exists: true` expression,
- `$gt`, `$gte`, `$lt`, `$lte` expressions,
- `$type` expressions,
- `$and` operator at the top-level only

For example, the following operation creates a compound index that indexes only the documents with a `rating` field greater than 5.

```
db.restaurants.createIndex(
   { cuisine: 1, name: 1 },
   { partialFilterExpression: { rating: { $gt: 5 } } }
)
```

You can specify a `partialFilterExpression` option for all MongoDB *index types* (page 516).

Behavior

Query Coverage MongoDB will not use the partial index for a query or sort operation if using the index results in an incomplete result set.

To use the partial index, a query must contain the filter expression (or a modified filter expression that specifies a subset of the filter expression) as part of its query condition.

For example, given the following index:

```
db.restaurants.createIndex(
   { cuisine: 1 },
   { partialFilterExpression: { rating: { $gt: 5 } } }
)
```

The following query can use the index since the query predicate includes the condition `rating: { $gte: 8 }` that matches a subset of documents matched by the index filter expression `ratings: { $gt: 5 }`:

```
db.restaurants.find( { cuisine: "Italian", rating: { $gte: 8 } } )
```

However, the following query cannot use the partial index on the `cuisine` field because using the index results in an incomplete result set. Specifically, the query predicate includes the condition `rating: { $lt: 8 }` while the index has the filter `rating: { $gt: 5 }`. That is, the query `{ cuisine: "Italian", rating: { $lt: 8 } }` matches more documents (e.g. an Italian restaurant with a rating equal to 1) than are indexed.

```
db.restaurants.find( { cuisine: "Italian", rating: { $lt: 8 } } )
```

Similarly, the following query cannot use the partial index because the query predicate does not include the filter expression and using the index would return an incomplete result set.

```
db.restaurants.find( { cuisine: "Italian" } )
```

Comparison with the `sparse` Index

Tip

Partial indexes represent a superset of the functionality offered by sparse indexes and should be preferred over sparse indexes.

Partial indexes offer a more expressive mechanism than *Sparse Indexes* (page 574) indexes to specify which documents are indexed.

Sparse indexes selects documents to index *solely* based on the existence of the indexed field, or for compound indexes, the existence of the indexed fields.

Partial indexes determine the index entries based on the specified filter. The filter can include fields other than the index keys and can specify conditions other than just an existence check. For example, a partial index can implement the same behavior as a sparse index:

```
db.contacts.createIndex(
   { name: 1 },
   { partialFilterExpression: { name: { $exists: true } } }
)
```

This partial index supports the same queries as a sparse index on the `name` field.

However, a partial index can also specify filter expressions on fields other than the index key. For example, the following operation creates a partial index, where the index is on the `name` field but the filter expression is on the `email` field:

```
db.contacts.createIndex(
   { name: 1 },
   { partialFilterExpression: { email: { $exists: true } } }
)
```

For the query optimizer to choose this partial index, the query predicate must include a non-null match on the `email` field as well as a condition on the `name` field.

For example, the following query can use the index:

```
db.contacts.find( { name: "xyz", email: { $regex: /\.org$/ } } )
```

However, the following query cannot use the index:

```
db.contacts.find( { name: "xyz", email: { $exists: false } } )
```

Restrictions

In MongoDB, you cannot create multiple versions of an index that differ only in the options. As such, you cannot create multiple partial indexes that differ only by the filter expression.

You cannot specify both the `partialFilterExpression` option and the `sparse` option.

Earlier versions of MongoDB do not support partial indexes. For sharded clusters or replica sets, all nodes must be version 3.2.

`_id` indexes cannot be partial indexes.

Shard key indexes cannot be partial indexes.

Examples

Create a Partial Index On A Collection Consider a collection `restaurants` containing documents that resemble the following

```
{
   "_id" : ObjectId("5641f6a7522545bc535b5dc9"),
   "address" : {
      "building" : "1007",
      "coord" : [
         -73.856077,
         40.848447
      ],
      "street" : "Morris Park Ave",
      "zipcode" : "10462"
   },
   "borough" : "Bronx",
   "cuisine" : "Bakery",
   "rating" : { "date" : ISODate("2014-03-03T00:00:00Z"),
                "grade" : "A",
                "score" : 2
              },
   "name" : "Morris Park Bake Shop",
   "restaurant_id" : "30075445"
}
```

You could add a partial index on the `borough` and `cuisine` fields choosing only to index documents where the `rating.grade` field is `A`:

```
db.restaurants.createIndex(
   { borough: 1, cuisine: 1 },
   { partialFilterExpression: { 'rating.grade': { $eq: "A" } } }
)
```

Then, the following query on the `restaurants` collection uses the partial index to return the restaurants in the Bronx with `rating.grade` equal to `A`:

```
db.restaurants.find( { borough: "Bronx", 'rating.grade': "A" } )
```

However, the following query cannot use the partial index because the query expression does not include the `rating.grade` field:

```
db.restaurants.find( { borough: "Bronx", cuisine: "Bakery" } )
```

Partial Index with Unique Constraint Partial indexes only index the documents in a collection that meet a specified filter expression. If you specify both the `partialFilterExpression` and a *unique constraint* (page 568), the unique constraint only applies to the documents that meet the filter expression. A partial index with a unique constraint does not prevent the insertion of documents that do not meet the unique constraint if the documents do not meet the filter criteria.

For example, a collection `users` contains the following documents:

```
{ "_id" : ObjectId("56424f1efa0358a27fa1f99a"), "username" : "david", "age" : 29 }
{ "_id" : ObjectId("56424f37fa0358a27fa1f99b"), "username" : "amanda", "age" : 35 }
{ "_id" : ObjectId("56424fe2fa0358a27fa1f99c"), "username" : "rajiv", "age" : 57 }
```

The following operation creates an index that specifies a *unique constraint* (page 568) on the `username` field and a partial filter expression `age: { $gte: 21 }`.

```
db.users.createIndex(
   { username: 1 },
   { unique: true, partialFilterExpression: { age: { $gte: 21 } } }
)
```

The index prevents the insertion of the following documents since documents already exist with the specified usernames and the `age` fields are greater than `21`:

```
db.users.insert( { username: "david", age: 27 } )
db.users.insert( { username: "amanda", age: 25 } )
db.users.insert( { username: "rajiv", age: 32 } )
```

However, the following documents with duplicate usernames are allowed since the unique constraint only applies to documents with `age` greater than or equal to 21.

```
db.users.insert( { username: "david", age: 20 } )
db.users.insert( { username: "amanda" } )
db.users.insert( { username: "rajiv", age: null } )
```

Sparse Indexes

On this page

- Create a Sparse Index (page 574)
- Behavior (page 575)
- Examples (page 575)

Sparse indexes only contain entries for documents that have the indexed field, even if the index field contains a null value. The index skips over any document that is missing the indexed field. The index is "sparse" because it does not include all documents of a collection. By contrast, non-sparse indexes contain all documents in a collection, storing null values for those documents that do not contain the indexed field.

Important: Changed in version 3.2: Starting in MongoDB 3.2, MongoDB provides the option to create *partial indexes* (page 570). Partial indexes offer a superset of the functionality of sparse indexes. If you are using MongoDB 3.2 or later, *partial indexes* (page 570) should be preferred over sparse indexes.

Create a Sparse Index

To create a `sparse` index, use the `db.collection.createIndex()` method with the `sparse` option set to `true`. For example, the following operation in the `mongo` shell creates a sparse index on the `xmpp_id` field of the `addresses` collection:

```
db.addresses.createIndex( { "xmpp_id": 1 }, { sparse: true } )
```

The index does not index documents that do not include the `xmpp_id` field.

Note: Do not confuse sparse indexes in MongoDB with block-level[14] indexes in other databases. Think of them as dense indexes with a specific filter.

---

[14] http://en.wikipedia.org/wiki/Database_index#Sparse_index

Behavior

`sparse` Index and Incomplete Results Changed in version 2.6.

If a sparse index would result in an incomplete result set for queries and sort operations, MongoDB will not use that index unless a `hint()` explicitly specifies the index.

For example, the query `{ x: { $exists: false } }` will not use a sparse index on the `x` field unless explicitly hinted. See *Sparse Index On A Collection Cannot Return Complete Results* (page 575) for an example that details the behavior.

Indexes that are `sparse` by Default *2dsphere (version 2)* (page 543), *2d* (page 557), `geoHaystack`, and *text* (page 533) indexes are always `sparse`.

`sparse` Compound Indexes Sparse *compound indexes* (page 522) that only contain ascending/descending index keys will index a document as long as the document contains at least one of the keys.

For sparse compound indexes that contain a geospatial key (i.e. *2dsphere* (page 543), *2d* (page 557), or `geoHaystack` index keys) along with ascending/descending index key(s), only the existence of the geospatial field(s) in a document determine whether the index references the document.

For sparse compound indexes that contain *text* (page 533) index keys along with ascending/descending index keys, only the existence of the `text` index field(s) determine whether the index references a document.

`sparse` and `unique` Properties An index that is both `sparse` and *unique* (page 568) prevents collection from having documents with duplicate values for a field but allows multiple documents that omit the key.

Examples

Create a Sparse Index On A Collection Consider a collection `scores` that contains the following documents:

```
{ "_id" : ObjectId("523b6e32fb408eea0eec2647"), "userid" : "newbie" }
{ "_id" : ObjectId("523b6e61fb408eea0eec2648"), "userid" : "abby", "score" : 82 }
{ "_id" : ObjectId("523b6e6ffb408eea0eec2649"), "userid" : "nina", "score" : 90 }
```

The collection has a sparse index on the field `score`:

```
db.scores.createIndex( { score: 1 } , { sparse: true } )
```

Then, the following query on the `scores` collection uses the sparse index to return the documents that have the `score` field less than (`$lt`) `90`:

```
db.scores.find( { score: { $lt: 90 } } )
```

Because the document for the userid `"newbie"` does not contain the `score` field and thus does not meet the query criteria, the query can use the sparse index to return the results:

```
{ "_id" : ObjectId("523b6e61fb408eea0eec2648"), "userid" : "abby", "score" : 82 }
```

Sparse Index On A Collection Cannot Return Complete Results Consider a collection `scores` that contains the following documents:

```
{ "_id" : ObjectId("523b6e32fb408eea0eec2647"), "userid" : "newbie" }
{ "_id" : ObjectId("523b6e61fb408eea0eec2648"), "userid" : "abby", "score" : 82 }
{ "_id" : ObjectId("523b6e6ffb408eea0eec2649"), "userid" : "nina", "score" : 90 }
```

The collection has a sparse index on the field `score`:

```
db.scores.createIndex( { score: 1 } , { sparse: true } )
```

Because the document for the userid `"newbie"` does not contain the `score` field, the sparse index does not contain an entry for that document.

Consider the following query to return **all** documents in the `scores` collection, sorted by the `score` field:

```
db.scores.find().sort( { score: -1 } )
```

Even though the sort is by the indexed field, MongoDB will **not** select the sparse index to fulfill the query in order to return complete results:

```
{ "_id" : ObjectId("523b6e6ffb408eea0eec2649"), "userid" : "nina", "score" : 90 }
{ "_id" : ObjectId("523b6e61fb408eea0eec2648"), "userid" : "abby", "score" : 82 }
{ "_id" : ObjectId("523b6e32fb408eea0eec2647"), "userid" : "newbie" }
```

To use the sparse index, explicitly specify the index with `hint()`:

```
db.scores.find().sort( { score: -1 } ).hint( { score: 1 } )
```

The use of the index results in the return of only those documents with the `score` field:

```
{ "_id" : ObjectId("523b6e6ffb408eea0eec2649"), "userid" : "nina", "score" : 90 }
{ "_id" : ObjectId("523b6e61fb408eea0eec2648"), "userid" : "abby", "score" : 82 }
```

See also:

`explain()` and *Analyze Query Performance* (page 159)

Sparse Index with Unique Constraint Consider a collection `scores` that contains the following documents:

```
{ "_id" : ObjectId("523b6e32fb408eea0eec2647"), "userid" : "newbie" }
{ "_id" : ObjectId("523b6e61fb408eea0eec2648"), "userid" : "abby", "score" : 82 }
{ "_id" : ObjectId("523b6e6ffb408eea0eec2649"), "userid" : "nina", "score" : 90 }
```

You could create an index with a *unique constraint* (page 568) and sparse filter on the `score` field using the following operation:

```
db.scores.createIndex( { score: 1 } , { sparse: true, unique: true } )
```

This index *would permit* the insertion of documents that had unique values for the `score` field *or* did not include a `score` field. As such, given the existing documents in the `scores` collection, the index permits the following *insert operations* (page 137):

```
db.scores.insert( { "userid": "AAAAAAA", "score": 43 } )
db.scores.insert( { "userid": "BBBBBBB", "score": 34 } )
db.scores.insert( { "userid": "CCCCCCC" } )
db.scores.insert( { "userid": "DDDDDDD" } )
```

However, the index *would not permit* the addition of the following documents since documents already exists with `score` value of `82` and `90`:

```
db.scores.insert( { "userid": "AAAAAAA", "score": 82 } )
db.scores.insert( { "userid": "BBBBBBB", "score": 90 } )
```

10.9.9 Index Build

On this page

- Background Construction (page 577)
- Index Names (page 578)
- View Index Build Operations (page 579)
- Terminate Index Build Operation (page 579)

By default, creating an index blocks all other operations on a database. When building an index on a collection, the database that holds the collection is unavailable for read or write operations until the index build completes. Any operation that requires a read or write lock on all databases (e.g. **listDatabases**) will wait for the foreground index build to complete.

Background Construction

For potentially long running index building operations, consider the `background` operation so that the MongoDB database remains available during the index building operation. For example, to create an index in the background of the `zipcode` field of the `people` collection, issue the following:

```
db.people.createIndex( { zipcode: 1}, {background: true} )
```

By default, `background` is `false` for building MongoDB indexes.

You can combine the background option with other options, as in the following:

```
db.people.createIndex( { zipcode: 1}, {background: true, sparse: true } )
```

Behavior

As of MongoDB version 2.4, a `mongod` instance can build more than one index in the background concurrently.

Changed in version 2.4: Before 2.4, a `mongod` instance could only build one background index per database at a time.

Background indexing operations run in the background so that other database operations can run while creating the index. However, the `mongo` shell session or connection where you are creating the index *will* block until the index build is complete. To continue issuing commands to the database, open another connection or `mongo` instance.

Queries will not use partially-built indexes: the index will only be usable once the index build is complete.

Note: If MongoDB is building an index in the background, you cannot perform other administrative operations involving that collection, including running `repairDatabase`, dropping the collection (i.e. `db.collection.drop()`), and running `compact`. These operations will return an error during background index builds.

Performance

The background index operation uses an incremental approach that is slower than the normal "foreground" index builds. If the index is larger than the available RAM, then the incremental process can take *much* longer than the foreground build.

If your application includes `createIndex()` operations, and an index *doesn't* exist for other operational concerns, building the index can have a severe impact on the performance of the database.

To avoid performance issues, make sure that your application checks for the indexes at start up using the `getIndexes()` method or the equivalent method for your driver[15] and terminates if the proper indexes do not exist. Always build indexes in production instances using separate application code, during designated maintenance windows.

Interrupted Index Builds

If a background index build is in progress when the `mongod` process terminates, when the instance restarts the index build will restart as foreground index build. If the index build encounters any errors, such as a duplicate key error, the `mongod` will exit with an error.

To start the `mongod` after a failed index build, use the `storage.indexBuildRetry` or `--noIndexBuildRetry` to skip the index build on start up.

Building Indexes on Secondaries

Changed in version 2.6: Secondary members can now build indexes in the background. Previously all index builds on secondaries were in the foreground.

Background index operations on a *replica set secondaries* begin after the *primary* completes building the index. If MongoDB builds an index in the background on the primary, the secondaries will then build that index in the background.

To build large indexes on secondaries the best approach is to restart one secondary at a time in *standalone* mode and build the index. After building the index, restart as a member of the replica set, allow it to catch up with the other members of the set, and then build the index on the next secondary. When all the secondaries have the new index, step down the primary, restart it as a standalone, and build the index on the former primary.

The amount of time required to build the index on a secondary must be within the window of the *oplog*, so that the secondary can catch up with the primary.

Indexes on secondary members in "recovering" mode are always built in the foreground to allow them to catch up as soon as possible.

See *Build Indexes on Replica Sets* (page 579) for a complete procedure for building indexes on secondaries.

Index Names

The default name for an index is the concatenation of the indexed keys and each key's direction in the index, 1 or -1.

Example

Issue the following command to create an index on `item` and `quantity`:

```
db.products.createIndex( { item: 1, quantity: -1 } )
```

The resulting index is named: `item_1_quantity_-1`.

Optionally, you can specify a name for an index instead of using the default name.

Example

Issue the following command to create an index on `item` and `quantity` and specify `inventory` as the index name:

[15] https://api.mongodb.org/

```
db.products.createIndex( { item: 1, quantity: -1 } , { name: "inventory" } )
```

The resulting index has the name `inventory`.

To view the name of an index, use the `getIndexes()` method.

View Index Build Operations

To see the status of an index build operation, you can use the `db.currentOp()` method in the `mongo` shell. To filter the current operations for index creation operations, see *currentOp-index-creation* for an example.

The `msg` field will include the percent of the build that is complete.

Terminate Index Build Operation

To terminate an ongoing index build, use the `db.killOp()` method in the `mongo` shell. For index builds, the effects of `db.killOp()` may not be immediate and may occur well after much of the index build operation has completed.

You cannot terminate a *replicated* index build on secondary members of a replica set. To minimize the impact of building an index on replica sets, see *Build Indexes on Replica Sets* (page 579).

Changed in version 2.4: Before MongoDB 2.4, you could *only* terminate *background* index builds. After 2.4, you can terminate both *background* index builds and foreground index builds.

See also:

`db.currentOp()`, `db.killOp()`

Build Indexes on Replica Sets

On this page

- Considerations (page 579)
- Procedure (page 580)

For replica sets, secondaries will begin building indexes *after* the *primary* finishes building the index. In *sharded clusters*, the `mongos` will send `createIndex()` to the primary members of the replica set for each shard, which then replicate to the secondaries after the primary finishes building the index.

To minimize the impact of building an index on your replica set, use the following procedure to build indexes.

Considerations

- Ensure that your *oplog* is large enough to permit the indexing or re-indexing operation to complete without falling too far behind to catch up. See the *oplog sizing* (page 657) documentation for additional information.
- This procedure *does* take one member out of the replica set at a time. However, this procedure will only affect one member of the set at a time rather than *all* secondaries at the same time.
- Before version 2.6 *Background index creation operations* (page 577) become *foreground* indexing operations on *secondary* members of replica sets. After 2.6, background index builds replicate as background index builds on the secondaries.

Procedure

Note: If you need to build an index in a *sharded cluster*, repeat the following procedure for each replica set that provides each *shard*.

Stop One Secondary Stop the `mongod` process on one secondary. Restart the `mongod` process *without* the `--replSet` option and running on a different port. [16] This instance is now in "standalone" mode.

For example, if your `mongod` *normally* runs with on the default port of `27017` with the `--replSet` option you would use the following invocation:

```
mongod --port 47017
```

Build the Index Create the new index using the `createIndex()` in the `mongo` shell, or comparable method in your driver. This operation will create or rebuild the index on this `mongod` instance

For example, to create an ascending index on the `username` field of the `records` collection, use the following `mongo` shell operation:

```
db.records.createIndex( { username: 1 } )
```

Restart the Program `mongod` When the index build completes, start the `mongod` instance with the `--replSet` option on its usual port:

```
mongod --port 27017 --replSet rs0
```

Modify the port number (e.g. `27017`) or the replica set name (e.g. `rs0`) as needed.

Allow replication to catch up on this member.

Build Indexes on all Secondaries Changed in version 2.6: Secondary members can now *build indexes in the background* (page 577). Previously all index builds on secondaries were in the foreground.

For each secondary in the set, build an index according to the following steps:

1. *Stop One Secondary* (page 580)
2. *Build the Index* (page 580)
3. *Restart the Program mongod* (page 580)

Build the Index on the Primary To build an index on the primary you can either:

1. *Build the index in the background* (page 577) on the primary.
2. Step down the primary using the `rs.stepDown()` method in the `mongo` shell to cause the current primary to become a secondary graceful and allow the set to elect another member as primary.

   Then repeat the index building procedure, listed below, to build the index on the primary:

   (a) *Stop One Secondary* (page 580)

   (b) *Build the Index* (page 580)

   (c) *Restart the Program mongod* (page 580)

[16] By running the `mongod` on a different port, you ensure that the other members of the replica set and all clients will not contact the member while you are building the index.

Building the index on the background, takes longer than the foreground index build and results in a less compact index structure. Additionally, the background index build may impact write performance on the primary. However, building the index in the background allows the set to be continuously up for write operations while MongoDB builds the index.

10.9.10 Index Intersection

On this page

- Index Prefix Intersection (page 581)
- Index Intersection and Compound Indexes (page 581)
- Index Intersection and Sort (page 582)

New in version 2.6.

MongoDB can use the intersection of multiple indexes to fulfill queries. [17] In general, each index intersection involves two indexes; however, MongoDB can employ multiple/nested index intersections to resolve a query.

To illustrate index intersection, consider a collection `orders` that has the following indexes:

```
{ qty: 1 }
{ item: 1 }
```

MongoDB can use the intersection of the two indexes to support the following query:

```
db.orders.find( { item: "abc123", qty: { $gt: 15 } } )
```

To determine if MongoDB used index intersection, run `explain()`; the results of *explain()* will include either an `AND_SORTED` stage or an `AND_HASH` stage.

Index Prefix Intersection

With index intersection, MongoDB can use an intersection of either the entire index or the index prefix. An index prefix is a subset of a compound index, consisting of one or more keys starting from the beginning of the index.

Consider a collection `orders` with the following indexes:

```
{ qty: 1 }
{ status: 1, ord_date: -1 }
```

To fulfill the following query which specifies a condition on both the `qty` field and the `status` field, MongoDB can use the intersection of the two indexes:

```
db.orders.find( { qty: { $gt: 10 } , status: "A" } )
```

Index Intersection and Compound Indexes

Index intersection does not eliminate the need for creating *compound indexes* (page 522). However, because both the list order (i.e. the order in which the keys are listed in the index) and the sort order (i.e. ascending or descending), matter in *compound indexes* (page 522), a compound index may not support a query condition that does not include the *index prefix keys* (page 524) or that specifies a different sort order.

For example, if a collection `orders` has the following compound index, with the `status` field listed before the `ord_date` field:

[17] In previous versions, MongoDB could use only a single index to fulfill most queries. The exception to this is queries with `$or` clauses, which could use a single index for each `$or` clause.

```
{ status: 1, ord_date: -1 }
```

The compound index can support the following queries:

```
db.orders.find( { status: { $in: ["A", "P" ] } } )
db.orders.find(
   {
     ord_date: { $gt: new Date("2014-02-01") },
     status: {$in:[ "P", "A" ] }
   }
)
```

But not the following two queries:

```
db.orders.find( { ord_date: { $gt: new Date("2014-02-01") } } )
db.orders.find( { } ).sort( { ord_date: 1 } )
```

However, if the collection has two separate indexes:

```
{ status: 1 }
{ ord_date: -1 }
```

The two indexes can, either individually or through index intersection, support all four aforementioned queries.

The choice between creating compound indexes that support your queries or relying on index intersection depends on the specifics of your system.

See also:

compound indexes (page 522), *Create Compound Indexes to Support Several Different Queries* (page 587)

Index Intersection and Sort

Index intersection does not apply when the `sort()` operation requires an index completely separate from the query predicate.

For example, the `orders` collection has the following indexes:

```
{ qty: 1 }
{ status: 1, ord_date: -1 }
{ status: 1 }
{ ord_date: -1 }
```

MongoDB cannot use index intersection for the following query with sort:

```
db.orders.find( { qty: { $gt: 10 } } ).sort( { status: 1 } )
```

That is, MongoDB does not use the `{ qty: 1 }` index for the query, and the separate `{ status: 1 }` or the `{ status: 1, ord_date: -1 }` index for the sort.

However, MongoDB can use index intersection for the following query with sort since the index `{ status: 1, ord_date: -1 }` can fulfill part of the query predicate.

```
db.orders.find( { qty: { $gt: 10 } , status: "A" } ).sort( { ord_date: -1 } )
```

10.9.11 Manage Indexes

On this page

- View Existing Indexes (page 583)
- Remove Indexes (page 583)
- Modify an Index (page 584)
- Rebuild Indexes (page 584)

The following procedures provides some common procedures for managing existing indexes. For instructions on creating indexes, refer to the specific index type pages.

View Existing Indexes

List all Indexes on a Collection

To return a list of all indexes on a collection, use the `db.collection.getIndexes()` method or a similar method for your driver[18].

For example, to view all indexes on the `people` collection:

```
db.people.getIndexes()
```

List all Indexes for a Database

To list all indexes on all collections in a database, you can use the following operation in the `mongo` shell:

```
db.getCollectionNames().forEach(function(collection) {
   indexes = db[collection].getIndexes();
   print("Indexes for " + collection + ":");
   printjson(indexes);
});
```

MongoDB 3.0 deprecates direct access to the `system.indexes` collection.

For MongoDB 3.0 deployments using the *WiredTiger* (page 595) storage engine, if you run `db.getCollectionNames()` and `db.collection.getIndexes()` from a version of the `mongo` shell before 3.0 or a version of the driver prior to *3.0 compatible version* (page 950), `db.getCollectionNames()` and `db.collection.getIndexes()` will return no data, even if there are existing collections and indexes. For more information, see *WiredTiger and Driver Version Compatibility* (page 946).

Remove Indexes

To remove an index from a collection, you can use the `db.collection.dropIndex()` method. To rebuild indexes, see *Rebuild Indexes* (page 584) instead.

Remove a Specific Index

To remove an index, use the `db.collection.dropIndex()` method.

For example, the following operation removes an ascending index on the `tax-id` field in the `accounts` collection:

[18] https://api.mongodb.org/

```
db.accounts.dropIndex( { "tax-id": 1 } )
```

The operation returns a document with the status of the operation:

```
{ "nIndexesWas" : 3, "ok" : 1 }
```

Where the value of `nIndexesWas` reflects the number of indexes *before* removing this index.

For *text* (page 533) indexes, pass the index name to the `db.collection.dropIndex()` method. See *Use the Index Name to Drop a text Index* (page 541) for details.

Remove All Indexes

You can also use the `db.collection.dropIndexes()` to remove *all* indexes, except for the *_id index* (page 515) from a collection.

These shell helpers provide wrappers around the `dropIndexes` *database command*. Your `client library` may have a different or additional interface for these operations.

Modify an Index

To modify an existing index, you need to drop and recreate the index with the exception of m TTL indexes.

If you need to rebuild indexes for a collection you can use the `db.collection.reIndex()` method to rebuild all indexes on a collection in a single operation. This operation drops all indexes, including the *_id index* (page 515), and then rebuilds all indexes.

Rebuild Indexes

If you need to rebuild indexes for a collection you can use the `db.collection.reIndex()` method to rebuild all indexes on a collection in a single operation. This operation drops all indexes for a collection, including the `_id` index, and then rebuilds all indexes.

Note: For replica sets, `db.collection.reIndex()` will not propagate from the *primary* to *secondaries*. `db.collection.reIndex()` will only affect a single `mongod` instance.

Important: `db.collection.reIndex()` will rebuild indexes in the *background* (page 577) *if the index was originally specified with this option*. However, `db.collection.reIndex()` will rebuild the `_id` index in the foreground, which takes the database's write lock.

```
db.accounts.reIndex()
```

This shell helper provides a wrapper around the `reIndex` *database command*. Your `client library` may have a different or additional interface for this operation.

To build or rebuild indexes for a *replica set*, see *Build Indexes on Replica Sets* (page 579).

10.9.12 Measure Index Use

On this page

- Get Index Access Information with `$indexStats` (page 585)
- Return Query Plan with `explain()` (page 585)
- Control Index Use with `hint()` (page 585)
- Instance Index Use Reporting (page 586)

Get Index Access Information with `$indexStats`

Use `$indexStats` to get usage statistics about an index.

Return Query Plan with `explain()`

Use the `db.collection.explain()` or the `cursor.explain()` method in *executionStats* mode to return statistics about the query process, including the index used, the number of documents scanned, and the time the query takes to process in milliseconds.

Run `db.collection.explain()` or the `cursor.explain()` method in *allPlansExecution* mode to view partial execution statistics collected during plan selection.

`db.collection.explain()` provides information on the execution of other operations, such as `db.collection.update()`. See `db.collection.explain()` for details.

Control Index Use with `hint()`

To *force* MongoDB to use a particular index for a `db.collection.find()` operation, specify the index with the `hint()` method. Append the `hint()` method to the `find()` method. Consider the following example:

```
db.people.find(
   { name: "John Doe", zipcode: { $gt: "63000" } }
).hint( { zipcode: 1 } )
```

To view the execution statistics for a specific index, append to the `db.collection.find()` the `hint()` method followed by `cursor.explain()`, e.g.:

```
db.people.find(
   { name: "John Doe", zipcode: { $gt: "63000" } }
).hint( { zipcode: 1 } ).explain("executionStats")
```

Or, append `hint()` method to `db.collection.explain().find()`:

```
db.people.explain("executionStats").find(
   { name: "John Doe", zipcode: { $gt: "63000" } }
).hint( { zipcode: 1 } )
```

Specify the `$natural` operator to the `hint()` method to prevent MongoDB from using *any* index:

```
db.people.find(
   { name: "John Doe", zipcode: { $gt: "63000" } }
).hint( { $natural: 1 } )
```

Instance Index Use Reporting

MongoDB provides a number of metrics of index use and operation that you may want to consider when analyzing index use for your database:

In the output of `serverStatus`:	`metrics.queryExecutor.scanned` `metrics.operation.scanAndOrder`
In the output of `collStats`:	`totalIndexSize` `indexSizes`
In the output of `dbStats`:	`dbStats.indexes` `dbStats.indexSize`

10.9.13 Indexing Strategies

The best indexes for your application must take a number of factors into account, including the kinds of queries you expect, the ratio of reads to writes, and the amount of free memory on your system.

When developing your indexing strategy you should have a deep understanding of your application's queries. Before you build indexes, map out the types of queries you will run so that you can build indexes that reference those fields. Indexes come with a performance cost, but are more than worth the cost for frequent queries on large data set. Consider the relative frequency of each query in the application and whether the query justifies an index.

The best overall strategy for designing indexes is to profile a variety of index configurations with data sets similar to the ones you'll be running in production to see which configurations perform best.Inspect the current indexes created for your collections to ensure they are supporting your current and planned queries. If an index is no longer used, drop the index.

Generally, MongoDB only uses *one* index to fulfill most queries. However, each clause of an `$or` query may use a different index, and starting in 2.6, MongoDB can use an *intersection* (page 581) of multiple indexes.

The following documents introduce indexing strategies:

Create Indexes to Support Your Queries **(page 586)** An index supports a query when the index contains all the fields scanned by the query. Creating indexes that supports queries results in greatly increased query performance.

Use Indexes to Sort Query Results **(page 587)** To support efficient queries, use the strategies here when you specify the sequential order and sort order of index fields.

Ensure Indexes Fit in RAM **(page 589)** When your index fits in RAM, the system can avoid reading the index from disk and you get the fastest processing.

Create Queries that Ensure Selectivity **(page 590)** Selectivity is the ability of a query to narrow results using the index. Selectivity allows MongoDB to use the index for a larger portion of the work associated with fulfilling the query.

Create Indexes to Support Your Queries

On this page

- Create a Single-Key Index if All Queries Use the Same, Single Key (page 587)
- Create Compound Indexes to Support Several Different Queries (page 587)

An index supports a query when the index contains all the fields scanned by the query. The query scans the index and not the collection. Creating indexes that support queries results in greatly increased query performance.

This document describes strategies for creating indexes that support queries.

Create a Single-Key Index if All Queries Use the Same, Single Key

If you only ever query on a single key in a given collection, then you need to create just one single-key index for that collection. For example, you might create an index on `category` in the `product` collection:

```
db.products.createIndex( { "category": 1 } )
```

Create Compound Indexes to Support Several Different Queries

If you sometimes query on only one key and at other times query on that key combined with a second key, then creating a compound index is more efficient than creating a single-key index. MongoDB will use the compound index for both queries. For example, you might create an index on both `category` and `item`.

```
db.products.createIndex( { "category": 1, "item": 1 } )
```

This allows you both options. You can query on just `category`, and you also can query on `category` combined with `item`. A single *compound index* (page 522) on multiple fields can support all the queries that search a "prefix" subset of those fields.

Example

The following index on a collection:

```
{ x: 1, y: 1, z: 1 }
```

Can support queries that the following indexes support:

```
{ x: 1 }
{ x: 1, y: 1 }
```

There are some situations where the prefix indexes may offer better query performance: for example if `z` is a large array.

The `{ x: 1, y: 1, z: 1 }` index can also support many of the same queries as the following index:

```
{ x: 1, z: 1 }
```

Also, `{ x: 1, z: 1 }` has an additional use. Given the following query:

```
db.collection.find( { x: 5 } ).sort( { z: 1} )
```

The `{ x: 1, z: 1 }` index supports both the query and the sort operation, while the `{ x: 1, y: 1, z: 1 }` index only supports the query. For more information on sorting, see *Use Indexes to Sort Query Results* (page 587).

Starting in version 2.6, MongoDB can use *index intersection* (page 581) to fulfill queries. The choice between creating compound indexes that support your queries or relying on index intersection depends on the specifics of your system. See *Index Intersection and Compound Indexes* (page 581) for more details.

Use Indexes to Sort Query Results

On this page

- Sort with a Single Field Index (page 588)
- Sort on Multiple Fields (page 588)

In MongoDB, sort operations can obtain the sort order by retrieving documents based on the ordering in an index. If the query planner cannot obtain the sort order from an index, it will sort the results in memory. Sort operations that use an index often have better performance than those that do not use an index. In addition, sort operations that do *not* use an index will abort when they use 32 megabytes of memory.

Sort with a Single Field Index

If an ascending or a descending index is on a single field, the sort operation on the field can be in either direction.

For example, create an ascending index on the field `a` for a collection `records`:

```
db.records.createIndex( { a: 1 } )
```

This index can support an ascending sort on `a`:

```
db.records.find().sort( { a: 1 } )
```

The index can also support the following descending sort on `a` by traversing the index in reverse order:

```
db.records.find().sort( { a: -1 } )
```

Sort on Multiple Fields

Create a *compound index* (page 522) to support sorting on multiple fields.

You can specify a sort on all the keys of the index or on a subset; however, the sort keys must be listed in the *same order* as they appear in the index. For example, an index key pattern `{ a: 1, b: 1 }` can support a sort on `{ a: 1, b: 1 }` but *not* on `{ b: 1, a: 1 }`.

The sort must specify the *same sort direction* (i.e.ascending/descending) for all its keys as the index key pattern or specify the *reverse sort direction* for all its keys as the index key pattern. For example, an index key pattern `{ a: 1, b: 1 }` can support a sort on `{ a: 1, b: 1 }` and `{ a: -1, b: -1 }` but *not* on `{ a: -1, b: 1 }`.

Sort and Index Prefix If the sort keys correspond to the index keys or an index *prefix*, MongoDB can use the index to sort the query results. A *prefix* of a compound index is a subset that consists of one or more keys at the start of the index key pattern.

For example, create a compound index on the `data` collection:

```
db.data.createIndex( { a:1, b: 1, c: 1, d: 1 } )
```

Then, the following are prefixes for that index:

```
{ a: 1 }
{ a: 1, b: 1 }
{ a: 1, b: 1, c: 1 }
```

The following query and sort operations use the index prefixes to sort the results. These operations do not need to sort the result set in memory.

Example	Index Prefix
`db.data.find().sort( { a: 1 } )`	`{ a: 1 }`
`db.data.find().sort( { a: -1 } )`	`{ a: 1 }`
`db.data.find().sort( { a: 1, b: 1 } )`	`{ a: 1, b: 1 }`
`db.data.find().sort( { a: -1, b: -1 } )`	`{ a: 1, b: 1 }`
`db.data.find().sort( { a: 1, b: 1, c: 1 } )`	`{ a: 1, b: 1, c: 1 }`
`db.data.find( { a: { $gt: 4 } } ).sort( { a: 1, b: 1 } )`	`{ a: 1, b: 1 }`

Consider the following example in which the prefix keys of the index appear in both the query predicate and the sort:

```
db.data.find( { a: { $gt: 4 } } ).sort( { a: 1, b: 1 } )
```

In such cases, MongoDB can use the index to retrieve the documents in order specified by the sort. As the example shows, the index prefix in the query predicate can be different from the prefix in the sort.

Sort and Non-prefix Subset of an Index An index can support sort operations on a non-prefix subset of the index key pattern. To do so, the query must include **equality** conditions on all the prefix keys that precede the sort keys.

For example, the collection `data` has the following index:

```
{ a: 1, b: 1, c: 1, d: 1 }
```

The following operations can use the index to get the sort order:

Example	Index Prefix
`db.data.find( { a: 5 } ).sort( { b: 1, c: 1 } )`	`{ a: 1 , b: 1, c: 1 }`
`db.data.find( { b: 3, a: 4 } ).sort( { c: 1 } )`	`{ a: 1, b: 1, c: 1 }`
`db.data.find( { a: 5, b: { $lt: 3} } ).sort( { b: 1 } )`	`{ a: 1, b: 1 }`

As the last operation shows, only the index fields *preceding* the sort subset must have the equality conditions in the query document; the other index fields may specify other conditions.

If the query does **not** specify an equality condition on an index prefix that precedes or overlaps with the sort specification, the operation will **not** efficiently use the index. For example, the following operations specify a sort document of `{ c: 1 }`, but the query documents do not contain equality matches on the preceding index fields `a` and `b`:

```
db.data.find( { a: { $gt: 2 } } ).sort( { c: 1 } )
db.data.find( { c: 5 } ).sort( { c: 1 } )
```

These operations **will not** efficiently use the index `{ a: 1, b: 1, c: 1, d: 1 }` and may not even use the index to retrieve the documents.

Ensure Indexes Fit in RAM

On this page

- Indexes that Hold Only Recent Values in RAM (page 590)

For the fastest processing, ensure that your indexes fit entirely in RAM so that the system can avoid reading the index from disk.

To check the size of your indexes, use the `db.collection.totalIndexSize()` helper, which returns data in bytes:

```
> db.collection.totalIndexSize()
4294976499
```

The above example shows an index size of almost 4.3 gigabytes. To ensure this index fits in RAM, you must not only have more than that much RAM available but also must have RAM available for the rest of the *working set*. Also remember:

If you have and use multiple collections, you must consider the size of all indexes on all collections. The indexes and the working set must be able to fit in memory at the same time.

There are some limited cases where indexes do not need to fit in memory. See *Indexes that Hold Only Recent Values in RAM* (page 590).

See also:

`collStats` and `db.collection.stats()`

Indexes that Hold Only Recent Values in RAM

Indexes do not have to fit *entirely* into RAM in all cases. If the value of the indexed field increments with every insert, and most queries select recently added documents; then MongoDB only needs to keep the parts of the index that hold the most recent or "right-most" values in RAM. This allows for efficient index use for read and write operations and minimize the amount of RAM required to support the index.

Create Queries that Ensure Selectivity

Selectivity is the ability of a query to narrow results using the index. Effective indexes are more selective and allow MongoDB to use the index for a larger portion of the work associated with fulfilling the query.

To ensure selectivity, write queries that limit the number of possible documents with the indexed field. Write queries that are appropriately selective relative to your indexed data.

Example

Suppose you have a field called `status` where the possible values are `new` and `processed`. If you add an index on `status` you've created a low-selectivity index. The index will be of little help in locating records.

A better strategy, depending on your queries, would be to create a *compound index* (page 522) that includes the low-selectivity field and another field. For example, you could create a compound index on `status` and `created_at`.

Another option, again depending on your use case, might be to use separate collections, one for each status.

Example

Consider an index `{ a : 1 }` (i.e. an index on the key `a` sorted in ascending order) on a collection where `a` has three values evenly distributed across the collection:

```
{ _id: ObjectId(), a: 1, b: "ab" }
{ _id: ObjectId(), a: 1, b: "cd" }
{ _id: ObjectId(), a: 1, b: "ef" }
{ _id: ObjectId(), a: 2, b: "jk" }
{ _id: ObjectId(), a: 2, b: "lm" }
{ _id: ObjectId(), a: 2, b: "no" }
{ _id: ObjectId(), a: 3, b: "pq" }
```

```
{ _id: ObjectId(), a: 3, b: "rs" }
{ _id: ObjectId(), a: 3, b: "tv" }
```

If you query for `{ a: 2, b: "no" }` MongoDB must scan 3 *documents* in the collection to return the one matching result. Similarly, a query for `{ a: { $gt: 1}, b: "tv" }` must scan 6 documents, also to return one result.

Consider the same index on a collection where `a` has *nine* values evenly distributed across the collection:

```
{ _id: ObjectId(), a: 1, b: "ab" }
{ _id: ObjectId(), a: 2, b: "cd" }
{ _id: ObjectId(), a: 3, b: "ef" }
{ _id: ObjectId(), a: 4, b: "jk" }
{ _id: ObjectId(), a: 5, b: "lm" }
{ _id: ObjectId(), a: 6, b: "no" }
{ _id: ObjectId(), a: 7, b: "pq" }
{ _id: ObjectId(), a: 8, b: "rs" }
{ _id: ObjectId(), a: 9, b: "tv" }
```

If you query for `{ a: 2, b: "cd" }`, MongoDB must scan only one document to fulfill the query. The index and query are more selective because the values of `a` are evenly distributed *and* the query can select a specific document using the index.

However, although the index on `a` is more selective, a query such as `{ a: { $gt: 5 }, b: "tv" }` would still need to scan 4 documents.

If overall selectivity is low, and if MongoDB must read a number of documents to return results, then some queries may perform faster without indexes. To determine performance, see *Measure Index Use* (page 584).

10.9.14 Indexing Reference

On this page

- Indexing Methods in the `mongo` Shell (page 592)
- Indexing Database Commands (page 592)
- Geospatial Query Selectors (page 592)
- Indexing Query Modifiers (page 593)

Indexing Methods in the mongo Shell

Name	Description
db.collection.createIndex()	Builds an index on a collection.
db.collection.dropIndex()	Removes a specified index on a collection.
db.collection.dropIndexes()	Removes all indexes on a collection.
db.collection.getIndexes()	Returns an array of documents that describe the existing indexes on a collection.
db.collection.reIndex()	Rebuilds all existing indexes on a collection.
db.collection.totalIndexSize()	Reports the total size used by the indexes on a collection. Provides a wrapper around the totalIndexSize field of the collStats output.
cursor.explain()	Reports on the query execution plan for a cursor.
cursor.hint()	Forces MongoDB to use a specific index for a query.
cursor.max()	Specifies an exclusive upper index bound for a cursor. For use with cursor.hint().
cursor.min()	Specifies an inclusive lower index bound for a cursor. For use with cursor.hint().
cursor.snapshot()	Forces the cursor to use the index on the _id field. Ensures that the cursor returns each document, with regards to the value of the _id field, only once.

Indexing Database Commands

Name	Description
createIndexes	Builds one or more indexes for a collection.
dropIndexes	Removes indexes from a collection.
compact	Defragments a collection and rebuilds the indexes.
reIndex	Rebuilds all indexes on a collection.
validate	Internal command that scans for a collection's data and indexes for correctness.
geoNear	Performs a geospatial query that returns the documents closest to a given point.
geoSearch	Performs a geospatial query that uses MongoDB's *haystack index* functionality.
checkShardingIndex	Internal command that validates index on shard key.

Geospatial Query Selectors

Name	Description
$geoWithin	Selects geometries within a bounding *GeoJSON geometry* (page 554). The *2dsphere* (page 543) and *2d* (page 557) indexes support $geoWithin.
$geoIntersects	Selects geometries that intersect with a *GeoJSON* geometry. The *2dsphere* (page 543) index supports $geoIntersects.
$near	Returns geospatial objects in proximity to a point. Requires a geospatial index. The *2dsphere* (page 543) and *2d* (page 557) indexes support $near.
$nearSphere	Returns geospatial objects in proximity to a point on a sphere. Requires a geospatial index. The *2dsphere* (page 543) and *2d* (page 557) indexes support $nearSphere.

Indexing Query Modifiers

Name	Description
`$explain`	Forces MongoDB to report on query execution plans. See `explain()`.
`$hint`	Forces MongoDB to use a specific index. See `hint()`
`$max`	Specifies an *exclusive* upper limit for the index to use in a query. See `max()`.
`$min`	Specifies an *inclusive* lower limit for the index to use in a query. See `min()`.
`$returnKey`	Forces the cursor to only return fields included in the index.
`$snapshot`	Guarantees that a query returns each document no more than once. See `snapshot()`.

10.10 Additional Resources

- Quick Reference Cards[19]

[19] https://www.mongodb.com/lp/misc/quick-reference-cards?jmp=docs

CHAPTER 11

Storage

The *storage engine* (page 595) is the primary component of MongoDB responsible for managing data. MongoDB provides a variety of storage engines, allowing you to choose one most suited to your application.

The *journal* is a log that helps the database recover in the event of a hard shutdown. There are several configurable options that allows the journal to strike a balance between performance and reliability that works for your particular use case.

GridFS (page 611) is a versatile storage system that is suited to handling large files, such as those exceeding the 16 MB document size limit.

11.1 Storage Engines

The *storage engine* is the component of the database that is responsible for managing how data is stored, both in memory and on disk. MongoDB supports multiple storage engines, as different engines perform better for specific workloads. Choosing the appropriate storage engine for your use case can significantly impact the performance of your applications.

WiredTiger (page 595) is the default storage engine starting in MongoDB 3.2. It is well-suited for most workloads and is recommended for new deployments. WiredTiger provides a document-level concurrency model, checkpointing, and compression, among other features. In MongoDB Enterprise, WiredTiger also supports *Encryption At Rest* (page 461).

MMAPv1 (page 603) is the original MongoDB storage engine and is the default storage engine for MongoDB versions before 3.2. It performs well on workloads with high volumes of reads and writes, as well as in-place updates.

The *In-Memory Storage Engine* (page 605) is available in MongoDB Enterprise. Rather than storing documents on-disk, it retains them in-memory for more predictable data latencies. This storage engine is in **beta** – do **not** use in production.

11.1.1 WiredTiger Storage Engine

On this page

- Document Level Concurrency (page 596)
- Snapshots and Checkpoints (page 596)
- Journal (page 596)
- Compression (page 597)
- Memory Use (page 597)

Starting in MongoDB 3.0, the WiredTiger storage engine is available in the 64-bit builds.

Changed in version 3.2: The WiredTiger storage engine is the default storage engine starting in MongoDB 3.2. For existing deployments, if you do not specify the `--storageEngine` or the `storage.engine` setting, MongoDB 3.2 can automatically determine the storage engine used to create the data files in the `--dbpath` or `storage.dbPath`. See *Default Storage Engine Change* (page 898).

Document Level Concurrency

WiredTiger uses *document-level* concurrency control for write operations. As a result, multiple clients can modify different documents of a collection at the same time.

For most read and write operations, WiredTiger uses optimistic concurrency control. WiredTiger uses only intent locks at the global, database and collection levels. When the storage engine detects conflicts between two operations, one will incur a write conflict causing MongoDB to transparently retry that operation.

Some global operations, typically short lived operations involving multiple databases, still require a global "instance-wide" lock. Some other operations, such as dropping a collection, still require an exclusive database lock.

Snapshots and Checkpoints

WiredTiger uses MultiVersion Concurrency Control (MVCC). At the start of an operation, WiredTiger provides a point-in-time snapshot of the data to the transaction. A snapshot presents a consistent view of the in-memory data.

When writing to disk, WiredTiger writes all the data in a snapshot to disk in a consistent way across all data files. The now-*durable* data act as a *checkpoint* in the data files. The *checkpoint* ensures that the data files are consistent up to and including the last checkpoint; i.e. checkpoints can act as recovery points.

MongoDB configures WiredTiger to create checkpoints (i.e. write the snapshot data to disk) at intervals of 60 seconds or 2 gigabytes of journal data.

During the write of a new checkpoint, the previous checkpoint is still valid. As such, even if MongoDB terminates or encounters an error while writing a new checkpoint, upon restart, MongoDB can recover from the last valid checkpoint.

The new checkpoint becomes accessible and permanent when WiredTiger's metadata table is atomically updated to reference the new checkpoint. Once the new checkpoint is accessible, WiredTiger frees pages from the old checkpoints.

Using WiredTiger, even without *journaling* (page 596), MongoDB can recover from the last checkpoint; however, to recover changes made after the last checkpoint, run with *journaling* (page 596).

Journal

WiredTiger uses a write-ahead transaction log in combination with *checkpoints* (page 596) to ensure data durability.

The WiredTiger journal persists all data modifications between checkpoints. If MongoDB exits between checkpoints, it uses the journal to replay all data modified since the last checkpoint. For information on the frequency with which MongoDB writes the journal data to disk, see *Journaling Process* (page 607).

WiredTiger journal is compressed using the *snappy* compression library. To specify an alternate compression algorithm or no compression, use the `storage.wiredTiger.engineConfig.journalCompressor` setting.

Note: Minimum log record size for WiredTiger is 128 bytes. If a log record is 128 bytes or smaller, WiredTiger does not compress that record.

You can disable journaling by setting `storage.journal.enabled` to `false`, which can reduce the overhead of maintaining the journal.

For *standalone* instances, not using the journal means that you will lose some data modifications when MongoDB exits unexpectedly between checkpoints. For members of *replica sets*, the replication process may provide sufficient durability guarantees.

See also:

Journaling with WiredTiger (page 606)

Compression

With WiredTiger, MongoDB supports compression for all collections and indexes. Compression minimizes storage use at the expense of additional CPU.

By default, WiredTiger uses block compression with the *snappy* compression library for all collections and *prefix compression* for all indexes.

For collections, block compression with *zlib* is also available. To specify an alternate compression algorithm or no compression, use the `storage.wiredTiger.collectionConfig.blockCompressor` setting.

For indexes, to disable *prefix compression*, use the `storage.wiredTiger.indexConfig.prefixCompression` setting.

Compression settings are also configurable on a per-collection and per-index basis during collection and index creation. See *create-collection-storage-engine-options* and *db.collection.createIndex() storageEngine option*.

For most workloads, the default compression settings balance storage efficiency and processing requirements.

The WiredTiger journal is also compressed by default. For information on journal compression, see *Journal* (page 596).

Memory Use

With WiredTiger, MongoDB utilizes both the WiredTiger cache and the filesystem cache.

Changed in version 3.2: Starting in MongoDB 3.2, the WiredTiger cache, by default, will use the larger of either:

- 60% of RAM minus 1 GB, or
- 1 GB.

For systems with up to 10 GB of RAM, the new default setting is less than or equal to the 3.0 default setting (For MongoDB 3.0, the WiredTiger cache uses either 1 GB or half of the installed physical RAM, whichever is larger).

For systems with more than 10 GB of RAM, the new default setting is greater than the 3.0 setting.

Via the filesystem cache, MongoDB automatically uses all free memory that is not used by the WiredTiger cache or by other processes. Data in the filesystem cache is compressed.

To adjust the size of the WiredTiger cache, see `storage.wiredTiger.engineConfig.cacheSizeGB` and `--wiredTigerCacheSizeGB`. Avoid increasing the WiredTiger cache size above its default value.

See also:

http://wiredtiger.com

Change Standalone to WiredTiger

New in version 3.0: The WiredTiger storage engine is available.

Changed in version 3.2: WiredTiger is the new default storage engine for MongoDB.

This tutorial gives an overview of changing the storage engine of a *standalone* MongoDB instance to *WiredTiger* (page 595).

Considerations This tutorial uses the `mongodump` and `mongorestore` utilities to export and import data. Ensure that these MongoDB package components are installed and updated on your system. In addition, make sure you have sufficient drive space available for the `mongodump` export file and the data files of your new `mongod` instance running with WiredTiger.

You must be using MongoDB version 3.0 or greater in order to use the WiredTiger storage engine. If upgrading from an earlier version of MongoDB, see the guides on *Upgrading to MongoDB 3.0* (page 953) or *Upgrading to MongoDB 3.2* (page 901) before proceeding with changing your storage engine.

Procedure

Step 1: Start the `mongod` you wish to change to WiredTiger. If `mongod` is already running, you can skip this step.

Step 2: Export data using `mongodump`.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 3: Create a data directory for the new `mongod` running with WiredTiger. Create a data directory for the new `mongod` instance that will run with the WiredTiger storage engine. `mongod` must have read and write permissions for this directory.

`mongod` with WiredTiger will not start with data files created with a different storage engine.

Step 4: Start `mongod` with WiredTiger. Start `mongod`, specifying `wiredTiger` as the *`--storageEngine`* and the newly created data directory for WiredTiger as the *`--dbpath`*. Specify additional options as appropriate.

```
mongod --storageEngine wiredTiger --dbpath <newWiredTigerDBPath>
```

You can also specify the options in a `configuration file`. To specify the storage engine, use the `storage.engine` setting.

Step 5: Upload the exported data using `mongorestore`.

```
mongorestore <exportDataDestination>
```

Specify additional options as appropriate. See `mongorestore` for available options.

Change Replica Set to WiredTiger

New in version 3.0: The WiredTiger storage engine is available. Also, replica sets may have members with different storage engines.

Changed in version 3.2: WiredTiger is the new default storage engine for MongoDB.

This tutorial gives an overview of changing the storage engine of a member of a *replica set* to *WiredTiger* (page 595).

Considerations Replica sets can have members with different storage engines. As such, you can update members to use the WiredTiger storage engine in a rolling fashion. Before changing all the members to use WiredTiger, you may prefer to run with mixed storage engines for some period. However, performance can vary according to workload.

You must be using MongoDB version 3.0 or greater in order to use the WiredTiger storage engine. If upgrading from an earlier version of MongoDB, see the guides on *Upgrading to MongoDB 3.0* (page 953) or *Upgrading to MongoDB 3.2* (page 901) before proceeding with changing your storage engine.

Before enabling the new WiredTiger storage engine, ensure that all replica set/sharded cluster members are running at least MongoDB version 2.6.8, and preferably version 3.0.0 or newer.

Procedure This procedure completely removes a *secondary* replica set member's data, starts `mongod` with WiredTiger, and performs an *initial sync* (page 699).

To update all members of the replica set to use WiredTiger, update the *secondary* members first. Then step down the *primary*, and update the stepped-down member.

Step 1: Shut down the secondary member. In the `mongo` shell, shut down the secondary `mongod` instance you wish to upgrade.

```
db.shutdownServer()
```

Step 2: Prepare a data directory for the new `mongod` running with WiredTiger. Prepare a data directory for the new `mongod` instance that will run with the WiredTiger storage engine. `mongod` must have read and write permissions for this directory. You can either delete the contents of the stopped secondary member's current data directory or create a new directory entirely.

`mongod` with WiredTiger will not start with data files created with a different storage engine.

Step 3: Start `mongod` with WiredTiger. Start `mongod`, specifying `wiredTiger` as the *`--storageEngine`* and the prepared data directory for WiredTiger as the *`--dbpath`*. Specify additional options as appropriate for this replica set member.

```
mongod --storageEngine wiredTiger --dbpath <newWiredTigerDBPath> --replSet <replSetName>
```

Since no data exists in the `--dbpath`, the `mongod` will perform an *initial sync* (page 699). The length of the initial sync process depends on the size of the database and network connection between members of the replica set.

You can also specify the options in a `configuration file`. To specify the storage engine, use the `storage.engine` setting.

Step 4: Repeat the procedure for other replica set secondaries you wish to upgrade. Perform this procedure again for the rest of the *secondary members* of the replica set you wish to use the WiredTiger storage engine.

Change Sharded Cluster to WiredTiger

New in version 3.0: The WiredTiger storage engine is available. Also, sharded clusters may have individual shards with different storage engine configurations.

Changed in version 3.2: WiredTiger is the new default storage engine for MongoDB.

This tutorial gives an overview of changing the storage engines of a component of a *sharded cluster* to *WiredTiger* (page 595).

Considerations This procedure may involve downtime, especially if one or more of your shards is a *standalone*. If you change the host or port of any *shard*, you must update the shard configuration as well.

You must be using MongoDB version 3.0 or greater in order to use the WiredTiger storage engine. If upgrading from an earlier version of MongoDB, see the guides on *Upgrading to MongoDB 3.0* (page 953) or *Upgrading to MongoDB 3.2* (page 901) before proceeding with changing your storage engine.

Before enabling the new WiredTiger storage engine, ensure that all replica set/sharded cluster members are running at least MongoDB version 2.6.8, and preferably version 3.0.0 or newer.

Change Shards to WiredTiger

Note: A sharded cluster **can** have mixed storage engines for its individual *shards*.

To change the storage engine for the *shards* to WiredTiger, refer to the appropriate procedure for each shard:

- If the shard is a *standalone*, see *Change Standalone to WiredTiger* (page 597).
- If the shard is a *replica set*, see *Change Replica Set to WiredTiger* (page 598).

Change Config Servers to WiredTiger To change the storage engines of the *config servers* of a sharded cluster, see *Change Config Servers to WiredTiger* (page 600).

You may safely **continue** to use *MMAPv1* (page 603) for the *config servers* even if the shards of the sharded cluster is using the WiredTiger storage engine. If you do choose to update the config servers to use WiredTiger, you must update **all three**.

See also:

Change Config Servers to WiredTiger (page 600)

Change Config Servers to WiredTiger New in version 3.0: The WiredTiger storage engine is available.

Changed in version 3.2: WiredTiger is the new default storage engine for MongoDB.

This tutorial gives an overview of changing the storage engine of the *config servers* in a *sharded cluster* to *WiredTiger* (page 595).

Considerations You may safely **continue** to use *MMAPv1* (page 603) for the *config servers* even if the shards of the sharded cluster is using the WiredTiger storage engine. If you do choose to update the config servers to use WiredTiger, you must update **all three**.

You must be using MongoDB version 3.0 or greater in order to use the WiredTiger storage engine. If upgrading from an earlier version of MongoDB, see the guides on *Upgrading to MongoDB 3.0* (page 953) or *Upgrading to MongoDB 3.2* (page 901) before proceeding with changing your storage engine.

Procedure This tutorial assumes that you have three config servers for this sharded cluster. The three servers are named **first**, **second**, and **third**, based on their position in the `mongos configDB` setting.

Important: During this process, at most only two config servers will be running at any given time to ensure that the sharded cluster's metadata is **read-only**.

Step 1: Disable the balancer.

```
sh.disableBalancer()
```

Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).

Step 2: Shut down the third config server to ensure read-only metadata. Connect a `mongo` shell to the **third** config server and use `db.shutdownServer()` to shut down the **third** config server.

The **third** config server is the last one listed in the `mongos` `configDB` setting.

```
db.shutdownServer()
```

Step 3: Export the data of the second config server with `mongodump`. While the **third** config server is down to ensure the config servers are read-only, prepare to upgrade the **second** config server to use WiredTiger. The **second** config server is the second server listed in the `mongos` setting `configDB`.

Export the data of the **second** config server with `mongodump`.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 4: For the second config server, create a new data directory for use with WiredTiger. Create a data directory in preparation for having the **second** config server run with WiredTiger. `mongod` will not start if the *`--dbpath`* directory contains data files created with a different storage engine.

`mongod` must have read and write permissions for the new directory.

Step 5: Stop the second config server. Connect a `mongo` shell to the **second** config server and use `db.shutdownServer()` to shut down the **second** config server.

```
db.shutdownServer()
```

Step 6: Start the second config server `mongod` with the WiredTiger storage engine option. Start `mongod` as a config server, specifying `wiredTiger` as the *`--storageEngine`* and the newly created data directory for WiredTiger as the *`--dbpath`*. Specify additional options as appropriate.

```
mongod --storageEngine wiredTiger --dbpath <newWiredTigerDBPath> --configsvr
```

You can also specify the options in a `configuration file`. To specify the storage engine, use the `storage.engine` setting.

Step 7: Upload the exported data using `mongorestore` to the second config server. Use `mongorestore` to upload the exported data. Specify additional options as appropriate. See `mongorestore` for available options.

```
mongorestore <exportDataDestination>
```

When the `mongorestore` finishes, the **second** config server upgrade to use WiredTiger is complete.

Step 8: Shut down the second config server to ensure read-only metadata. When the **second** config server upgrade is complete, shut down the **second** config server in preparation to upgrade the other config servers. This is necessary to maintain at most only two active config servers and keep the sharded cluster's metadata read-only.

Connect a `mongo` shell to the **second** config server and use `db.shutdownServer()` to shut down the **second** config server.

```
db.shutdownServer()
```

Step 9: Restart the third config server to prepare for its upgrade. Restart the **third** config server with its original startup options. **Do not** change its options to use the WiredTiger storage engine at this point.

```
mongod --configsvr --dbpath <existingDBPath>
```

Include any other options in use for the third config server.

Step 10: Export the data of the third config server with `mongodump`.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 11: For the third config server, create a new data directory for use with WiredTiger. Create a data directory in preparation for having the **third** config server run with WiredTiger. `mongod` will not start if the *`--dbpath`* directory contains data files created with a different storage engine.

`mongod` must have read and write permissions for the new directory.

Step 12: Stop the third config server. Connect a `mongo` shell to the **third** config server and use `db.shutdownServer()` to shut down the **third** config server.

```
db.shutdownServer()
```

Step 13: Start the third config server with the WiredTiger storage engine option. Start `mongod` as a config server, specifying `wiredTiger` as the *`--storageEngine`* and the newly created data directory for WiredTiger as the *`--dbpath`*. Specify additional options as appropriate.

```
mongod --storageEngine wiredTiger --dbpath <newWiredTigerDBPath> --configsvr
```

You can also specify the options in a `configuration file`. To specify the storage engine, use the `storage.engine` setting.

Step 14: Upload the exported data using `mongorestore` to the third config server. Use `mongorestore` to upload the exported data. Specify additional options as appropriate. See `mongorestore` for available options.

```
mongorestore <exportDataDestination>
```

When the `mongorestore` finishes, the **third** config server upgrade to use WiredTiger is complete.

Step 15: Export data of the first config server with `mongodump`. To prepare for the upgrade of the **first** config server to use WiredTiger, export the data of the **first** config server with `mongodump`.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 16: For the first config server, create a new data directory for use with WiredTiger. Create a data directory in preparation for having the **first** config server run with WiredTiger. `mongod` will not start if the *`--dbpath`* directory contains data files created with a different storage engine.

`mongod` must have read and write permissions for the new directory.

Step 17: Stop the first config server. Connect a `mongo` shell to the **first** config server and use `db.shutdownServer()` to shut down the **first** config server.

```
db.shutdownServer()
```

Step 18: Start the first config server with the WiredTiger storage engine option. Start `mongod` as a config server, specifying `wiredTiger` as the *`--storageEngine`* and the newly created data directory for WiredTiger as the *`--dbpath`*. Specify additional options as appropriate.

```
mongod --storageEngine wiredTiger --dbpath <newWiredTigerDBPath> --configsvr
```

You can also specify the options in a `configuration file`. To specify the storage engine, use the `storage.engine` setting.

Step 19: Upload the exported data using `mongorestore` to the first config server. Use `mongorestore` to upload the exported data. Specify additional options as appropriate. See `mongorestore` for available options.

```
mongorestore <exportDataDestination>
```

When the `mongorestore` finishes, the **first** config server upgrade to use WiredTiger is complete.

Step 20: Restart the second config server to enable writes to the sharded cluster's metadata. Restart the **second** config server, specifying WiredTiger as the *`--storageEngine`* and the newly created WiredTiger data directory as the *`--dbpath`*. Specify additional options as appropriate.

```
mongod --storageEngine wiredTiger --dbpath <newWiredTigerDBPath> --configsvr
```

You can also specify the options in a `configuration file`. To specify the storage engine, use the `storage.engine` setting.

Once all three config servers are up, the sharded cluster's metadata is available for writes.

Step 21: Re-enable the balancer. Once all three config servers are up and running with WiredTiger, *re-enable the balancer* (page 803).

```
sh.startBalancer()
```

11.1.2 MMAPv1 Storage Engine

On this page

- Journal (page 604)
- Record Storage Characteristics (page 604)
- Record Allocation Strategies (page 604)
- Memory Use (page 605)

MMAPv1 is MongoDB's original storage engine based on memory mapped files. It excels at workloads with high volume inserts, reads, and in-place updates.

Changed in version 3.2: Starting in MongoDB 3.2, the MMAPv1 is no longer the default storage engine; instead, the *WiredTiger* (page 595) storage engine is the default storage engine . See *Default Storage Engine Change* (page 898).

Journal

In order to ensure that all modifications to a MongoDB data set are durably written to disk, MongoDB, by default, records all modifications to an on-disk journal. MongoDB writes more frequently to the journal than it writes the data files.

In the default configuration for the *MMAPv1 storage engine* (page 603), MongoDB writes to the data files on disk every 60 seconds and writes to the *journal* files roughly every 100 milliseconds.

To change the interval for writing to the data files, use the `storage.syncPeriodSecs` setting. For the journal files, see `storage.journal.commitIntervalMs` setting.

These values represent the *maximum* amount of time between the completion of a write operation and when MongoDB writes to the data files or to the journal files. In many cases MongoDB and the operating system flush data to disk more frequently, so that the above values represents a theoretical maximum.

The journal allows MongoDB to successfully recover data from data files after a `mongod` instance exits without flushing all changes. See *Journaling* (page 606) for more information about the journal in MongoDB.

Record Storage Characteristics

All records are contiguously located on disk, and when a document becomes larger than the allocated record, MongoDB must allocate a new record. New allocations require MongoDB to move a document and update all indexes that refer to the document, which takes more time than in-place updates and leads to storage fragmentation.

Changed in version 3.0.0.

By default, MongoDB uses *Power of 2 Sized Allocations* (page 604) so that every document in MongoDB is stored in a *record* which contains the document itself and extra space, or *padding*. Padding allows the document to grow as the result of updates while minimizing the likelihood of reallocations.

Record Allocation Strategies

MongoDB supports multiple record allocation strategies that determine how `mongod` adds padding to a document when creating a record. Because documents in MongoDB may grow after insertion and all records are contiguous on disk, the padding can reduce the need to relocate documents on disk following updates. Relocations are less efficient than in-place updates and can lead to storage fragmentation. As a result, all padding strategies trade additional space for increased efficiency and decreased fragmentation.

Different allocation strategies support different kinds of workloads: the *power of 2 allocations* (page 604) are more efficient for insert/update/delete workloads; while *exact fit allocations* (page 605) is ideal for collections *without* update and delete workloads.

Power of 2 Sized Allocations

Changed in version 3.0.0.

MongoDB 3.0 uses the power of 2 sizes allocation as the default record allocation strategy for MMAPv1. With the power of 2 sizes allocation strategy, each record has a size in bytes that is a power of 2 (e.g. 32, 64, 128, 256, 512 ... 2 MB). For documents larger than 2 MB, the allocation is rounded up to the nearest multiple of 2 MB.

The power of 2 sizes allocation strategy has the following key properties:

- Can efficiently reuse freed records to reduce fragmentation. Quantizing record allocation sizes into a fixed set of sizes increases the probability that an insert will fit into the free space created by an earlier document deletion or relocation.

- Can reduce moves. The added padding space gives a document room to grow without requiring a move. In addition to saving the cost of moving, this results in less updates to indexes. Although the power of 2 sizes strategy can minimize moves, it does not eliminate them entirely.

No Padding Allocation Strategy

Changed in version 3.0.0.

For collections whose workloads do not change the document sizes, such as workloads that consist of insert-only operations or update operations that do not increase document size (such as incrementing a counter), you can disable the *power of 2 allocation* (page 604) using the `collMod` command with the `noPadding` flag or the `db.createCollection()` method with the `noPadding` option.

Prior to version 3.0.0, MongoDB used an allocation strategy that included a dynamically calculated `padding` as a factor of the document size.

Memory Use

With MMAPv1, MongoDB automatically uses all free memory on the machine as its cache. System resource monitors show that MongoDB uses a lot of memory, but its usage is dynamic. If another process suddenly needs half the server's RAM, MongoDB will yield cached memory to the other process.

Technically, the operating system's virtual memory subsystem manages MongoDB's memory. This means that MongoDB will use as much free memory as it can, swapping to disk as needed. Deployments with enough memory to fit the application's working data set in RAM will achieve the best performance.

11.1.3 In-Memory Storage Engine

On this page

- Specify In-Memory Storage Engine (page 605)
- Concurrency (page 606)
- Durability (page 606)

Warning: The in-memory storage engine is currently in **beta**. Do not use in production.

Starting in MongoDB Enterprise 3.2, an in-memory storage engine is available in the 64-bit builds for beta-testing purposes. Other than some metadata and diagnostic data, the in-memory storage engine does not maintain any on-disk data. By avoiding disk I/O, the in-memory storage engine allows for more predictable latency of database operations.

Specify In-Memory Storage Engine

To select the in-memory storage engine, specify:

- `inMemory` for the `--storageEngine` option, or the `storage.engine` setting if using a configuration file.
- `--dbpath`, or `storage.dbPath` if using a configuration file. Although the in-memory storage engine does not write data to the filesystem, it maintains in the `--dbpath` small metadata files and diagnostic data as well temporary files for building large indexes.

For example, from the command line:

```
mongod --storageEngine inMemory --dbpath <path>
```

Or, if using the `YAML configuration file format`:

```
storage:
   engine: inMemory
   dbPath: <path>
```

See *cli-mongod-inmemory* for configuration options specific to this storage engine.

Warning: The in-memory storage engine does not persist data after process shutdown.

Concurrency

The in-memory storage engine uses *document-level* concurrency control for write operations. As a result, multiple clients can modify different documents of a collection at the same time.

Durability

The in-memory storage engine is non-persistent and does not write data to a persistent storage. As such, the concept of *journal* or waiting for data to become *durable* does not apply to the in-memory storage engine.

Write operations that specify a write concern `journaled` (page 181) are acknowledged immediately. When an `mongod` instance shuts down, either as result of the `shutdown` command or due to a system error, recovery of in-memory data is impossible.

11.2 Journaling

On this page

- Journaling and the WiredTiger Storage Engine (page 606)
- Journaling and the MMAPv1 Storage Engine (page 607)
- Journaling and the In-Memory Storage Engine (page 609)

To provide durability in the event of a failure, MongoDB uses *write ahead logging* to on-disk *journal* files.

11.2.1 Journaling and the WiredTiger Storage Engine

Important: The *log* mentioned in this section refers to the WiredTiger write-ahead log (i.e. the journal) and not the MongoDB log file.

WiredTiger (page 595) uses *checkpoints* (page 596) to provide a consistent view of data on disk and allow MongoDB to recover from the last checkpoint. However, if MongoDB exits unexpectedly in between checkpoints, journaling is required to recover information that occurred after the last checkpoint.

With journaling, the recovery process:

1. Looks in the data files to find the identifier of the last checkpoint.
2. Searches in the journal files for the record that matches the identifier of the last checkpoint.

3. Apply the operations in the journal files since the last checkpoint.

Journaling Process

Changed in version 3.2.

With journaling, WiredTiger creates one journal record for each client initiated write operation. The journal record includes any internal write operations caused by the initial write. For example, an update to a document in a collection may result in modifications to the indexes; WiredTiger creates a single journal record that includes both the update operation and its associated index modifications.

MongoDB configures WiredTiger to use in-memory buffering for storing the journal records. Threads coordinate to allocate and copy into their portion of the buffer. All journal records up to 128 kB are buffered.

WiredTiger syncs the buffered journal records to disk according to the following intervals or conditions:

- New in version 3.2: Every 50 milliseconds.
- MongoDB sets checkpoints to occur in WiredTiger on user data at an interval of 60 seconds or when 2 GB of journal data has been written, whichever occurs first.
- If the write operation includes a write concern of `j: true` (page 181), WiredTiger forces a sync of the WiredTiger journal files.
- Because MongoDB uses a journal file size limit of 100 MB, WiredTiger creates a new journal file approximately every 100 MB of data. When WiredTiger creates a new journal file, WiredTiger syncs the previous journal file.

Important: In between write operations, while the journal records remain in the WiredTiger buffers, updates can be lost following a hard shutdown of `mongod`.

See also:

The `serverStatus` command returns information on the WiredTiger journal statistics in the `wiredTiger.log` field.

Journal Files

For the journal files, MongoDB creates a subdirectory named `journal` under the `dbPath` directory. WiredTiger journal files have names with the following format `WiredTigerLog.<sequence>` where `<sequence>` is a zero-padded number starting from `0000000001`.

Journal files contain a record per each write operation. Each record has a unique identifier.

MongoDB configures WiredTiger to use snappy compression for the journaling data.

Minimum log record size for WiredTiger is 128 bytes. If a log record is 128 bytes or smaller, WiredTiger does not compress that record.

WiredTiger journal files for MongoDB have a maximum size limit of approximately 100 MB. Once the file exceeds that limit, WiredTiger creates a new journal file.

WiredTiger automatically removes old journal files to maintain only the files needed to recover from last checkpoint.

WiredTiger will pre-allocate journal files.

11.2.2 Journaling and the MMAPv1 Storage Engine

With *MMAPv1* (page 603), when a write operation occurs, MongoDB updates the in-memory view. With journaling enabled, MongoDB writes the in-memory changes first to on-disk journal files. If MongoDB should terminate or

encounter an error before committing the changes to the data files, MongoDB can use the journal files to apply the write operation to the data files and maintain a consistent state.

Journaling Process

With journaling, MongoDB's storage layer has two internal views of the data set: the *private view*, used to write to the journal files, and the *shared view*, used to write to the data files:

1. MongoDB first applies write operations to the private view.
2. MongoDB then applies the changes in the private view to the on-disk *journal files* (page 608) in the `journal` directory roughly every 100 milliseconds. MongoDB records the write operations to the on-disk journal files in batches called *group commits*. Grouping the commits help minimize the performance impact of journaling since these commits must block all writers during the commit. Writes to the journal are atomic, ensuring the consistency of the on-disk journal files. For information on the frequency of the commit interval, see `storage.journal.commitIntervalMs`.
3. Upon a journal commit, MongoDB applies the changes from the journal to the shared view.
4. Finally, MongoDB applies the changes in the shared view to the data files. More precisely, at default intervals of 60 seconds, MongoDB asks the operating system to flush the shared view to the data files. The operating system may choose to flush the shared view to disk at a higher frequency than 60 seconds, particularly if the system is low on free memory. To change the interval for writing to the data files, use the `storage.syncPeriodSecs` setting.

If the `mongod` instance were to crash without having applied the writes to the data files, the journal could replay the writes to the shared view for eventual write to the data files.

When MongoDB flushes write operations to the data files, MongoDB notes which journal writes have been flushed. Once a journal file contains only flushed writes, it is no longer needed for recovery and MongoDB can recycle it for a new journal file.

Once the journal operations have been applied to the shared view and flushed to disk (i.e. pages in the shared view and private view are in sync), MongoDB asks the operating system to remap the shared view to the private view in order to save physical RAM. MongoDB routinely asks the operating system to remap the shared view to the private view in order to save physical RAM. Upon a new remapping, the operating system knows that physical memory pages can be shared between the shared view and the private view mappings.

Note: The interaction between the shared view and the on-disk data files is similar to how MongoDB works *without* journaling. Without journaling, MongoDB asks the operating system to flush in-memory changes to the data files every 60 seconds.

Journal Files

With journaling enabled, MongoDB creates a subdirectory named `journal` under the `dbPath` directory. The `journal` directory contains journal files named `j._<sequence>` where `<sequence>` is an integer starting from `0` and a "last sequence number" file `lsn`.

Journal files contain the write ahead logs; each journal entry describes the bytes the write operation changed in the data files. Journal files are append-only files. When a journal file holds 1 gigabyte of data, MongoDB creates a new journal file. If you use the `storage.smallFiles` option when starting `mongod`, you limit the size of each journal file to 128 megabytes.

The `lsn` file contains the last time MongoDB flushed the changes to the data files.

Once MongoDB applies all the write operations in a particular journal file to the data files, MongoDB can recycle it for a new journal file.

Unless you write *many* bytes of data per second, the `journal` directory should contain only two or three journal files.

A clean shutdown removes all the files in the journal directory. A dirty shutdown (crash) leaves files in the journal directory; these are used to automatically recover the database to a consistent state when the mongod process is restarted.

Journal Directory

To speed the frequent sequential writes that occur to the current journal file, you can ensure that the journal directory is on a different filesystem from the database data files.

Important: If you place the journal on a different filesystem from your data files, you *cannot* use a filesystem snapshot alone to capture valid backups of a `dbPath` directory. In this case, use `fsyncLock()` to ensure that database files are consistent before the snapshot and `fsyncUnlock()` once the snapshot is complete.

Preallocation Lag

MongoDB may preallocate journal files if the `mongod` process determines that it is more efficient to preallocate journal files than create new journal files as needed.

Depending on your filesystem, you might experience a preallocation lag the first time you start a `mongod` instance with journaling enabled. The amount of time required to pre-allocate files might last several minutes; during this time, you will not be able to connect to the database. This is a one-time preallocation and does not occur with future invocations.

To avoid preallocation lag, see *Avoid Preallocation Lag for MMAPv1* (page 610).

11.2.3 Journaling and the In-Memory Storage Engine

Warning: The in-memory storage engine is currently in **beta**. Do not use in production.

The *In-Memory Storage Engine* (page 605) is available in MongoDB Enterprise 3.2 and later. Because its data is kept in memory, there is no separate journal. Write operations with a write concern of `j: true` (page 181) are immediately acknowledged.

See also:

In-Memory Storage Engine: Durability (page 606)

Manage Journaling

On this page

- Procedures (page 610)

MongoDB uses *write ahead logging* to an on-disk *journal* to guarantee *write operation* (page 114) durability. The MMAPv1 storage engine also requires the *journal* in order to provide crash resiliency.

The WiredTiger storage engine does not require journaling to guarantee a consistent state after a crash. The database will be restored to the last consistent *checkpoint* (page 596) during recovery. However, if MongoDB exits unexpectedly in between checkpoints, journaling is required to recover writes that occurred after the last checkpoint.

With journaling enabled, if `mongod` stops unexpectedly, the program can recover everything written to the journal. MongoDB will re-apply the write operations on restart and maintain a consistent state. By default, the greatest extent of lost writes, i.e., those not made to the journal, are those made in the last 100 milliseconds, plus the time it takes to perform the actual journal writes. See `commitIntervalMs` for more information on the default.

Procedures

Enable Journaling For 64-bit builds of `mongod`, journaling is enabled by default.

To enable journaling, start `mongod` with the `--journal` command line option.

Disable Journaling

> **Warning:** Do not disable journaling on production systems. When using the MMAPv1 storage engine *withou* journal, if your `mongod` instance stops without shutting down cleanly unexpectedly for any reason, (e.g. pow failure) and you are not running with journaling, then you must recover from an unaffected *replica set* member backup, as described in *repair* (page 366).

To disable journaling, start `mongod` with the `--nojournal` command line option.

Get Commit Acknowledgment You can get commit acknowledgment with the *Write Concern* (page 179) and the `j` (page 181) option. For details, see *Write Concern* (page 179).

Avoid Preallocation Lag for MMAPv1 With the *MMAPv1 storage engine* (page 603), MongoDB may preallocate journal files if the `mongod` process determines that it is more efficient to preallocate journal files than create new journal files as needed.

Depending on your filesystem, you might experience a preallocation lag the first time you start a `mongod` instance with journaling enabled. The amount of time required to pre-allocate files might last several minutes; during this time, you will not be able to connect to the database. This is a one-time preallocation and does not occur with future invocations.

To avoid *preallocation lag* (page 608), you can preallocate files in the journal directory by copying them from another instance of `mongod`.

Preallocated files do not contain data. It is safe to later remove them. But if you restart `mongod` with journaling, `mongod` will create them again.

Example

The following sequence preallocates journal files for an instance of `mongod` running on port `27017` with a database path of `/data/db`.

For demonstration purposes, the sequence starts by creating a set of journal files in the usual way.

1. Create a temporary directory into which to create a set of journal files:

   ```
   mkdir ~/tmpDbpath
   ```

2. Create a set of journal files by staring a `mongod` instance that uses the temporary directory:

   ```
   mongod --port 10000 --dbpath ~/tmpDbpath --journal
   ```

3. When you see the following log output, indicating `mongod` has the files, press CONTROL+C to stop the `mongod` instance:

```
[initandlisten] waiting for connections on port 10000
```

4. Preallocate journal files for the new instance of `mongod` by moving the journal files from the data directory of the existing instance to the data directory of the new instance:

   ```
   mv ~/tmpDbpath/journal /data/db/
   ```

5. Start the new `mongod` instance:

   ```
   mongod --port 27017 --dbpath /data/db --journal
   ```

Monitor Journal Status Use the following commands and methods to monitor journal status:

- `serverStatus`

  The `serverStatus` command returns database status information that is useful for assessing performance.

- `journalLatencyTest`

  Use `journalLatencyTest` to measure how long it takes on your volume to write to the disk in an append-only fashion. You can run this command on an idle system to get a baseline sync time for journaling. You can also run this command on a busy system to see the sync time on a busy system, which may be higher if the journal directory is on the same volume as the data files.

  The `journalLatencyTest` command also provides a way to check if your disk drive is buffering writes in its local cache. If the number is very low (i.e., less than 2 milliseconds) and the drive is non-SSD, the drive is probably buffering writes. In that case, enable cache write-through for the device in your operating system, unless you have a disk controller card with battery backed RAM.

Change the Group Commit Interval for MMAPv1 For the *MMAPv1 storage engine* (page 603), you can set the group commit interval using the *`--journalCommitInterval`* command line option. The allowed range is `2` to `300` milliseconds.

Lower values increase the durability of the journal at the expense of disk performance.

Recover Data After Unexpected Shutdown On a restart after a crash, MongoDB replays all journal files in the journal directory before the server becomes available. If MongoDB must replay journal files, `mongod` notes these events in the log output.

There is no reason to run `repairDatabase` in these situations.

11.3 GridFS

On this page

- When to Use GridFS (page 612)
- Use GridFS (page 612)
- GridFS Collections (page 612)
- GridFS Indexes (page 614)
- Additional Resources (page 615)

GridFS is a specification for storing and retrieving files that exceed the *BSON*-document *size limit* of 16 MB.

Instead of storing a file in a single document, GridFS divides the file into parts, or chunks [1], and stores each chunk as a separate document. By default, GridFS uses a chunk size of 255 kB; that is, GridFS divides a file into chunks of 255 kB with the exception of the last chunk. The last chunk is only as large as necessary. Similarly, files that are no larger than the chunk size only have a final chunk, using only as much space as needed plus some additional metadata.

GridFS uses two collections to store files. One collection stores the file chunks, and the other stores file metadata. The section *GridFS Collections* (page 612) describes each collection in detail.

When you query GridFS for a file, the driver will reassemble the chunks as needed. You can perform range queries on files stored through GridFS. You can also access information from arbitrary sections of files, such as to "skip" to the middle of a video or audio file.

GridFS is useful not only for storing files that exceed 16 MB but also for storing any files for which you want access without having to load the entire file into memory. See also *When to Use GridFS* (page 612).

Changed in version 2.4.10: The default chunk size changed from 256 kB to 255 kB.

11.3.1 When to Use GridFS

In MongoDB, use *GridFS* for storing files larger than 16 MB.

In some situations, storing large files may be more efficient in a MongoDB database than on a system-level filesystem.

- If your filesystem limits the number of files in a directory, you can use GridFS to store as many files as needed.
- When you want to access information from portions of large files without having to load whole files into memory, you can use GridFS to recall sections of files without reading the entire file into memory.
- When you want to keep your files and metadata automatically synced and deployed across a number of systems and facilities, you can use GridFS. When using *geographically distributed replica sets* (page 642), MongoDB can distribute files and their metadata automatically to a number of `mongod` instances and facilities.

Do not use GridFS if you need to update the content of the entire file atomically. As an alternative you can store multiple versions of each file and specify the current version of the file in the metadata. You can update the metadata field that indicates "latest" status in an atomic update after uploading the new version of the file, and later remove previous versions if needed.

Furthermore, if your files are all smaller the 16 MB `BSON Document Size` limit, consider storing the file manually within a single document instead of using GridFS. You may use the BinData data type to store the binary data. See your `drivers` documentation for details on using BinData.

11.3.2 Use GridFS

To store and retrieve files using *GridFS*, use either of the following:

- A MongoDB driver. See the `drivers` documentation for information on using GridFS with your driver.
- The `mongofiles` command-line tool. See the `mongofiles` reference for documentation.

11.3.3 GridFS Collections

GridFS stores files in two collections:

- `chunks` stores the binary chunks. For details, see *The chunks Collection* (page 613).

[1] The use of the term *chunks* in the context of GridFS is not related to the use of the term *chunks* in the context of sharding.

- `files` stores the file's metadata. For details, see *The files Collection* (page 613).

GridFS places the collections in a common bucket by prefixing each with the bucket name. By default, GridFS uses two collections with a bucket named `fs`:

- `fs.files`
- `fs.chunks`

You can choose a different bucket name, as well as create multiple buckets in a single database. The full collection name, which includes the bucket name, is subject to the `namespace length limit`.

The `chunks` Collection

Each document in the `chunks` [1] collection represents a distinct chunk of a file as represented in *GridFS*. Documents in this collection have the following form:

```
{
  "_id" : <ObjectId>,
  "files_id" : <ObjectId>,
  "n" : <num>,
  "data" : <binary>
}
```

A document from the `chunks` collection contains the following fields:

`chunks.`**`_id`**
: The unique *ObjectId* of the chunk.

`chunks.`**`files_id`**
: The `_id` of the "parent" document, as specified in the `files` collection.

`chunks.`**`n`**
: The sequence number of the chunk. GridFS numbers all chunks, starting with 0.

`chunks.`**`data`**
: The chunk's payload as a *BSON* `Binary` type.

The `files` Collection

Each document in the `files` collection represents a file in *GridFS*. Consider a document in the `files` collection, which has the following form:

```
{
  "_id" : <ObjectId>,
  "length" : <num>,
  "chunkSize" : <num>,
  "uploadDate" : <timestamp>,
  "md5" : <hash>,
  "filename" : <string>,
  "contentType" : <string>,
  "aliases" : <string array>,
  "metadata" : <dataObject>,
}
```

Documents in the `files` collection contain some or all of the following fields:

`files.`**`_id`**
: The unique identifier for this document. The `_id` is of the data type you chose for the original document. The default type for MongoDB documents is *BSON ObjectId*.

`files.length`
: The size of the document in bytes.

`files.chunkSize`
: The size of each chunk in **bytes**. GridFS divides the document into chunks of size `chunkSize`, except for the last, which is only as large as needed. The default size is 255 kilobytes (kB).

  Changed in version 2.4.10: The default chunk size changed from 256 kB to 255 kB.

`files.uploadDate`
: The date the document was first stored by GridFS. This value has the `Date` type.

`files.md5`
: An MD5 hash of the complete file returned by the `filemd5` command. This value has the `String` type.

`files.filename`
: Optional. A human-readable name for the GridFS file.

`files.contentType`
: Optional. A valid MIME type for the GridFS file.

`files.aliases`
: Optional. An array of alias strings.

`files.metadata`
: Optional. Any additional information you want to store.

Applications may create additional arbitrary fields.

11.3.4 GridFS Indexes

GridFS uses indexes on each of the `chunks` and `files` collections for efficiency. `Drivers` that conform to the GridFS specification[2] automatically create these indexes for convenience. You can also create any additional indexes as desired to suit your application's needs.

The `chunks` Index

GridFS uses a *unique*, *compound* index on the `chunks` collection using the `files_id` and `n` fields. This allows for efficient retrieval of chunks, as demonstrated in the following example:

```
db.fs.chunks.find( { files_id: myFileID } ).sort( { n: 1 } )
```

`Drivers` that conform to the GridFS specification[3] will automatically ensure that this index exists before read and write operations. See the relevant driver documentation for the specific behavior of your GridFS application.

If this index does not exist, you can issue the following operation to create it using the `mongo` shell:

```
db.fs.chunks.createIndex( { files_id: 1, n: 1 }, { unique: true } );
```

The `files` Index

GridFS uses an *index* on the `files` collection using the `filename` and `uploadDate` fields. This index allows for efficient retrieval of files, as shown in this example:

[2] https://github.com/mongodb/specifications/blob/master/source/gridfs/gridfs-spec.rst
[3] https://github.com/mongodb/specifications/blob/master/source/gridfs/gridfs-spec.rst

```
db.fs.files.find( { filename: myFileName } ).sort( { uploadDate: 1 } )
```

Drivers that conform to the GridFS specification[4] will automatically ensure that this index exists before read and write operations. See the relevant driver documentation for the specific behavior of your GridFS application.

If this index does not exist, you can issue the following operation to create it using the mongo shell:

```
db.fs.files.createIndex( { filename: 1, uploadDate: 1 } );
```

11.3.5 Additional Resources

- Building MongoDB Applications with Binary Files Using GridFS: Part 1[5]
- Building MongoDB Applications with Binary Files Using GridFS: Part 2[6]

11.4 FAQ: MongoDB Storage

On this page

- Storage Engine Fundamentals (page 850)
- Can you mix storage engines in a replica set? (page 850)
- WiredTiger Storage Engine (page 851)
- MMAPv1 Storage Engine (page 852)
- Can I manually pad documents to prevent moves during updates? (page 855)
- Data Storage Diagnostics (page 855)

This document addresses common questions regarding MongoDB's storage system.

11.4.1 Storage Engine Fundamentals

What is a storage engine?

A storage engine is the part of a database that is responsible for managing how data is stored, both in memory and on disk. Many databases support multiple storage engines, where different engines perform better for specific workloads. For example, one storage engine might offer better performance for read-heavy workloads, and another might support a higher-throughput for write operations.

See also:

Storage Engines (page 595)

11.4.2 Can you mix storage engines in a replica set?

Yes. You can have a replica set members that use different storage engines.

When designing these multi-storage engine deployments consider the following:

[4] https://github.com/mongodb/specifications/blob/master/source/gridfs/gridfs-spec.rst
[5] http://www.mongodb.com/blog/post/building-mongodb-applications-binary-files-using-gridfs-part-1?jmp=docs
[6] http://www.mongodb.com/blog/post/building-mongodb-applications-binary-files-using-gridfs-part-2?jmp=docs

- the oplog on each member may need to be sized differently to account for differences in throughput between different storage engines.
- recovery from backups may become more complex if your backup captures data files from MongoDB: you may need to maintain backups for each storage engine.

11.4.3 WiredTiger Storage Engine

Can I upgrade an existing deployment to a WiredTiger?

Yes. See:

- *Change Standalone to WiredTiger* (page 597)
- *Change Replica Set to WiredTiger* (page 598)
- *Change Sharded Cluster to WiredTiger* (page 599)

How much compression does WiredTiger provide?

The ratio of compressed data to uncompressed data depends on your data and the compression library used. By default, collection data in WiredTiger use *Snappy block compression*; *zlib* compression is also available. Index data use *prefix compression* by default.

To what size should I set the WiredTiger cache?

With WiredTiger, MongoDB utilizes both the WiredTiger cache and the filesystem cache.

Changed in version 3.2: Starting in MongoDB 3.2, the WiredTiger cache, by default, will use the larger of either:

- 60% of RAM minus 1 GB, or
- 1 GB.

For systems with up to 10 GB of RAM, the new default setting is less than or equal to the 3.0 default setting (For MongoDB 3.0, the WiredTiger cache uses either 1 GB or half of the installed physical RAM, whichever is larger).

For systems with more than 10 GB of RAM, the new default setting is greater than the 3.0 setting.

Via the filesystem cache, MongoDB automatically uses all free memory that is not used by the WiredTiger cache or by other processes. Data in the filesystem cache is compressed.

To adjust the size of the WiredTiger cache, see `storage.wiredTiger.engineConfig.cacheSizeGB` and `--wiredTigerCacheSizeGB`. Avoid increasing the WiredTiger cache size above its default value.

Note: The `storage.wiredTiger.engineConfig.cacheSizeGB` only limits the size of the WiredTiger cache, not the total amount of memory used by `mongod`. The WiredTiger cache is only one component of the RAM used by MongoDB. MongoDB also automatically uses all free memory on the machine via the filesystem cache (data in the filesystem cache is compressed).

In addition, the operating system will use any free RAM to buffer filesystem blocks.

To accommodate the additional consumers of RAM, you may have to decrease WiredTiger cache size.

The default WiredTiger cache size value assumes that there is a single `mongod` instance per machine. If a single machine contains multiple MongoDB instances, then you should decrease the setting to accommodate the other `mongod` instances.

If you run `mongod` in a container (e.g. `lxc`, `cgroups`, Docker, etc.) that does *not* have access to all of the RAM available in a system, you must set `storage.wiredTiger.engineConfig.cacheSizeGB` to a value less than the amount of RAM available in the container. The exact amount depends on the other processes running in the container.

To view statistics on the cache and eviction rate, see the `wiredTiger.cache` field returned from the `serverStatus` command.

How frequently does WiredTiger write to disk?

MongoDB configures WiredTiger to create checkpoints (i.e. write the snapshot data to disk) at intervals of 60 seconds or 2 gigabytes of journal data.

For journal data, MongoDB writes to disk according to the following intervals or condition:

- New in version 3.2: Every 50 milliseconds.
- MongoDB sets checkpoints to occur in WiredTiger on user data at an interval of 60 seconds or when 2 GB of journal data has been written, whichever occurs first.
- If the write operation includes a write concern of `j: true` (page 181), WiredTiger forces a sync of the WiredTiger journal files.
- Because MongoDB uses a journal file size limit of 100 MB, WiredTiger creates a new journal file approximately every 100 MB of data. When WiredTiger creates a new journal file, WiredTiger syncs the previous journal file.

11.4.4 MMAPv1 Storage Engine

What are memory mapped files?

A memory-mapped file is a file with data that the operating system places in memory by way of the `mmap()` system call. `mmap()` thus *maps* the file to a region of virtual memory. Memory-mapped files are the critical piece of the MMAPv1 storage engine in MongoDB. By using memory mapped files, MongoDB can treat the contents of its data files as if they were in memory. This provides MongoDB with an extremely fast and simple method for accessing and manipulating data.

How do memory mapped files work?

MongoDB uses memory mapped files for managing and interacting with all data.

Memory mapping assigns files to a block of virtual memory with a direct byte-for-byte correlation. MongoDB memory maps data files to memory as it accesses documents. Unaccessed data is *not* mapped to memory.

Once mapped, the relationship between file and memory allows MongoDB to interact with the data in the file as if it were memory.

How frequently does MMAPv1 write to disk?

In the default configuration for the *MMAPv1 storage engine* (page 603), MongoDB writes to the data files on disk every 60 seconds and writes to the *journal* files roughly every 100 milliseconds.

To change the interval for writing to the data files, use the `storage.syncPeriodSecs` setting. For the journal files, see `storage.journal.commitIntervalMs` setting.

These values represent the *maximum* amount of time between the completion of a write operation and when MongoDB writes to the data files or to the journal files. In many cases MongoDB and the operating system flush data to disk more frequently, so that the above values represents a theoretical maximum.

Why are the files in my data directory larger than the data in my database?

The data files in your data directory, which is the `/data/db` directory in default configurations, might be larger than the data set inserted into the database. Consider the following possible causes:

Preallocated data files

MongoDB preallocates its data files to avoid filesystem fragmentation, and because of this, the size of these files do not necessarily reflect the size of your data.

The `storage.mmapv1.smallFiles` option will reduce the size of these files, which may be useful if you have many small databases on disk.

The `oplog`

If this `mongod` is a member of a replica set, the data directory includes the *oplog.rs* file, which is a preallocated *capped collection* in the `local` database.

The default allocation is approximately 5% of disk space on 64-bit installations. In most cases, you should not need to resize the oplog. See *Oplog Sizing* (page 657) for more information.

The `journal`

The data directory contains the journal files, which store write operations on disk before MongoDB applies them to databases. See *Journaling* (page 606).

Empty records

MongoDB maintains lists of empty records in data files as it deletes documents and collections. MongoDB can reuse this space, but will not, by default, return this space to the operating system.

To allow MongoDB to more effectively reuse the space, you can de-fragment your data. To de-fragment, use the `compact` command. The `compact` requires up to 2 gigabytes of extra disk space to run. Do not use `compact` if you are critically low on disk space. For more information on its behavior and other considerations, see `compact`.

`compact` only removes fragmentation from MongoDB data files within a collection and does not return any disk space to the operating system. To return disk space to the operating system, see *How do I reclaim disk space?* (page 853).

How do I reclaim disk space?

The following provides some options to consider when reclaiming disk space.

Note: You do not need to reclaim disk space for MongoDB to reuse freed space. See *Empty records* (page 853) for information on reuse of freed space.

---

repairDatabase

You can use `repairDatabase` on a database to rebuilds the database, de-fragmenting the associated storage in the process.

`repairDatabase` requires free disk space equal to the size of your current data set plus 2 gigabytes. If the volume that holds dbpath lacks sufficient space, you can mount a separate volume and use that for the repair. For additional information and considerations, see `repairDatabase`.

> **Warning:** Do not use `repairDatabase` if you are critically low on disk space.
> `repairDatabase` will block all other operations and may take a long time to complete.

You can only run `repairDatabase` on a standalone `mongod` instance.

You can also run the `repairDatabase` operation for all databases on the server by restarting your `mongod` standalone instance with the *`--repair`* and *`--repairpath`* options. All databases on the server will be unavailable during this operation.

Resync the Member of the Replica Set

For a secondary member of a replica set, you can perform a *resync of the member* (page 699) by: stopping the secondary member to resync, deleting all data and subdirectories from the member's data directory, and restarting.

For details, see *Resync a Member of a Replica Set* (page 699).

What is the working set?

Working set represents the total body of data that the application uses in the course of normal operation. Often this is a subset of the total data size, but the specific size of the working set depends on actual moment-to-moment use of the database.

If you run a query that requires MongoDB to scan every document in a collection, the working set will expand to include every document. Depending on physical memory size, this may cause documents in the working set to "page out," or to be removed from physical memory by the operating system. The next time MongoDB needs to access these documents, MongoDB may incur a hard page fault.

For best performance, the majority of your *active* set should fit in RAM.

What are page faults?

With the MMAPv1 storage engine, page faults can occur as MongoDB reads from or writes data to parts of its data files that are not currently located in physical memory. In contrast, operating system page faults happen when physical memory is exhausted and pages of physical memory are swapped to disk.

If there is free memory, then the operating system can find the page on disk and load it to memory directly. However, if there is no free memory, the operating system must:

- find a page in memory that is stale or no longer needed, and write the page to disk.
- read the requested page from disk and load it into memory.

This process, on an active system, can take a long time, particularly in comparison to reading a page that is already in memory.

See *Page Faults* (page 311) for more information.

What is the difference between soft and hard page faults?

Page faults occur when MongoDB, with the MMAP storage engine, needs access to data that isn't currently in active memory. A "hard" page fault refers to situations when MongoDB must access a disk to access the data. A "soft" page fault, by contrast, merely moves memory pages from one list to another, such as from an operating system file cache.

See *Page Faults* (page 311) for more information.

11.4.5 Can I manually pad documents to prevent moves during updates?

Changed in version 3.0.0.

With the *MMAPv1 storage engine* (page 603), an update can cause a document to move on disk if the document grows in size. To *minimize* document movements, MongoDB uses *padding*.

You should not have to pad manually because by default, MongoDB uses *Power of 2 Sized Allocations* (page 604) to add *padding automatically* (page 604). The *Power of 2 Sized Allocations* (page 604) ensures that MongoDB allocates document space in sizes that are powers of 2, which helps ensure that MongoDB can efficiently reuse free space created by document deletion or relocation as well as reduce the occurrences of reallocations in many cases.

However, *if you must* pad a document manually, you can add a temporary field to the document and then `$unset` the field, as in the following example.

Warning: Do not manually pad documents in a capped collection. Applying manual padding to a document in a capped collection can break replication. Also, the padding is not preserved if you re-sync the MongoDB instance.

```
var myTempPadding = [ "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa",
                      "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa",
                      "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa",
                      "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa"];

db.myCollection.insert( { _id: 5, paddingField: myTempPadding } );

db.myCollection.update( { _id: 5 },
                        { $unset: { paddingField: "" } }
                      )

db.myCollection.update( { _id: 5 },
                        { $set: { realField: "Some text that I might have needed padding for" } }
                      )
```

See also:

Record Allocation Strategies (page 604)

11.4.6 Data Storage Diagnostics

How can I check the size of a collection?

To view the statistics for a collection, including the data size, use the `db.collection.stats()` method from the `mongo` shell. The following example issues `db.collection.stats()` for the `orders` collection:

```
db.orders.stats();
```

MongoDB also provides the following methods to return specific sizes for the collection:

- `db.collection.dataSize()` to return data size in bytes for the collection.

- `db.collection.storageSize()` to return allocation size in bytes, including unused space.
- `db.collection.totalSize()` to return the data size plus the index size in bytes.
- `db.collection.totalIndexSize()` to return the index size in bytes.

The following script prints the statistics for each database:

```
db._adminCommand("listDatabases").databases.forEach(function (d) {
   mdb = db.getSiblingDB(d.name);
   printjson(mdb.stats());
})
```

The following script prints the statistics for each collection in each database:

```
db._adminCommand("listDatabases").databases.forEach(function (d) {
   mdb = db.getSiblingDB(d.name);
   mdb.getCollectionNames().forEach(function(c) {
      s = mdb[c].stats();
      printjson(s);
   })
})
```

How can I check the size of indexes for a collection?

To view the size of the data allocated for an index, use the `db.collection.stats()` method and check the `indexSizes` field in the returned document.

How can I get information on the storage use of a database?

The `db.stats()` method in the `mongo` shell returns the current state of the "active" database. For the description of the returned fields, see *dbStats Output*.

CHAPTER 12

Replication

A *replica set* in MongoDB is a group of `mongod` processes that maintain the same data set. Replica sets provide redundancy and high availability, and are the basis for all production deployments. This section introduces replication in MongoDB as well as the components and architecture of replica sets. The section also provides tutorials for common tasks related to replica sets.

Replication Introduction **(page 623)** An introduction to replica sets, their behavior, operation, and use.

Replication Concepts **(page 627)** The core documentation of replica set operations, configurations, architectures and behaviors.

- ***Replica Set Members*** **(page 628)** Introduces the components of replica sets.
- ***Replica Set Deployment Architectures*** **(page 636)** Introduces architectural considerations related to replica sets deployment planning.
- ***Replica Set High Availability*** **(page 644)** Presents the details of the automatic failover and recovery process with replica sets.
- ***Replica Set Read and Write Semantics*** **(page 648)** Presents the semantics for targeting read and write operations to the replica set, with an awareness of location and set configuration.

Replica Set Tutorials **(page 665)** Tutorials for common tasks related to the use and maintenance of replica sets.

Replication Reference **(page 716)** Reference for functions and operations related to replica sets.

12.1 Replication Introduction

On this page

- Redundancy and Data Availability (page 623)
- Replication in MongoDB (page 624)
- Additional Resources (page 627)

Replication is the process of synchronizing data across multiple servers.

12.1.1 Redundancy and Data Availability

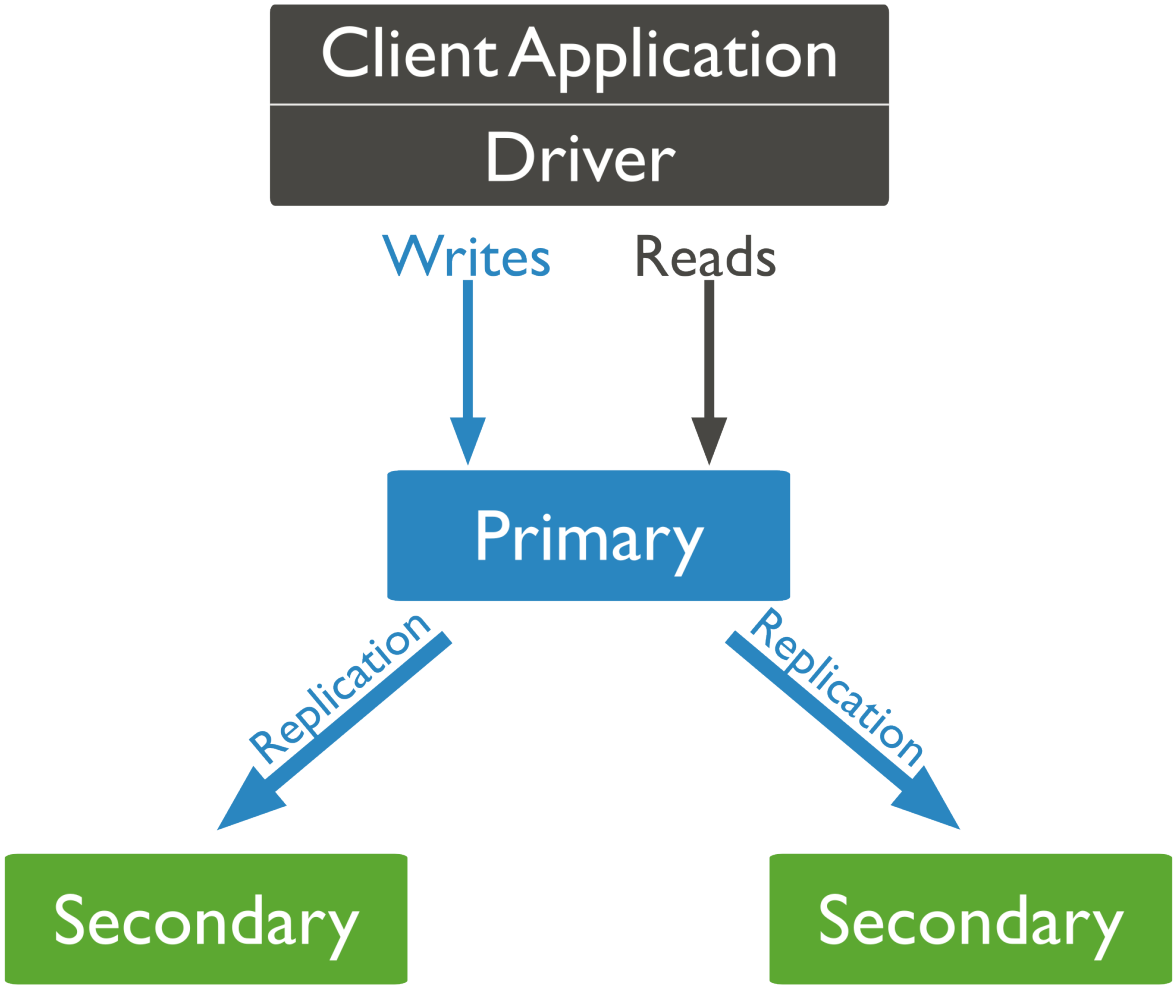
Replication provides redundancy and increases data availability. With multiple copies of data on different database servers, replication provides a level of fault tolerance against the loss of a single database server.

In some cases, replication can provide increased read capacity as clients can send read operations to different servers. Maintaining copies of data in different data centers can increase data locality and availability for distributed applications. You can also maintain additional copies for dedicated purposes, such as disaster recovery, reporting, or backup.

12.1.2 Replication in MongoDB

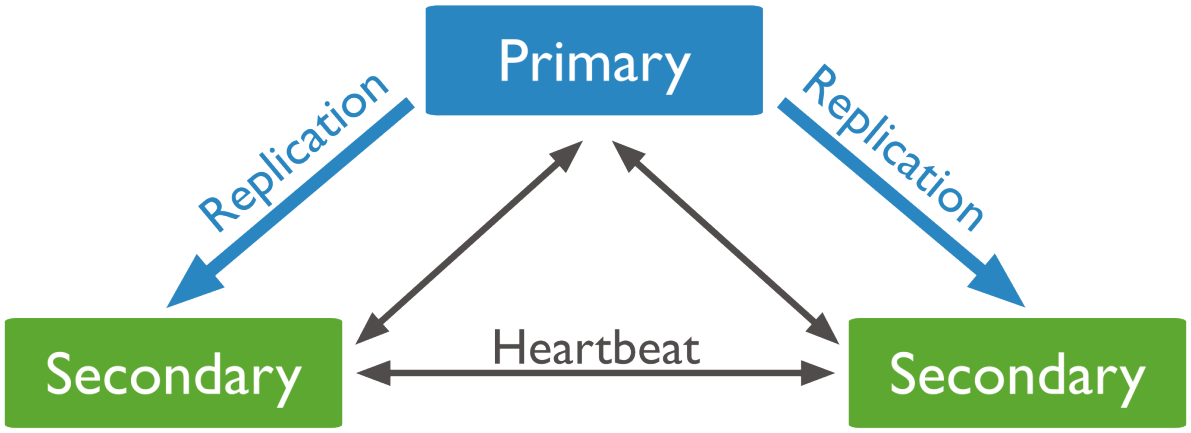
A replica set is a group of `mongod` instances that maintain the same data set. A replica set contains several data bearing nodes and optionally one arbiter node. Of the data bearing nodes, one and only one member is deemed the primary node, while the other nodes are deemed secondary nodes.

The *primary node* (page 628) receives all write operations. A replica set can have only one primary capable of confirming writes with `{ w: "majority" }` (page 180) write concern; although in some circumstances, another mongod instance may transiently believe itself to also be primary. [1] The primary records all changes to its data sets in its operation log, i.e. *oplog* (page 656). For more information on primary node operation, see *Replica Set Primary* (page 628).

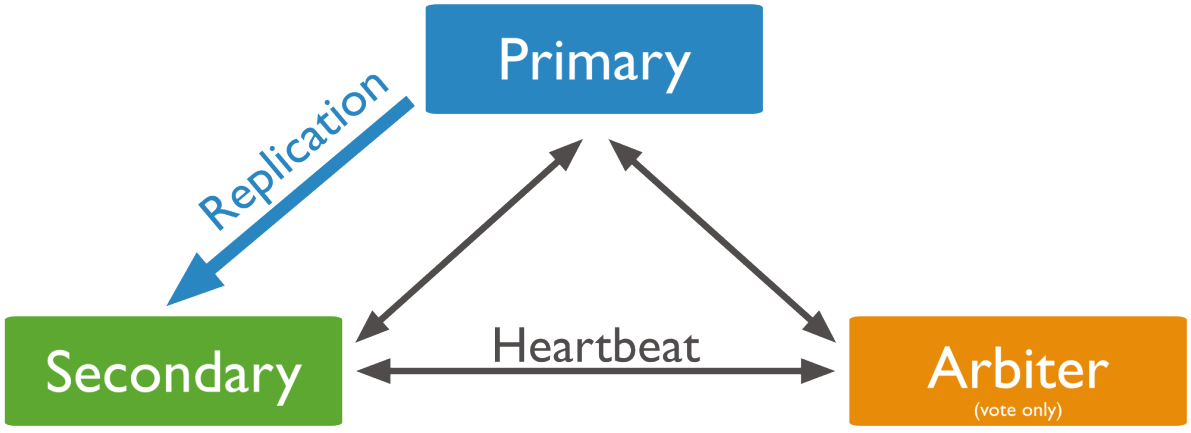


[1] In *some circumstances* (page 729), two nodes in a replica set may *transiently* believe that they are the primary, but at most, one of them will be able to complete writes with `{ w: "majority" }` (page 180) write concern. The node that can complete `{ w: "majority" }` (page 180) writes is the current primary, and the other node is a former primary that has not yet recognized its demotion, typically due to a *network partition*. When this occurs, clients that connect to the former primary may observe stale data despite having requested read preference `primary` (page 728), and new writes to the former primary will eventually roll back.

The *secondaries* (page 628) replicate the primary's oplog and apply the operations to their data sets such that the secondaries' data sets reflect the primary's data set. If the primary is unavailable, an eligible secondary will hold an election to elect itself the new primary. For more information on secondary members, see *Replica Set Secondary Members* (page 628).



You may add an extra `mongod` instance to a replica set as an *arbiter* (page 635). Arbiters do not maintain a data set. The purpose of an arbiter is to maintain a quorum in a replica set by responding to heartbeat and election requests by other replica set members. Because they do not store a data set, arbiters can be a good way to provide replica set quorum functionality with a cheaper resource cost than a fully functional replica set member with a data set. If your replica set has an even number of members, add an arbiter to obtain a majority of votes in an election for primary. Arbiters do not require dedicated hardware. For more information on arbiters, see *Replica Set Arbiter* (page 635).



An *arbiter* (page 635) will always be an arbiter whereas a *primary* (page 628) may step down and become a *secondary* (page 628) and a *secondary* (page 628) may become the primary during an election.

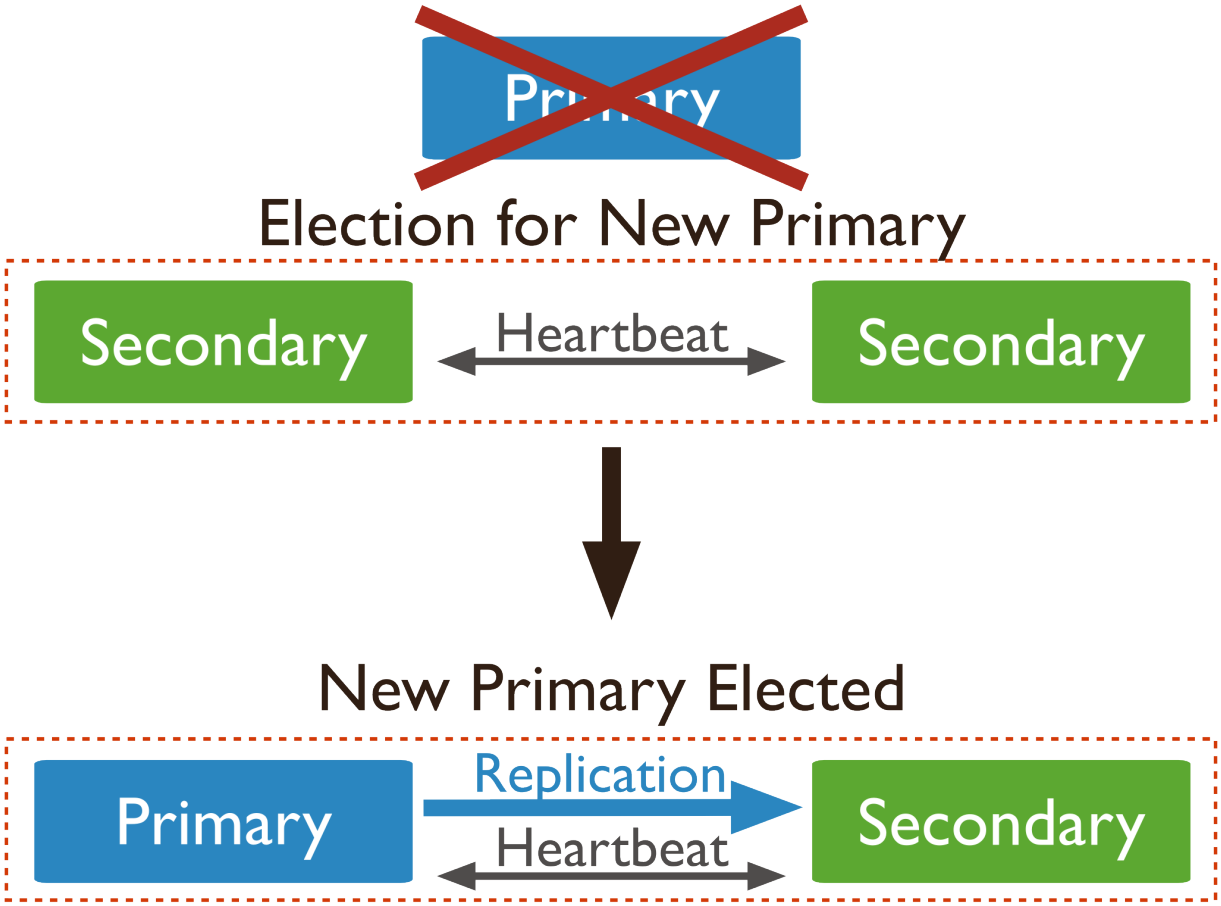
Asynchronous Replication

Secondaries apply operations from the primary asynchronously. By applying operations after the primary, sets can continue to function despite the failure of one or more members. For more information on replication mechanics, see *Replica Set Oplog* (page 656) and *Replica Set Data Synchronization* (page 658).

Automatic Failover

When a primary does not communicate with the other members of the set for more than 10 seconds, an eligible secondary will hold an election to elect itself the new primary. The first secondary to hold an election and receive a majority of the members' votes becomes primary.

New in version 3.2: MongoDB introduces a version 1 of the replication protocol (`protocolVersion: 1` (page 718)) to reduce replica set failover time and accelerates the detection of multiple simultaneous primaries. New replica sets will, by default, use `protocolVersion: 1` (page 718). Previous versions of MongoDB use version 0 of the protocol.



See *Replica Set Elections* (page 644) and *Rollbacks During Replica Set Failover* (page 647) for more information.

Read Operations

By default, clients read from the primary [1]; however, clients can specify a *read preference* (page 651) to send read operations to secondaries. *Asynchronous replication* (page 625) to secondaries means that reads from secondaries may return data that does not reflect the state of the data on the primary. For information on reading from replica sets, see *Read Preference* (page 651).

In MongoDB, clients can see the results of writes before the writes are *durable*:

- Regardless of *write concern* (page 179), other clients using `"local"` (page 182) (i.e. the default) readConcern can see the result of a write operation before the write operation is acknowledged to the issuing client.

- Clients using `"local"` (page 182) (i.e. the default) readConcern can read data which may be subsequently *rolled back* (page 647).

For more information on read isolations, consistency and recency for MongoDB, see *Read Isolation, Consistency, and Recency* (page 133).

Additional Features

Replica sets provide a number of options to support application needs. For example, you may deploy a replica set with *members in multiple data centers* (page 642), or control the outcome of elections by adjusting the `members[n].priority` (page 720) of some members. Replica sets also support dedicated members for reporting, disaster recovery, or backup functions.

See *Priority 0 Replica Set Members* (page 631), *Hidden Replica Set Members* (page 633) and *Delayed Replica Set Members* (page 634) for more information.

12.1.3 Additional Resources

- Quick Reference Cards[2]
- Webinar: Managing Your Mission Critical App - Ensuring Zero Downtime[3]

12.2 Replication Concepts

These documents describe and provide examples of replica set operation, configuration, and behavior. For an overview of replication, see *Replication Introduction* (page 623). For documentation of the administration of replica sets, see *Replica Set Tutorials* (page 665). The *Replication Reference* (page 716) documents commands and operations specific to replica sets.

Replica Set Members **(page 628)** Introduces the components of replica sets.

> ***Replica Set Primary*** **(page 628)** The primary is the only member of a replica set that accepts write operations.
>
> ***Replica Set Secondary Members*** **(page 628)** Secondary members replicate the primary's data set and accept read operations. If the set has no primary, a secondary can become primary.
>
> ***Priority 0 Replica Set Members*** **(page 631)** Priority 0 members are secondaries that cannot become the primary.
>
> ***Hidden Replica Set Members*** **(page 633)** Hidden members are secondaries that are invisible to applications. These members support dedicated workloads, such as reporting or backup.
>
> ***Replica Set Arbiter*** **(page 635)** An arbiter does not maintain a copy of the data set but participate in elections.

Replica Set Deployment Architectures **(page 636)** Introduces architectural considerations related to replica sets deployment planning.

Replica Set High Availability **(page 644)** Presents the details of the automatic failover and recovery process with replica sets.

> ***Replica Set Elections*** **(page 644)** Elections occur when the primary becomes unavailable and the replica set members autonomously select a new primary.
>
> ***Read Preference*** **(page 651)** Read preference specifies where (i.e. which members of the replica set) the drivers should direct the read operations.

[2] https://www.mongodb.com/lp/misc/quick-reference-cards?jmp=docs
[3] http://www.mongodb.com/webinar/managing-mission-critical-app-downtime?jmp=docs

Replication Processes **(page 656)** Mechanics of the replication process and related topics.

Master Slave Replication **(page 659)** Master-slave replication provided redundancy in early versions of MongoDB. Replica sets replace master-slave for most use cases.

12.2.1 Replica Set Members

A *replica set* in MongoDB is a group of `mongod` processes that provide redundancy and high availability. The members of a replica set are:

***Primary* (page ??).** The primary receives all write operations.

***Secondaries* (page ??).** Secondaries replicate operations from the primary to maintain an identical data set. Secondaries may have additional configurations for special usage profiles. For example, secondaries may be *non-voting* (page 646) or *priority 0* (page 631).

You can also maintain an *arbiter* (page **??**) as part of a replica set. Arbiters do not keep a copy of the data. However, arbiters play a role in the elections that select a primary if the current primary is unavailable.

The minimum requirements for a replica set are: A *primary* (page **??**), a *secondary* (page **??**), and an *arbiter* (page **??**). Most deployments, however, will keep three members that store data: A *primary* (page **??**) and two *secondary members* (page **??**).

Changed in version 3.0.0: A replica set can have up to *50 members* (page 942) but only 7 voting members. [4] In previous versions, replica sets can have up to 12 members.

Replica Set Primary

The primary is the only member in the replica set that receives write operations. MongoDB applies write operations on the *primary* and then records the operations on the primary's *oplog* (page 656). *Secondary* (page **??**) members replicate this log and apply the operations to their data sets.

In the following three-member replica set, the primary accepts all write operations. Then the secondaries replicate the oplog to apply to their data sets.

All members of the replica set can accept read operations. However, by default, an application directs its read operations to the primary member. See *Read Preference* (page 651) for details on changing the default read behavior.

The replica set can have at most one primary. [5] If the current primary becomes unavailable, an election determines the new primary. See *Replica Set Elections* (page 644) for more details.

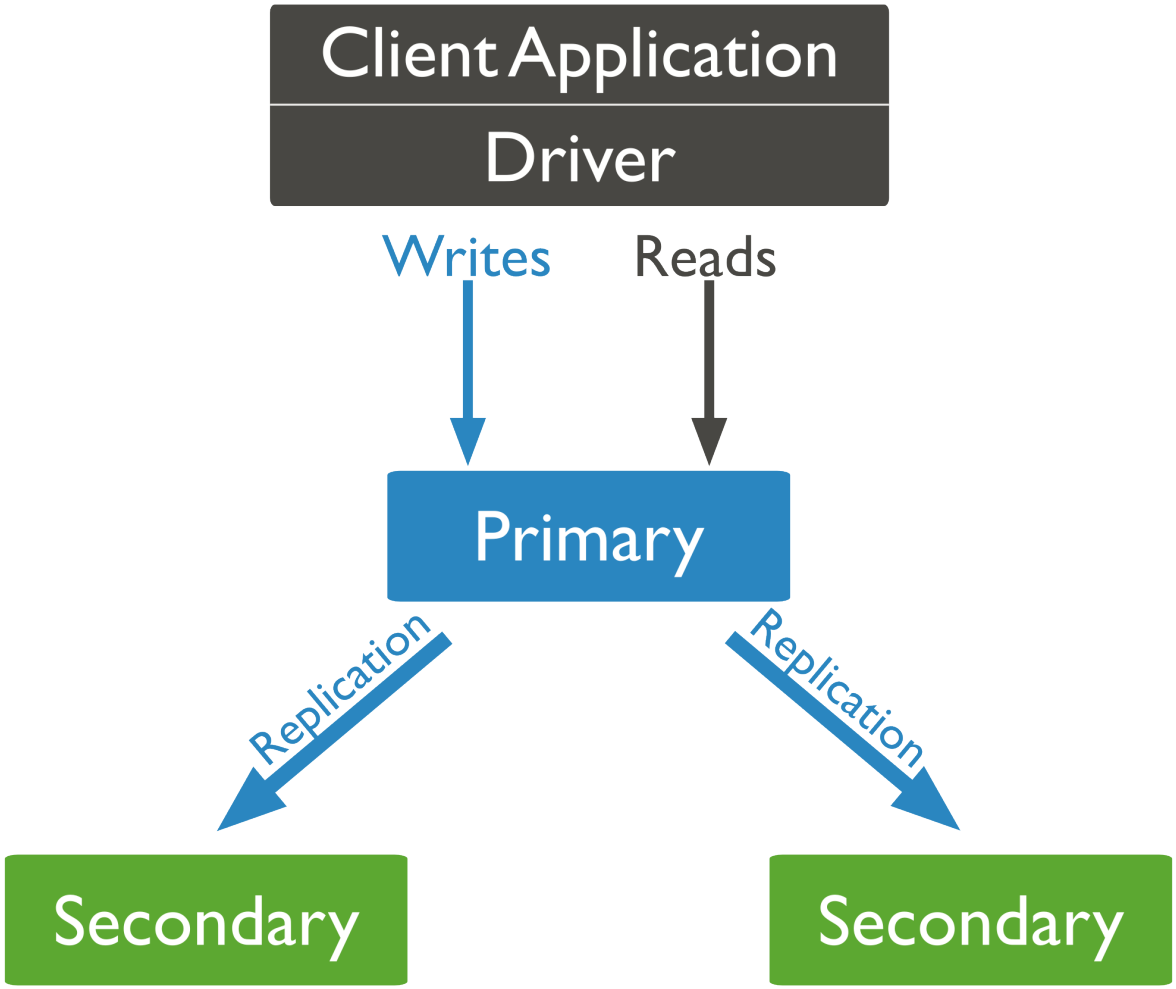
In the following 3-member replica set, the primary becomes unavailable. This triggers an election which selects one of the remaining secondaries as the new primary.

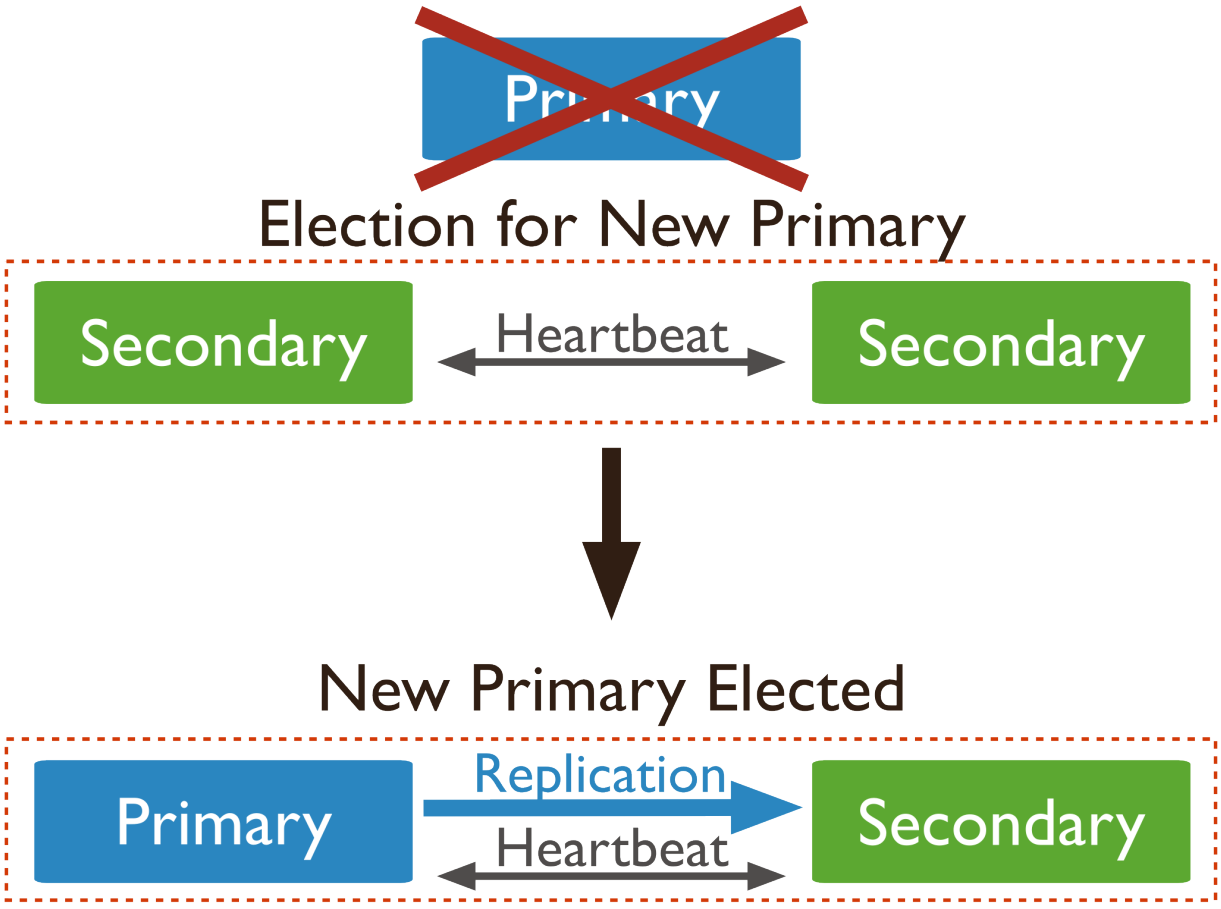
Replica Set Secondary Members

A secondary maintains a copy of the *primary's* data set. To replicate data, a secondary applies operations from the primary's *oplog* (page 656) to its own data set in an asynchronous process. A replica set can have one or more secondaries.

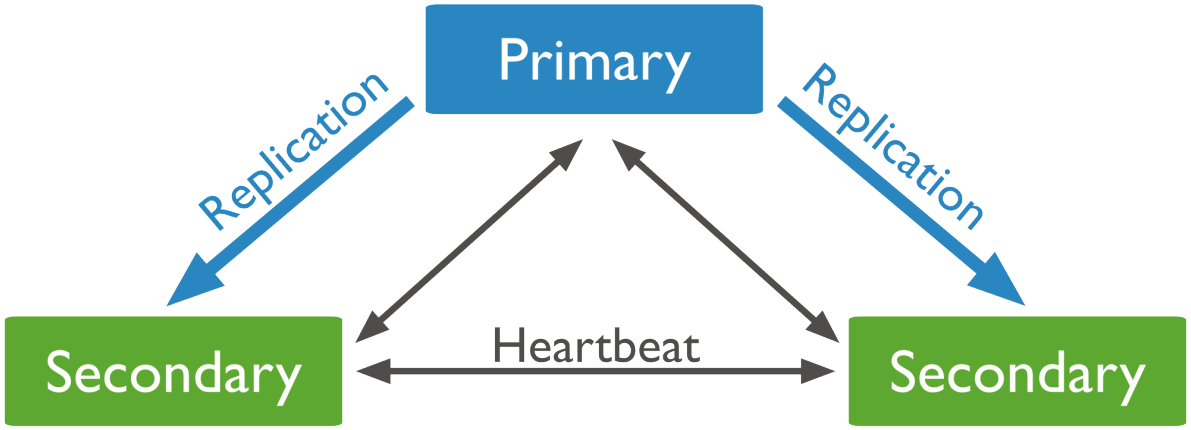
[4] While replica sets are the recommended solution for production, a replica set can support up to `50 members` in total. If your deployment requires more than 50 members, you'll need to use *master-slave* (page 659) replication. However, master-slave replication lacks the automatic failover capabilities.

[5] In *some circumstances* (page 729), two nodes in a replica set may *transiently* believe that they are the primary, but at most, one of them will be able to complete writes with `{ w: "majority" }` (page 180) write concern. The node that can complete `{ w: "majority" }` (page 180) writes is the current primary, and the other node is a former primary that has not yet recognized its demotion, typically due to a *network partition*. When this occurs, clients that connect to the former primary may observe stale data despite having requested read preference `primary` (page 728), and new writes to the former primary will eventually roll back.


Client Application
Driver
Writes
Reads
Primary
Replication
Replication
Secondary
Secondary


Election for New Primary
Secondary
Heartbeat
Secondary
New Primary Elected
Primary
Replication
Heartbeat
Secondary

The following three-member replica set has two secondary members. The secondaries replicate the primary's oplog and apply the operations to their data sets.



Although clients cannot write data to secondaries, clients can read data from secondary members. See *Read Preference* (page 651) for more information on how clients direct read operations to replica sets.

A secondary can become a primary. If the current primary becomes unavailable, the replica set holds an *election* to choose which of the secondaries becomes the new primary.

In the following three-member replica set, the primary becomes unavailable. This triggers an election where one of the remaining secondaries becomes the new primary.

See *Replica Set Elections* (page 644) for more details.

You can configure a secondary member for a specific purpose. You can configure a secondary to:

- Prevent it from becoming a primary in an election, which allows it to reside in a secondary data center or to serve as a cold standby. See *Priority 0 Replica Set Members* (page 631).
- Prevent applications from reading from it, which allows it to run applications that require separation from normal traffic. See *Hidden Replica Set Members* (page 633).
- Keep a running "historical" snapshot for use in recovery from certain errors, such as unintentionally deleted databases. See *Delayed Replica Set Members* (page 634).

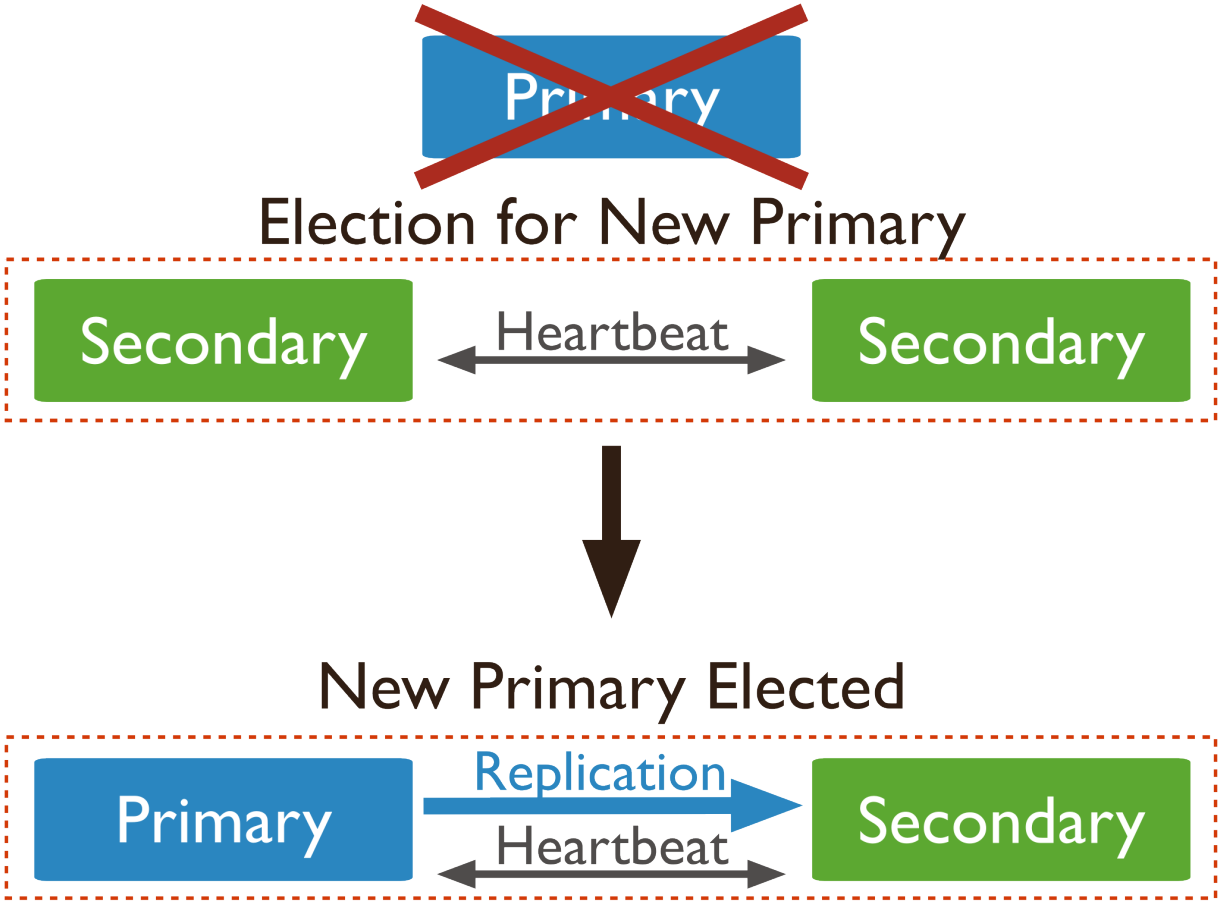
Priority 0 Replica Set Members

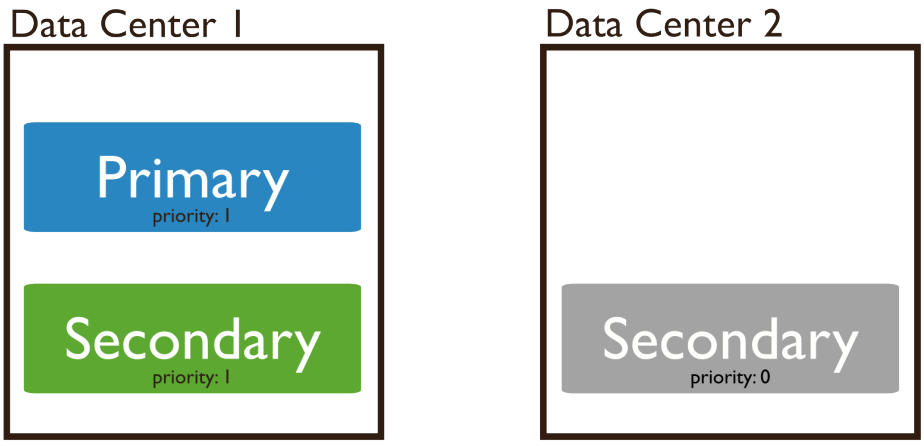
On this page

- Priority 0 Members as Standbys (page 633)
- Priority 0 Members and Failover (page 633)
- Configuration (page 633)

A *priority 0* member is a secondary that **cannot** become *primary*. *Priority 0* members cannot *trigger elections*. Otherwise these members function as normal secondaries. A *priority 0* member maintains a copy of the data set, accepts read operations, and votes in elections. Configure a *priority 0* member to prevent *secondaries* from becoming primary, which is particularly useful in multi-data center deployments.

In a three-member replica set, in one data center hosts the primary and a secondary. A second data center hosts one *priority 0* member that cannot become primary.


Election for New Primary
Secondary
Heartbeat
Secondary
New Primary Elected
Primary
Replication
Heartbeat
Secondary


Data Center 1
Primary
priority: 1
Secondary
priority: 1
Data Center 2
Secondary
priority: 0

Priority 0 Members as Standbys A *priority 0* member can function as a standby. In some replica sets, it might not be possible to add a new member in a reasonable amount of time. A standby member keeps a current copy of the data to be able to replace an unavailable member.

In many cases, you need not set standby to *priority 0*. However, in sets with varied hardware or *geographic distribution* (page 642), a *priority 0* standby ensures that only qualified members become primary.

A *priority 0* standby may also be valuable for some members of a set with different hardware or workload profiles. In these cases, deploy a member with *priority 0* so it can't become primary. Also consider using an *hidden member* (page 633) for this purpose.

If your set already has seven voting members, also configure the member as *non-voting* (page 646).

Priority 0 Members and Failover When configuring a *priority 0* member, consider potential failover patterns, including all possible network partitions. Always ensure that your main data center contains both a quorum of voting members and contains members that are eligible to be primary.

Configuration To configure a *priority 0* member, see *Prevent Secondary from Becoming Primary* (page 686).

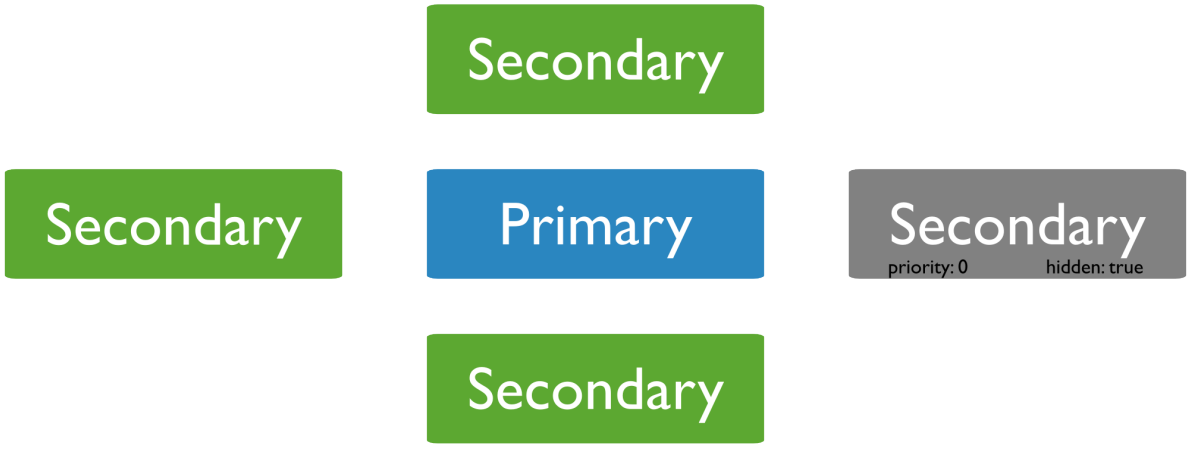
Hidden Replica Set Members

On this page

- Behavior (page 633)
- Further Reading (page 634)

A hidden member maintains a copy of the *primary's* data set but is **invisible** to client applications. Hidden members are good for workloads with different usage patterns from the other members in the *replica set*. Hidden members must always be *priority 0 members* (page 631) and so **cannot become primary**. The `db.isMaster()` method does not display hidden members. Hidden members, however, **may vote** in *elections* (page 644).

In the following five-member replica set, all four secondary members have copies of the primary's data set, but one of the secondary members is hidden.



Behavior

Read Operations Clients will not distribute reads with the appropriate *read preference* (page 651) to hidden members. As a result, these members receive no traffic other than basic replication. Use hidden members for dedicated tasks such as reporting and backups. *Delayed members* (page 634) should be hidden.

In a sharded cluster, `mongos` do not interact with hidden members.

Voting Hidden members *may* vote in replica set elections. If you stop a voting hidden member, ensure that the set has an active majority or the *primary* will step down.

For the purposes of backups,

- If using the MMAPv1 storage engine, you can avoid stopping a hidden member with the `db.fsyncLock()` and `db.fsyncUnlock()` operations to flush all writes and lock the `mongod` instance for the duration of the backup operation.
- Changed in version 3.2: Starting in MongoDB 3.2, `db.fsyncLock()` can ensure that the data files do not change for MongoDB instances using either the MMAPv1 or the WiredTiger storage engine, thus providing consistency for the purposes of creating backups.

  In previous MongoDB version, `db.fsyncLock()` *cannot* guarantee a consistent set of files for low-level backups (e.g. via file copy `cp`, `scp`, `tar`) for WiredTiger.

Further Reading For more information about backing up MongoDB databases, see *MongoDB Backup Methods* (page 282). To configure a hidden member, see *Configure a Hidden Replica Set Member* (page 687).

Delayed Replica Set Members

On this page

- Considerations (page 634)
- Example (page 635)
- Configuration (page 635)

Delayed members contain copies of a *replica set's* data set. However, a delayed member's data set reflects an earlier, or delayed, state of the set. For example, if the current time is 09:52 and a member has a delay of an hour, the delayed member has no operation more recent than 08:52.

Because delayed members are a "rolling backup" or a running "historical" snapshot of the data set, they may help you recover from various kinds of human error. For example, a delayed member can make it possible to recover from unsuccessful application upgrades and operator errors including dropped databases and collections.

Considerations

Requirements Delayed members:

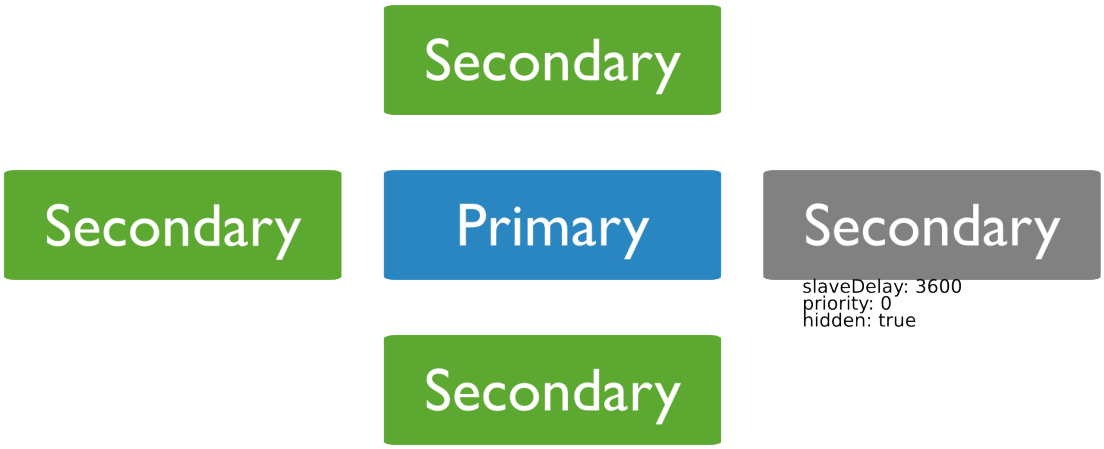
- **Must be** *priority 0* (page 631) members. Set the priority to 0 to prevent a delayed member from becoming primary.
- **Should be** *hidden* (page 633) members. Always prevent applications from seeing and querying delayed members.
- *do* vote in *elections* for primary, if `members[n].votes` (page 721) is set to 1.

Behavior Delayed members copy and apply operations from the source *oplog* on a delay. When choosing the amount of delay, consider that the amount of delay:

- must be equal to or greater than your expected maintenance window durations.
- must be *smaller* than the capacity of the oplog. For more information on oplog size, see *Oplog Size* (page 657).

Sharding In sharded clusters, delayed members have limited utility when the *balancer* is enabled. Because delayed members replicate chunk migrations with a delay, the state of delayed members in a sharded cluster are not useful for recovering to a previous state of the sharded cluster if any migrations occur during the delay window.

Example In the following 5-member replica set, the primary and all secondaries have copies of the data set. One member applies operations with a delay of 3600 seconds (one hour). This delayed member is also *hidden* and is a *priority 0 member*.



Configuration A delayed member has its `members[n].priority` (page 720) equal to `0`, `members[n].hidden` (page 720) equal to `true`, and its `members[n].slaveDelay` (page 721) equal to the number of seconds of delay:

```
{
   "_id" : <num>,
   "host" : <hostname:port>,
   "priority" : 0,
   "slaveDelay" : <seconds>,
   "hidden" : true
}
```

To configure a delayed member, see *Configure a Delayed Replica Set Member* (page 689).

Replica Set Arbiter

On this page

- Example (page 636)
- Security (page 636)

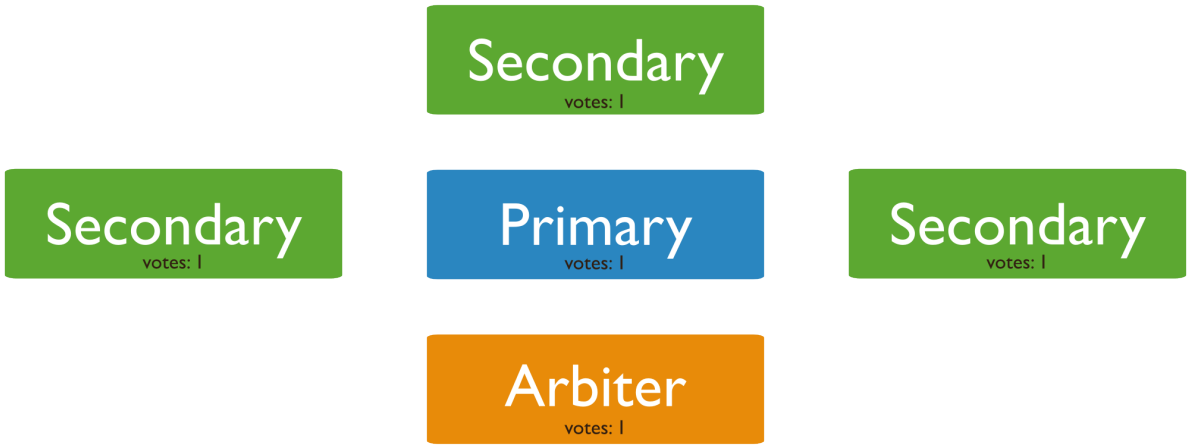
An arbiter does **not** have a copy of data set and **cannot** become a primary. Replica sets may have arbiters to add a vote in *elections of for primary* (page 644). Arbiters *always* have exactly `1` vote election, and thus allow replica sets to have an uneven number of members, without the overhead of a member that replicates data.

Important: Do not run an arbiter on systems that also host the primary or the secondary members of the replica set.

Only add an arbiter to sets with even numbers of members. If you add an arbiter to a set with an odd number of members, the set may suffer from tied *elections*. To add an arbiter, see *Add an Arbiter to Replica Set* (page 677).

Example

For example, in the following replica set, an arbiter allows the set to have an odd number of votes for elections:



Security

Authentication When running with `authorization`, arbiters exchange credentials with other members of the set to authenticate. MongoDB encrypts the authentication process. The MongoDB authentication exchange is cryptographically secure.

Arbiters use `keyfiles` to authenticate to the replica set.

Communication The only communication between arbiters and other set members are: votes during elections, heartbeats, and configuration data. These exchanges are not encrypted.

However, if your MongoDB deployment uses TLS/SSL, MongoDB will encrypt *all* communication between replica set members. See *Configure mongod and mongos for TLS/SSL* (page 451) for more information.

As with all MongoDB components, run arbiters in trusted network environments.

12.2.2 Replica Set Deployment Architectures

On this page

- Strategies (page 637)
- Replica Set Naming (page 638)
- Deployment Patterns (page 638)

The architecture of a *replica set* affects the set's capacity and capability. This document provides strategies for replica set deployments and describes common architectures.

The standard replica set deployment for production system is a three-member replica set. These sets provide redundancy and fault tolerance. Avoid complexity when possible, but let your application requirements dictate the architecture.

Strategies

Determine the Number of Members

Add members in a replica set according to these strategies.

Maximum Number of Voting Members A replica set can have up to `50 members`, but only `7 voting members`. [6] If the replica set already has 7 voting members, additional members must be *non-voting members* (page 646).

Deploy an Odd Number of Members Ensure that the replica set has an odd number of voting members. If you have an *even* number of voting members, deploy an *arbiter* (page **??**) so that the set has an odd number of voting members.

An *arbiter* does not store a copy of the data and requires fewer resources. As a result, you may run an arbiter on an application server or other shared process. With no copy of the data, it may be possible to place an arbiter into environments that you would not place other members of the replica set. Consult your security policies.

> **Warning:** In general, avoid deploying more than one arbiter per replica set.

Consider Fault Tolerance *Fault tolerance* for a replica set is the number of members that can become unavailable and still leave enough members in the set to elect a primary. In other words, it is the difference between the number of members in the set and the majority of voting members needed to elect a primary. Without a primary, a replica set cannot accept write operations. Fault tolerance is an effect of replica set size, but the relationship is not direct. See the following table:

Number of Members	Majority Required to Elect a New Primary	Fault Tolerance
3	2	1
4	3	1
5	3	2
6	4	2

Adding a member to the replica set does not *always* increase the fault tolerance. However, in these cases, additional members can provide support for dedicated functions, such as backups or reporting.

[6] While replica sets are the recommended solution for production, a replica set can support up to `50 members` in total. If your deployment requires more than 50 members, you'll need to use *master-slave* (page 659) replication. However, master-slave replication lacks the automatic failover capabilities.

Use Hidden and Delayed Members for Dedicated Functions Add *hidden* (page 633) or *delayed* (page 634) members to support dedicated functions, such as backup or reporting.

Load Balance on Read-Heavy Deployments In a deployment with *very* high read traffic, you can improve read throughput by distributing reads to secondary members. As your deployment grows, add or move members to alternate data centers to improve redundancy and availability.

Always ensure that the main facility is able to elect a primary.

Add Capacity Ahead of Demand The existing members of a replica set must have spare capacity to support adding a new member. Always add new members before the current demand saturates the capacity of the set.

Distribute Members Geographically

To protect your data in case of a data center failure, keep at least one member in an alternate data center. If possible, use an odd number of data centers, and choose a distribution of members that maximizes the likelihood that even with a loss of a data center, the remaining replica set members can form a majority or at minimum, provide a copy of your data.

To ensure that the members in your main data center be elected primary before the members in the alternate data center, set the `members[n].priority` (page 720) of the members in the alternate data center to be lower than that of the members in the primary data center.

For more information, see *Replica Sets Distributed Across Two or More Data Centers* (page 642)

Target Operations with Tag Sets

Use *replica set tag sets* (page 700) to target read operations to specific members or to customize write concern to request acknowledgement from specific members.

See also:

Data Center Awareness (page 308) and *Operational Segregation in MongoDB Deployments* (page 308).

Use Journaling to Protect Against Power Failures

MongoDB enables *journaling* (page 606) by default. Journaling protects against data loss in the event of service interruptions, such as power failures and unexpected reboots.

Replica Set Naming

If your application connects to more than one replica set, each set should have a distinct name. Some drivers group replica set connections by replica set name.

Deployment Patterns

The following documents describe common replica set deployment patterns. Other patterns are possible and effective depending on the application's requirements. If needed, combine features of each architecture in your own deployment:

Three Member Replica Sets **(page 639)** Three-member replica sets provide the minimum recommended architecture for a replica set.

Replica Sets Distributed Across Two or More Data Centers **(page 642)** Geographically distributed sets include members in multiple locations to protect against facility-specific failures, such as power outages.

Three Member Replica Sets

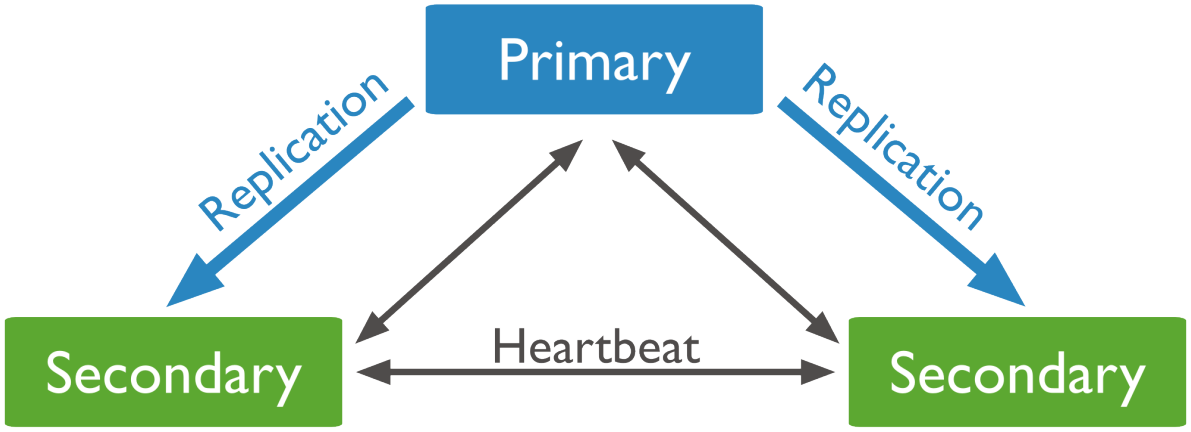
On this page

- Primary with Two Secondary Members (page 639)
- Primary with a Secondary and an Arbiter (page 639)

The minimum architecture of a replica set has three members. A three member replica set can have either three members that hold data, or two members that hold data and an arbiter.

Primary with Two Secondary Members A replica set with three members that store data has:

- One *primary* (page 628).
- Two *secondary* (page 628) members. Both secondaries can become the primary in an *election* (page 644).



These deployments provide two complete copies of the data set at all times in addition to the primary. These replica sets provide additional fault tolerance and *high availability* (page 644). If the primary is unavailable, the replica set elects a secondary to be primary and continues normal operation. The old primary rejoins the set when available.

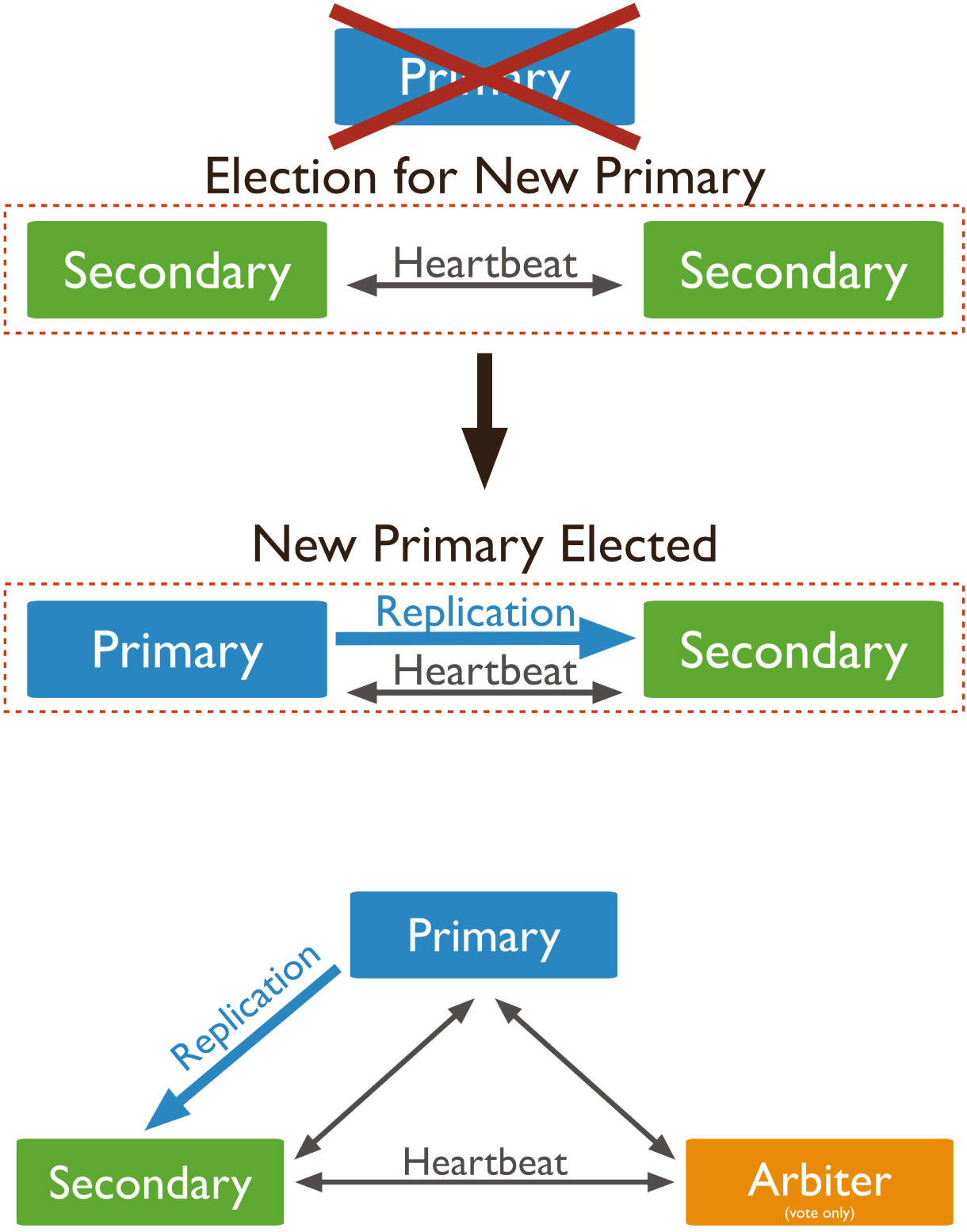
Primary with a Secondary and an Arbiter A three member replica set with a two members that store data has:

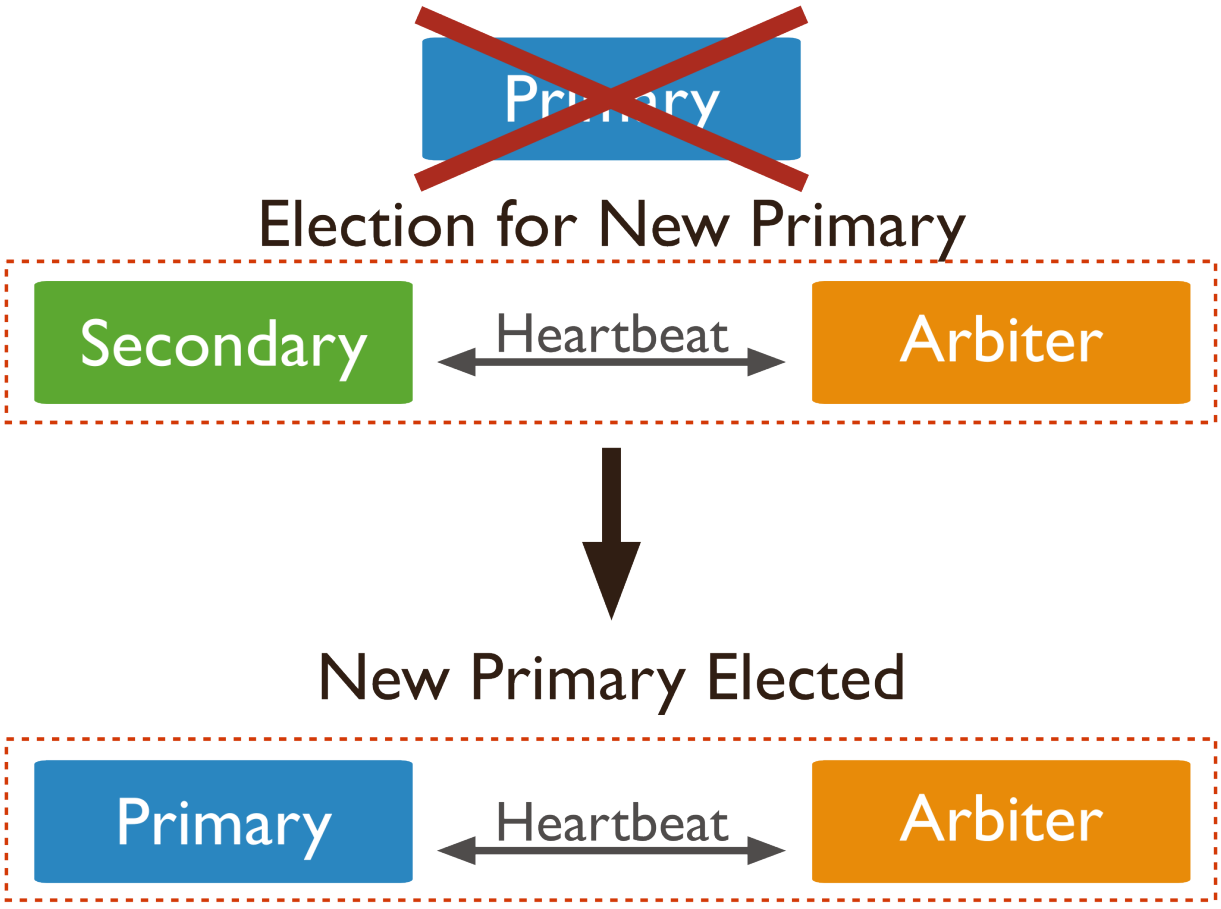
- One *primary* (page 628).
- One *secondary* (page 628) member. The secondary can become primary in an *election* (page 644).
- One *arbiter* (page 635). The arbiter only votes in elections.

Since the arbiter does not hold a copy of the data, these deployments provides only one complete copy of the data. Arbiters require fewer resources, at the expense of more limited redundancy and fault tolerance.

However, a deployment with a primary, secondary, and an arbiter ensures that a replica set remains available if the primary *or* the secondary is unavailable. If the primary is unavailable, the replica set will elect the secondary to be primary.

See also:


Election for New Primary
Secondary
Heartbeat
Secondary
New Primary Elected
Primary
Replication
Heartbeat
Secondary
Primary
Replication
Secondary
Heartbeat
Arbiter
(vote only)


Election for New Primary
Secondary
Heartbeat
Arbiter
New Primary Elected
Primary
Heartbeat
Arbiter

Deploy a Replica Set (page 667).

Replica Sets Distributed Across Two or More Data Centers

On this page

- Overview (page 642)
- Distribution of the Members (page 642)
- Electability of Members (page 643)
- Connectivity (page 643)
- Additional Resource (page 644)

Overview While *replica sets* provide basic protection against single-instance failure, replica sets whose members are all located in a single data center are susceptible to data center failures. Power outages, network interruptions, and natural disasters are all issues that can affect replica sets whose members are located in a single facility.

Distributing replica set members across geographically distincts data centers adds redundancy and provides fault tolerance if one of the data centers is unavailable.

Distribution of the Members To protect your data in case of a data center failure, keep at least one member in an alternate data center. If possible, use an odd number of data centers, and choose a distribution of members that maximizes the likelihood that even with a loss of a data center, the remaining replica set members can form a majority or at minimum, provide a copy of your data.

Examples

Three-member Replica Set For example, for a three-member replica set, some possible distributions of members include:

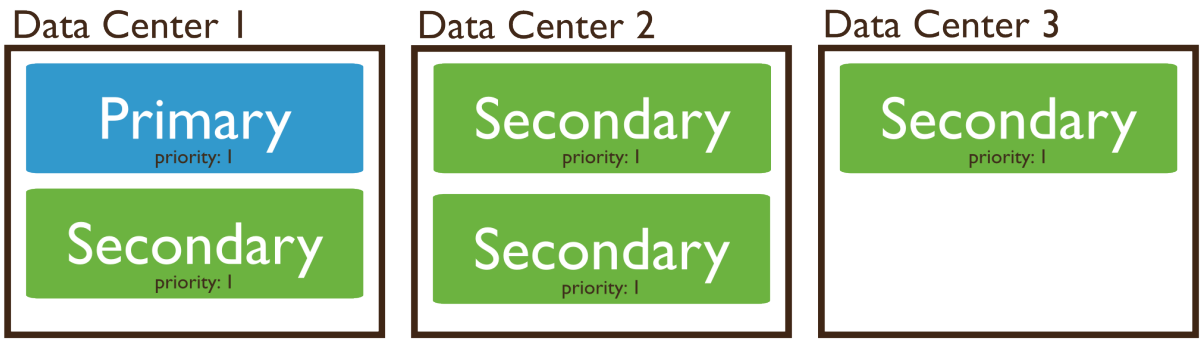
- Two data centers: two members to Data Center 1 and one member to Data Center 2. If one of the members of the replica set is an arbiter, distribute the arbiter to Data Center 1 with a data-bearing member.
  - If Data Center 1 goes down, the replica set becomes read-only.
  - If Data Center 2 goes down, the replica set remains writeable as the members in Data Center 1 can hold an election.
- Three data centers: one members to Data Center 1, one member to Data Center 2, and one member to Data Center 3.
  - If any Data Center goes down, the replica set remains writeable as the remaining members can hold an election.

Five-member Replica Set For a replica set with 5 members, some possible distributions of members include:

- Two data centers: three members to Data Center 1 and two members to Data Center 2.
  - If Data Center 1 goes down, the replica set becomes read-only.
  - If Data Center 2 goes down, the replica set remains writeable as the members in Data Center 1 can create a majority.

- Three data centers: two member to Data Center 1, two members to Data Center 2, and one member to site Data Center 3.
  - If any Data Center goes down, the replica set remains writeable as the remaining members can hold an election.

For example, the following 5 member replica set distributes its members across three data centers.



Electability of Members Some members of the replica set, such as members that have networking restraint or limited resources, should not be able to become primary in a *failover*. Configure members that should not become primary to have *priority 0* (page 631).

In some cases, you may prefer that the members in one data center be elected primary before the members in the other data centers. You can modify the `priority` (page 720) of the members such that the members in the one data center has higher `priority` (page 720) than the members in the other data centers.

In the following example, the replica set members in Data Center 1 have a higher priority than the members in Data Center 2 and 3; the members in Data Center 2 have a higher priority than the member in Data Center 3:

Connectivity Verify that your network configuration allows communication among all members; i.e. each member must be able to connect to every other member.

See also:

Deploy a Geographically Redundant Replica Set (page 672), *Deploy a Replica Set* (page 667), *Add an Arbiter to Replica Set* (page 677), and *Add Members to a Replica Set* (page 679).

Additional Resource

- Whitepaper: MongoDB Multi-Data Center Deployments[7]
- Webinar: Multi-Data Center Deployment[8]

12.2.3 Replica Set High Availability

Replica sets provide high availability using automatic *failover*. Failover allows a *secondary* member to become *primary* if the current primary becomes unavailable.

Changed in version 3.2: MongoDB introduces a version 1 of the replication protocol (`protocolVersion: 1` (page 718)) to reduce replica set failover time and accelerates the detection of multiple simultaneous primaries. New replica sets will, by default, use `protocolVersion: 1` (page 718). Previous versions of MongoDB use version 0 of the protocol. To upgrade existing replica sets to use `protocolVersion: 1` (page 718), see *Upgrade a Replica Set to 3.2* (page 902).

Replica set members keep the same data set but are otherwise independent. If the primary becomes unavailable, an eligible secondary holds an *election* (page 644) to elect itself as a new primary. In some situations, the failover process may undertake a *rollback* (page 647). [9]

Replica Set Elections

On this page

- Factors and Conditions that Affect Elections (page 644)
- Non-Voting Members (page 646)

Replica sets use elections to determine which set member will become *primary*. Elections occur after initiating a replica set, and also any time the primary becomes unavailable. The primary is the only member in the set that can accept write operations. If a primary becomes unavailable, elections allow the set to recover normal operations without manual intervention. Elections are part of the *failover process* (page 644).

In the following three-member replica set, the primary is unavailable. One of the remaining secondaries holds an election to elect itself as a new primary.

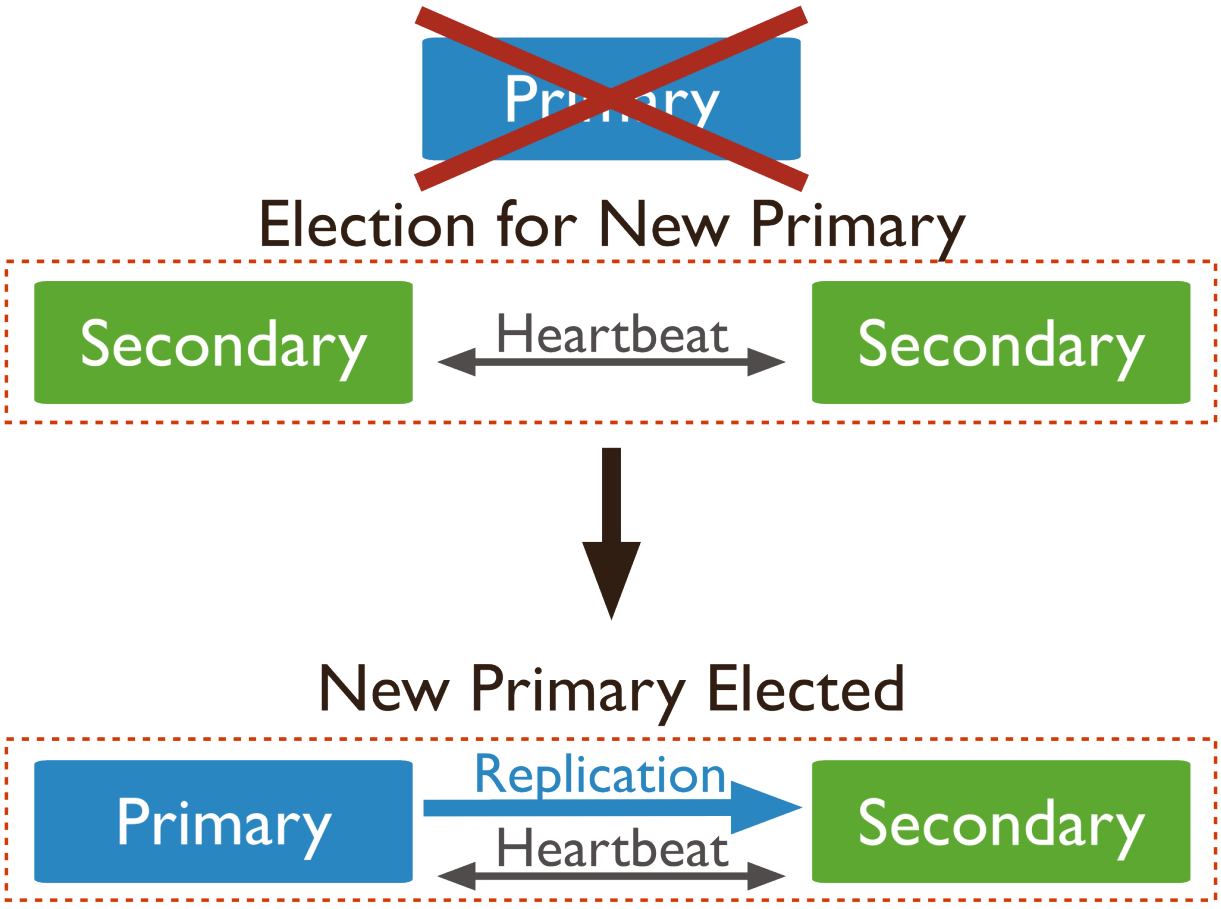
Elections are essential for independent operation of a replica set; however, elections take time to complete. While an election is in process, the replica set has no primary and cannot accept writes and all remaining members become read-only. MongoDB avoids elections unless necessary.

If a majority of the replica set is inaccessible or unavailable to the current primary, the primary will step down and become a secondary. The replica set cannot accept writes after this occurs, but remaining members can continue to serve read queries if such queries are configured to run on secondaries.

Factors and Conditions that Affect Elections

Replication Election Protocol New in version 3.2: MongoDB introduces a version 1 of the replication protocol (`protocolVersion: 1` (page 718)) to reduce replica set failover time and accelerates the detection of multiple simultaneous primaries. New replica sets will, by default, use `protocolVersion: 1` (page 718). Previous versions of MongoDB use version 0 of the protocol.

[7] http://www.mongodb.com/lp/white-paper/multi-dc?jmp=docs
[8] https://www.mongodb.com/presentations/webinar-multi-data-center-deployment?jmp=docs
[9] Replica sets remove "rollback" data when needed without intervention. Administrators must apply or discard rollback data manually.


Election for New Primary
Secondary
Heartbeat
Secondary
New Primary Elected
Primary
Replication
Heartbeat
Secondary

Heartbeats Replica set members send heartbeats (pings) to each other every two seconds. If a heartbeat does not return within 10 seconds, the other members mark the delinquent member as inaccessible.

Member Priority After a replica set has a stable primary, the election algorithm will make a "best-effort" attempt to have the secondary with the highest `priority` (page 720) available call an election. Higher priority secondaries call elections relatively sooner than lower priority secondaries; however, a lower priority node can still be elected as primary for brief periods of time, even if a higher priority secondary is available. Replica set members will continue to call elections until the highest priority available member becomes primary.

Members with a priority value of `0` cannot become primary and do not seek election. For details, see *Priority 0 Replica Set Members* (page 631).

Loss of a Data Center With a distributed replica set, the loss of a data center may affect the ability of the remaining members in other data center or data centers to elect a primary.

If possible, distribute the replica set members across data centers to maximize the likelihood that even with a loss of a data center, one of the remaining replica set members can become the new primary.

See also:

Replica Sets Distributed Across Two or More Data Centers (page 642)

Network Partition A *network partition* may segregate a primary into a partition with a minority of nodes. When the primary detects that it can only see a minority of nodes in the replica set, the primary steps down as primary and becomes a secondary. Independently, a member in the partition that can communicate with a majority of the nodes (including itself) holds an election to become the new primary.

Vetoes in Elections Changed in version 3.2: The `protocolVersion: 1` (page 718) obviates the need for vetos. The following veto discussion applies to replica sets that use the older `protocolVersion: 0` (page 718).

For replica sets using `protocolVersion: 0` (page 718), all members of a replica set can veto an election, including *non-voting members* (page 646). A member will veto an election:

- If the member seeking an election is not a member of the voter's set.
- If the current primary has more recent operations (i.e. a higher `optime`) than the member seeking election, from the perspective of another voting member.
- If the current primary has the same or more recent operations (i.e. a higher or equal `optime`) than the member seeking election.
- If a *priority 0 member* (page 631) [10] is the most current member at the time of the election. In this case, another eligible member of the set will catch up to the state of the *priority 0 member* (page 631) member and then attempt to become primary.
- If the member seeking an election has a lower priority than another member in the set that is also eligible for election.

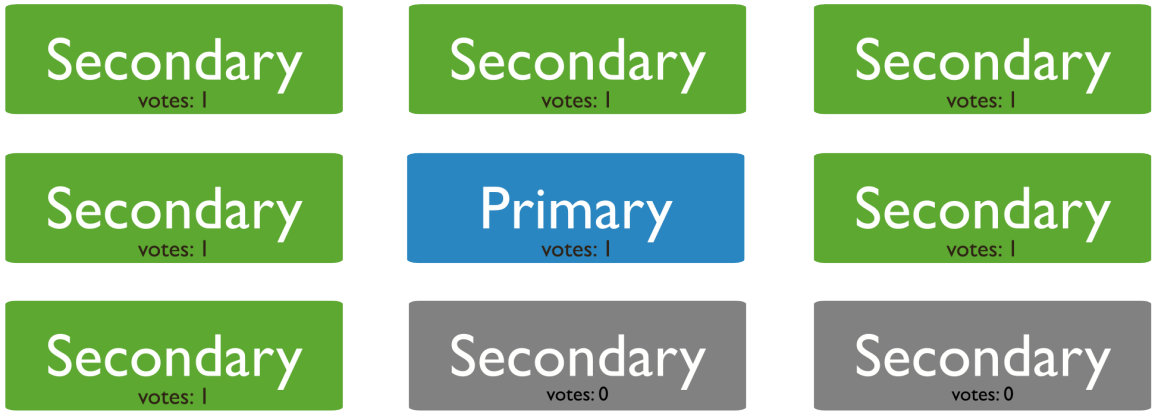
Non-Voting Members

Although non-voting members do not vote in elections, these members hold copies of the replica set's data and can accept read operations from client applications.

[10] *Hidden* (page 633) and *delayed* (page 634) imply *priority 0* (page 631) configuration.

Because a replica set can have up to `50 members`, but only `7 voting members`, non-voting members allow a replica set to have more than seven members.

For instance, the following nine-member replica set has seven voting members and two non-voting members.



A non-voting member has a `members[n].votes` (page 721) setting equal to `0` in its member configuration:

```
{
  "_id" : <num>
  "host" : <hostname:port>,
  "votes" : 0
}
```

Important: Do **not** alter the number of votes to control which members will become primary. Instead, modify the `members[n].priority` (page 720) option. *Only* alter the number of votes in exceptional cases. For example, to permit more than seven members.

To configure a non-voting member, see *Configure Non-Voting Replica Set Member* (page 690).

Rollbacks During Replica Set Failover

On this page

- Collect Rollback Data (page 648)
- Avoid Replica Set Rollbacks (page 648)
- Rollback Limitations (page 648)

A rollback reverts write operations on a former *primary* when the member rejoins its *replica set* after a *failover*. A rollback is necessary only if the primary had accepted write operations that the *secondaries* had **not** successfully replicated before the primary stepped down. When the primary rejoins the set as a secondary, it reverts, or "rolls back," its write operations to maintain database consistency with the other members.

MongoDB attempts to avoid rollbacks, which should be rare. When a rollback does occur, it is often the result of a network partition. Secondaries that can not keep up with the throughput of operations on the former primary, increase the size and impact of the rollback.

A rollback does *not* occur if the write operations replicate to another member of the replica set before the primary steps down *and* if that member remains available and accessible to a majority of the replica set.

Collect Rollback Data

When a rollback does occur, MongoDB writes the rollback data to *BSON* files in the `rollback/` folder under the database's `dbPath` directory. The names of rollback files have the following form:

```
<database>.<collection>.<timestamp>.bson
```

For example:

```
records.accounts.2011-05-09T18-10-04.0.bson
```

To read the contents of the rollback files, use `bsondump`. Based on the content and the knowledge of their applications, administrators can decide the next course of action to take.

Avoid Replica Set Rollbacks

For replica sets, the default *write concern {w: 1}* (page 179) only provides acknowledgement of write operations on the primary. With the default write concern, data may be rolled back if the primary steps down before the write operations have replicated to any of the secondaries.

To prevent rollbacks of data that have been acknowledged to the client, use *w: majority write concern* (page 180) to guarantee that the write operations propagate to a majority of the replica set nodes before returning with acknowledgement to the issuing client.

Note:

- Regardless of *write concern* (page 179), other clients using `"local"` (page 182) (i.e. the default) readConcern can see the result of a write operation before the write operation is acknowledged to the issuing client.
- Clients using `"local"` (page 182) (i.e. the default) readConcern can read data which may be subsequently *rolled back* (page 647).

Rollback Limitations

A `mongod` instance will not rollback more than 300 megabytes of data. If your system must rollback more than 300 megabytes, you must manually intervene to recover the data. If this is the case, the following line will appear in your `mongod` log:

```
[replica set sync] replSet syncThread: 13410 replSet too much data to roll back
```

In this situation, save the data directly or force the member to perform an initial sync. To force initial sync, sync from a "current" member of the set by deleting the content of the `dbPath` directory for the member that requires a larger rollback.

See also:

Replica Set High Availability (page 644) and *Replica Set Elections* (page 644).

12.2.4 Replica Set Read and Write Semantics

From the perspective of a client application, whether a MongoDB instance is running as a single server (i.e. "standalone") or a *replica set* is transparent. However, MongoDB provides additional read and write configurations for replica sets.

Note: *Sharded clusters* where the shards are also replica sets provide the same operational semantics with regards to

write and read operations.

Write Concern for Replica Sets **(page 649)** Write concern describes the level of acknowledgement requested from MongoDB for write operations.

Read Preference **(page 651)** Read preference specifies where (i.e. which members of the replica set) the drivers should direct the read operations.

Read Preference Processes **(page 654)** Describes the mechanics of read preference.

Write Concern for Replica Sets

On this page

- Verify Write Operations to Replica Sets (page 649)
- Modify Default Write Concern (page 649)
- Custom Write Concerns (page 651)

From the perspective of a client application, whether a MongoDB instance is running as a single server (i.e. "standalone") or a *replica set* is transparent. However, replica sets offer some configuration options for write. [11]

Verify Write Operations to Replica Sets

For a replica set, the default *write concern* (page 179) requests acknowledgement only from the primary. You can, however, override this default write concern, such as to confirm write operations on a specified number of the replica set members.

To override the default write concern, specify a write concern with each write operation. For example, the following method includes a write concern that specifies that the method return only after the write propagates to the primary and at least one secondary or the method times out after 5 seconds.

```
db.products.insert(
   { item: "envelopes", qty : 100, type: "Clasp" },
   { writeConcern: { w: 2, wtimeout: 5000 } }
)
```

You can include a timeout threshold for a write concern. This prevents write operations from blocking indefinitely if the write concern is unachievable. For example, if the write concern requires acknowledgement from 4 members of the replica set and the replica set has only available 3 members, the operation blocks until those members become available. See *wtimeout* (page 181).

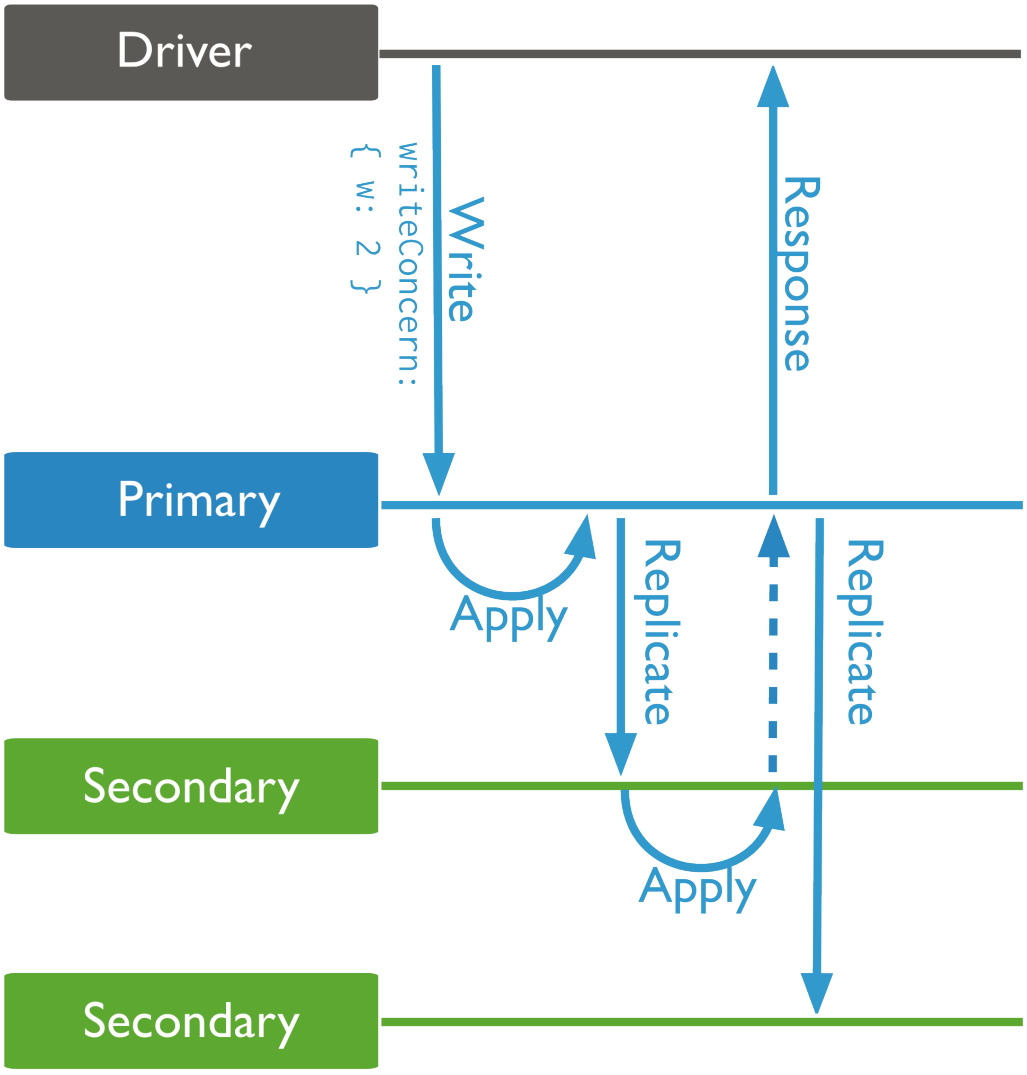
See also:

Write Method Acknowledgements (page 1002)

Modify Default Write Concern

You can modify the default write concern for a replica set by setting the `settings.getLastErrorDefaults` (page 722) setting in the *replica set configuration* (page 717). The following sequence of commands creates a configuration that waits for the write operation to complete on a majority of the voting members before returning:

[11] *Sharded clusters* where the shards are also replica sets provide the same configuration options with regards to write and read operations.


Driver
Write
writeConcern:
{ w: 2 }
Response
Primary
Apply
Replicate
Replicate
Secondary
Apply
Secondary

```
cfg = rs.conf()
cfg.settings = {}
cfg.settings.getLastErrorDefaults = { w: "majority", wtimeout: 5000 }
rs.reconfig(cfg)
```

If you issue a write operation with a specific write concern, the write operation uses its own write concern instead of the default.

See also:

Write Concern (page 179)

Custom Write Concerns

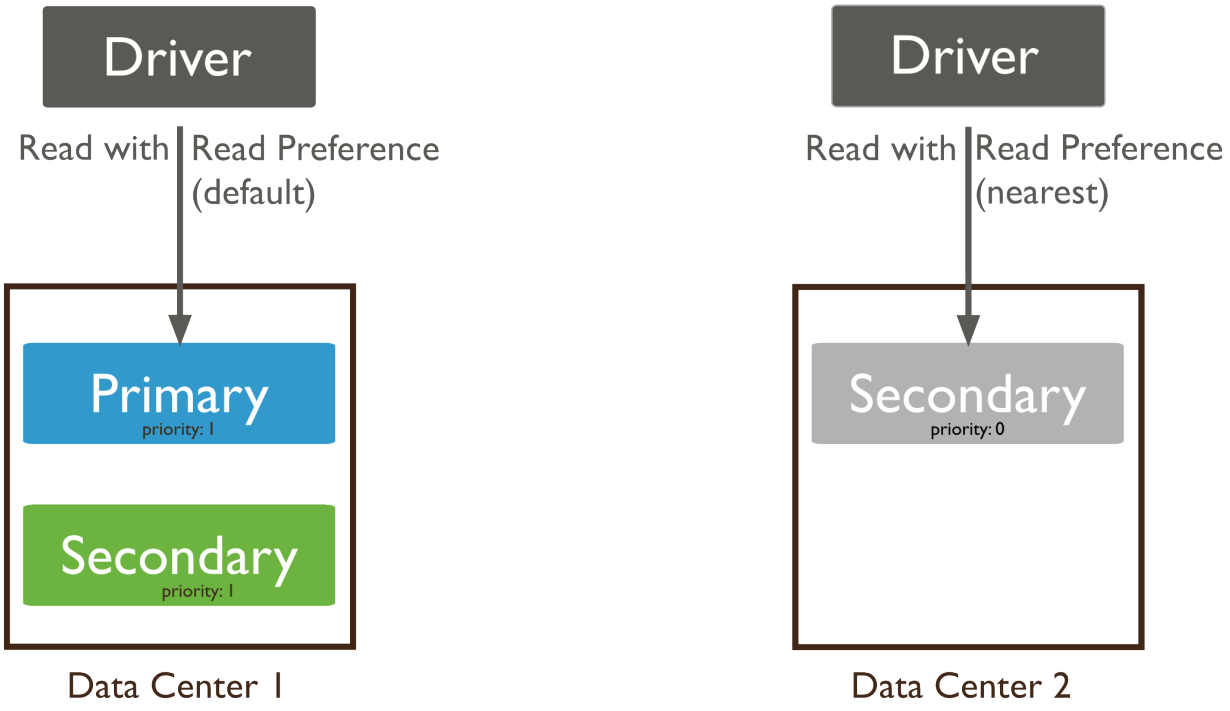
You can *tag* (page 700) the members of replica sets and use the resulting tag sets to create custom write concerns. See *Configure Replica Set Tag Sets* (page 700) for information on configuring custom write concerns using tag sets.

Read Preference

On this page

- Use Cases (page 652)
- Read Preference Modes (page 653)
- Tag Sets (page 653)

Read preference describes how MongoDB clients route read operations to the members of a *replica set*.



By default, an application directs its read operations to the *primary* member in a *replica set*.

In MongoDB, in a replica set with one primary member [12],

- With `"local"` (page 182) `readConcern`, reads from the primary reflect the latest writes in absence of a failover;
- With `"majority"` (page 182) `readConcern`, read operations from the primary or the secondaries have *eventual consistency*.

Important: Exercise care when specifying read preferences: Modes other than `primary` (page 728) may return stale data because with *asynchronous replication* (page 625), data in the secondary may not reflect the most recent write operations. [1]

Note: The read preference does not affect the visibility of data; i.e, clients can see the results of writes before they are acknowledged or have propagated to a majority of replica set members:

- Regardless of *write concern* (page 179), other clients using `"local"` (page 182) (i.e. the default) readConcern can see the result of a write operation before the write operation is acknowledged to the issuing client.
- Clients using `"local"` (page 182) (i.e. the default) readConcern can read data which may be subsequently *rolled back* (page 647).

Use Cases

Indications The following are common use cases for using non-`primary` (page 728) read preference modes:

- Running systems operations that do not affect the front-end application.

  Note: Read preferences aren't relevant to direct connections to a single `mongod` instance. However, in order to perform read operations on a direct connection to a secondary member of a replica set, you must set a read preference, such as *secondary*.

- Providing local reads for geographically distributed applications.

  If you have application servers in multiple data centers, you may consider having a *geographically distributed replica set* (page 642) and using a non primary read preference or the `nearest` (page 729). This allows the client to read from the lowest-latency members, rather than always reading from the primary.

- Maintaining availability during a failover.

  Use `primaryPreferred` (page 728) if you want an application to read from the primary under normal circumstances, but to allow stale reads from secondaries when the primary is unavailable. This provides a "read-only mode" for your application during a failover.

Counter-Indications In general, do *not* use `secondary` (page 728) and `secondaryPreferred` (page 729) to provide extra capacity for reads, because:

- All members of a replica have roughly equivalent write traffic; as a result, secondaries will service reads at roughly the same rate as the primary.

[12] In *some circumstances* (page 729), two nodes in a replica set may *transiently* believe that they are the primary, but at most, one of them will be able to complete writes with `{ w: "majority" }` (page 180) write concern. The node that can complete `{ w: "majority" }` (page 180) writes is the current primary, and the other node is a former primary that has not yet recognized its demotion, typically due to a *network partition*. When this occurs, clients that connect to the former primary may observe stale data despite having requested read preference `primary` (page 728), and new writes to the former primary will eventually roll back.

- Replication is asynchronous and there is some amount of delay between a successful write operation and its replication to secondaries. Reading from a secondary can return out-of-date data; reading from different secondaries may result in non-monotonic reads.
- Distributing read operations to secondaries can compromise availability if *any* members of the set become unavailable because the remaining members of the set will need to be able to handle all application requests.
- For queries of sharded collections, for clusters with the *balancer* (page 758) active, secondaries may return stale results with missing or duplicated data because of incomplete or terminated chunk migrations.

Sharding (page 733) increases read and write capacity by distributing read and write operations across a group of machines, and is often a better strategy for adding capacity.

See *Read Preference Processes* (page 654) for more information about the internal application of read preferences.

Read Preference Modes

Important: All read preference modes except `primary` (page 728) may return stale data because *secondaries* replicate operations from the primary with some delay. [1] Ensure that your application can tolerate stale data if you choose to use a non-`primary` (page 728) mode.

MongoDB `drivers` support five read preference modes.

Read Preference Mode	Description
`primary` (page 728)	Default mode. All operations read from the current replica set *primary*.
`primaryPreferred` (page 728)	In most situations, operations read from the *primary* but if it is unavailable, operations read from *secondary* members.
`secondary` (page 728)	All operations read from the *secondary* members of the replica set.
`secondaryPreferred` (page 729)	In most situations, operations read from *secondary* members but if no *secondary* members are available, operations read from the *primary*.
`nearest` (page 729)	Operations read from member of the *replica set* with the least network latency, irrespective of the member's type.

The syntax for specifying the read preference mode is specific to the driver and to the idioms of the host language[13].

Read preference modes are also available to clients connecting to a *sharded cluster* through a `mongos`. The `mongos` instance obeys specified read preferences when connecting to the *replica set* that provides each *shard* in the cluster.

In the `mongo` shell, the `readPref()` cursor method provides access to read preferences.

For more information, see *read preference background* (page 651) and *read preference behavior* (page 654). See also the documentation for your driver[14].

Tag Sets

Tag sets allow you to target read operations to specific members of a replica set.

Custom read preferences and write concerns evaluate tag sets in different ways. Read preferences consider the value of a tag when selecting a member to read from. Write concerns ignore the value of a tag to when selecting a member, *except* to consider whether or not the value is unique.

You can specify tag sets with the following read preference modes:

[13] https://api.mongodb.org/
[14] https://api.mongodb.org/

- primaryPreferred (page 728)
- secondary (page 728)
- secondaryPreferred (page 729)
- nearest (page 729)

Tags are not compatible with mode primary (page 728) and, in general, only apply when *selecting* (page 654) a *secondary* member of a set for a read operation. However, the nearest (page 729) read mode, when combined with a tag set, selects the matching member with the lowest network latency. This member may be a primary or secondary.

All interfaces use the same *member selection logic* (page 654) to choose the member to which to direct read operations, basing the choice on read preference mode and tag sets.

For information on configuring tag sets, see the *Configure Replica Set Tag Sets* (page 700) tutorial.

For more information on how read preference *modes* (page 728) interact with tag sets, see the *documentation for each read preference mode* (page 727).

Read Preference Processes

On this page

- Member Selection (page 654)
- Request Association (page 655)
- Auto-Retry (page 655)
- Read Preference in Sharded Clusters (page 656)

MongoDB drivers use the following procedures to direct operations to replica sets and sharded clusters. To determine how to route their operations, applications periodically update their view of the replica set's state, identifying which members are up or down, which member is *primary*, and verifying the latency to each mongod instance.

Member Selection

Clients, by way of their drivers, and mongos instances for sharded clusters, periodically update their view of the replica set's state.

When you select non-primary (page 728) read preference, the driver will determine which member to target using the following process:

1. Assembles a list of suitable members, taking into account member type (i.e. secondary, primary, or all members).
2. Excludes members not matching the tag sets, if specified.
3. Determines which suitable member is the closest to the client in absolute terms.
4. Builds a list of members that are within a defined ping distance (in milliseconds) of the "absolute nearest" member.

   Applications can configure the threshold used in this stage. The default "acceptable latency" is 15 milliseconds, which you can override in the drivers with their own secondaryAcceptableLatencyMS option. For mongos you can use the *--localThreshold* or localPingThresholdMs runtime options to set this value.
5. Selects a member from these hosts at random. The member receives the read operation.

Drivers can then associate the thread or connection with the selected member. This *request association* (page 655) is configurable by the application. See your `driver` documentation about request association configuration and default behavior.

Request Association

Important: *Request association* is configurable by the application. See your `driver` documentation about request association configuration and default behavior.

Because *secondary* members of a *replica set* may lag behind the current *primary* by different amounts, reads for *secondary* members may reflect data at different points in time. To prevent sequential reads from jumping around in time, the driver **can** associate application threads to a specific member of the set after the first read, thereby preventing reads from other members. The thread will continue to read from the same member until:

- The application performs a read with a different read preference,
- The thread terminates, or
- The client receives a socket exception, as is the case when there's a network error or when the `mongod` closes connections during a *failover*. This triggers a *retry* (page 655), which may be transparent to the application.

When using request association, if the client detects that the set has elected a new *primary*, the driver will discard all associations between threads and members.

Auto-Retry

Connections between MongoDB drivers and `mongod` instances in a *replica set* must balance two concerns:

1. The client should attempt to prefer current results, and any connection should read from the same member of the replica set as much as possible. Requests should prefer *request association* (page 655) (e.g. *pinning*).
2. The client should minimize the amount of time that the database is inaccessible as the result of a connection issue, networking problem, or *failover* in a replica set.

As a result, MongoDB drivers:

- Reuse a connection to a specific `mongod` for as long as possible after establishing a connection to that instance. This connection is *pinned* to this `mongod`.
- Attempt to reconnect to a new member, obeying existing *read preference modes* (page 728), if the connection to `mongod` is lost.

  Reconnections are transparent to the application itself. If the connection permits reads from *secondary* members, after reconnecting, the application can receive two sequential reads returning from different secondaries. Depending on the state of the individual secondary member's replication, the documents can reflect the state of your database at different moments.
- Return an error *only* after attempting to connect to three members of the set that match the *read preference mode* (page 728) and *tag set* (page 653). If there are fewer than three members of the set, the client will error after connecting to all existing members of the set.

  After this error, the driver selects a new member using the specified read preference mode. In the absence of a specified read preference, the driver uses `primary` (page 728).
- After detecting a failover situation, [15] the driver attempts to refresh the state of the replica set as quickly as possible.

[15] When a *failover* occurs, all members of the set close all client connections that produce a socket error in the driver. This behavior prevents or minimizes *rollback*.

Changed in version 3.0.0: `mongos` instances take a slightly different approach. `mongos` instances return connections to secondaries to the connection pool after every request. As a result, the `mongos` reevaluates read preference for every operation.

Read Preference in Sharded Clusters

In most *sharded clusters*, each shard consists of a *replica set*. As such, read preferences are also applicable. With regard to read preference, read operations in a sharded cluster are identical to unsharded replica sets.

Unlike simple replica sets, in sharded clusters, all interactions with the shards pass from the clients to the `mongos` instances that are actually connected to the set members. `mongos` is then responsible for the application of read preferences, which is transparent to applications.

There are no configuration changes required for full support of read preference modes in sharded environments, as long as the `mongos` is at least version 2.2. All `mongos` maintain their own connection pool to the replica set members. As a result:

- A request without a specified preference has `primary` (page 728), the default, unless, the `mongos` reuses an existing connection that has a different mode set.

  To prevent confusion, always explicitly set your read preference mode.

- All `nearest` (page 729) and latency calculations reflect the connection between the `mongos` and the `mongod` instances, not the client and the `mongod` instances.

  This produces the desired result, because all results must pass through the `mongos` before returning to the client.

12.2.5 Replication Processes

Members of a *replica set* replicate data continuously. First, a member uses *initial sync* to capture the data set. Then the member continuously records and applies every operation that modifies the data set. Every member records operations in its *oplog* (page 656), which is a *capped collection*.

Replica Set Oplog **(page 656)** The oplog records all operations that modify the data in the replica set.

Replica Set Data Synchronization **(page 658)** Secondaries must replicate all changes accepted by the primary. This process is the basis of replica set operations.

Replica Set Oplog

On this page

- Oplog Size (page 657)
- Workloads that Might Require a Larger Oplog Size (page 657)
- Oplog Status (page 658)

The *oplog* (operations log) is a special *capped collection* that keeps a rolling record of all operations that modify the data stored in your databases. MongoDB applies database operations on the *primary* and then records the operations on the primary's oplog. The *secondary* members then copy and apply these operations in an asynchronous process. All replica set members contain a copy of the oplog, in the `local.oplog.rs` (page 724) collection, which allows them to maintain the current state of the database.

To facilitate replication, all replica set members send heartbeats (pings) to all other members. Any member can import oplog entries from any other member.

Whether applied once or multiple times to the target dataset, each operation in the oplog produces the same results, i.e. each operation in the oplog is *idempotent*. For proper replication operations, entries in the oplog must be idempotent:

- initial sync
- post-rollback catch-up
- sharding chunk migrations

Oplog Size

When you start a replica set member for the first time, MongoDB creates an oplog of a default size. The size depends on the architectural details of your operating system.

In most cases, the default oplog size is sufficient. For example, if an oplog is 5% of free disk space and fills up in 24 hours of operations, then secondaries can stop copying entries from the oplog for up to 24 hours without becoming too stale to continue replicating. However, most replica sets have much lower operation volumes, and their oplogs can hold much higher numbers of operations.

Before `mongod` creates an oplog, you can specify its size with the `oplogSizeMB` option. However, after you have started a replica set member for the first time, you can only change the size of the oplog using the *Change the Size of the Oplog* (page 693) procedure.

By default, the size of the oplog is as follows:

- For 64-bit Linux, Solaris, FreeBSD, and Windows systems, MongoDB allocates 5% of the available free disk space, but will always allocate at least 1 gigabyte and never more than 50 gigabytes.
- For 64-bit OS X systems, MongoDB allocates 183 megabytes of space to the oplog.
- For 32-bit systems, MongoDB allocates about 48 megabytes of space to the oplog.

Workloads that Might Require a Larger Oplog Size

If you can predict your replica set's workload to resemble one of the following patterns, then you might want to create an oplog that is larger than the default. Conversely, if your application predominantly performs reads with a minimal amount of write operations, a smaller oplog may be sufficient.

The following workloads might require a larger oplog size.

Updates to Multiple Documents at Once The oplog must translate multi-updates into individual operations in order to maintain *idempotency*. This can use a great deal of oplog space without a corresponding increase in data size or disk use.

Deletions Equal the Same Amount of Data as Inserts If you delete roughly the same amount of data as you insert, the database will not grow significantly in disk use, but the size of the operation log can be quite large.

Significant Number of In-Place Updates If a significant portion of the workload is updates that do not increase the size of the documents, the database records a large number of operations but does not change the quantity of data on disk.

Oplog Status

To view oplog status, including the size and the time range of operations, issue the `rs.printReplicationInfo()` method. For more information on oplog status, see *Check the Size of the Oplog* (page 714).

Under various exceptional situations, updates to a *secondary's* oplog might lag behind the desired performance time. Use `db.getReplicationInfo()` from a secondary member and the `replication status` output to assess the current state of replication and determine if there is any unintended replication delay.

See *Replication Lag* (page 712) for more information.

Replica Set Data Synchronization

On this page

- Initial Sync (page 658)
- Replication (page 658)

In order to maintain up-to-date copies of the shared data set, secondary members of a replica set *sync* or replicate data from other members. MongoDB uses two forms of data synchronization: *initial sync* (page 658) to populate new members with the full data set, and replication to apply ongoing changes to the entire data set.

Initial Sync

Initial sync copies all the data from one member of the replica set to another member. A member uses initial sync when the member has no data, such as when the member is new, or when the member has data but is missing a history of the set's replication.

When you perform an initial sync, MongoDB:

1. Clones all databases. To clone, the `mongod` queries every collection in each source database and inserts all data into its own copies of these collections. At this time, `_id` indexes are also built. The clone process only copies valid data, omitting invalid documents.
2. Applies all changes to the data set. Using the oplog from the source, the `mongod` updates its data set to reflect the current state of the replica set.
3. Builds all indexes on all collections (except `_id` indexes, which were already completed).

   When the `mongod` finishes building all index builds, the member can transition to a normal state, i.e. *secondary*.

Changed in version 3.0: When the clone process omits an invalid document from the sync, MongoDB writes a message to the logs that begins with `Cloner: found corrupt document in <collection>`.

To perform an initial sync, see *Resync a Member of a Replica Set* (page 699).

Replication

Secondary members replicate data continuously after the initial sync. Secondary members copy the *oplog* (page 656) from their *sync from* source and apply these operations in an asynchronous process.

In most cases, secondaries sync from the primary. Secondaries may automatically change their *sync from* source if needed based on changes in the ping time and state of other members' replication.

Changed in version 3.2: MongoDB 3.2 replica set members with `1 vote` (page 721) cannot sync from members with `0 votes` (page 721).

Secondaries avoid syncing from *delayed members* (page 634) and *hidden members* (page 633).

If a secondary member has `members[n].buildIndexes` (page 719) set to `true`, it can only sync from other members where `buildIndexes` (page 719) is `true`. Members where `buildIndexes` (page 719) is `false` can sync from any other member, barring other sync restrictions. `buildIndexes` (page 719) is `true` by default.

Multithreaded Replication MongoDB applies write operations in batches using multiple threads to improve concurrency. MongoDB groups batches by namespace (*MMAPv1* (page 603)) or by document id (*WiredTiger* (page 595)) and simultaneously applies each group of operations using a different thread. MongoDB always applies write operations to a given document in their original write order.

While applying a batch, MongoDB blocks all read operations. As a result, secondary read queries can never return data that reflect a state that never existed on the primary.

Pre-Fetching Indexes to Improve Replication Throughput

Note: Applies to MMAPv1 only.

With the *MMAPv1* (page 603) storage engine, MongoDB fetches memory pages that hold affected data and indexes to help improve the performance of applying oplog entries. This *pre-fetch* stage minimizes the amount of time MongoDB holds write locks while applying oplog entries. By default, secondaries will pre-fetch all *Indexes* (page 515).

Optionally, you can disable all pre-fetching or only pre-fetch the index on the `_id` field. See the `secondaryIndexPrefetch` setting for more information.

12.2.6 Master Slave Replication

On this page

- Fundamental Operations (page 660)
- Run time Master-Slave Configuration (page 661)
- Security (page 661)
- Ongoing Administration and Operation of Master-Slave Deployments (page 662)

Important: *Replica sets* (page 627) replace *master-slave* replication for most use cases. If possible, use replica sets rather than master-slave replication for all new production deployments. This documentation remains to support legacy deployments and for archival purposes only.

In addition to providing all the functionality of master-slave deployments, replica sets are also more robust for production use. Master-slave replication preceded replica sets and made it possible to have a large number of non-master (i.e. slave) nodes, as well as to restrict replicated operations to only a single database; however, master-slave replication provides less redundancy and does not automate failover. See *Deploy Master-Slave Equivalent using Replica Sets* (page 662) for a replica set configuration that is equivalent to master-slave replication. If you wish to convert an existing master-slave deployment to a replica set, see *Convert a Master-Slave Deployment to a Replica Set* (page 662).

Fundamental Operations

Initial Deployment

To configure a *master-slave* deployment, start two `mongod` instances: one in master mode, and the other in slave mode.

To start a `mongod` instance in master mode, invoke `mongod` as follows:

```
mongod --master --dbpath /data/masterdb/
```

With the *`--master`* option, the `mongod` will create a `local.oplog.$main` (page 725) collection, which the "operation log" that queues operations that the slaves will apply to replicate operations from the master. The *`--dbpath`* is optional.

To start a `mongod` instance in slave mode, invoke `mongod` as follows:

```
mongod --slave --source <masterhostname><:<port>> --dbpath /data/slavedb/
```

Specify the hostname and port of the master instance to the *`--source`* argument. The *`--dbpath`* is optional.

For slave instances, MongoDB stores data about the source server in the `local.sources` (page 725) collection.

Configuration Options for Master-Slave Deployments

As an alternative to specifying the *`--source`* run-time option, can add a document to `local.sources` (page 725) specifying the master instance, as in the following operation in the `mongo` shell:

```
use local
db.sources.find()
db.sources.insert( { host: <masterhostname> <,only: <databasename>> } );
```

In line 1, you switch context to the `local` database. In line 2, the `find()` operation should return no documents, to ensure that there are no documents in the `sources` collection. Finally, line 3 uses `db.collection.insert()` to insert the source document into the `local.sources` (page 725) collection. The model of the `local.sources` (page 725) document is as follows:

`host`
: The host field specifies the master `mongod` instance, and holds a resolvable hostname, i.e. IP address, or a name from a `host` file, or preferably a fully qualified domain name.

  You can append `<:port>` to the host name if the `mongod` is not running on the default `27017` port.

`only`
: Optional. Specify a name of a database. When specified, MongoDB will only replicate the indicated database.

Operational Considerations for Replication with Master Slave Deployments

Master instances store operations in an *oplog* which is a *capped collection* (page 6). As a result, if a slave falls too far behind the state of the master, it cannot "catchup" and must re-sync from scratch. Slave may become out of sync with a master if:

- The slave falls far behind the data updates available from that master.
- The slave stops (i.e. shuts down) and restarts later after the master has overwritten the relevant operations from the master.

When slaves are out of sync, replication stops. Administrators must intervene manually to restart replication. Use the `resync` command. Alternatively, the *`--autoresync`* allows a slave to restart replication automatically, after ten second pause, when the slave falls out of sync with the master. With *`--autoresync`* specified, the slave will only attempt to re-sync once in a ten minute period.

To prevent these situations you should specify a larger oplog when you start the `master` instance, by adding the *`--oplogSize`* option when starting `mongod`. If you do not specify *`--oplogSize`*, `mongod` will allocate 5% of available disk space on start up to the oplog, with a minimum of 1 GB for 64-bit machines and 50 MB for 32-bit machines.

Run time Master-Slave Configuration

MongoDB provides a number of command line options for `mongod` instances in *master-slave* deployments. See the *Master-Slave Replication Command Line Options* for options.

Diagnostics

On a *master* instance, issue the following operation in the `mongo` shell to return replication status from the perspective of the master:

```
rs.printReplicationInfo()
```

New in version 2.6: `rs.printReplicationInfo()`. For previous versions, use `db.printReplicationInfo()`.

On a *slave* instance, use the following operation in the `mongo` shell to return the replication status from the perspective of the slave:

```
rs.printSlaveReplicationInfo()
```

New in version 2.6: `rs.printSlaveReplicationInfo()`. For previous versions, use `db.printSlaveReplicationInfo()`.

Use the `serverStatus` as in the following operation, to return status of the replication:

```
db.serverStatus( { repl: 1 } )
```

See *server status repl fields* for documentation of the relevant section of output.

Security

When running with `authorization` enabled, in *master-slave* deployments configure a `keyFile` so that slave `mongod` instances can authenticate and communicate with the master `mongod` instance.

To enable authentication and configure the `keyFile` add the following option to your configuration file:

```
keyFile = /srv/mongodb/keyfile
```

Note: You may chose to set these run-time configuration options using the *`--keyFile`* option on the command line.

Setting `keyFile` enables authentication and specifies a key file for the `mongod` instances to use when authenticating to each other. The content of the key file is arbitrary but must be the same on all members of the deployment can connect to each other.

The key file must be less one kilobyte in size and may only contain characters in the base64 set. The key file must not have group or "world" permissions on UNIX systems. Use the following command to use the OpenSSL package to generate "random" content for use in a key file:

```
openssl rand -base64 741
```

See also:

Security (page 391) for more information about security in MongoDB

Ongoing Administration and Operation of Master-Slave Deployments

Deploy Master-Slave Equivalent using Replica Sets

If you want a replication configuration that resembles *master-slave* replication, using *replica sets* replica sets, consider the following replica configuration document. In this deployment hosts `<master>` and `<slave>` [16] provide replication that is roughly equivalent to a two-instance master-slave deployment:

```
{
  _id : 'setName',
  members : [
    { _id : 0, host : "<master>", priority : 1 },
    { _id : 1, host : "<slave>", priority : 0, votes : 0 }
  ]
}
```

See *Replica Set Configuration* (page 717) for more information about replica set configurations.

Convert a Master-Slave Deployment to a Replica Set

To convert a master-slave deployment to a replica set, restart the current master as a one-member replica set. Then remove the data directories from previous secondaries and add them as new secondaries to the new replica set.

1. To confirm that the current instance is master, run:

   ```
   db.isMaster()
   ```

   This should return a document that resembles the following:

   ```
   {
           "ismaster" : true,
           "maxBsonObjectSize" : 16777216,
           "maxMessageSizeBytes" : 48000000,
           "localTime" : ISODate("2013-07-08T20:15:13.664Z"),
           "ok" : 1
   }
   ```

2. Shut down the `mongod` processes on the master and all slave(s), using the following command while connected to each instance:

   ```
   db.adminCommand({shutdown : 1, force : true})
   ```

3. Back up your `/data/db` directories, in case you need to revert to the master-slave deployment.

4. Start the former master with the *`--replSet`* option, as in the following:

[16] In replica set configurations, the `members[n].host` (page 719) field must hold a resolvable hostname.

```
mongod --replSet <setname>
```

5. Connect to the `mongod` with the `mongo` shell, and initiate the replica set with the following command:

   ```
   rs.initiate()
   ```

   When the command returns, you will have successfully deployed a one-member replica set. You can check the status of your replica set at any time by running the following command:

   ```
   rs.status()
   ```

You can now follow the *convert a standalone to a replica set* (page 678) tutorial to deploy your replica set, picking up from the *Expand the Replica Set* (page 679) section.

Failing over to a Slave (Promotion)

To permanently failover from a unavailable or damaged *master* (`A` in the following example) to a *slave* (`B`):

1. Shut down `A`.
2. Stop `mongod` on `B`.
3. Back up and move all data files that begin with `local` on `B` from the `dbPath`.

   Warning: Removing `local.*` is irrevocable and cannot be undone. Perform this step with extreme caution.

4. Restart `mongod` on `B` with the `--master` option.

Note: This is a one time operation, and is not reversible. `A` cannot become a slave of `B` until it completes a full resync.

Inverting Master and Slave

If you have a *master* (`A`) and a *slave* (`B`) and you would like to reverse their roles, follow this procedure. The procedure assumes `A` is healthy, up-to-date and available.

If `A` is not healthy but the hardware is okay (power outage, server crash, etc.), skip steps 1 and 2 and in step 8 replace all of `A`'s files with `B`'s files in step 8.

If `A` is not healthy and the hardware is not okay, replace `A` with a new machine. Also follow the instructions in the previous paragraph.

To invert the master and slave in a deployment:

1. Halt writes on `A` using the *fsync* command.
2. Make sure `B` is up to date with the state of `A`.
3. Shut down `B`.
4. Back up and move all data files that begin with `local` on `B` from the `dbPath` to remove the existing `local.sources` data.

   Warning: Removing `local.*` is irrevocable and cannot be undone. Perform this step with extreme caution.

5. Start `B` with the `--master` option.

6. Do a write on B, which primes the *oplog* to provide a new sync start point.
7. Shut down B. B will now have a new set of data files that start with local.
8. Shut down A and replace all files in the dbPath of A that start with local with a copy of the files in the dbPath of B that begin with local.

   Considering compressing the local files from B while you copy them, as they may be quite large.
9. Start B with the *--master* option.
10. Start A with all the usual slave options, but include *fastsync*.

Creating a Slave from an Existing Master's Disk Image

If you can stop write operations to the *master* for an indefinite period, you can copy the data files from the master to the new *slave* and then start the slave with *--fastsync*.

> **Warning:** Be careful with *--fastsync*. If the data on both instances is **not** identical, a discrepancy will exist forever.

fastsync is a way to start a slave by starting with an existing master disk image/backup. This option declares that the administrator guarantees the image is correct and completely up-to-date with that of the master. If you have a full and complete copy of data from a master you can use this option to avoid a full synchronization upon starting the slave.

Creating a Slave from an Existing Slave's Disk Image

You can just copy the other *slave's* data file snapshot without any special options. Only take data snapshots when:

- a mongod process is down, or
- when the mongod is locked using db.fsyncLock() for MMAPv1 or WiredTiger storage engine.

Changed in version 3.2: Starting in MongoDB 3.2, db.fsyncLock() can ensure that the data files do not change for MongoDB instances using either the MMAPv1 or the WiredTiger storage engine, thus providing consistency for the purposes of creating backups.

In previous MongoDB version, db.fsyncLock() *cannot* guarantee a consistent set of files for low-level backups (e.g. via file copy cp, scp, tar) for WiredTiger.

Resyncing a Slave that is too Stale to Recover

Slaves asynchronously apply write operations from the *master* that the slaves poll from the master's *oplog*. The oplog is finite in length, and if a slave is too far behind, a full resync will be necessary. To resync the slave, connect to a slave using the mongo and issue the resync command:

```
use admin
db.runCommand( { resync: 1 } )
```

This forces a full resync of all data (which will be very slow on a large database). You can achieve the same effect by stopping mongod on the slave, deleting the entire content of the dbPath on the slave, and restarting the mongod.

Slave Chaining

Slaves cannot be "chained." They must all connect to the *master* directly.

If a slave attempts "slave from" another slave you will see the following line in the `mongod` long of the shell:

```
assertion 13051 tailable cursor requested on non capped collection ns:local.oplog.$main
```

Correcting a Slave's Source

To change a *slave's* source, manually modify the slave's `local.sources` (page 725) collection.

Example

Consider the following: If you accidentally set an incorrect hostname for the slave's *source*, as in the following example:

```
mongod --slave --source prod.mississippi
```

You can correct this, by restarting the slave without the *--slave* and *--source* arguments:

```
mongod
```

Connect to this `mongod` instance using the `mongo` shell and update the `local.sources` (page 725) collection, with the following operation sequence:

```
use local

db.sources.update( { host : "prod.mississippi" },
                   { $set : { host : "prod.mississippi.example.net" } } )
```

Restart the slave with the correct command line arguments or with no *--source* option. After configuring `local.sources` (page 725) the first time, the *--source* will have no subsequent effect. Therefore, both of the following invocations are correct:

```
mongod --slave --source prod.mississippi.example.net
```

or

```
mongod --slave
```

The slave now polls data from the correct *master*.

12.3 Replica Set Tutorials

The administration of *replica sets* includes the initial deployment of the set, adding and removing members to a set, and configuring the operational parameters and properties of the set. Administrators generally need not intervene in failover or replication processes as MongoDB automates these functions. In the exceptional situations that require manual interventions, the tutorials in these sections describe processes such as resyncing a member. The tutorials in this section form the basis for all replica set administration.

Replica Set Deployment Tutorials **(page 666)** Instructions for deploying replica sets, as well as adding and removing members from an existing replica set.

Deploy a Replica Set **(page 667)** Configure a three-member replica set for production systems.

Convert a Standalone to a Replica Set **(page 678)** Convert an existing standalone `mongod` instance into a three-member replica set.

Add Members to a Replica Set **(page 679)** Add a new member to an existing replica set.

Remove Members from Replica Set **(page 682)** Remove a member from a replica set.

Continue reading from *Replica Set Deployment Tutorials* (page 666) for additional tutorials of related to setting up replica set deployments.

Member Configuration Tutorials **(page 684)** Tutorials that describe the process for configuring replica set members.

Adjust Priority for Replica Set Member **(page 685)** Change the precedence given to a replica set members in an election for primary.

Prevent Secondary from Becoming Primary **(page 686)** Make a secondary member ineligible for election as primary.

Configure a Hidden Replica Set Member **(page 687)** Configure a secondary member to be invisible to applications in order to support significantly different usage, such as a dedicated backups.

Continue reading from *Member Configuration Tutorials* (page 684) for more tutorials that describe replica set configuration.

Replica Set Maintenance Tutorials **(page 693)** Procedures and tasks for common operations on active replica set deployments.

Change the Size of the Oplog **(page 693)** Increase the size of the *oplog* which logs operations. In most cases, the default oplog size is sufficient.

Resync a Member of a Replica Set **(page 699)** Sync the data on a member. Either perform initial sync on a new member or resync the data on an existing member that has fallen too far behind to catch up by way of normal replication.

Force a Member to Become Primary **(page 697)** Force a replica set member to become primary.

Change Hostnames in a Replica Set **(page 706)** Update the replica set configuration to reflect changes in members' hostnames.

Continue reading from *Replica Set Maintenance Tutorials* (page 693) for descriptions of additional replica set maintenance procedures.

Troubleshoot Replica Sets **(page 711)** Describes common issues and operational challenges for replica sets. For additional diagnostic information, see *FAQ: MongoDB Diagnostics* (page 856).

12.3.1 Replica Set Deployment Tutorials

The following tutorials provide information in deploying replica sets.

Deploy a Replica Set **(page 667)** Configure a three-member replica set for production systems.

Deploy a Replica Set for Testing and Development **(page 669)** Configure a three-member replica set for either development or testing systems.

Deploy a Geographically Redundant Replica Set **(page 672)** Create a geographically redundant replica set to protect against location-centered availability limitations (e.g. network and power interruptions).

Add an Arbiter to Replica Set **(page 677)** Add an arbiter give a replica set an odd number of voting members to prevent election ties.

Convert a Standalone to a Replica Set **(page 678)** Convert an existing standalone `mongod` instance into a three-member replica set.

Add Members to a Replica Set **(page 679)** Add a new member to an existing replica set.

Remove Members from Replica Set **(page 682)** Remove a member from a replica set.

Replace a Replica Set Member **(page 684)** Update the replica set configuration when the hostname of a member's corresponding `mongod` instance has changed.

Deploy a Replica Set

On this page

- Overview (page 667)
- Requirements (page 667)
- Considerations When Deploying a Replica Set (page 667)
- Procedure (page 668)

This tutorial describes how to create a three-member *replica set* from three existing `mongod` instances running with *access control* (page 433) disabled.

To deploy a replica set with enabled *access control* (page 433), see *Deploy New Replica Set with Access Control* (page 427). If you wish to deploy a replica set from a single MongoDB instance, see *Convert a Standalone to a Replica Set* (page 678). For more information on replica set deployments, see the *Replication* (page 623) and *Replica Set Deployment Architectures* (page 636) documentation.

Overview

Three member *replica sets* provide enough redundancy to survive most network partitions and other system failures. These sets also have sufficient capacity for many distributed read operations. Replica sets should always have an odd number of members. This ensures that *elections* (page 644) will proceed smoothly. For more about designing replica sets, see *the Replication overview* (page 623).

The basic procedure is to start the `mongod` instances that will become members of the replica set, configure the replica set itself, and then add the `mongod` instances to it.

Requirements

For production deployments, you should maintain as much separation between members as possible by hosting the `mongod` instances on separate machines. When using virtual machines for production deployments, you should place each `mongod` instance on a separate host server serviced by redundant power circuits and redundant network paths.

Before you can deploy a replica set, you must install MongoDB on each system that will be part of your *replica set*. If you have not already installed MongoDB, see the *installation tutorials* (page 21).

Before creating your replica set, you should verify that your network configuration allows communication among all members; i.e. each member must be able to connect to every other member. For instructions on how to check your connection, see *Test Connections Between all Members* (page 713).

Considerations When Deploying a Replica Set

Architecture In a production, deploy each member of the replica set to its own machine and if possible bind to the standard MongoDB port of `27017`. Use the `bind_ip` option to ensure that MongoDB listens for connections from applications on configured addresses.

See *Replica Set Deployment Architectures* (page 636) for more information.

Connectivity Ensure that network traffic can pass between all members of the set and all clients in the network securely and efficiently. Consider the following:

- Establish a virtual private network. Ensure that your network topology routes all traffic between members within a single site over the local area network.
- Configure access control to prevent connections from unknown clients to the replica set.
- Configure networking and firewall rules so that incoming and outgoing packets are permitted only on the default MongoDB port and only from within your deployment.

Finally ensure that each member of a replica set is accessible by way of resolvable DNS or hostnames. You should either configure your DNS names appropriately or set up your systems' `/etc/hosts` file to reflect this configuration.

Configuration Specify the run time configuration on each system in a `configuration file` stored in `/etc/mongod.conf` or a related location. Create the directory where MongoDB stores data files before deploying MongoDB.

For more information about the run time options used above and other configuration options, see `https://docs.mongodb.org/manual/reference/configuration-options`.

Procedure

The following procedure outlines the steps to deploy a replica set when access control is disabled.

Step 1: Start each member of the replica set with the appropriate options. For each member, start a `mongod` and specify the replica set name through the `replSet` option. Specify any other parameters specific to your deployment. For replication-specific parameters, see *cli-mongod-replica-set*.

If your application connects to more than one replica set, each set should have a distinct name. Some drivers group replica set connections by replica set name.

The following example specifies the replica set name through the *`--replSet`* command-line option:

```
mongod --replSet "rs0"
```

You can also specify the `replica set name` in the `configuration file`. To start `mongod` with a configuration file, specify the file with the *`--config`* option:

```
mongod --config $HOME/.mongodb/config
```

In production deployments, you can configure a *init script* to manage this process. Init scripts are beyond the scope of this document.

Step 2: Connect a `mongo` shell to a replica set member. For example, to connect to a `mongod` running on localhost on the default port of `27017`, simply issue:

```
mongo
```

Step 3: Initiate the replica set. Use `rs.initiate()` on *one and only one* member of the replica set:

```
rs.initiate()
```

MongoDB initiates a set that consists of the current member and that uses the default replica set configuration.

Step 4: Verify the initial replica set configuration. Use `rs.conf()` to display the *replica set configuration object* (page 717):

```
rs.conf()
```

The replica set configuration object resembles the following:

```
{
   "_id" : "rs0",
   "version" : 1,
   "members" : [
      {
         "_id" : 1,
         "host" : "mongodb0.example.net:27017"
      }
   ]
}
```

Step 5: Add the remaining members to the replica set. Add the remaining members with the `rs.add()` method. You must be connected to the *primary* to add members to a replica set.

`rs.add()` can, in some cases, trigger an election. If the *mongod* you are connected to becomes a *secondary*, you need to connect the *mongo* shell to the new primary to continue adding new replica set members. Use `rs.status()` to identify the primary in the replica set.

The following example adds two members:

```
rs.add("mongodb1.example.net")
rs.add("mongodb2.example.net")
```

When complete, you have a fully functional replica set. The new replica set will elect a *primary*.

Step 6: Check the status of the replica set. Use the `rs.status()` operation:

```
rs.status()
```

See also:

Deploy New Replica Set with Access Control (page 427)

Deploy a Replica Set for Testing and Development

On this page

- Overview (page 670)
- Requirements (page 670)
- Considerations (page 670)
- Procedure (page 670)

This procedure describes deploying a replica set in a development or test environment. For a production deployment, refer to the *Deploy a Replica Set* (page 667) tutorial.

This tutorial describes how to create a three-member *replica set* from three existing `mongod` instances running with *access control* (page 433) disabled.

To deploy a replica set with enabled *access control* (page 433), see *Deploy New Replica Set with Access Control* (page 427). If you wish to deploy a replica set from a single MongoDB instance, see *Convert a Standalone to a Replica Set* (page 678). For more information on replica set deployments, see the *Replication* (page 623) and *Replica Set Deployment Architectures* (page 636) documentation.

Overview

Three member *replica sets* provide enough redundancy to survive most network partitions and other system failures. These sets also have sufficient capacity for many distributed read operations. Replica sets should always have an odd number of members. This ensures that *elections* (page 644) will proceed smoothly. For more about designing replica sets, see *the Replication overview* (page 623).

The basic procedure is to start the `mongod` instances that will become members of the replica set, configure the replica set itself, and then add the `mongod` instances to it.

Requirements

For test and development systems, you can run your `mongod` instances on a local system, or within a virtual instance.

Before you can deploy a replica set, you must install MongoDB on each system that will be part of your *replica set*. If you have not already installed MongoDB, see the *installation tutorials* (page 21).

Before creating your replica set, you should verify that your network configuration allows all possible connections between each member. For a successful replica set deployment, every member must be able to connect to every other member. For instructions on how to check your connection, see *Test Connections Between all Members* (page 713).

Considerations

Replica Set Naming

Important: These instructions should only be used for test or development deployments.

The examples in this procedure create a new replica set named `rs0`.

If your application connects to more than one replica set, each set should have a distinct name. Some drivers group replica set connections by replica set name.

You will begin by starting three `mongod` instances as members of a replica set named `rs0`.

Procedure

1. Create the necessary data directories for each member by issuing a command similar to the following:

   ```
   mkdir -p /srv/mongodb/rs0-0 /srv/mongodb/rs0-1 /srv/mongodb/rs0-2
   ```

   This will create directories called “rs0-0”, “rs0-1”, and “rs0-2”, which will contain the instances’ database files.

2. Start your `mongod` instances in their own shell windows by issuing the following commands:

   First member:

   ```
   mongod --port 27017 --dbpath /srv/mongodb/rs0-0 --replSet rs0 --smallfiles --oplogSize 128
   ```

   Second member:

   ```
   mongod --port 27018 --dbpath /srv/mongodb/rs0-1 --replSet rs0 --smallfiles --oplogSize 128
   ```

Third member:

```
mongod --port 27019 --dbpath /srv/mongodb/rs0-2 --replSet rs0 --smallfiles --oplogSize 128
```

This starts each instance as a member of a replica set named `rs0`, each running on a distinct port, and specifies the path to your data directory with the *`--dbpath`* setting. If you are already using the suggested ports, select different ports.

The *`--smallfiles`* and *`--oplogSize`* settings reduce the disk space that each `mongod` instance uses. This is ideal for testing and development deployments as it prevents overloading your machine. For more information on these and other configuration options, see `https://docs.mongodb.org/manual/reference/configuration-options`.

3. Connect to one of your `mongod` instances through the `mongo` shell. You will need to indicate which instance by specifying its port number. For the sake of simplicity and clarity, you may want to choose the first one, as in the following command;

   ```
   mongo --port 27017
   ```

4. In the `mongo` shell, use `rs.initiate()` to initiate the replica set. You can create a replica set configuration object in the `mongo` shell environment, as in the following example:

   ```
   rsconf = {
              _id: "rs0",
              members: [
                         {
                          _id: 0,
                          host: "<hostname>:27017"
                         }
                        ]
            }
   ```

   replacing `<hostname>` with your system's hostname, and then pass the `rsconf` file to `rs.initiate()` as follows:

   ```
   rs.initiate( rsconf )
   ```

5. Display the current *replica configuration* (page 717) by issuing the following command:

   ```
   rs.conf()
   ```

   The replica set configuration object resembles the following

   ```
   {
      "_id" : "rs0",
      "version" : 4,
      "members" : [
         {
            "_id" : 1,
            "host" : "localhost:27017"
         }
      ]
   }
   ```

6. In the `mongo` shell connected to the *primary*, add the second and third `mongod` instances to the replica set using the `rs.add()` method. Replace `<hostname>` with your system's hostname in the following examples:

   ```
   rs.add("<hostname>:27018")
   rs.add("<hostname>:27019")
   ```

   When complete, you should have a fully functional replica set. The new replica set will elect a *primary*.

Check the status of your replica set at any time with the `rs.status()` operation.

See also:

The documentation of the following shell functions for more information:

- `rs.initiate()`
- `rs.conf()`
- `rs.reconfig()`
- `rs.add()`

You may also consider the simple setup script[17] as an example of a basic automatically-configured replica set.

Refer to *Replica Set Read and Write Semantics* (page 648) for a detailed explanation of read and write semantics in MongoDB.

Deploy a Geographically Redundant Replica Set

On this page

- Overview (page 672)
- Considerations (page 672)
- Prerequisites (page 673)
- Procedures (page 673)

Overview

This tutorial outlines the process for deploying a *replica set* with *members in multiple locations* (page 642). The tutorial addresses three-member replica sets and five-member replica sets. If you have an even number of replica set members, add an arbiter to deploy an odd number replica set.

For more information on distributed replica sets, see *Replica Sets Distributed Across Two or More Data Centers* (page 642). See also *Replica Set Deployment Architectures* (page 636) and see *Replication* (page 623).

Considerations

Architecture In a production, deploy each member of the replica set to its own machine and if possible bind to the standard MongoDB port of `27017`. Use the `bind_ip` option to ensure that MongoDB listens for connections from applications on configured addresses.

See *Replica Set Deployment Architectures* (page 636) for more information.

Connectivity Ensure that network traffic can pass between all members of the set and all clients in the network securely and efficiently. Consider the following:

- Establish a virtual private network. Ensure that your network topology routes all traffic between members within a single site over the local area network.
- Configure access control to prevent connections from unknown clients to the replica set.

[17] https://github.com/mongodb/mongo-snippets/blob/master/replication/simple-setup.py

- Configure networking and firewall rules so that incoming and outgoing packets are permitted only on the default MongoDB port and only from within your deployment.

Finally ensure that each member of a replica set is accessible by way of resolvable DNS or hostnames. You should either configure your DNS names appropriately or set up your systems' `/etc/hosts` file to reflect this configuration.

Configuration Specify the run time configuration on each system in a `configuration file` stored in `/etc/mongod.conf` or a related location. Create the directory where MongoDB stores data files before deploying MongoDB.

For more information about the run time options used above and other configuration options, see `https://docs.mongodb.org/manual/reference/configuration-options`.

Distribution of the Members If possible, use an odd number of data centers, and choose a distribution of members that maximizes the likelihood that even with a loss of a data center, the remaining replica set members can form a majority or at minimum, provide a copy of your data.

Voting Members Never deploy more than seven voting members.

Prerequisites

For all configurations in this tutorial, deploy each replica set member on a separate system. Although you may deploy more than one replica set member on a single system, doing so reduces the redundancy and capacity of the replica set. Such deployments are typically for testing purposes.

This tutorial assumes you have installed MongoDB on each system that will be part of your replica set. If you have not already installed MongoDB, see the *installation tutorials* (page 21).

Procedures

Deploy a Geographically Redundant Three-Member Replica Set For a geographically redundant three-member replica set deployment, you must decide how to distribute your system. Some possible distributions for the three members are:

- Across Three Data Centers: One members to each site.
- Across Two Data Centers: Two members to Site A and one member to Site B. If one of the members of the replica set is an arbiter, distribute the arbiter to Site A with a data-bearing member.

Step 1: Start each member of the replica set with the appropriate options. For each member, start a `mongod` and specify the replica set name through the `replSet` option. Specify any other parameters specific to your deployment. For replication-specific parameters, see *cli-mongod-replica-set*.

If your application connects to more than one replica set, each set should have a distinct name. Some drivers group replica set connections by replica set name.

The following example specifies the replica set name through the *`--replSet`* command-line option:

```
mongod --replSet "rs0"
```

You can also specify the `replica set name` in the `configuration file`. To start `mongod` with a configuration file, specify the file with the *`--config`* option:

```
mongod --config $HOME/.mongodb/config
```

In production deployments, you can configure a *init script* to manage this process. Init scripts are beyond the scope of this document.

Step 2: Connect a mongo shell to a replica set member. For example, to connect to a `mongod` running on localhost on the default port of `27017`, simply issue:

```
mongo
```

Step 3: Initiate the replica set. Use `rs.initiate()` on *one and only one* member of the replica set:

```
rs.initiate()
```

MongoDB initiates a set that consists of the current member and that uses the default replica set configuration.

Step 4: Verify the initial replica set configuration. Use `rs.conf()` to display the *replica set configuration object* (page 717):

```
rs.conf()
```

The replica set configuration object resembles the following:

```
{
   "_id" : "rs0",
   "version" : 1,
   "members" : [
      {
         "_id" : 1,
         "host" : "mongodb0.example.net:27017"
      }
   ]
}
```

Step 5: Add the remaining members to the replica set. Add the remaining members with the `rs.add()` method. You must be connected to the *primary* to add members to a replica set.

`rs.add()` can, in some cases, trigger an election. If the *mongod* you are connected to becomes a *secondary*, you need to connect the *mongo* shell to the new primary to continue adding new replica set members. Use `rs.status()` to identify the primary in the replica set.

The following example adds two members:

```
rs.add("mongodb1.example.net")
rs.add("mongodb2.example.net")
```

When complete, you have a fully functional replica set. The new replica set will elect a *primary*.

Step 6: Optional. Configure the member eligibility for becoming primary. In some cases, you may prefer that the members in one data center be elected primary before the members in the other data centers. You can modify the `priority` (page 720) of the members such that the members in the one data center has higher `priority` (page 720) than the members in the other data centers.

Some members of the replica set, such as members that have networking restraint or limited resources, should not be able to become primary in a *failover*. Configure members that should not become primary to have *priority 0* (page 631).

For example, to lower the relative eligibility of the the member located in one of the sites (in this example, `mongodb2.example.net`), set the member's priority to `0.5`.

1. View the replica set configuration to determine the `members` (page 719) array position for the member. Keep in mind the array position is not the same as the `_id`:

   ```
   rs.conf()
   ```

2. Copy the replica set configuration object to a variable (to `cfg` in the example below). Then, in the variable, set the correct priority for the member. Then pass the variable to `rs.reconfig()` to update the replica set configuration.

   For example, to set priority for the third member in the array (i.e., the member at position 2), issue the following sequence of commands:

   ```
   cfg = rs.conf()
   cfg.members[2].priority = 0.5
   rs.reconfig(cfg)
   ```

   Note: The `rs.reconfig()` shell method can force the current primary to step down, causing an election. When the primary steps down, all clients will disconnect. This is the intended behavior. While most elections complete within a minute, always make sure any replica configuration changes occur during scheduled maintenance periods.

After these commands return, you have a geographically redundant three-member replica set.

Step 7: Check the status of the replica set. Use the `rs.status()` operation:

```
rs.status()
```

Deploy a Geographically Redundant Five-Member Replica Set For a geographically redundant five-member replica set deployment, you must decide how to distribute your system. Some possible distributions for the five members are:

- Across Three Data Centers: Two members in Site A, two members in Site B, one member in Site C.
- Across Four Data Centers: Two members in one site, and one member in the other three sites.
- Across Five Data Centers: One members in each site.
- Across Two Data Centers: Three members in Site A and two members in Site B.

The following five-member replica set includes an arbiter.

Step 1: Start each member of the replica set with the appropriate options. For each member, start a `mongod` and specify the replica set name through the `replSet` option. Specify any other parameters specific to your deployment. For replication-specific parameters, see *cli-mongod-replica-set*.

If your application connects to more than one replica set, each set should have a distinct name. Some drivers group replica set connections by replica set name.

The following example specifies the replica set name through the *`--replSet`* command-line option:

```
mongod --replSet "rs0"
```

You can also specify the `replica set name` in the `configuration file`. To start `mongod` with a configuration file, specify the file with the *`--config`* option:

```
mongod --config $HOME/.mongodb/config
```

In production deployments, you can configure a *init script* to manage this process. Init scripts are beyond the scope of this document.

Step 2: Connect a `mongo` shell to a replica set member. For example, to connect to a `mongod` running on localhost on the default port of `27017`, simply issue:

```
mongo
```

Step 3: Initiate the replica set. Use `rs.initiate()` on *one and only one* member of the replica set:

```
rs.initiate()
```

MongoDB initiates a set that consists of the current member and that uses the default replica set configuration.

Step 4: Verify the initial replica set configuration. Use `rs.conf()` to display the *replica set configuration object* (page 717):

```
rs.conf()
```

The replica set configuration object resembles the following:

```
{
   "_id" : "rs0",
   "version" : 1,
   "members" : [
      {
         "_id" : 1,
         "host" : "mongodb0.example.net:27017"
      }
   ]
}
```

Step 5: Add the remaining secondary members to the replica set. Use `rs.add()` in a `mongo` shell connected to the current primary. The commands should resemble the following:

```
rs.add("mongodb1.example.net")
rs.add("mongodb2.example.net")
rs.add("mongodb3.example.net")
```

When complete, you should have a fully functional replica set. The new replica set will elect a *primary*.

Step 6: Add the arbiter. In the same shell session, issue the following command to add the arbiter (e.g. `mongodb4.example.net`):

```
rs.addArb("mongodb4.example.net")
```

Step 7: Optional. Configure the member eligibility for becoming primary. In some cases, you may prefer that the members in one data center be elected primary before the members in the other data centers. You can modify the `priority` (page 720) of the members such that the members in the one data center has higher `priority` (page 720) than the members in the other data centers.

Some members of the replica set, such as members that have networking restraint or limited resources, should not be able to become primary in a *failover*. Configure members that should not become primary to have *priority 0* (page 631).

For example, to lower the relative eligibility of the the member located in one of the sites (in this example, `mongodb2.example.net`), set the member's priority to `0.5`.

1. View the replica set configuration to determine the `members` (page 719) array position for the member. Keep in mind the array position is not the same as the `_id`:

   ```
   rs.conf()
   ```

2. Copy the replica set configuration object to a variable (to `cfg` in the example below). Then, in the variable, set the correct priority for the member. Then pass the variable to `rs.reconfig()` to update the replica set configuration.

   For example, to set priority for the third member in the array (i.e., the member at position 2), issue the following sequence of commands:

   ```
   cfg = rs.conf()
   cfg.members[2].priority = 0.5
   rs.reconfig(cfg)
   ```

   Note: The `rs.reconfig()` shell method can force the current primary to step down, causing an election. When the primary steps down, all clients will disconnect. This is the intended behavior. While most elections complete within a minute, always make sure any replica configuration changes occur during scheduled maintenance periods.

After these commands return, you have a geographically redundant five-member replica set.

Step 8: Check the status of the replica set. Use the `rs.status()` operation:

```
rs.status()
```

Add an Arbiter to Replica Set

On this page

- Considerations (page 678)
- Add an Arbiter (page 678)

Arbiters are `mongod` instances that are part of a *replica set* but do not hold data. Arbiters participate in *elections* (page 644) in order to break ties. If a replica set has an even number of members, add an arbiter.

Arbiters have minimal resource requirements and do not require dedicated hardware. You can deploy an arbiter on an application server or a monitoring host.

Important: Do not run an arbiter on the same system as a member of the replica set.

Considerations

An arbiter does not store data, but until the arbiter's `mongod` process is added to the replica set, the arbiter will act like any other `mongod` process and start up with a set of data files and with a full-sized *journal*.

To minimize the default creation of data, set the following in the arbiter's `configuration file`:

- `storage.journal.enabled` to `false`

  Warning: Never set `storage.journal.enabled` to `false` on a data-bearing node.

- For MMAPv1 storage engine, `storage.mmapv1.smallFiles` to `true`

These settings are specific to arbiters. Do not set `storage.journal.enabled` to `false` on a data-bearing node. Similarly, do not set `storage.mmapv1.smallFiles` unless specifically indicated.

Add an Arbiter

1. Create a data directory (e.g. `storage.dbPath`) for the arbiter. The `mongod` instance uses the directory for configuration data. The directory *will not* hold the data set. For example, create the `/data/arb` directory:

   ```
   mkdir /data/arb
   ```

2. Start the arbiter, specifying the data directory and the replica set name. The following starts an arbiter using the `/data/arb` as the `dbPath` and `rs` for the replica set name:

   ```
   mongod --port 30000 --dbpath /data/arb --replSet rs
   ```

3. Connect to the primary and add the arbiter to the replica set. Use the `rs.addArb()` method, as in the following example:

   ```
   rs.addArb("m1.example.net:30000")
   ```

   This operation adds the arbiter running on port `30000` on the `m1.example.net` host.

Convert a Standalone to a Replica Set

On this page

- Procedure (page 678)

This tutorial describes the process for converting a *standalone* `mongod` instance into a three-member *replica set*. Use standalone instances for testing and development, but always use replica sets in production. To install a standalone instance, see the *installation tutorials* (page 21).

To deploy a replica set without using a pre-existing `mongod` instance, see *Deploy a Replica Set* (page 667).

Procedure

1. Shut down the *standalone* `mongod` instance.

2. Restart the instance. Use the `--replSet` option to specify the name of the new replica set.

   For example, the following command starts a standalone instance as a member of a new replica set named `rs0`. The command uses the standalone's existing database path of `/srv/mongodb/db0`:

```
mongod --port 27017 --dbpath /srv/mongodb/db0 --replSet rs0
```

If your application connects to more than one replica set, each set should have a distinct name. Some drivers group replica set connections by replica set name.

For more information on configuration options, see `https://docs.mongodb.org/manual/reference/configurati` and the `mongod` manual page.

3. Connect to the `mongod` instance.
4. Use `rs.initiate()` to initiate the new replica set:

   ```
   rs.initiate()
   ```

   The replica set is now operational.

   To view the replica set configuration, use `rs.conf()`. To check the status of the replica set, use `rs.status()`.

Expand the Replica Set Add additional replica set members by doing the following:

1. On two distinct systems, start two new standalone `mongod` instances. For information on starting a standalone instance, see the *installation tutorial* (page 21) specific to your environment.
2. On your connection to the original `mongod` instance (the former standalone instance), issue a command in the following form for each new instance to add to the replica set:

   ```
   rs.add("<hostname><:port>")
   ```

   Replace `<hostname>` and `<port>` with the resolvable hostname and port of the `mongod` instance to add to the set. For more information on adding a host to a replica set, see *Add Members to a Replica Set* (page 679).

Sharding Considerations If the new replica set is part of a *sharded cluster*, change the shard host information in the *config database* by doing the following:

1. Connect to one of the sharded cluster's `mongos` instances and issue a command in the following form:

   ```
   db.getSiblingDB("config").shards.save( {_id: "<name>", host: "<replica-set>/<member,><member,><.
   ```

   Replace `<name>` with the name of the shard. Replace `<replica-set>` with the name of the replica set. Replace `<member,><member,><>` with the list of the members of the replica set.

2. Restart all `mongos` instances. If possible, restart all components of the replica sets (i.e., all `mongos` and all shard `mongod` instances).

Add Members to a Replica Set

On this page

- Overview (page 680)
- Requirements (page 680)
- Procedures (page 680)

Overview

This tutorial explains how to add an additional member to an existing *replica set*. For background on replication deployment patterns, see the *Replica Set Deployment Architectures* (page 636) document.

Maximum Voting Members A replica set can have a maximum of seven *voting members* (page 644). To add a member to a replica set that already has seven voting members, you must either add the member as a *non-voting member* (page 646) or remove a vote from an `existing member`.

Init Scripts In production deployments you can configure a *init script* to manage member processes.

Existing Members You can use these procedures to add new members to an existing set. You can also use the same procedure to "re-add" a removed member. If the removed member's data is still relatively recent, it can recover and catch up easily.

Data Files If you have a backup or snapshot of an existing member, you can move the data files (e.g. the `dbPath` directory) to a new system and use them to quickly initiate a new member. The files must be:

- A valid copy of the data files from a member of the same replica set. See *Backup and Restore with Filesystem Snapshots* (page 343) document for more information.

  Important: Always use filesystem snapshots to create a copy of a member of the existing replica set. **Do not** use `mongodump` and `mongorestore` to seed a new replica set member.

- More recent than the oldest operation in the *primary's oplog*. The new member must be able to become current by applying operations from the primary's oplog.

Requirements

1. An active replica set.
2. A new MongoDB system capable of supporting your data set, accessible by the active replica set through the network.

Otherwise, use the MongoDB *installation tutorial* (page 21) and the *Deploy a Replica Set* (page 667) tutorials.

Procedures

Prepare the Data Directory Before adding a new member to an existing *replica set*, prepare the new member's *data directory* using one of the following strategies:

- Make sure the new member's data directory *does not* contain data. The new member will copy the data from an existing member.

  If the new member is in a *recovering* state, it must exit and become a *secondary* before MongoDB can copy all data as part of the replication process. This process takes time but does not require administrator intervention.

- Manually copy the data directory from an existing member. The new member becomes a secondary member and will catch up to the current state of the replica set. Copying the data over may shorten the amount of time for the new member to become current.

Ensure that you can copy the data directory to the new member and begin replication within the *window allowed by the oplog* (page 657). Otherwise, the new instance will have to perform an initial sync, which completely resynchronizes the data, as described in *Resync a Member of a Replica Set* (page 699).

Use `rs.printReplicationInfo()` to check the current state of replica set members with regards to the oplog.

For background on replication deployment patterns, see the *Replica Set Deployment Architectures* (page 636) document.

Add a Member to an Existing Replica Set

1. Start the new `mongod` instance. Specify the data directory and the replica set name. The following example specifies the `/srv/mongodb/db0` data directory and the `rs0` replica set:

   ```
   mongod --dbpath /srv/mongodb/db0 --replSet rs0
   ```

   Take note of the host name and port information for the new `mongod` instance.

   For more information on configuration options, see the `mongod` manual page.

   Optional

   You can specify the data directory and replica set in the `mongod.conf configuration file`, and start the `mongod` with the following command:

   ```
   mongod --config /etc/mongod.conf
   ```

   ---

2. Connect to the replica set's primary.

   You can only add members while connected to the primary. If you do not know which member is the primary, log into any member of the replica set and issue the `db.isMaster()` command.

3. Use `rs.add()` to add the new member to the replica set. For example, to add a member at host `mongodb3.example.net`, issue the following command:

   ```
   rs.add("mongodb3.example.net")
   ```

   You can include the port number, depending on your setup:

   ```
   rs.add("mongodb3.example.net:27017")
   ```

4. Verify that the member is now part of the replica set. Call the `rs.conf()` method, which displays the *replica set configuration* (page 717):

   ```
   rs.conf()
   ```

   To view replica set status, issue the `rs.status()` method. For a description of the status fields, see `https://docs.mongodb.org/manual/reference/command/replSetGetStatus`.

Configure and Add a Member You can add a member to a replica set by passing to the `rs.add()` method a `members` (page 719) document. The document must be in the form of a `members` (page 719) document. These documents define a replica set member in the same form as the *replica set configuration document* (page 717).

Important: Specify a value for the `_id` field of the `members` (page 719) document. MongoDB does not automatically populate the `_id` field in this case. Finally, the `members` (page 719) document must declare the `host` value. All other fields are optional.

Example

To add a member with the following configuration:

- an `_id` of `1`.
- a `hostname and port number` of `mongodb3.example.net:27017`.
- a `priority` value within the replica set of `0`.
- a configuration as `hidden`,

Issue the following:

```
rs.add({_id: 1, host: "mongodb3.example.net:27017", priority: 0, hidden: true})
```

Remove Members from Replica Set

On this page

- Remove a Member Using `rs.remove()` (page 682)
- Remove a Member Using `rs.reconfig()` (page 682)

To remove a member of a *replica set* use either of the following procedures.

Remove a Member Using `rs.remove()`

1. Shut down the `mongod` instance for the member you wish to remove. To shut down the instance, connect using the `mongo` shell and the `db.shutdownServer()` method.
2. Connect to the replica set's current *primary*. To determine the current primary, use `db.isMaster()` while connected to any member of the replica set.
3. Use `rs.remove()` in either of the following forms to remove the member:

   ```
   rs.remove("mongod3.example.net:27017")
   rs.remove("mongod3.example.net")
   ```

   MongoDB disconnects the shell briefly as the replica set elects a new primary. The shell then automatically reconnects. The shell displays a `DBClientCursor::init call() failed` error even though the command succeeds.

Remove a Member Using `rs.reconfig()`

To remove a member you can manually edit the *replica set configuration document* (page 717), as described here.

1. Shut down the `mongod` instance for the member you wish to remove. To shut down the instance, connect using the `mongo` shell and the `db.shutdownServer()` method.
2. Connect to the replica set's current *primary*. To determine the current primary, use `db.isMaster()` while connected to any member of the replica set.
3. Issue the `rs.conf()` method to view the current configuration document and determine the position in the `members` array of the member to remove:

   Example

`mongod_C.example.net` is in position `2` of the following configuration file:

```
{
    "_id" : "rs",
    "version" : 7,
    "members" : [
        {
            "_id" : 0,
            "host" : "mongod_A.example.net:27017"
        },
        {
            "_id" : 1,
            "host" : "mongod_B.example.net:27017"
        },
        {
            "_id" : 2,
            "host" : "mongod_C.example.net:27017"
        }
    ]
}
```

---

4. Assign the current configuration document to the variable `cfg`:

   ```
   cfg = rs.conf()
   ```

5. Modify the `cfg` object to remove the member.

   Example

   To remove `mongod_C.example.net:27017` use the following JavaScript operation:

   ```
   cfg.members.splice(2,1)
   ```

   ---

6. Overwrite the replica set configuration document with the new configuration by issuing the following:

   ```
   rs.reconfig(cfg)
   ```

   As a result of `rs.reconfig()` the shell will disconnect while the replica set renegotiates which member is primary. The shell displays a `DBClientCursor::init call() failed` error even though the command succeeds, and will automatically reconnected.

7. To confirm the new configuration, issue `rs.conf()`.

   For the example above the output would be:

   ```
   {
       "_id" : "rs",
       "version" : 8,
       "members" : [
           {
               "_id" : 0,
               "host" : "mongod_A.example.net:27017"
           },
           {
               "_id" : 1,
               "host" : "mongod_B.example.net:27017"
           }
       ]
   }
   ```

Replace a Replica Set Member

On this page

- Operation (page 684)
- Example (page 684)

If you need to change the hostname of a replica set member without changing the configuration of that member or the set, you can use the operation outlined in this tutorial. For example if you must re-provision systems or rename hosts, you can use this pattern to minimize the scope of that change.

Operation

To change the hostname for a replica set member modify the `members[n].host` (page 719) field. The value of `members[n]._id` (page 719) field will not change when you reconfigure the set.

See *Replica Set Configuration* (page 717) and `rs.reconfig()` for more information.

Note: Any replica set configuration change can trigger the current *primary* to step down, which forces an *election* (page 644). During the election, the current shell session and clients connected to this replica set disconnect, which produces an error even when the operation succeeds.

Example

To change the hostname to `mongo2.example.net` for the replica set member configured at `members[0]`, issue the following sequence of commands:

```
cfg = rs.conf()
cfg.members[0].host = "mongo2.example.net"
rs.reconfig(cfg)
```

12.3.2 Member Configuration Tutorials

The following tutorials provide information in configuring replica set members to support specific operations, such as to provide dedicated backups, to support reporting, or to act as a cold standby.

Adjust Priority for Replica Set Member **(page 685)** Change the precedence given to a replica set members in an election for primary.

Prevent Secondary from Becoming Primary **(page 686)** Make a secondary member ineligible for election as primary.

Configure a Hidden Replica Set Member **(page 687)** Configure a secondary member to be invisible to applications in order to support significantly different usage, such as a dedicated backups.

Configure a Delayed Replica Set Member **(page 689)** Configure a secondary member to keep a delayed copy of the data set in order to provide a rolling backup.

Configure Non-Voting Replica Set Member **(page 690)** Create a secondary member that keeps a copy of the data set but does not vote in an election.

Convert a Secondary to an Arbiter **(page 691)** Convert a secondary to an arbiter.

Adjust Priority for Replica Set Member

On this page

- Overview (page 685)
- Considerations (page 685)
- Procedure (page 685)

Overview

The priority settings of replica set members affect the outcomes of *elections* (page 644) for primary. Use this setting to ensure that some members are more likely to become primary and that others can never become primary.

The value of the member's `members[n].priority` (page 720) setting determines the member's priority in elections. The higher the number, the higher the priority.

Considerations

To modify priorities, you update the `members` (page 719) array in the replica configuration object. The array index begins with `0`. Do **not** confuse this index value with the value of the replica set member's `members[n]._id` (page 719) field in the array.

The value of `members[n].priority` (page 720) can be any floating point (i.e. decimal) number between `0` and `1000`. The default value for the `members[n].priority` (page 720) field is `1`.

To block a member from seeking election as primary, assign it a priority of `0`. *Hidden members* (page 633) and *delayed members* (page 634) have `members[n].priority` (page 720) set to `0`.

For *arbiters* (page 635), the default `members[n].priority` (page 720) value is `1`; however, arbiters cannot become primary regardless of the configured value.

Adjust priority during a scheduled maintenance window. Reconfiguring priority can force the current primary to step down, leading to an election. Before an election the primary closes all open *client* connections.

Procedure

Step 1: Copy the replica set configuration to a variable. In the `mongo` shell, use `rs.conf()` to retrieve the replica set configuration and assign it to a variable. For example:

```
cfg = rs.conf()
```

Step 2: Change each member's priority value. Change each member's `members[n].priority` (page 720) value, as configured in the `members` (page 719) array.

```
cfg.members[0].priority = 0.5
cfg.members[1].priority = 2
cfg.members[2].priority = 2
```

This sequence of operations modifies the value of `cfg` to set the priority for the first three members defined in the `members` (page 719) array.

Step 3: Assign the replica set the new configuration. Use `rs.reconfig()` to apply the new configuration.

```
rs.reconfig(cfg)
```

This operation updates the configuration of the replica set using the configuration defined by the value of `cfg`.

Prevent Secondary from Becoming Primary

On this page

- Overview (page 686)
- Considerations (page 686)
- Procedure (page 686)
- Related Documents (page 687)

Overview

In a replica set, by default all *secondary* members are eligible to become primary through the election process. You can use the `priority` to affect the outcome of these elections by making some members more likely to become primary and other members less likely or unable to become primary.

Secondaries that cannot become primary are also unable to trigger elections. In all other respects these secondaries are identical to other secondaries.

To prevent a *secondary* member from ever becoming a *primary* in a *failover*, assign the secondary a priority of `0`, as described here. For a detailed description of secondary-only members and their purposes, see *Priority 0 Replica Set Members* (page 631).

Considerations

When updating the replica configuration object, access the replica set members in the `members` (page 719) array with the **array index**. The array index begins with `0`. Do **not** confuse this index value with the value of the `members[n]._id` (page 719) field in each document in the `members` (page 719) array.

Note: MongoDB does not permit the current *primary* to have a priority of `0`. To prevent the current primary from again becoming a primary, you must first step down the current primary using `rs.stepDown()`.

Procedure

This tutorial uses a sample replica set with 5 members.

Warning:

- The `rs.reconfig()` shell method can force the current primary to step down, which causes an *election* (page 644). When the primary steps down, the `mongod` closes all client connections. While this typically takes 10-20 seconds, try to make these changes during scheduled maintenance periods.
- To successfully reconfigure a replica set, a majority of the members must be accessible. If your replica set has an even number of members, add an *arbiter* (page 677) to ensure that members can quickly obtain a majority of votes in an election for primary.

Step 1: Retrieve the current replica set configuration. The `rs.conf()` method returns a *replica set configuration document* (page 717) that contains the current configuration for a replica set.

In a `mongo` shell connected to a primary, run the `rs.conf()` method and assign the result to a variable:

```
cfg = rs.conf()
```

The returned document contains a `members` (page 719) field which contains an array of member configuration documents, one document for each member of the replica set.

Step 2: Assign priority value of `0`. To prevent a secondary member from becoming a primary, update the secondary member's `members[n].priority` (page 720) to `0`.

To assign a priority value to a member of the replica set, access the member configuration document using the array index. In this tutorial, the secondary member to change corresponds to the configuration document found at position `2` of the `members` (page 719) array.

```
cfg.members[2].priority = 0
```

The configuration change does not take effect until you reconfigure the replica set.

Step 3: Reconfigure the replica set. Use `rs.reconfig()` method to reconfigure the replica set with the updated replica set configuration document.

Pass the `cfg` variable to the `rs.reconfig()` method:

```
rs.reconfig(cfg)
```

Related Documents

- `members[n].priority` (page 720)
- *Adjust Priority for Replica Set Member* (page 685)
- *Replica Set Reconfiguration*
- *Replica Set Elections* (page 644)

Configure a Hidden Replica Set Member

On this page

- Considerations (page 688)
- Examples (page 688)
- Related Documents (page 689)

Hidden members are part of a *replica set* but cannot become *primary* and are invisible to client applications. Hidden members may vote in *elections* (page 644). For a more information on hidden members and their uses, see *Hidden Replica Set Members* (page 633).

Considerations

The most common use of hidden nodes is to support *delayed members* (page 634). If you only need to prevent a member from becoming primary, configure a *priority 0 member* (page 631).

If the `settings.chainingAllowed` (page 721) setting allows secondary members to sync from other secondaries, MongoDB by default prefers non-hidden members over hidden members when selecting a sync target. MongoDB will only choose hidden members as a last resort. If you want a secondary to sync from a hidden member, use the `replSetSyncFrom` database command to override the default sync target. See the documentation for `replSetSyncFrom` before using the command.

See also:

Manage Chained Replication (page 705)

Examples

Member Configuration Document To configure a secondary member as hidden, set its `members[n].priority` (page 720) value to `0` and set its `members[n].hidden` (page 720) value to `true` in its member configuration:

```
{
  "_id" : <num>
  "host" : <hostname:port>,
  "priority" : 0,
  "hidden" : true
}
```

Configuration Procedure The following example hides the secondary member currently at the index `0` in the `members` (page 719) array. To configure a *hidden member*, use the following sequence of operations in a `mongo` shell connected to the primary, specifying the member to configure by its array index in the `members` (page 719) array:

```
cfg = rs.conf()
cfg.members[0].priority = 0
cfg.members[0].hidden = true
rs.reconfig(cfg)
```

After re-configuring the set, this secondary member has a priority of `0` so that it cannot become primary and is hidden. The other members in the set will not advertise the hidden member in the `isMaster` or `db.isMaster()` output.

When updating the replica configuration object, access the replica set members in the `members` (page 719) array with the **array index**. The array index begins with `0`. Do **not** confuse this index value with the value of the `members[n]._id` (page 719) field in each document in the `members` (page 719) array.

Warning:

- The `rs.reconfig()` shell method can force the current primary to step down, which causes an *election* (page 644). When the primary steps down, the `mongod` closes all client connections. While this typically takes 10-20 seconds, try to make these changes during scheduled maintenance periods.
- To successfully reconfigure a replica set, a majority of the members must be accessible. If your replica set has an even number of members, add an *arbiter* (page 677) to ensure that members can quickly obtain a majority of votes in an election for primary.

Related Documents

- *Replica Set Reconfiguration*
- *Replica Set Elections* (page 644)
- *Read Preference* (page 651)

Configure a Delayed Replica Set Member

On this page

- Example (page 689)
- Related Documents (page 690)

To configure a delayed secondary member, set its `members[n].priority` (page 720) value to `0`, its `members[n].hidden` (page 720) value to `true`, and its `members[n].slaveDelay` (page 721) value to the number of seconds to delay.

Important: The length of the secondary `members[n].slaveDelay` (page 721) must fit within the window of the oplog. If the oplog is shorter than the `members[n].slaveDelay` (page 721) window, the delayed member cannot successfully replicate operations.

When you configure a delayed member, the delay applies both to replication and to the member's *oplog*. For details on delayed members and their uses, see *Delayed Replica Set Members* (page 634).

Example

The following example sets a 1-hour delay on a secondary member currently at the index `0` in the `members` (page 719) array. To set the delay, issue the following sequence of operations in a `mongo` shell connected to the primary:

```
cfg = rs.conf()
cfg.members[0].priority = 0
cfg.members[0].hidden = true
cfg.members[0].slaveDelay = 3600
rs.reconfig(cfg)
```

After the replica set reconfigures, the delayed secondary member cannot become *primary* and is hidden from applications. The `members[n].slaveDelay` (page 721) value delays both replication and the member's *oplog* by 3600 seconds (1 hour).

When updating the replica configuration object, access the replica set members in the `members` (page 719) array with the **array index**. The array index begins with `0`. Do **not** confuse this index value with the value of the `members[n]._id` (page 719) field in each document in the `members` (page 719) array.

Warning:

- The `rs.reconfig()` shell method can force the current primary to step down, which causes an *election* (page 644). When the primary steps down, the `mongod` closes all client connections. While this typically takes 10-20 seconds, try to make these changes during scheduled maintenance periods.
- To successfully reconfigure a replica set, a majority of the members must be accessible. If your replica set has an even number of members, add an *arbiter* (page 677) to ensure that members can quickly obtain a majority of votes in an election for primary.

Related Documents

- members[n].slaveDelay (page 721)
- *Replica Set Reconfiguration*
- *Oplog Size* (page 657)
- *Change the Size of the Oplog* (page 693) tutorial
- *Replica Set Elections* (page 644)

Configure Non-Voting Replica Set Member

On this page

- Example (page 690)
- Related Documents (page 691)

Non-voting members allow you to add additional members for read distribution beyond the maximum seven voting members. To configure a member as non-voting, set its members[n].votes (page 721) value to 0.

Example

To disable the ability to vote in elections for the fourth, fifth, and sixth replica set members, use the following command sequence in the mongo shell connected to the primary. You identify each replica set member by its array index in the members (page 719) array:

```
cfg = rs.conf()
cfg.members[3].votes = 0
cfg.members[4].votes = 0
cfg.members[5].votes = 0
rs.reconfig(cfg)
```

This sequence gives 0 votes to the fourth, fifth, and sixth members of the set according to the order of the members (page 719) array in the output of rs.conf(). This setting allows the set to elect these members as *primary* but does not allow them to vote in elections. Place voting members so that your designated primary or primaries can reach a majority of votes in the event of a network partition.

When updating the replica configuration object, access the replica set members in the members (page 719) array with the **array index**. The array index begins with 0. Do **not** confuse this index value with the value of the members[n]._id (page 719) field in each document in the members (page 719) array.

Warning:

- The rs.reconfig() shell method can force the current primary to step down, which causes an *election* (page 644). When the primary steps down, the mongod closes all client connections. While this typically takes 10-20 seconds, try to make these changes during scheduled maintenance periods.
- To successfully reconfigure a replica set, a majority of the members must be accessible. If your replica set has an even number of members, add an *arbiter* (page 677) to ensure that members can quickly obtain a majority of votes in an election for primary.

In general and when possible, all members should have only 1 vote. This prevents intermittent ties, deadlocks, or the wrong members from becoming primary. Use members[n].priority (page 720) to control which members are more likely to become primary.

Related Documents

- `members[n].votes` (page 721)
- *Replica Set Reconfiguration*
- *Replica Set Elections* (page 644)

Convert a Secondary to an Arbiter

On this page

- Convert Secondary to Arbiter and Reuse the Port Number (page 691)
- Convert Secondary to Arbiter Running on a New Port Number (page 692)

If you have a *secondary* in a *replica set* that no longer needs to hold data but that needs to remain in the set to ensure that the set can *elect a primary* (page 644), you may convert the secondary to an *arbiter* (page **??**) using either procedure in this tutorial. Both procedures are operationally equivalent:

- You may operate the arbiter on the same port as the former secondary. In this procedure, you must shut down the secondary and remove its data before restarting and reconfiguring it as an arbiter.

  For this procedure, see *Convert Secondary to Arbiter and Reuse the Port Number* (page 691).

- Run the arbiter on a new port. In this procedure, you can reconfigure the server as an arbiter before shutting down the instance running as a secondary.

  For this procedure, see *Convert Secondary to Arbiter Running on a New Port Number* (page 692).

Convert Secondary to Arbiter and Reuse the Port Number

1. If your application is connecting directly to the secondary, modify the application so that MongoDB queries don't reach the secondary.
2. Shut down the secondary.
3. Remove the *secondary* from the *replica set* by calling the `rs.remove()` method. Perform this operation while connected to the current *primary* in the `mongo` shell:

   ```
   rs.remove("<hostname><:port>")
   ```

4. Verify that the replica set no longer includes the secondary by calling the `rs.conf()` method in the `mongo` shell:

   ```
   rs.conf()
   ```

5. Move the secondary's data directory to an archive folder. For example:

   ```
   mv /data/db /data/db-old
   ```

   Optional

   You may remove the data instead.

6. Create a new, empty data directory to point to when restarting the `mongod` instance. You can reuse the previous name. For example:

```
mkdir /data/db
```

7. Restart the `mongod` instance for the secondary, specifying the port number, the empty data directory, and the replica set. You can use the same port number you used before. Issue a command similar to the following:

   ```
   mongod --port 27021 --dbpath /data/db --replSet rs
   ```

8. In the `mongo` shell convert the secondary to an arbiter using the `rs.addArb()` method:

   ```
   rs.addArb("<hostname><:port>")
   ```

9. Verify the arbiter belongs to the replica set by calling the `rs.conf()` method in the `mongo` shell.

   ```
   rs.conf()
   ```

   The arbiter member should include the following:

   ```
   "arbiterOnly" : true
   ```

Convert Secondary to Arbiter Running on a New Port Number

1. If your application is connecting directly to the secondary or has a connection string referencing the secondary, modify the application so that MongoDB queries don't reach the secondary.
2. Create a new, empty data directory to be used with the new port number. For example:

   ```
   mkdir /data/db-temp
   ```

3. Start a new `mongod` instance on the new port number, specifying the new data directory and the existing replica set. Issue a command similar to the following:

   ```
   mongod --port 27021 --dbpath /data/db-temp --replSet rs
   ```

4. In the `mongo` shell connected to the current primary, convert the new `mongod` instance to an arbiter using the `rs.addArb()` method:

   ```
   rs.addArb("<hostname><:port>")
   ```

5. Verify the arbiter has been added to the replica set by calling the `rs.conf()` method in the `mongo` shell.

   ```
   rs.conf()
   ```

   The arbiter member should include the following:

   ```
   "arbiterOnly" : true
   ```

6. Shut down the secondary.
7. Remove the *secondary* from the *replica set* by calling the `rs.remove()` method in the `mongo` shell:

   ```
   rs.remove("<hostname><:port>")
   ```

8. Verify that the replica set no longer includes the old secondary by calling the `rs.conf()` method in the `mongo` shell:

   ```
   rs.conf()
   ```

9. Move the secondary's data directory to an archive folder. For example:

```
mv /data/db /data/db-old
```

Optional

You may remove the data instead.

12.3.3 Replica Set Maintenance Tutorials

The following tutorials provide information in maintaining existing replica sets.

Change the Size of the Oplog **(page 693)** Increase the size of the *oplog* which logs operations. In most cases, the default oplog size is sufficient.

Perform Maintenance on Replica Set Members **(page 695)** Perform maintenance on a member of a replica set while minimizing downtime.

Force a Member to Become Primary **(page 697)** Force a replica set member to become primary.

Resync a Member of a Replica Set **(page 699)** Sync the data on a member. Either perform initial sync on a new member or resync the data on an existing member that has fallen too far behind to catch up by way of normal replication.

Configure Replica Set Tag Sets **(page 700)** Assign tags to replica set members for use in targeting read and write operations to specific members.

Reconfigure a Replica Set with Unavailable Members **(page 704)** Reconfigure a replica set when a majority of replica set members are down or unreachable.

Manage Chained Replication **(page 705)** Disable or enable chained replication. Chained replication occurs when a secondary replicates from another secondary instead of the primary.

Change Hostnames in a Replica Set **(page 706)** Update the replica set configuration to reflect changes in members' hostnames.

Configure a Secondary's Sync Target **(page 710)** Specify the member that a secondary member synchronizes from.

Change the Size of the Oplog

On this page

- Overview (page 694)
- Procedure (page 694)

The *oplog* exists internally as a *capped collection*, so you cannot modify its size in the course of normal operations. In most cases the *default oplog size* (page 657) is an acceptable size; however, in some situations you may need a larger or smaller oplog. For example, you might need to change the oplog size if your applications perform large numbers of multi-updates or deletes in short periods of time.

This tutorial describes how to resize the oplog. For a detailed explanation of oplog sizing, see *Oplog Size* (page 657). For details how oplog size affects *delayed members* and affects *replication lag*, see *Delayed Replica Set Members* (page 634).

Overview

To change the size of the oplog, you must perform maintenance on each member of the replica set in turn. The procedure requires: stopping the `mongod` instance and starting as a standalone instance, modifying the oplog size, and restarting the member.

Important: Always start rolling replica set maintenance with the secondaries, and finish with the maintenance on primary member.

Procedure

- Restart the member in standalone mode.

  Tip

  Always use `rs.stepDown()` to force the primary to become a secondary, before stopping the server. This facilitates a more efficient election process.

- Recreate the oplog with the new size and with an old oplog entry as a seed.
- Restart the `mongod` instance as a member of the replica set.

Restart a Secondary in Standalone Mode on a Different Port Shut down the `mongod` instance for one of the non-primary members of your replica set. For example, to shut down, use the `db.shutdownServer()` method:

```
db.shutdownServer()
```

Restart this `mongod` as a standalone instance running on a different port and *without* the *`--replSet`* parameter. Use a command similar to the following:

```
mongod --port 37017 --dbpath /srv/mongodb
```

Create a Backup of the Oplog (Optional) Optionally, backup the existing oplog on the standalone instance, as in the following example:

```
mongodump --db local --collection 'oplog.rs' --port 37017
```

Recreate the Oplog with a New Size and a Seed Entry Save the last entry from the oplog. For example, connect to the instance using the `mongo` shell, and enter the following command to switch to the `local` database:

```
use local
```

In `mongo` shell scripts you can use the following operation to set the `db` object:

```
db = db.getSiblingDB('local')
```

Ensure that the `temp` temporary collection is empty by dropping the collection:

```
db.temp.drop()
```

Use the `db.collection.save()` method and a sort on reverse *natural order* to find the last entry and save it to a temporary collection:

```
db.temp.save( db.oplog.rs.find( { }, { ts: 1, h: 1 } ).sort( {$natural : -1} ).limit(1).next() )
```

To see this oplog entry, use the following operation:

```
db.temp.find()
```

Remove the Existing Oplog Collection Drop the old `oplog.rs` collection in the `local` database. Use the following command:

```
db = db.getSiblingDB('local')
db.oplog.rs.drop()
```

This returns `true` in the shell.

Create a New Oplog Use the `create` command to create a new oplog of a different size. Specify the `size` argument in bytes. A value of `2 * 1024 * 1024 * 1024` will create a new oplog that's 2 gigabytes:

```
db.runCommand( { create: "oplog.rs", capped: true, size: (2 * 1024 * 1024 * 1024) } )
```

Upon success, this command returns the following status:

```
{ "ok" : 1 }
```

Insert the Last Entry of the Old Oplog into the New Oplog Insert the previously saved last entry from the old oplog into the new oplog. For example:

```
db.oplog.rs.save( db.temp.findOne() )
```

To confirm the entry is in the new oplog, use the following operation:

```
db.oplog.rs.find()
```

Restart the Member Restart the `mongod` as a member of the replica set on its usual port. For example:

```
db.shutdownServer()
mongod --replSet rs0 --dbpath /srv/mongodb
```

The replica set member will recover and "catch up" before it is eligible for election to primary.

Repeat Process for all Members that may become Primary Repeat this procedure for all members you want to change the size of the oplog. Repeat the procedure for the primary as part of the following step.

Change the Size of the Oplog on the Primary To finish the rolling maintenance operation, step down the primary with the `rs.stepDown()` method and repeat the oplog resizing procedure above.

Perform Maintenance on Replica Set Members

On this page

- Overview (page 696)
- Procedure (page 696)

Overview

Replica sets allow a MongoDB deployment to remain available during the majority of a maintenance window.

This document outlines the basic procedure for performing maintenance on each of the members of a replica set. Furthermore, this particular sequence strives to minimize the amount of time that the *primary* is unavailable and controlling the impact on the entire deployment.

Use these steps as the basis for common replica set operations, particularly for procedures such as *upgrading to the latest version of MongoDB* (page 335) and *changing the size of the oplog* (page 693).

Procedure

For each member of a replica set, starting with a secondary member, perform the following sequence of events, ending with the primary:

- Restart the `mongod` instance as a standalone.
- Perform the task on the standalone instance.
- Restart the `mongod` instance as a member of the replica set.

Step 1: Stop a secondary. In the `mongo` shell, shut down the `mongod` instance:

```
db.shutdownServer()
```

Step 2: Restart the secondary as a standalone on a different port. At the operating system shell prompt, restart `mongod` as a standalone instance running on a different port and *without* the `--replSet` parameter:

```
mongod --port 37017 --dbpath /srv/mongodb
```

Always start `mongod` with the same user, even when restarting a replica set member as a standalone instance.

Step 3: Perform maintenance operations on the secondary. While the member is a standalone, use the `mongo` shell to perform maintenance:

```
mongo --port 37017
```

Step 4: Restart `mongod` as a member of the replica set. After performing all maintenance tasks, use the following procedure to restart the `mongod` as a member of the replica set on its usual port.

From the `mongo` shell, shut down the standalone server after completing the maintenance:

```
db.shutdownServer()
```

Restart the `mongod` instance as a member of the replica set using its normal command-line arguments or configuration file.

The secondary takes time to *catch up to the primary* (page 658). From the `mongo` shell, use the following command to verify that the member has caught up from the `RECOVERING` (page 726) state to the `SECONDARY` (page 726) state.

```
rs.status()
```

Step 5: Perform maintenance on the primary last. To perform maintenance on the primary after completing maintenance tasks on all secondaries, use `rs.stepDown()` in the `mongo` shell to step down the primary and allow one of the secondaries to be elected the new primary. Specify a 300 second waiting period to prevent the member from being elected primary again for five minutes:

```
rs.stepDown(300)
```

After the primary steps down, the replica set will elect a new primary. See *Replica Set Elections* (page 644) for more information about replica set elections.

Force a Member to Become Primary

On this page

- Overview (page 697)
- Consideration (page 697)
- Procedures (page 697)

Overview

You can force a *replica set* member to become *primary* by giving it a higher `members[n].priority` (page 720) value than any other member in the set.

Optionally, you also can force a member never to become primary by setting its `members[n].priority` (page 720) value to `0`, which means the member can never seek *election* (page 644) as primary. For more information, see *Priority 0 Replica Set Members* (page 631).

For more information on priorities, see `members[n].priority` (page 720).

Consideration

A majority of the configured members of a replica set *must* be available for a set to reconfigure a set or elect a primary. See *Replica Set Elections* (page 644) for more information.

Procedures

Force a Member to be Primary by Setting its Priority High This procedure assumes your current *primary* is `m1.example.net` and that you'd like to instead make `m3.example.net` primary. The procedure also assumes you have a three-member *replica set* with the configuration below. For more information on configurations, see *Replica Set Configuration Use*.

This procedure assumes this configuration:

```
{
    "_id" : "rs",
    "version" : 7,
    "members" : [
        {
            "_id" : 0,
            "host" : "m1.example.net:27017"
        },
        {
```

```
            "_id" : 1,
            "host" : "m2.example.net:27017"
        },
        {
            "_id" : 2,
            "host" : "m3.example.net:27017"
        }
    ]
}
```

1. In a `mongo` shell connected to the primary, use the following sequence of operations to make `m3.example.net` the primary:

   ```
   cfg = rs.conf()
   cfg.members[0].priority = 0.5
   cfg.members[1].priority = 0.5
   cfg.members[2].priority = 1
   rs.reconfig(cfg)
   ```

   The last statement calls `rs.reconfig()` with the modified configuration document to configure `m3.example.net` to have a higher `members[n].priority` (page 720) value than the other `mongod` instances.

   The following sequence of events occur:

   - `m3.example.net` and `m2.example.net` sync with `m1.example.net` (typically within 10 seconds).
   - `m1.example.net` sees that it no longer has highest priority and, in most cases, steps down. `m1.example.net` *does not* step down if `m3.example.net`'s sync is far behind. In that case, `m1.example.net` waits until `m3.example.net` is within 10 seconds of its optime and then steps down. This minimizes the amount of time with no primary following failover.
   - The step down forces on election in which `m3.example.net` becomes primary based on its `priority` setting.

2. Optionally, if `m3.example.net` is more than 10 seconds behind `m1.example.net`'s optime, and if you don't need to have a primary designated within 10 seconds, you can force `m1.example.net` to step down by running:

   ```
   db.adminCommand({replSetStepDown: 86400, force: 1})
   ```

   This prevents `m1.example.net` from being primary for 86,400 seconds (24 hours), even if there is no other member that can become primary. When `m3.example.net` catches up with `m1.example.net` it will become primary.

   If you later want to make `m1.example.net` primary again while it waits for `m3.example.net` to catch up, issue the following command to make `m1.example.net` seek election again:

   ```
   rs.freeze()
   ```

   The `rs.freeze()` provides a wrapper around the `replSetFreeze` database command.

Force a Member to be Primary Using Database Commands Changed in version 1.8.

Consider a *replica set* with the following members:

- `mdb0.example.net` - the current *primary*.
- `mdb1.example.net` - a *secondary*.

- `mdb2.example.net` - a secondary .

To force a member to become primary use the following procedure:

1. In a `mongo` shell, run `rs.status()` to ensure your replica set is running as expected.
2. In a `mongo` shell connected to the `mongod` instance running on `mdb2.example.net`, freeze `mdb2.example.net` so that it does not attempt to become primary for 120 seconds.

   ```
   rs.freeze(120)
   ```

3. In a `mongo` shell connected the `mongod` running on `mdb0.example.net`, step down this instance that the `mongod` is not eligible to become primary for 120 seconds:

   ```
   rs.stepDown(120)
   ```

   `mdb1.example.net` becomes primary.

   Note: During the transition, there is a short window where the set does not have a primary.

For more information, consider the `rs.freeze()` and `rs.stepDown()` methods that wrap the `replSetFreeze` and `replSetStepDown` commands.

Resync a Member of a Replica Set

On this page

- Procedures (page 699)

A *replica set* member becomes "stale" when its replication process falls so far behind that the *primary* overwrites oplog entries the member has not yet replicated. The member cannot catch up and becomes "stale." When this occurs, you must completely resynchronize the member by removing its data and performing an *initial sync* (page 658).

This tutorial addresses both resyncing a stale member and to creating a new member using seed data from another member. When syncing a member, choose a time when the system has the bandwidth to move a large amount of data. Schedule the synchronization during a time of low usage or during a maintenance window.

MongoDB provides two options for performing an initial sync:

- Restart the `mongod` with an empty data directory and let MongoDB's normal initial syncing feature restore the data. This is the more simple option but may take longer to replace the data.

  See *Procedures* (page 699).

- Restart the machine with a copy of a recent data directory from another member in the replica set. This procedure can replace the data more quickly but requires more manual steps.

  See *Sync by Copying Data Files from Another Member* (page 700).

Procedures

Automatically Sync a Member

Warning: During initial sync, `mongod` will remove the content of the `dbPath`.

This procedure relies on MongoDB's regular process for *initial sync* (page 658). This will store the current data on the member. For an overview of MongoDB initial sync process, see the *Replication Processes* (page 656) section.

If the instance has no data, you can simply follow the *Add Members to a Replica Set* (page 679) or *Replace a Replica Set Member* (page 684) procedure to add a new member to a replica set.

You can also force a `mongod` that is already a member of the set to to perform an initial sync by restarting the instance without the content of the `dbPath` as follows:

1. Stop the member's `mongod` instance. To ensure a clean shutdown, use the `db.shutdownServer()` method from the `mongo` shell or on Linux systems, the *`mongod --shutdown`* option.
2. Delete all data and sub-directories from the member's data directory. By removing the data `dbPath`, MongoDB will perform a complete resync. Consider making a backup first.

At this point, the `mongod` will perform an initial sync. The length of the initial sync process depends on the size of the database and network connection between members of the replica set.

Initial sync operations can impact the other members of the set and create additional traffic to the primary and can only occur if another member of the set is accessible and up to date.

Sync by Copying Data Files from Another Member This approach "seeds" a new or stale member using the data files from an existing member of the replica set. The data files **must** be sufficiently recent to allow the new member to catch up with the *oplog*. Otherwise the member would need to perform an initial sync.

Copy the Data Files You can capture the data files as either a snapshot or a direct copy. However, in most cases you cannot copy data files from a running `mongod` instance to another because the data files will change during the file copy operation.

Important: If copying data files, you must copy the content of the `local` database.

You *cannot* use a `mongodump` backup for the data files, **only a snapshot backup**. For approaches to capturing a consistent snapshot of a running `mongod` instance, see the *MongoDB Backup Methods* (page 282) documentation.

Sync the Member After you have copied the data files from the "seed" source, start the `mongod` instance and allow it to apply all operations from the oplog until it reflects the current state of the replica set.

Configure Replica Set Tag Sets

On this page

- Differences Between Read Preferences and Write Concerns (page 701)
- Add Tag Sets to a Replica Set (page 701)
- Custom Multi-Datacenter Write Concerns (page 702)
- Configure Tag Sets for Functional Segregation of Read and Write Operations (page 703)

Tag sets let you customize *write concern* and *read preferences* for a *replica set*. MongoDB stores tag sets in the replica set configuration object, which is the document returned by `rs.conf()`, in the `members[n].tags` (page 720) embedded document.

This section introduces the configuration of tag sets. For an overview on tag sets and their use, see `w: <tag set>` and *Tag Sets* (page 653).

Differences Between Read Preferences and Write Concerns

Custom read preferences and write concerns evaluate tags sets in different ways:

- Read preferences consider the value of a tag when selecting a member to read from.
- Write concerns do not use the value of a tag to select a member except to consider whether or not the value is unique.

For example, a tag set for a read operation may resemble the following document:

```
{ "disk": "ssd", "use": "reporting" }
```

To fulfill such a read operation, a member would need to have both of these tags. Any of the following tag sets would satisfy this requirement:

```
{ "disk": "ssd", "use": "reporting" }
{ "disk": "ssd", "use": "reporting", "rack": "a" }
{ "disk": "ssd", "use": "reporting", "rack": "d" }
{ "disk": "ssd", "use": "reporting", "mem": "r"}
```

The following tag sets would *not* be able to fulfill this query:

```
{ "disk": "ssd" }
{ "use": "reporting" }
{ "disk": "ssd", "use": "production" }
{ "disk": "ssd", "use": "production", "rack": "k" }
{ "disk": "spinning", "use": "reporting", "mem": "32" }
```

Add Tag Sets to a Replica Set

Given the following replica set configuration:

```
{
    "_id" : "rs0",
    "version" : 1,
    "members" : [
            {
                    "_id" : 0,
                    "host" : "mongodb0.example.net:27017"
            },
            {
                    "_id" : 1,
                    "host" : "mongodb1.example.net:27017"
            },
            {
                    "_id" : 2,
                    "host" : "mongodb2.example.net:27017"
            }
    ]
}
```

You could add tag sets to the members of this replica set with the following command sequence in the `mongo` shell:

```
conf = rs.conf()
conf.members[0].tags = { "dc": "east", "use": "production"  }
conf.members[1].tags = { "dc": "east", "use": "reporting"  }
conf.members[2].tags = { "use": "production"  }
rs.reconfig(conf)
```

After this operation the output of `rs.conf()` would resemble the following:

```
{
    "_id" : "rs0",
    "version" : 2,
    "members" : [
            {
                    "_id" : 0,
                    "host" : "mongodb0.example.net:27017",
                    "tags" : {
                            "dc": "east",
                            "use": "production"
                    }
            },
            {
                    "_id" : 1,
                    "host" : "mongodb1.example.net:27017",
                    "tags" : {
                            "dc": "east",
                            "use": "reporting"
                    }
            },
            {
                    "_id" : 2,
                    "host" : "mongodb2.example.net:27017",
                    "tags" : {
                            "use": "production"
                    }
            }
     ]
}
```

Important: In tag sets, all tag values must be strings.

Custom Multi-Datacenter Write Concerns

Given a five member replica set with members in two data centers:

1. a facility `VA` tagged `dc_va`
2. a facility `GTO` tagged `dc_gto`

Create a custom write concern to require confirmation from two data centers using replica set tags, using the following sequence of operations in the `mongo` shell:

1. Create a replica set configuration JavaScript object `conf`:

   ```
   conf = rs.conf()
   ```

2. Add tags to the replica set members reflecting their locations:

   ```
   conf.members[0].tags = { "dc_va": "rack1"}
   conf.members[1].tags = { "dc_va": "rack2"}
   conf.members[2].tags = { "dc_gto": "rack1"}
   conf.members[3].tags = { "dc_gto": "rack2"}
   conf.members[4].tags = { "dc_va": "rack1"}
   rs.reconfig(conf)
   ```

3. Create a custom `settings.getLastErrorModes` (page 722) setting to ensure that the write operation will propagate to at least one member of each facility:

   ```
   conf.settings = { getLastErrorModes: { MultipleDC : { "dc_va": 1, "dc_gto": 1 } } }
   ```

4. Reconfigure the replica set using the modified `conf` configuration object:

   ```
   rs.reconfig(conf)
   ```

To ensure that a write operation propagates to at least one member of the set in both data centers, use the `MultipleDC` write concern mode as follows:

```
db.users.insert( { id: "xyz", status: "A" }, { writeConcern: { w: "MultipleDC" } } )
```

Alternatively, if you want to ensure that each write operation propagates to at least 2 racks in each facility, reconfigure the replica set as follows in the `mongo` shell:

1. Create a replica set configuration object `conf`:

   ```
   conf = rs.conf()
   ```

2. Redefine the `settings.getLastErrorModes` (page 722) value to require two different values of both `dc_va` and `dc_gto`:

   ```
   conf.settings = { getLastErrorModes: { MultipleDC : { "dc_va": 2, "dc_gto": 2}}
   ```

3. Reconfigure the replica set using the modified `conf` configuration object:

   ```
   rs.reconfig(conf)
   ```

Now, the following write operation will only return after the write operation propagates to at least two different racks in the each facility:

Changed in version 2.6: A new protocol for *write operations* (page 995) integrates write concerns with the write operations. Previous versions used the `getLastError` command to specify the write concerns.

```
db.users.insert( { id: "xyz", status: "A" }, { writeConcern: { w: "MultipleDC" } } )
```

Configure Tag Sets for Functional Segregation of Read and Write Operations

Given a replica set with tag sets that reflect:

- data center facility,
- physical rack location of instance, and
- storage system (i.e. disk) type.

Where each member of the set has a tag set that resembles one of the following: [18]

```
{"dc_va": "rack1", disk:"ssd", ssd: "installed" }
{"dc_va": "rack2", disk:"raid"}
{"dc_gto": "rack1", disk:"ssd", ssd: "installed" }
{"dc_gto": "rack2", disk:"raid"}
{"dc_va": "rack1", disk:"ssd", ssd: "installed" }
```

To target a read operation to a member of the replica set with a disk type of `ssd`, you could use the following tag set:

[18] Since read preferences and write concerns use the value of fields in tag sets differently, larger deployments may have some redundancy.

```
{ disk: "ssd" }
```

However, to create comparable write concern modes, you would specify a different set of `settings.getLastErrorModes` (page 722) configuration. Consider the following sequence of operations in the `mongo` shell:

1. Create a replica set configuration object `conf`:

   ```
   conf = rs.conf()
   ```

2. Redefine the `settings.getLastErrorModes` (page 722) value to configure two write concern modes:

   ```
   conf.settings = {
                     "getLastErrorModes" : {
                             "ssd" : {
                                        "ssd" : 1
                             },
                             "MultipleDC" : {
                                     "dc_va" : 1,
                                    "dc_gto" : 1
                             }
                     }
                   }
   ```

3. Reconfigure the replica set using the modified `conf` configuration object:

   ```
   rs.reconfig(conf)
   ```

Now you can specify the `MultipleDC` write concern mode, as in the following, to ensure that a write operation propagates to each data center.

Changed in version 2.6: A new protocol for *write operations* (page 995) integrates write concerns with the write operations. Previous versions used the `getLastError` command to specify the write concerns.

```
db.users.insert( { id: "xyz", status: "A" }, { writeConcern: { w: "MultipleDC" } } )
```

Additionally, you can specify the `ssd` write concern mode to ensure that a write operation propagates to at least one instance with an SSD.

Reconfigure a Replica Set with Unavailable Members

On this page

- Reconfigure by Forcing the Reconfiguration (page 704)

To reconfigure a *replica set* when a **majority** of members are available, use the `rs.reconfig()` operation on the current *primary*, following the example in the *Replica Set Reconfiguration Procedure*.

This document provides steps for re-configuring a replica set when *only* a **minority** of members are accessible.

You may need to use the procedure, for example, in a geographically distributed replica set, where *no* local group of members can reach a majority. See *Replica Set Elections* (page 644) for more information on this situation.

Reconfigure by Forcing the Reconfiguration

This procedure lets you recover while a majority of *replica set* members are down or unreachable. You connect to any surviving member and use the `force` option to the `rs.reconfig()` method.

The `force` option forces a new configuration onto the member. Use this procedure only to recover from catastrophic interruptions. Do not use `force` every time you reconfigure. Also, do not use the `force` option in any automatic scripts and do not use `force` when there is still a *primary*.

To force reconfiguration:

1. Back up a surviving member.
2. Connect to a surviving member and save the current configuration. Consider the following example commands for saving the configuration:

   ```
   cfg = rs.conf()

   printjson(cfg)
   ```

3. On the same member, remove the down and unreachable members of the replica set from the `members` (page 719) array by setting the array equal to the surviving members alone. Consider the following example, which uses the `cfg` variable created in the previous step:

   ```
   cfg.members = [cfg.members[0] , cfg.members[4] , cfg.members[7]]
   ```

4. On the same member, reconfigure the set by using the `rs.reconfig()` command with the `force` option set to `true`:

   ```
   rs.reconfig(cfg, {force : true})
   ```

   This operation forces the secondary to use the new configuration. The configuration is then propagated to all the surviving members listed in the `members` array. The replica set then elects a new primary.

   Note: When you use `force : true`, the version number in the replica set configuration increases significantly, by tens or hundreds of thousands. This is normal and designed to prevent set version collisions if you accidentally force re-configurations on both sides of a network partition and then the network partitioning ends.

   ---

5. If the failure or partition was only temporary, shut down or decommission the removed members as soon as possible.

See also:

Resync a Member of a Replica Set (page 699)

Manage Chained Replication

On this page

- Disable Chained Replication (page 706)
- Re-enable Chained Replication (page 706)

Starting in version 2.0, MongoDB supports chained replication. A chained replication occurs when a *secondary* member replicates from another secondary member instead of from the *primary*. This might be the case, for example, if a secondary selects its replication target based on ping time and if the closest member is another secondary.

Chained replication can reduce load on the primary. But chained replication can also result in increased replication lag, depending on the topology of the network.

You can use the `settings.chainingAllowed` (page 721) setting in *Replica Set Configuration* (page 717) to disable chained replication for situations where chained replication is causing lag.

MongoDB enables chained replication by default. This procedure describes how to disable it and how to re-enable it.

Note: If chained replication is disabled, you still can use `replSetSyncFrom` to specify that a secondary replicates from another secondary. But that configuration will last only until the secondary recalculates which member to sync from.

Disable Chained Replication

To disable chained replication, set the `settings.chainingAllowed` (page 721) field in *Replica Set Configuration* (page 717) to `false`.

You can use the following sequence of commands to set `settings.chainingAllowed` (page 721) to `false`:

1. Copy the configuration settings into the `cfg` object:

   ```
   cfg = rs.config()
   ```

2. Take note of whether the current configuration settings contain the `settings` embedded document. If they do, skip this step.

   Warning: To avoid data loss, skip this step if the configuration settings contain the `settings` embedded document.

   If the current configuration settings **do not** contain the `settings` embedded document, create the embedded document by issuing the following command:

   ```
   cfg.settings = { }
   ```

3. Issue the following sequence of commands to set `settings.chainingAllowed` (page 721) to `false`:

   ```
   cfg.settings.chainingAllowed = false
   rs.reconfig(cfg)
   ```

Re-enable Chained Replication

To re-enable chained replication, set `settings.chainingAllowed` (page 721) to `true`. You can use the following sequence of commands:

```
cfg = rs.config()
cfg.settings.chainingAllowed = true
rs.reconfig(cfg)
```

Change Hostnames in a Replica Set

On this page

- Overview (page 707)
- Assumptions (page 707)
- Change Hostnames while Maintaining Replica Set Availability (page 708)
- Change All Hostnames at the Same Time (page 709)

For most *replica sets*, the hostnames in the `members[n].host` (page 719) field never change. However, if organizational needs change, you might need to migrate some or all host names.

Note: Always use resolvable hostnames for the value of the `members[n].host` (page 719) field in the replica set configuration to avoid confusion and complexity.

Overview

This document provides two separate procedures for changing the hostnames in the `members[n].host` (page 719) field. Use either of the following approaches:

- *Change hostnames without disrupting availability* (page 708). This approach ensures your applications will always be able to read and write data to the replica set, but the approach can take a long time and may incur downtime at the application layer.

  If you use the first procedure, you must configure your applications to connect to the replica set at both the old and new locations, which often requires a restart and reconfiguration at the application layer and which may affect the availability of your applications. Re-configuring applications is beyond the scope of this document.

- *Stop all members running on the old hostnames at once* (page 709). This approach has a shorter maintenance window, but the replica set will be unavailable during the operation.

See also:

Replica Set Reconfiguration Process, *Deploy a Replica Set* (page 667), and *Add Members to a Replica Set* (page 679).

Assumptions

Given a *replica set* with three members:

- `database0.example.com:27017` (the *primary*)
- `database1.example.com:27017`
- `database2.example.com:27017`

And with the following `rs.conf()` output:

```
{
    "_id" : "rs",
    "version" : 3,
    "members" : [
        {
            "_id" : 0,
            "host" : "database0.example.com:27017"
        },
        {
            "_id" : 1,
            "host" : "database1.example.com:27017"
        },
        {
            "_id" : 2,
            "host" : "database2.example.com:27017"
        }
    ]
}
```

The following procedures change the members' hostnames as follows:

- `mongodb0.example.net:27017` (the primary)

- `mongodb1.example.net:27017`
- `mongodb2.example.net:27017`

Use the most appropriate procedure for your deployment.

Change Hostnames while Maintaining Replica Set Availability

This procedure uses the above *assumptions* (page 707).

1. For each *secondary* in the replica set, perform the following sequence of operations:
   (a) Stop the secondary.
   (b) Restart the secondary at the new location.
   (c) Open a `mongo` shell connected to the replica set's primary. In our example, the primary runs on port `27017` so you would issue the following command:

   ```
   mongo --port 27017
   ```

   (d) Use `rs.reconfig()` to update the *replica set configuration document* (page 717) with the new hostname.

   For example, the following sequence of commands updates the hostname for the secondary at the array index `1` of the `members` array (i.e. `members[1]`) in the replica set configuration document:

   ```
   cfg = rs.conf()
   cfg.members[1].host = "mongodb1.example.net:27017"
   rs.reconfig(cfg)
   ```

   For more information on updating the configuration document, see *replica-set-reconfiguration-usage*.
   (e) Make sure your client applications are able to access the set at the new location and that the secondary has a chance to catch up with the other members of the set.

   Repeat the above steps for each non-primary member of the set.
2. Open a `mongo` shell connected to the primary and step down the primary using the `rs.stepDown()` method:

   ```
   rs.stepDown()
   ```

   The replica set elects another member to the become primary.
3. When the step down succeeds, shut down the old primary.
4. Start the `mongod` instance that will become the new primary in the new location.
5. Connect to the current primary, which was just elected, and update the *replica set configuration document* (page 717) with the hostname of the node that is to become the new primary.

   For example, if the old primary was at position `0` and the new primary's hostname is `mongodb0.example.net:27017`, you would run:

   ```
   cfg = rs.conf()
   cfg.members[0].host = "mongodb0.example.net:27017"
   rs.reconfig(cfg)
   ```
6. Open a `mongo` shell connected to the new primary.
7. To confirm the new configuration, call `rs.conf()` in the `mongo` shell.

   Your output should resemble:

```
{
    "_id" : "rs",
    "version" : 4,
    "members" : [
        {
            "_id" : 0,
            "host" : "mongodb0.example.net:27017"
        },
        {
            "_id" : 1,
            "host" : "mongodb1.example.net:27017"
        },
        {
            "_id" : 2,
            "host" : "mongodb2.example.net:27017"
        }
    ]
}
```

Change All Hostnames at the Same Time

This procedure uses the above *assumptions* (page 707).

1. Stop all members in the *replica set*.
2. Restart each member *on a different port* and *without* using the *--replSet* run-time option. Changing the port number during maintenance prevents clients from connecting to this host while you perform maintenance. Use the member's usual *--dbpath*, which in this example is `/data/db1`. Use a command that resembles the following:

   ```
   mongod --dbpath /data/db1/ --port 37017
   ```

3. For each member of the replica set, perform the following sequence of operations:

   (a) Open a `mongo` shell connected to the `mongod` running on the new, temporary port. For example, for a member running on a temporary port of `37017`, you would issue this command:

   ```
   mongo --port 37017
   ```

   (b) Edit the replica set configuration manually. The replica set configuration is the only document in the `system.replset` collection in the `local` database. Edit the replica set configuration with the new hostnames and correct ports for all the members of the replica set. Consider the following sequence of commands to change the hostnames in a three-member set:

   ```
   use local

   cfg = db.system.replset.findOne( { "_id": "rs" } )

   cfg.members[0].host = "mongodb0.example.net:27017"

   cfg.members[1].host = "mongodb1.example.net:27017"

   cfg.members[2].host = "mongodb2.example.net:27017"

   db.system.replset.update( { "_id": "rs" } , cfg )
   ```

   (c) Stop the `mongod` process on the member.

4. After re-configuring all members of the set, start each `mongod` instance in the normal way: use the usual port number and use the *`--replSet`* option. For example:

   ```
   mongod --dbpath /data/db1/ --port 27017 --replSet rs
   ```

5. Connect to one of the `mongod` instances using the `mongo` shell. For example:

   ```
   mongo --port 27017
   ```

6. To confirm the new configuration, call `rs.conf()` in the `mongo` shell.

   Your output should resemble:

   ```
   {
       "_id" : "rs",
       "version" : 4,
       "members" : [
           {
               "_id" : 0,
               "host" : "mongodb0.example.net:27017"
           },
           {
               "_id" : 1,
               "host" : "mongodb1.example.net:27017"
           },
           {
               "_id" : 2,
               "host" : "mongodb2.example.net:27017"
           }
       ]
   }
   ```

Configure a Secondary's Sync Target

On this page

- Overview (page 710)
- Considerations (page 711)
- Procedure (page 711)

Overview

Secondaries capture data from the primary member to maintain an up to date copy of the sets' data. However, by default secondaries may automatically change their sync targets to secondary members based on changes in the ping time between members and the state of other members' replication. See *Replica Set Data Synchronization* (page 658) and *Manage Chained Replication* (page 705) for more information.

For some deployments, implementing a custom replication sync topology may be more effective than the default sync target selection logic. MongoDB provides the ability to specify a host to use as a sync target.

To temporarily override the default sync target selection logic, you may manually configure a *secondary* member's sync target to temporarily pull *oplog* entries. The following provide access to this functionality:

- `replSetSyncFrom` command, or
- `rs.syncFrom()` helper in the `mongo` shell

Considerations

Sync Logic Only modify the default sync logic as needed, and always exercise caution. `replSetSyncFrom`, or `rs.syncFrom()`, will not affect an in-progress initial sync operation. To affect the sync target for the initial sync, run `replSetSyncFrom`, or `rs.syncFrom()`, operation *before* initial sync.

If you run `replSetSyncFrom`, or `rs.syncFrom()`, during initial sync, MongoDB produces no error messages, but the sync target will not change until after the initial sync operation.

Target The member to sync from must be a valid source for data in the set. To sync from a member, the member must:

- Have data. It cannot be an arbiter, in startup or recovering mode, and must be able to answer data queries.
- Be accessible.
- Be a member of the same set in the replica set configuration.
- Build indexes with the `members[n].buildIndexes` (page 719) setting.
- A different member of the set, to prevent syncing from itself.

If you attempt to replicate from a member that is more than 10 seconds behind the current member, `mongod` will log a warning but will still replicate from the lagging member.

If you run `replSetSyncFrom`, or `rs.syncFrom()`, during initial sync, MongoDB produces no error messages, but the sync target will not change until after the initial sync operation.

Persistence `replSetSyncFrom`, or `rs.syncFrom()`, provide a temporary override of default behavior. `mongod` will revert to the default sync behavior in the following situations:

- The `mongod` instance restarts.
- The connection between the `mongod` and the sync target closes.
- If the sync target falls more than 30 seconds behind another member of the replica set.

Procedure

To use the `replSetSyncFrom` command in the `mongo` shell:

```
db.adminCommand( { replSetSyncFrom: "hostname<:port>" } );
```

To use the `rs.syncFrom()` helper in the `mongo` shell:

```
rs.syncFrom("hostname<:port>");
```

12.3.4 Troubleshoot Replica Sets

On this page

- Check Replica Set Status (page 712)
- Check the Replication Lag (page 712)
- Test Connections Between all Members (page 713)
- Socket Exceptions when Rebooting More than One Secondary (page 714)
- Check the Size of the Oplog (page 714)
- Oplog Entry Timestamp Error (page 715)
- Duplicate Key Error on `local.slaves` (page 716)

This section describes common strategies for troubleshooting *replica set* deployments.

Check Replica Set Status

To display the current state of the replica set and current state of each member, run the `rs.status()` method in a `mongo` shell connected to the replica set's *primary*. For descriptions of the information displayed by `rs.status()`, see `https://docs.mongodb.org/manual/reference/command/replSetGetStatus`.

Note: The `rs.status()` method is a wrapper that runs the `replSetGetStatus` database command.

Check the Replication Lag

Replication lag is a delay between an operation on the *primary* and the application of that operation from the *oplog* to the *secondary*. Replication lag can be a significant issue and can seriously affect MongoDB *replica set* deployments. Excessive replication lag makes "lagged" members ineligible to quickly become primary and increases the possibility that distributed read operations will be inconsistent.

To check the current length of replication lag:

- In a `mongo` shell connected to the primary, call the `rs.printSlaveReplicationInfo()` method.

  Returns the `syncedTo` value for each member, which shows the time when the last oplog entry was written to the secondary, as shown in the following example:

  ```
  source: m1.example.net:27017
      syncedTo: Thu Apr 10 2014 10:27:47 GMT-0400 (EDT)
      0 secs (0 hrs) behind the primary
  source: m2.example.net:27017
      syncedTo: Thu Apr 10 2014 10:27:47 GMT-0400 (EDT)
      0 secs (0 hrs) behind the primary
  ```

  A *delayed member* (page 634) may show as `0` seconds behind the primary when the inactivity period on the primary is greater than the `members[n].slaveDelay` (page 721) value.

  Note: The `rs.status()` method is a wrapper around the `replSetGetStatus` database command.

- Monitor the rate of replication by watching the oplog time in the "replica" graph in the MongoDB Cloud Manager[19] and in Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[20]. For more information see the MongoDB Cloud Manager documentation[21] and Ops Manager documentation[22].

[19] https://cloud.mongodb.com/?jmp=docs
[20] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[21] https://docs.cloud.mongodb.com/
[22] https://docs.opsmanager.mongodb.com/current/

Possible causes of replication lag include:

- **Network Latency**

  Check the network routes between the members of your set to ensure that there is no packet loss or network routing issue.

  Use tools including `ping` to test latency between set members and `traceroute` to expose the routing of packets network endpoints.

- **Disk Throughput**

  If the file system and disk device on the secondary is unable to flush data to disk as quickly as the primary, then the secondary will have difficulty keeping state. Disk-related issues are incredibly prevalent on multi-tenant systems, including virtualized instances, and can be transient if the system accesses disk devices over an IP network (as is the case with Amazon's EBS system.)

  Use system-level tools to assess disk status, including `iostat` or `vmstat`.

- **Concurrency**

  In some cases, long-running operations on the primary can block replication on secondaries. For best results, configure *write concern* (page 180) to require confirmation of replication to secondaries. This prevents write operations from returning if replication cannot keep up with the write load.

  Use the *database profiler* to see if there are slow queries or long-running operations that correspond to the incidences of lag.

- **Appropriate Write Concern**

  If you are performing a large data ingestion or bulk load operation that requires a large number of writes to the primary, particularly with `unacknowledged write concern`, the secondaries will not be able to read the oplog fast enough to keep up with changes.

  To prevent this, request *write acknowledgment write concern* (page 179) after every 100, 1,000, or an another interval to provide an opportunity for secondaries to catch up with the primary.

  For more information see:

  - *Write Concern* (page 180)
  - *Replica Set Write Concern* (page 126)
  - *Oplog Size* (page 657)

Test Connections Between all Members

All members of a *replica set* must be able to connect to every other member of the set to support replication. Always verify connections in both "directions." Networking topologies and firewall configurations can prevent normal and required connectivity, which can block replication.

Consider the following example of a bidirectional test of networking:

Example

Given a replica set with three members running on three separate hosts:

- `m1.example.net`
- `m2.example.net`
- `m3.example.net`

1. Test the connection from `m1.example.net` to the other hosts with the following operation set `m1.example.net`:

   ```
   mongo --host m2.example.net --port 27017

   mongo --host m3.example.net --port 27017
   ```

2. Test the connection from `m2.example.net` to the other two hosts with the following operation set from `m2.example.net`, as in:

   ```
   mongo --host m1.example.net --port 27017

   mongo --host m3.example.net --port 27017
   ```

   You have now tested the connection between `m2.example.net` and `m1.example.net` in both directions.

3. Test the connection from `m3.example.net` to the other two hosts with the following operation set from the `m3.example.net` host, as in:

   ```
   mongo --host m1.example.net --port 27017

   mongo --host m2.example.net --port 27017
   ```

If any connection, in any direction fails, check your networking and firewall configuration and reconfigure your environment to allow these connections.

Socket Exceptions when Rebooting More than One Secondary

When you reboot members of a replica set, ensure that the set is able to elect a primary during the maintenance. This means ensuring that a majority of the set's `members[n].votes` (page 721) are available.

When a set's active members can no longer form a majority, the set's *primary* steps down and becomes a *secondary*. The former primary closes all open connections to client applications. Clients attempting to write to the former primary receive socket exceptions and *Connection reset* errors until the set can elect a primary.

Example

Given a three-member replica set where every member has one vote, the set can elect a primary if at least two members can connect to each other. If you reboot the two secondaries at once, the primary steps down and becomes a secondary. Until at least another secondary becomes available, i.e. at least one of the rebooted secondaries also becomes available, the set has no primary and cannot elect a new primary.

For more information on votes, see *Replica Set Elections* (page 644). For related information on connection errors, see *Does TCP keepalive time affect MongoDB Deployments?* (page 857).

Check the Size of the Oplog

A larger *oplog* can give a replica set a greater tolerance for lag, and make the set more resilient.

To check the size of the oplog for a given *replica set* member, connect to the member in a `mongo` shell and run the `rs.printReplicationInfo()` method.

The output displays the size of the oplog and the date ranges of the operations contained in the oplog. In the following example, the oplog is about 10 MB and is able to fit about 26 hours (94400 seconds) of operations:

```
configured oplog size:   10.10546875MB
log length start to end: 94400 (26.22hrs)
oplog first event time:  Mon Mar 19 2012 13:50:38 GMT-0400 (EDT)
oplog last event time:   Wed Oct 03 2012 14:59:10 GMT-0400 (EDT)
now:                     Wed Oct 03 2012 15:00:21 GMT-0400 (EDT)
```

The oplog should be long enough to hold all transactions for the longest downtime you expect on a secondary. At a minimum, an oplog should be able to hold minimum 24 hours of operations; however, many users prefer to have 72 hours or even a week's work of operations.

For more information on how oplog size affects operations, see:

- *Oplog Size* (page 657),
- *Delayed Replica Set Members* (page 634), and
- *Check the Replication Lag* (page 712).

Note: You normally want the oplog to be the same size on all members. If you resize the oplog, resize it on all members.

To change oplog size, see the *Change the Size of the Oplog* (page 693) tutorial.

Oplog Entry Timestamp Error

Consider the following error in `mongod` output and logs:

```
replSet error fatal couldn't query the local local.oplog.rs collection.  Terminating mongod after 30
<timestamp> [rsStart] bad replSet oplog entry?
```

Often, an incorrectly typed value in the `ts` field in the last *oplog* entry causes this error. The correct data type is Timestamp.

Check the type of the `ts` value using the following two queries against the oplog collection:

```
db = db.getSiblingDB("local")
db.oplog.rs.find().sort({$natural:-1}).limit(1)
db.oplog.rs.find({ts:{$type:17}}).sort({$natural:-1}).limit(1)
```

The first query returns the last document in the oplog, while the second returns the last document in the oplog where the `ts` value is a Timestamp. The `$type` operator allows you to select *BSON type* 17, is the Timestamp data type.

If the queries don't return the same document, then the last document in the oplog has the wrong data type in the `ts` field.

Example

If the first query returns this as the last oplog entry:

```
{ "ts" : {t: 1347982456000, i: 1},
  "h" : NumberLong("8191276672478122996"),
  "op" : "n",
  "ns" : "",
  "o" : { "msg" : "Reconfig set", "version" : 4 } }
```

And the second query returns this as the last entry where `ts` has the `Timestamp` type:

```
{ "ts" : Timestamp(1347982454000, 1),
  "h" : NumberLong("6188469075153256465"),
  "op" : "n",
```

```
  "ns" : "",
  "o" : { "msg" : "Reconfig set", "version" : 3 } }
```

Then the value for the `ts` field in the last oplog entry is of the wrong data type.

To set the proper type for this value and resolve this issue, use an update operation that resembles the following:

```
db.oplog.rs.update( { ts: { t:1347982456000, i:1 } },
                    { $set: { ts: new Timestamp(1347982456000, 1)}})
```

Modify the timestamp values as needed based on your oplog entry. This operation may take some period to complete because the update must scan and pull the entire oplog into memory.

Duplicate Key Error on `local.slaves`

Changed in version 3.0.0.

MongoDB 3.0.0 removes the `local.slaves` (page 724) collection. For `local.slaves` error in earlier versions of MongoDB, refer to the appropriate version of the MongoDB Manual.

12.4 Replication Reference

On this page

- Replication Methods in the `mongo` Shell (page 716)
- Replication Database Commands (page 717)
- Replica Set Reference Documentation (page 717)

12.4.1 Replication Methods in the `mongo` Shell

Name	Description
`rs.add()`	Adds a member to a replica set.
`rs.addArb()`	Adds an *arbiter* to a replica set.
`rs.conf()`	Returns the replica set configuration document.
`rs.freeze()`	Prevents the current member from seeking election as primary for a period of time.
`rs.help()`	Returns basic help text for *replica set* functions.
`rs.initiate()`	Initializes a new replica set.
`rs.printReplicationInfo()`	Prints a report of the status of the replica set from the perspective of the primary.
`rs.printSlaveReplicationInfo()`	Prints a report of the status of the replica set from the perspective of the secondaries.
`rs.reconfig()`	Re-configures a replica set by applying a new replica set configuration object.
`rs.remove()`	Remove a member from a replica set.
`rs.slaveOk()`	Sets the `slaveOk` property for the current connection. Deprecated. Use `readPref()` and `Mongo.setReadPref()` to set *read preference*.
`rs.status()`	Returns a document with information about the state of the replica set.
`rs.stepDown()`	Causes the current *primary* to become a secondary which forces an *election*.
`rs.syncFrom()`	Sets the member that this replica set member will sync from, overriding the default sync target selection logic.

12.4.2 Replication Database Commands

Name	Description
`replSetFreeze`	Prevents the current member from seeking election as *primary* for a period of time.
`replSetGetStatus`	Returns a document that reports on the status of the replica set.
`replSetInitiate`	Initializes a new replica set.
`replSetMaintenance`	Enables or disables a maintenance mode, which puts a *secondary* node in a `RECOVERING` state.
`replSetReconfig`	Applies a new configuration to an existing replica set.
`replSetStepDown`	Forces the current *primary* to *step down* and become a *secondary*, forcing an election.
`replSetSyncFrom`	Explicitly override the default logic for selecting a member to replicate from.
`resync`	Forces a `mongod` to re-synchronize from the *master*. For master-slave replication only.
`applyOps`	Internal command that applies *oplog* entries to the current data set.
`isMaster`	Displays information about this member's role in the replica set, including whether it is the master.
`replSetGetConfig`	Returns the replica set's configuration object.

12.4.3 Replica Set Reference Documentation

Replica Set Configuration **(page 717)** Complete documentation of the *replica set* configuration object returned by `rs.conf()`.

The local Database **(page 723)** Complete documentation of the content of the `local` database that `mongod` instances use to support replication.

Replica Set Member States **(page 725)** Reference for the replica set member states.

Read Preference Reference **(page 727)** Complete documentation of the five read preference modes that the MongoDB drivers support.

Replica Set Configuration

On this page

- Example Output (page 717)
- Replica Set Configuration Fields (page 718)

You can access the configuration of a *replica set* using the `rs.conf()` method or the `replSetGetConfig` command.

To modify the configuration for a replica set, use the `rs.reconfig()` method, passing a configuration document to the method. See `rs.reconfig()` for more information.

Example Output

The following document provides a representation of a replica set configuration document. The configuration of your replica set may include only a subset of these settings:

```
{
  _id: <string>,
  version: <int>,
  protocolVersion: <number>,
  members: [
```

```
    {
      _id: <int>,
      host: <string>,
      arbiterOnly: <boolean>,
      buildIndexes: <boolean>,
      hidden: <boolean>,
      priority: <number>,
      tags: <document>,
      slaveDelay: <int>,
      votes: <number>
    },
    ...
  ],
  settings: {
    chainingAllowed : <boolean>,
    heartbeatIntervalMillis : <int>,
    heartbeatTimeoutSecs: <int>,
    electionTimeoutMillis : <int>,
    getLastErrorModes : <document>,
    getLastErrorDefaults : <document>
  }
}
```

Replica Set Configuration Fields

_id

Type: string

The name of the replica set. Once set, you cannot change the name of a replica set.

`_id` (page 718) *must* be identical to the `replication.replSetName` or the value of *–replSet* specified to `mongod` on the command line.

See

`replSetName` or *`--replSet`* for information on setting the replica set name.

---

version

Type: int

An incrementing number used to distinguish revisions of the replica set configuration object from previous iterations of the configuration.

configsvr

New in version 3.2.

Type: boolean

Default: false

Indicates whether the replica set is used for a sharded cluster's config servers. Set to `true` if the replica set is for a sharded cluster's config servers.

See also:

Sharded Cluster Enhancements (page 889)

protocolVersion

New in version 3.2.

Type: number

Default: 1 for new replica sets

Version of the *replica set election protocol* (page 888).

Set to `1` to enable the *replication election enhancements* (page 888) introduced in MongoDB 3.2.

By default, new replica sets in MongoDB 3.2 use `protocolVersion: 1`. Previous versions of MongoDB use version 0 of the protocol and cannot run as members of a replica set configuration that specifies `protocolVersion` 1.

members
members

Type: array

An array of member configuration documents, one for each member of the replica set. The `members` (page 719) array is a zero-indexed array.

Each member-specific configuration document can contain the following fields:

`members[n]._id`
Type: integer

An integer identifier of every member in the replica set. Values must be between 0 and 255 inclusive. Each replica set member must have a unique `_id<members[n]._id>`. Once set, you cannot change the `_id` (page 719) of a member.

Note: When updating the replica configuration object, access the replica set members in the `members` (page 719) array with the **array index**. The array index begins with `0`. Do **not** confuse this index value with the value of the `members[n]._id` (page 719) field in each document in the `members` (page 719) array.

`members[n].host`
Type: string

The hostname and, if specified, the port number, of the set member.

The hostname name must be resolvable for every host in the replica set.

> **Warning:** `members[n].host` (page 719) cannot hold a value that resolves to `localhost` or the local interface unless *all* members of the set are on hosts that resolve to `localhost`.

`members[n].arbiterOnly`
Optional.

Type: boolean

Default: false

A boolean that identifies an arbiter. A value of `true` indicates that the member is an arbiter.

When using the `rs.addArb()` method to add an arbiter, the method automatically sets `members[n].arbiterOnly` (page 719) to `true` for the added member.

`members[n].buildIndexes`
Optional.

Type: boolean

Default: true

A boolean that indicates whether the `mongod` builds *indexes* on this member. You can only set this value when adding a member to a replica set. You cannot change `members[n].buildIndexes` (page 719) field after the member has been added to the set. To add a member, see `rs.add()` and `rs.reconfig()`.

Do not set to `false` for `mongod` instances that receive queries from clients.

Setting `buildIndexes` to `false` may be useful if **all** the following conditions are true:

- you are only using this instance to perform backups using `mongodump`, *and*
- this member will receive no queries, *and*
- index creation and maintenance overburdens the host system.

Even if set to `false`, secondaries *will* build indexes on the `_id` field in order to facilitate operations required for replication.

> **Warning:** If you set `members[n].buildIndexes` (page 719) to `false`, you must also set `members[n].priority` (page 720) to `0`. If `members[n].priority` (page 720) is not `0`, MongoDB will return an error when attempting to add a member with `members[n].buildIndexes` (page 719) equal to `false`.
> To ensure the member receives no queries, you should make all instances that do not build indexes hidden.
> Other secondaries cannot replicate from a member where `members[n].buildIndexes` (page 719) is false.

`members[n].hidden`

Optional.

Type: boolean

Default: false

When this value is `true`, the replica set hides this instance and does not include the member in the output of `db.isMaster()` or `isMaster`. This prevents read operations (i.e. queries) from ever reaching this host by way of secondary *read preference*.

See also:

Hidden Replica Set Members (page 633)

`members[n].priority`

Optional.

Type: Number, between 0 and 1000.

Default: 1.0

A number that indicates the relative eligibility of a member to become a *primary*.

Specify higher values to make a member *more* eligible to become *primary*, and lower values to make the member *less* eligible. A member with a `members[n].priority` (page 720) of `0` is ineligible to become primary.

Changing the balance of priority in a replica set will trigger one or more elections. If a lower priority secondary is elected over a higher priority secondary, replica set members will continue to call elections until the highest priority available member becomes primary.

See also:

Replica Set Elections (page 644).

members[n].**tags**
Optional.

Type: document

Default: none

A *tag set* document containing mappings of arbitrary keys and values. These documents describe replica set members in order to customize *write concern* (page 179) and *read preference* (page 727) and thereby allow configurable data center awareness.

This field is only present if there are tags assigned to the member. See *Configure Replica Set Tag Sets* (page 700) for more information.

Use `replicaset.members[n].tags` to configure write concerns in conjunction with `settings.getLastErrorModes` (page 722) and `settings.getLastErrorDefaults` (page 722).

Important: In tag sets, all tag values must be strings.

For more information on configuring tag sets for read preference and write concern, see *Configure Replica Set Tag Sets* (page 700).

members[n].**slaveDelay**
Optional.

Type: integer

Default: 0

The number of seconds "behind" the primary that this replica set member should "lag".

Use this option to create *delayed members* (page 634). Delayed members maintain a copy of the data that reflects the state of the data at some time in the past.

See also:

Delayed Replica Set Members (page 634)

members[n].**votes**
Optional.

Type: integer

Default: 1

The number of votes a server will cast in a *replica set election* (page 644). The number of votes each member has is either `1` or `0`, and *arbiters* (page **??**) always have exactly `1` vote.

A replica set can have up to `50 members` but only 7 voting members. If you need more than 7 members in one replica set, set `members[n].votes` (page 721) to `0` for the additional non-voting members.

Changed in version 3.0.0: Members cannot have `members[n].votes` (page 721) greater than `1`. For details, see *Replica Set Configuration Validation* (page 947).

settings
settings
Optional.

Type: document

A document that contains configuration options that apply to the whole replica set.

The `settings` (page 721) document contain the following fields:

settings.**chainingAllowed**
Optional.

Type: boolean

Default: true

When settings.chainingAllowed (page 721) is true, the replica set allows *secondary* members to replicate from other secondary members. When settings.chainingAllowed (page 721) is false, secondaries can replicate only from the *primary*.

See also:

Manage Chained Replication (page 705)

settings.**getLastErrorDefaults**
Optional.

Type: document

A document that specifies the *write concern* (page 649) for the replica set. The replica set will use this write concern only when *write operations* (page 1002) or getLastError specify no other write concern.

If settings.getLastErrorDefaults (page 722) is not set, the default write concern for the replica set only requires confirmation from the primary.

settings.**getLastErrorModes**
Optional.

Type: document

A document used to define an extended *write concern* through the use of members[n].tags (page 720). The extended *write concern* can provide *data-center awareness*.

For example, the following document defines an extended write concern named eastCoast and associates with a write to a member that has the east tag.

```
{ getLastErrorModes: { eastCoast: { "east": 1 } } }
```

Write operations to the replica set can use the extended write concern, e.g. { w: "eastCoast" }.

See *Configure Replica Set Tag Sets* (page 700) for more information and example.

settings.**heartbeatTimeoutSecs**
Optional.

Type: int

Default: 10

Number of seconds that the replica set members wait for a successful heartbeat from each other. If a member does not respond in time, other members mark the delinquent member as inaccessible.

settings.**electionTimeoutMillis**
New in version 3.2.

Optional.

Type: int

Default: 10000 (10 seconds)

The time limit in milliseconds for detecting when a replica set's primary is unreachable:

- Higher values result in slower failovers but decreased sensitivity to primary node or network slowness or spottiness.

- Lower values result in faster failover, but increased sensitivity to primary node or network slowness or spottiness.

The setting only applies when using `protocolVersion: 1`.

`settings.`**`heartbeatIntervalMillis`**
New in version 3.2.

Internal use only.

The frequency in milliseconds of the heartbeats.

The `local` Database

On this page

- Overview (page 723)
- Collection on all `mongod` Instances (page 723)
- Collections on Replica Set Members (page 724)
- Collections used in Master/Slave Replication (page 725)

Overview

Every `mongod` instance has its own `local` database, which stores data used in the replication process, and other instance-specific data. The `local` database is invisible to replication: collections in the `local` database are not replicated.

In replication, the `local` database store stores internal replication data for each member of a *replica set*. The `local` stores the following collections:

Changed in version 2.4: When running with authentication (i.e. `authorization`), authenticating to the `local` database is **not** equivalent to authenticating to the `admin` database. In previous versions, authenticating to the `local` database provided access to all databases.

Collection on all `mongod` Instances

`local.`**`startup_log`**
On startup, each `mongod` instance inserts a document into `startup_log` (page 723) with diagnostic information about the `mongod` instance itself and host information. `startup_log` (page 723) is a capped collection. This information is primarily useful for diagnostic purposes.

Example

Consider the following prototype of a document from the `startup_log` (page 723) collection:

```
{
  "_id" : "<string>",
  "hostname" : "<string>",
  "startTime" : ISODate("<date>"),
  "startTimeLocal" : "<string>",
  "cmdLine" : {
        "dbpath" : "<path>",
        "<option>" : <value>
  },
```

```
    "pid" : <number>,
    "buildinfo" : {
          "version" : "<string>",
          "gitVersion" : "<string>",
          "sysInfo" : "<string>",
          "loaderFlags" : "<string>",
          "compilerFlags" : "<string>",
          "allocator" : "<string>",
          "versionArray" : [ <num>, <num>, <...> ],
          "javascriptEngine" : "<string>",
          "bits" : <number>,
          "debug" : <boolean>,
          "maxBsonObjectSize" : <number>
    }
}
```

Documents in the `startup_log` (page 723) collection contain the following fields:

`local.startup_log._id`
: Includes the system hostname and a millisecond epoch value.

`local.startup_log.hostname`
: The system's hostname.

`local.startup_log.startTime`
: A UTC *ISODate* value that reflects when the server started.

`local.startup_log.startTimeLocal`
: A string that reports the `startTime` (page 724) in the system's local time zone.

`local.startup_log.cmdLine`
: An embedded document that reports the `mongod` runtime options and their values.

`local.startup_log.pid`
: The process identifier for this process.

`local.startup_log.buildinfo`
: An embedded document that reports information about the build environment and settings used to compile this `mongod`. This is the same output as `buildInfo`. See `buildInfo`.

Collections on Replica Set Members

`local.system.replset`
: `local.system.replset` (page 724) holds the replica set's configuration object as its single document. To view the object's configuration information, issue `rs.conf()` from the `mongo` shell. You can also query this collection directly.

`local.oplog.rs`
: `local.oplog.rs` (page 724) is the capped collection that holds the *oplog*. You set its size at creation using the `oplogSizeMB` setting. To resize the oplog after replica set initiation, use the *Change the Size of the Oplog* (page 693) procedure. For additional information, see the *Oplog Size* (page 657) section.

`local.replset.minvalid`
: This contains an object used internally by replica sets to track replication status.

`local.slaves`
: *Removed in version 3.0:* Replica set members no longer mirror replication status of the set to the `local.slaves` (page 724) collection. Use `rs.status()` instead.

Collections used in Master/Slave Replication

In *master/slave* replication, the `local` database contains the following collections:

- On the master:

  `local.oplog.$main`
  This is the oplog for the master-slave configuration.

  `local.slaves`
  Removed in version 3.0: MongoDB no longer stores information about each slave in the `local.slaves` (page 724) collection. Use `db.serverStatus( { repl: 1 } )` instead.

- On each slave:

  `local.sources`
  This contains information about the slave's master server.

Replica Set Member States

On this page

- States (page 725)

Each member of a replica set has a state that reflects its disposition within the set.

Number	Name	State Description
0	`STARTUP` (page 726)	Not yet an active member of any set. All members start up in this state. The `mongod` parses the *replica set configuration document* (page 684) while in `STARTUP` (page 726).
1	`PRIMARY` (page 725)	The member in state *primary* (page 628) is the only member that can accept write operations.
2	`SECONDARY` (page 726)	A member in state *secondary* (page 628) is replicating the data store. Data is available for reads, although they may be stale.
3	`RECOVERING` (page 726)	Can vote. Members either perform startup self-checks, or transition from completing a *rollback* (page 647) or *resync* (page 699).
5	`STARTUP2` (page 726)	The member has joined the set and is running an initial sync.
6	`UNKNOWN` (page 726)	The member's state, as seen from another member of the set, is not yet known.
7	`ARBITER` (page 726)	*Arbiters* (page **??**) do not replicate data and exist solely to participate in elections.
8	`DOWN` (page 726)	The member, as seen from another member of the set, is unreachable.
9	`ROLLBACK` (page 727)	This member is actively performing a *rollback* (page 647). Data is not available for reads.
10	`REMOVED` (page 726)	This member was once in a replica set but was subsequently removed.

States

Core States

PRIMARY
Members in PRIMARY (page 725) state accept write operations. A replica set has at most one primary at a time. A SECONDARY (page 726) member becomes primary after an *election* (page 644). Members in the PRIMARY (page 725) state are eligible to vote.

SECONDARY
Members in SECONDARY (page 726) state replicate the primary's data set and can be configured to accept read operations. Secondaries are eligible to vote in elections, and may be elected to the PRIMARY (page 725) state if the primary becomes unavailable.

ARBITER
Members in ARBITER (page 726) state do not replicate data or accept write operations. They are eligible to vote, and exist solely to break a tie during elections. Replica sets should only have a member in the ARBITER (page 726) state if the set would otherwise have an even number of members, and could suffer from tied elections. There should only be at most one arbiter configured in any replica set.

See *Replica Set Members* (page 628) for more information on core states.

Other States

STARTUP
Each member of a replica set starts up in STARTUP (page 726) state. mongod then loads that member's replica set configuration, and transitions the member's state to STARTUP2 (page 726). Members in STARTUP (page 726) are not eligible to vote, as they are not yet a recognized member of any replica set.

STARTUP2
Each member of a replica set enters the STARTUP2 (page 726) state as soon as mongod finishes loading that member's configuration, at which time it becomes an active member of the replica set. The member then decides whether or not to undertake an initial sync. If a member begins an initial sync, the member remains in STARTUP2 (page 726) until all data is copied and all indexes are built. Afterwards, the member transitions to RECOVERING (page 726).

RECOVERING
A member of a replica set enters RECOVERING (page 726) state when it is not ready to accept reads. The RECOVERING (page 726) state can occur during normal operation, and doesn't necessarily reflect an error condition. Members in the RECOVERING (page 726) state are eligible to vote in elections, but are not eligible to enter the PRIMARY (page 725) state.

A member transitions from RECOVERING (page 726) to SECONDARY (page 726) after replicating enough data to guarantee a consistent view of the data for client reads. The only difference between RECOVERING (page 726) and SECONDARY (page 726) states is that RECOVERING (page 726) prohibits client reads and SECONDARY (page 726) permits them. SECONDARY (page 726) state does not guarantee anything about the staleness of the data with respect to the primary.

Due to overload, a *secondary* may fall far enough behind the other members of the replica set such that it may need to *resync* (page 699) with the rest of the set. When this happens, the member enters the RECOVERING (page 726) state and requires manual intervention.

Error States

Members in any error state can't vote.

UNKNOWN
Members that have never communicated status information to the replica set are in the UNKNOWN (page 726) state.

DOWN
Members that lose their connection to the replica set are seen as DOWN (page 726) by the remaining members of the set.

REMOVED

Members that are removed from the replica set enter the `REMOVED` (page 726) state. When members enter the `REMOVED` (page 726) state, the logs will mark this event with a `replSet REMOVED` message entry.

ROLLBACK

Whenever the replica set replaces a *primary* in an election, the old primary may contain documents that did not replicate to the *secondary* members. In this case, the old primary member reverts those writes. During *rollback* (page 647), the member will have `ROLLBACK` (page 727) state.

FATAL

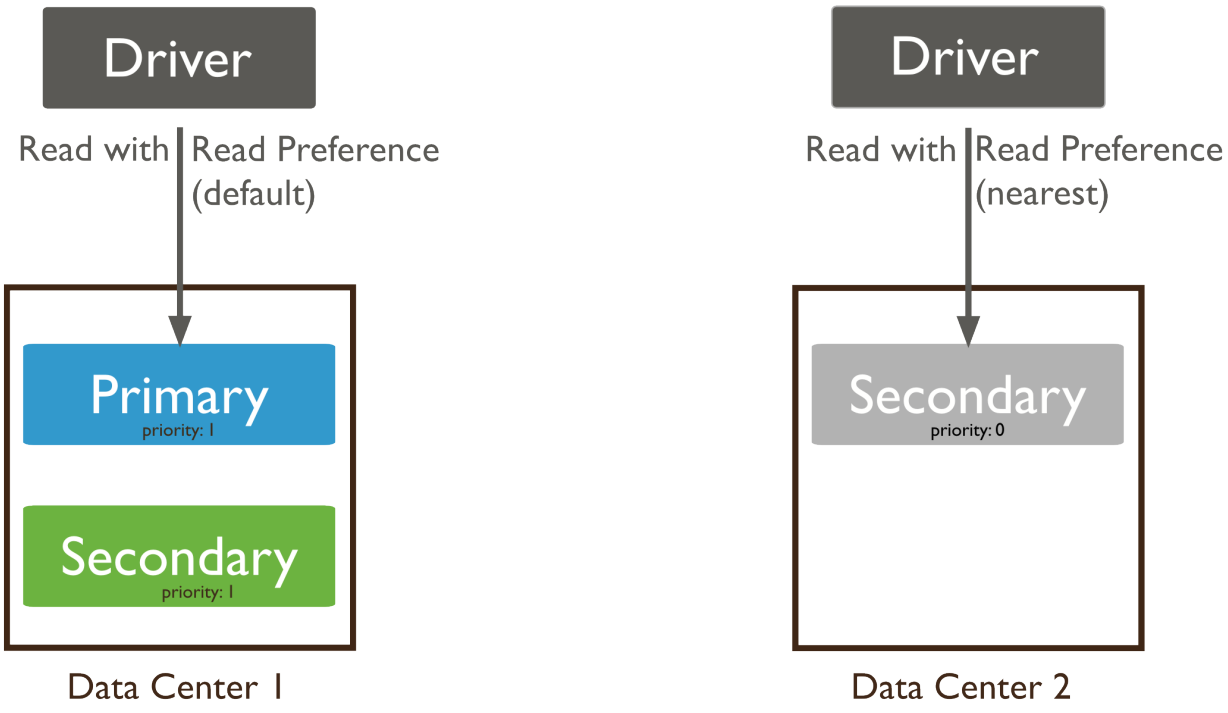
Removed in version 3.0: A member in `FATAL` (page 727) encountered an unrecoverable error. The member must be shut down and restarted; a resync may be required as well.

Read Preference Reference

On this page

- Read Preference Modes (page 728)
- Use Cases (page 729)
- Read Preferences for Database Commands (page 730)

Read preference describes how MongoDB clients route read operations to the members of a *replica set*.



By default, an application directs its read operations to the *primary* member in a *replica set*.

In MongoDB, in a replica set with one primary member [23],

[23] In *some circumstances* (page 729), two nodes in a replica set may *transiently* believe that they are the primary, but at most, one of them will be able to complete writes with `{ w: "majority" }` (page 180) write concern. The node that can complete `{ w: "majority" }` (page 180) writes is the current primary, and the other node is a former primary that has not yet recognized its demotion, typically due to a *network*

- With "local" (page 182) readConcern, reads from the primary reflect the latest writes in absence of a failover;
- With "majority" (page 182) readConcern, read operations from the primary or the secondaries have *eventual consistency*.

Read Preference Mode	Description
primary (page 728)	Default mode. All operations read from the current replica set *primary*.
primaryPreferred (page 728)	In most situations, operations read from the *primary* but if it is unavailable, operations read from *secondary* members.
secondary (page 728)	All operations read from the *secondary* members of the replica set.
secondaryPreferred (page 729)	In most situations, operations read from *secondary* members but if no *secondary* members are available, operations read from the *primary*.
nearest (page 729)	Operations read from member of the *replica set* with the least network latency, irrespective of the member's type.

Note: The read preference does not affect the visibility of data; i.e, clients can see the results of writes before they are made *durable*:

- Regardless of *write concern* (page 179), other clients using "local" (page 182) (i.e. the default) readConcern can see the result of a write operation before the write operation is acknowledged to the issuing client.
- Clients using "local" (page 182) (i.e. the default) readConcern can read data which may be subsequently *rolled back* (page 647).

For more information on read isolation level in MongoDB, see *Read Isolation, Consistency, and Recency* (page 133).

Read Preference Modes

primary

All read operations use only the current replica set *primary*. [6] This is the default read mode. If the primary is unavailable, read operations produce an error or throw an exception.

The primary (page 728) read preference mode is not compatible with read preference modes that use *tag sets* (page 653). If you specify a tag set with primary (page 728), the driver will produce an error.

primaryPreferred

In most situations, operations read from the *primary* member of the set. However, if the primary is unavailable, as is the case during *failover* situations, operations read from secondary members.

When the read preference includes a *tag set* (page 653), the client reads first from the primary, if available, and then from *secondaries* that match the specified tags. If no secondaries have matching tags, the read operation produces an error.

Since the application may receive data from a secondary, read operations using the primaryPreferred (page 728) mode may return stale data in some situations.

secondary

Operations read *only* from the *secondary* members of the set. If no secondaries are available, then this read operation produces an error or exception.

Most sets have at least one secondary, but there are situations where there may be no available secondary. For example, a set with a primary, a secondary, and an *arbiter* may not have any secondaries if a member is in recovering state or unavailable.

partition. When this occurs, clients that connect to the former primary may observe stale data despite having requested read preference primary (page 728), and new writes to the former primary will eventually roll back.

When the read preference includes a *tag set* (page 653), the client attempts to find secondary members that match the specified tag set and directs reads to a random secondary from among the *nearest group* (page 654). If no secondaries have matching tags, the read operation produces an error. [24]

Read operations using the `secondary` (page 728) mode may return stale data.

`secondaryPreferred`

In most situations, operations read from *secondary* members, but in situations where the set consists of a single *primary* (and no other members), the read operation will use the set's primary.

When the read preference includes a *tag set* (page 653), the client attempts to find a secondary member that matches the specified tag set and directs reads to a random secondary from among the *nearest group* (page 654). If no secondaries have matching tags, the client ignores tags and reads from the primary.

Read operations using the `secondaryPreferred` (page 729) mode may return stale data.

`nearest`

The driver reads from the *nearest* member of the *set* according to the *member selection* (page 654) process. Reads in the `nearest` (page 729) mode do not consider the member's *type*. Reads in `nearest` (page 729) mode may read from both primaries and secondaries.

Set this mode to minimize the effect of network latency on read operations without preference for current or stale data.

If you specify a *tag set* (page 653), the client attempts to find a replica set member that matches the specified tag set and directs reads to an arbitrary member from among the *nearest group* (page 654).

Read operations using the `nearest` (page 729) mode may return stale data.

Note: All operations read from a member of the nearest group of the replica set that matches the specified read preference mode. The `nearest` (page 729) mode prefers low latency reads over a member's *primary* or *secondary* status.

For `nearest` (page 729), the client assembles a list of acceptable hosts based on tag set and then narrows that list to the host with the shortest ping time and all other members of the set that are within the "local threshold," or acceptable latency. See *Member Selection* (page 654) for more information.

Use Cases

Depending on the requirements of an application, you can configure different applications to use different read preferences, or use different read preferences for different queries in the same application. Consider the following applications for different read preference strategies.

Maximize Consistency To avoid *stale* reads, use `primary` (page 728) read preference and `"majority"` (page 182) `readConcern`. If the primary is unavailable, e.g. during elections or when a majority of the replica set is not accessible, read operations using `primary` (page 728) read preference produce an error or throw an exception. In some circumstances, it may be possible for a replica set to temporarily have two primaries; however, only one primary will be capable of confirming writes with the `"majority"` (page 180) write concern.

- A partial *network partition* may segregate a primary (P_{old}) into a partition with a minority of the nodes, while the other side of the partition contains a majority of nodes. The partition with the majority will elect a new primary (P_{new}), but for a brief period, the old primary (P_{old}) may still continue to serve reads and writes, as it has not yet

[24] If your set has more than one secondary, and you use the `secondary` (page 728) read preference mode, consider the following effect. If you have a *three member replica set* (page 639) with a primary and two secondaries, and one secondary becomes unavailable, all `secondary` (page 728) queries must target the remaining secondary. This will double the load on this secondary. Plan and provide capacity to support this as needed.

detected that it can only see a minority of nodes in the replica set. During this period, if the old primary (p_{old}) is still visible to clients as a primary, reads from this primary may reflect stale data.

- A primary (p_{old}) may become unresponsive, which will trigger an election and a new primary (P_{new}) can be elected, serving reads and writes. If the unresponsive primary (p_{old}) starts responding again, two primaries will be visible for a brief period. The brief period will end when p_{old} steps down. However, during the brief period, clients might read from the old primary p_{old}, which can provide stale data.

To increase consistency, you can disable automatic *failover*; however, disabling automatic failover sacrifices availability.

Maximize Availability To permit read operations when possible, use `primaryPreferred` (page 728). When there's a primary you will get consistent reads [6], but if there is no primary you can still query *secondaries*. However, when using this read mode, consider the situation described in *secondary vs secondaryPreferred* (page 730).

Minimize Latency To always read from a low-latency node, use `nearest` (page 729). The driver or `mongos` will read from the nearest member and those no more than 15 milliseconds [25] further away than the nearest member.

`nearest` (page 729) does *not* guarantee consistency. If the nearest member to your application server is a secondary with some replication lag, queries could return stale data. `nearest` (page 729) only reflects network distance and does not reflect I/O or CPU load.

Query From Geographically Distributed Members If the members of a replica set are geographically distributed, you can create replica tags based that reflect the location of the instance and then configure your application to query the members nearby.

For example, if members in "east" and "west" data centers are *tagged* (page 700) `{'dc': 'east'}` and `{'dc': 'west'}`, your application servers in the east data center can read from nearby members with the following read preference:

```
db.collection.find().readPref( { mode: 'nearest',
                                 tags: [ {'dc': 'east'} ] } )
```

Although `nearest` (page 729) already favors members with low network latency, including the tag makes the choice more predictable.

`secondary` vs `secondaryPreferred` For specific dedicated queries (e.g. ETL, reporting), you may shift the read load from the primary by using the `secondary` (page 728) read preference mode. For this use case, the `secondary` (page 728) mode is preferable to the `secondaryPreferred` (page 729) mode because `secondaryPreferred` (page 729) risks the following situation: if all secondaries are unavailable and your replica set has enough *arbiters* [26] to prevent the primary from stepping down, then the primary will receive all traffic from the clients. If the primary is unable to handle this load, the queries will compete with the writes. For this reason, use read preference `secondary` (page 728) to distribute these specific dedicated queries instead of `secondaryPreferred` (page 729).

Read Preferences for Database Commands

Because some *database commands* read and return data from the database, all of the official drivers support full *read preference mode semantics* (page 728) for the following commands:

- `group`

[25] This threshold is configurable. See `localPingThresholdMs` for `mongos` or your driver documentation for the appropriate setting.
[26] In general, avoid deploying more than one arbiter per replica set.

- mapReduce [27]
- aggregate [28]
- collStats
- dbStats
- count
- distinct
- geoNear
- geoSearch
- parallelCollectionScan

New in version 2.4: mongos adds support for routing commands to shards using read preferences. Previously mongos sent all commands to shards' primaries.

[27] Only "inline" mapReduce operations that do not write data support read preference, otherwise these operations must run on the *primary* members.

[28] Using the $out pipeline operator forces the aggregation pipeline to run on the primary.

CHAPTER 13

Sharding

Sharding is the process of storing data records across multiple machines and is MongoDB's approach to meeting the demands of data growth. As the size of the data increases, a single machine may not be sufficient to store the data nor provide an acceptable read and write throughput. Sharding solves the problem with horizontal scaling. With sharding, you add more machines to support data growth and the demands of read and write operations.

13.1 Sharding Introduction

On this page

- Purpose of Sharding (page 733)
- Sharding in MongoDB (page 735)
- Data Partitioning (page 736)
- Maintaining a Balanced Data Distribution (page 737)
- Additional Resources (page 739)

Sharding is a method for storing data across multiple machines. MongoDB uses sharding to support deployments with very large data sets and high throughput operations.

13.1.1 Purpose of Sharding

Database systems with large data sets and high throughput applications can challenge the capacity of a single server. High query rates can exhaust the CPU capacity of the server. Larger data sets exceed the storage capacity of a single machine. Finally, working set sizes larger than the system's RAM stress the I/O capacity of disk drives.

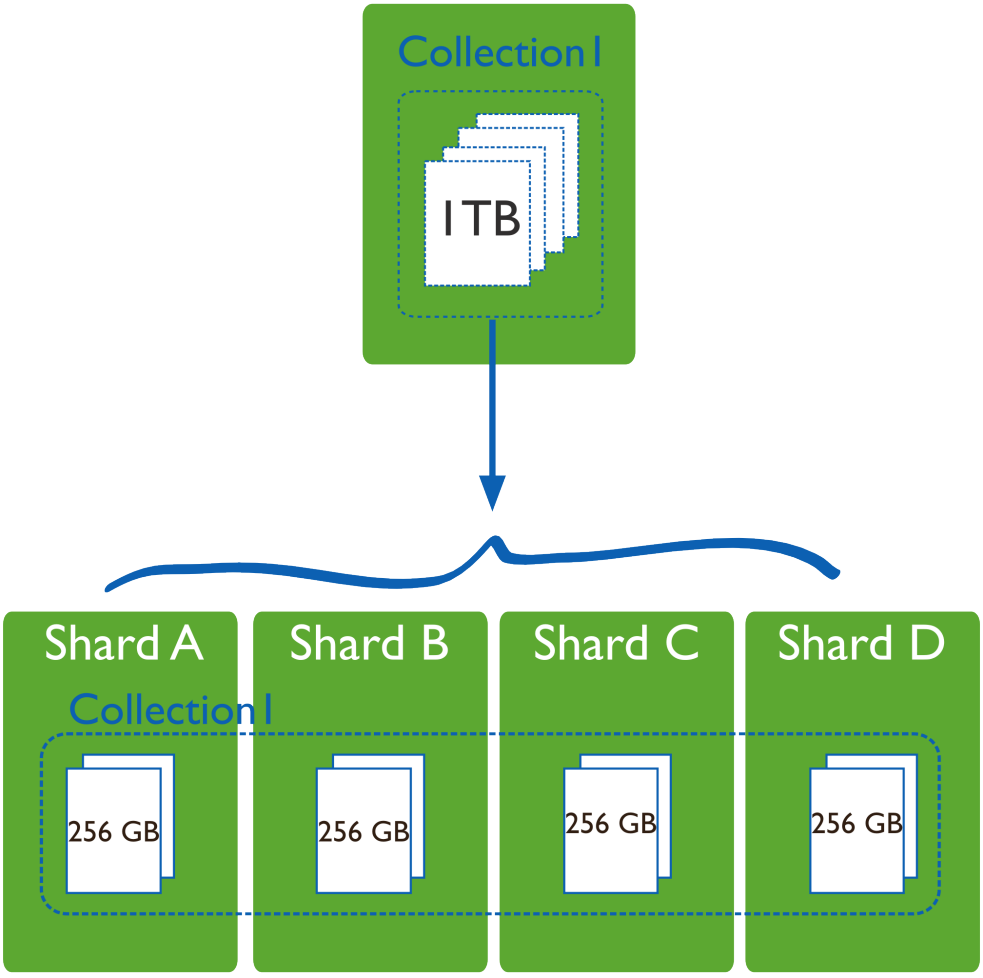
To address these issues of scales, database systems have two basic approaches: **vertical scaling** and **sharding**.

Vertical scaling adds more CPU and storage resources to increase capacity. Scaling by adding capacity has limitations: high performance systems with large numbers of CPUs and large amount of RAM are disproportionately *more expensive* than smaller systems. Additionally, cloud-based providers may only allow users to provision smaller instances. As a result there is a *practical maximum* capability for vertical scaling.

Sharding, or *horizontal scaling*, by contrast, divides the data set and distributes the data over multiple servers, or **shards**. Each shard is an independent database, and collectively, the shards make up a single logical database.

Sharding addresses the challenge of scaling to support high throughput and large data sets:

- Sharding reduces the number of operations each shard handles. Each shard processes fewer operations as the cluster grows. As a result, a cluster can increase capacity and throughput *horizontally*.


Collection I
1 TB
Shard A
Collection I
256 GB
Shard B
256 GB
Shard C
256 GB
Shard D
256 GB

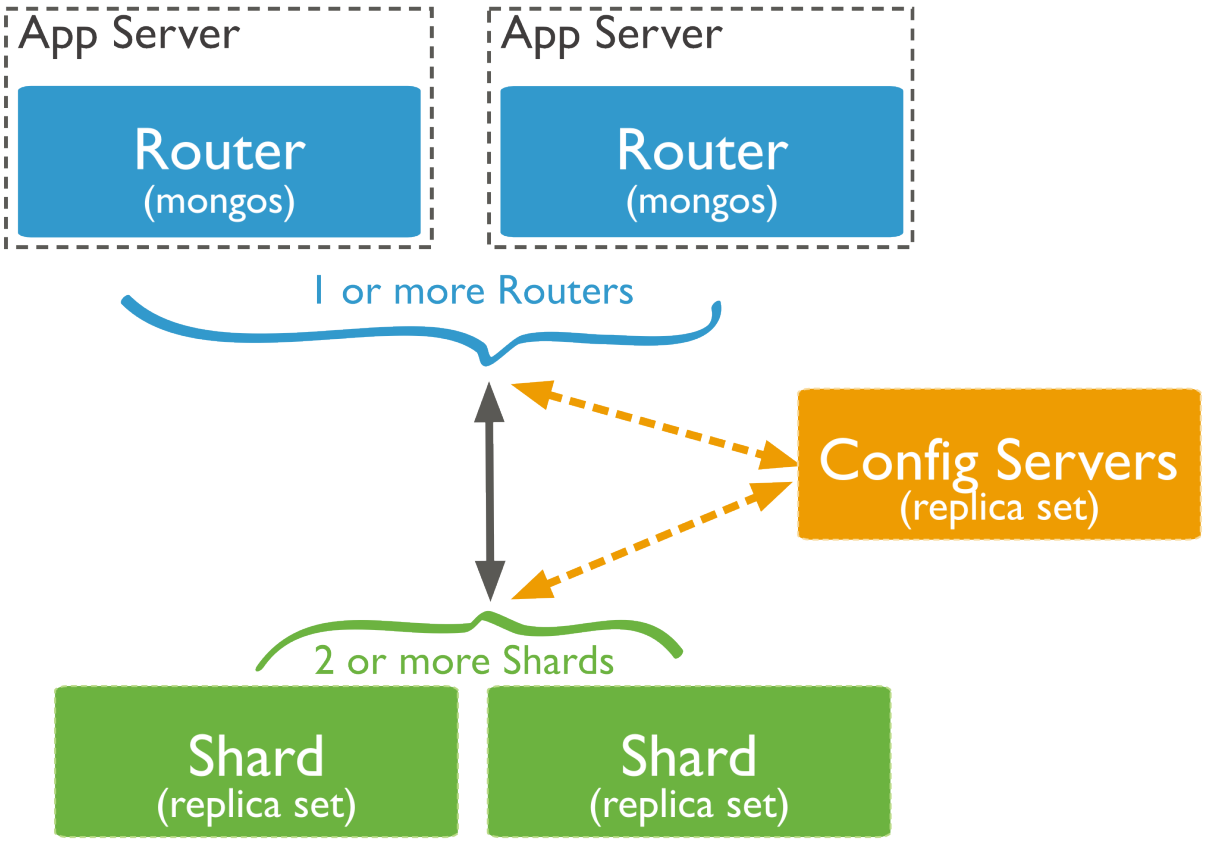
For example, to insert data, the application only needs to access the shard responsible for that record.

- Sharding reduces the amount of data that each server needs to store. Each shard stores less data as the cluster grows.

  For example, if a database has a 1 terabyte data set, and there are 4 shards, then each shard might hold only 256 GB of data. If there are 40 shards, then each shard might hold only 25 GB of data.

13.1.2 Sharding in MongoDB

MongoDB supports sharding through the configuration of a *sharded clusters*.



Sharded cluster has the following components: *shards*, *query routers* and *config servers*.

Shards store the data. To provide high availability and data consistency, in a production sharded cluster, each shard is a *replica set* [1]. For more information on replica sets, see *Replica Sets* (page 627).

Query Routers, or `mongos` instances, interface with client applications and direct operations to the appropriate shard or shards. A client sends requests to a `mongos`, which then routes the operations to the shards and returns the results to the clients. A sharded cluster can contain more than one `mongos` to divide the client request load, and most sharded clusters have more than one `mongos` for this reason.

Config servers store the cluster's metadata. This data contains a mapping of the cluster's data set to the shards. The query router uses this metadata to target operations to specific shards.

[1] For development and testing purposes only, each **shard** can be a single `mongod` instead of a replica set.

Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

13.1.3 Data Partitioning

MongoDB distributes data, or shards, at the collection level. Sharding partitions a collection's data by the **shard key**.

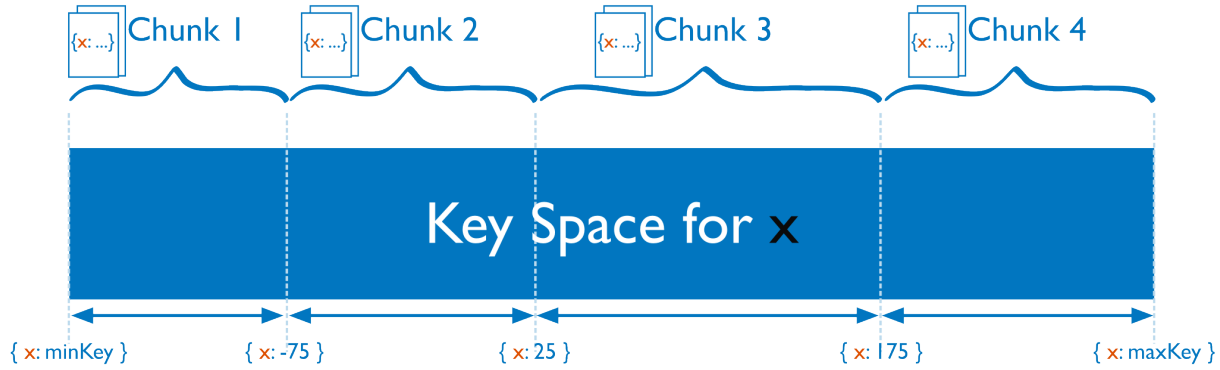
Shard Keys

To shard a collection, you need to select a **shard key**. A *shard key* is either an indexed field or an indexed compound field that exists in every document in the collection. MongoDB divides the shard key values into **chunks** and distributes the *chunks* evenly across the shards. To divide the shard key values into chunks, MongoDB uses either **range based partitioning** or **hash based partitioning**. See the *Shard Key* (page 747) documentation for more information.

Range Based Sharding

For *range-based sharding*, MongoDB divides the data set into ranges determined by the shard key values to provide **range based partitioning**. Consider a numeric shard key: If you visualize a number line that goes from negative infinity to positive infinity, each value of the shard key falls at some point on that line. MongoDB partitions this line into smaller, non-overlapping ranges called **chunks** where a chunk is range of values from some minimum value to some maximum value.

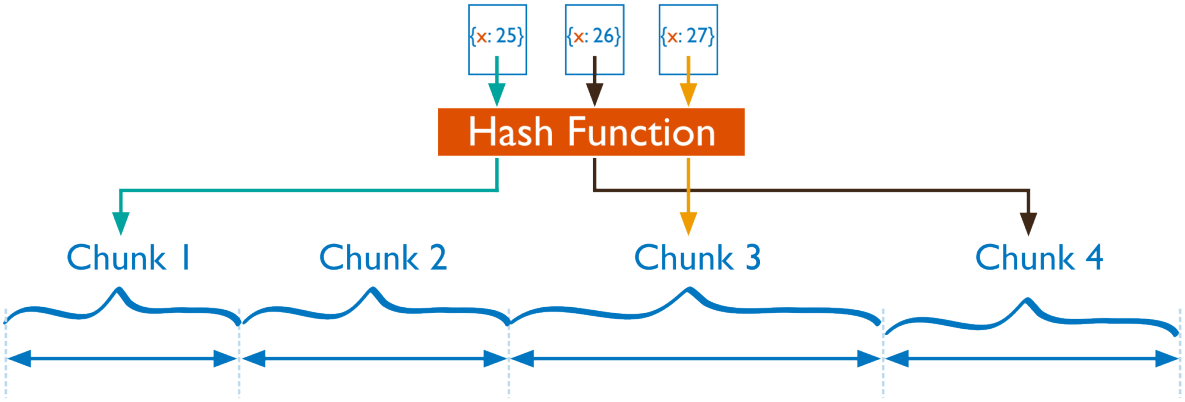
Given a range based partitioning system, documents with "close" shard key values are likely to be in the same chunk, and therefore on the same shard.



Hash Based Sharding

For *hash based partitioning*, MongoDB computes a hash of a field's value, and then uses these hashes to create chunks.

With hash based partitioning, two documents with "close" shard key values are *unlikely* to be part of the same chunk. This ensures a more random distribution of a collection in the cluster.



Performance Distinctions between Range and Hash Based Partitioning

Range based partitioning supports more efficient range queries. Given a range query on the shard key, the query router can easily determine which chunks overlap that range and route the query to only those shards that contain these chunks.

However, range based partitioning can result in an uneven distribution of data, which may negate some of the benefits of sharding. For example, if the shard key is a linearly increasing field, such as time, then all requests for a given time range will map to the same chunk, and thus the same shard. In this situation, a small set of shards may receive the majority of requests and the system would not scale very well.

Hash based partitioning, by contrast, ensures an even distribution of data at the expense of efficient range queries. Hashed key values results in random distribution of data across chunks and therefore shards. But random distribution makes it more likely that a range query on the shard key will not be able to target a few shards but would more likely query every shard in order to return a result.

Customized Data Distribution with Tag Aware Sharding

MongoDB allows administrators to direct the balancing policy using **tag aware sharding**. Administrators create and associate tags with ranges of the shard key, and then assign those tags to the shards. Then, the balancer migrates tagged data to the appropriate shards and ensures that the cluster always enforces the distribution of data that the tags describe.

Tags are the primary mechanism to control the behavior of the balancer and the distribution of chunks in a cluster. Most commonly, tag aware sharding serves to improve the locality of data for sharded clusters that span multiple data centers.

See *Tag Aware Sharding* (page 756) for more information.

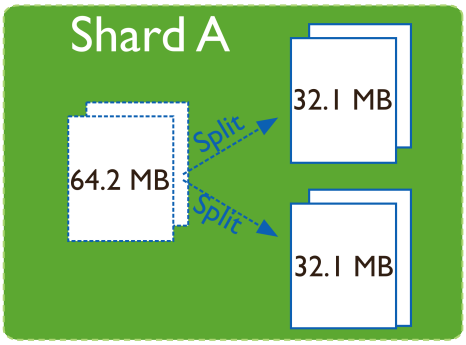
13.1.4 Maintaining a Balanced Data Distribution

The addition of new data or the addition of new servers can result in data distribution imbalances within the cluster, such as a particular shard contains significantly more chunks than another shard or a size of a chunk is significantly greater than other chunk sizes.

MongoDB ensures a balanced cluster using two background process: splitting and the balancer.

Splitting

Splitting is a background process that keeps chunks from growing too large. When a chunk grows beyond a *specified chunk size* (page 762), MongoDB splits the chunk in half. Inserts and updates triggers splits. Splits are an efficient meta-data change. To create splits, MongoDB does *not* migrate any data or affect the shards.



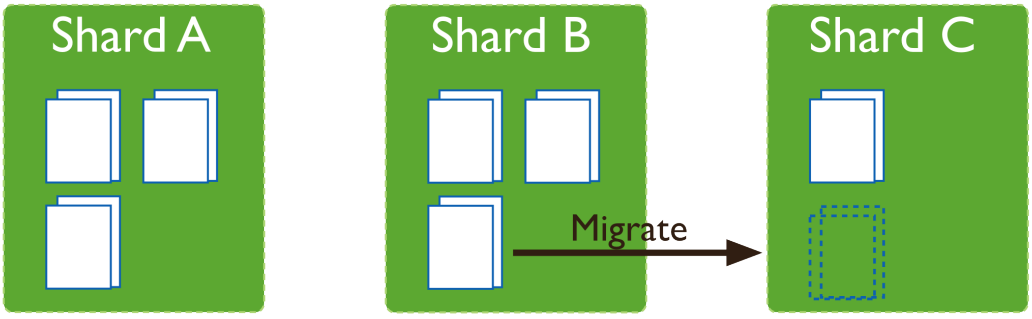
Balancing

The *balancer* (page 758) is a background process that manages chunk migrations. The balancer can run from any of the `mongos` instances in a cluster.

When the distribution of a sharded collection in a cluster is uneven, the balancer process migrates chunks from the shard that has the largest number of chunks to the shard with the least number of chunks until the collection balances. For example: if collection `users` has 100 chunks on *shard 1* and 50 chunks on *shard 2*, the balancer will migrate chunks from *shard 1* to *shard 2* until the collection achieves balance.

The shards manage *chunk migrations* as a background operation between an *origin shard* and a *destination shard*. During a chunk migration, the *destination shard* is sent all the current documents in the chunk from the *origin shard*. Next, the destination shard captures and applies all changes made to the data during the migration process. Finally, the metadata regarding the location of the chunk on *config server* is updated.

If there's an error during the migration, the balancer aborts the process leaving the chunk unchanged on the origin shard. MongoDB removes the chunk's data from the origin shard **after** the migration completes successfully.



Adding and Removing Shards from the Cluster

Adding a shard to a cluster creates an imbalance since the new shard has no chunks. While MongoDB begins migrating data to the new shard immediately, it can take some time before the cluster balances.

When removing a shard, the balancer migrates all chunks from a shard to other shards. After migrating all data and updating the meta data, you can safely remove the shard.

13.1.5 Additional Resources

- Sharding Methods for MongoDB (Presentation)[2]
- Everything You Need to Know About Sharding (Presentation)[3]
- MongoDB for Time Series Data: Sharding[4]
- MongoDB Operations Best Practices White Paper[5]
- Talk to a MongoDB Expert About Scaling[6]
- MongoDB Consulting Package[7]
- Quick Reference Cards[8]

13.2 Sharding Concepts

These documents present the details of sharding in MongoDB. These include the components, the architectures, and the behaviors of MongoDB sharded clusters. For an overview of sharding and sharded clusters, see *Sharding Introduction* (page 733).

Sharded Cluster Components **(page 740)** A sharded cluster consists of shards, config servers, and `mongos` instances.

> ***Shards*** **(page 740)** A shard is a single server or replica set that holds a part of the sharded collection.
>
> ***Config Servers*** **(page 742)** Config servers hold the metadata about the cluster, such as the shard location of the data.

Sharded Cluster Architectures **(page 744)** Outlines the requirements for sharded clusters, and provides examples of several possible architectures for sharded clusters.

> ***Sharded Cluster Requirements*** **(page 744)** Discusses the requirements for sharded clusters in MongoDB.
>
> ***Production Cluster Architecture*** **(page 745)** Outlines the components required to deploy a redundant and highly available sharded cluster.
>
> Continue reading from *Sharded Cluster Architectures* (page 744) for additional descriptions of sharded cluster deployments.

Sharded Cluster Behavior **(page 747)** Discusses the operations of sharded clusters with regards to the automatic balancing of data in a cluster and other related availability and security considerations.

> ***Shard Keys*** **(page 747)** MongoDB uses the shard key to divide a collection's data across the cluster's shards.

[2] http://www.mongodb.com/presentations/webinar-sharding-methods-mongodb?jmp=docs
[3] http://www.mongodb.com/presentations/webinar-everything-you-need-know-about-sharding?jmp=docs
[4] http://www.mongodb.com/presentations/mongodb-time-series-data-part-3-sharding?jmp=docs
[5] http://www.mongodb.com/lp/white-paper/ops-best-practices?jmp=docs
[6] http://www.mongodb.com/lp/contact/planning-for-scale?jmp=docs
[7] https://www.mongodb.com/products/consulting?jmp=docs
[8] https://www.mongodb.com/lp/misc/quick-reference-cards?jmp=docs

Sharded Cluster High Availability **(page 750)** Sharded clusters provide ways to address some availability concerns.

Sharded Cluster Query Routing **(page 752)** The cluster's routers, or `mongos` instances, send reads and writes to the relevant shard or shards.

Sharding Mechanics **(page 758)** Discusses the internal operation and behavior of sharded clusters, including chunk migration, balancing, and the cluster metadata.

Sharded Collection Balancing **(page 758)** Balancing distributes a sharded collection's data cluster to all of the shards.

Sharded Cluster Metadata **(page 764)** The cluster maintains internal metadata that reflects the location of data within the cluster.

Continue reading from *Sharding Mechanics* (page 758) for more documentation of the behavior and operation of sharded clusters.

13.2.1 Sharded Cluster Components

Sharded clusters implement *sharding*. A sharded cluster consists of the following components:

Shards A shard is a MongoDB instance that holds a subset of a collection's data. Each shard is either a single `mongod` instance or a *replica set*. In production, all shards are replica sets. For more information see *Shards* (page 740).

Config Servers *Config servers* (page 742) hold metadata about the sharded cluster. The metadata maps *chunks* to shards.

Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

For more information, see *Config Servers* (page 742).

`mongos` Instances `mongos` instances route the reads and writes from applications to the shards. Applications do not access the shards directly. For more information see *Sharded Cluster Query Routing* (page 752).

To deploy a sharded cluster, see *Deploy a Sharded Cluster* (page 765).

Shards

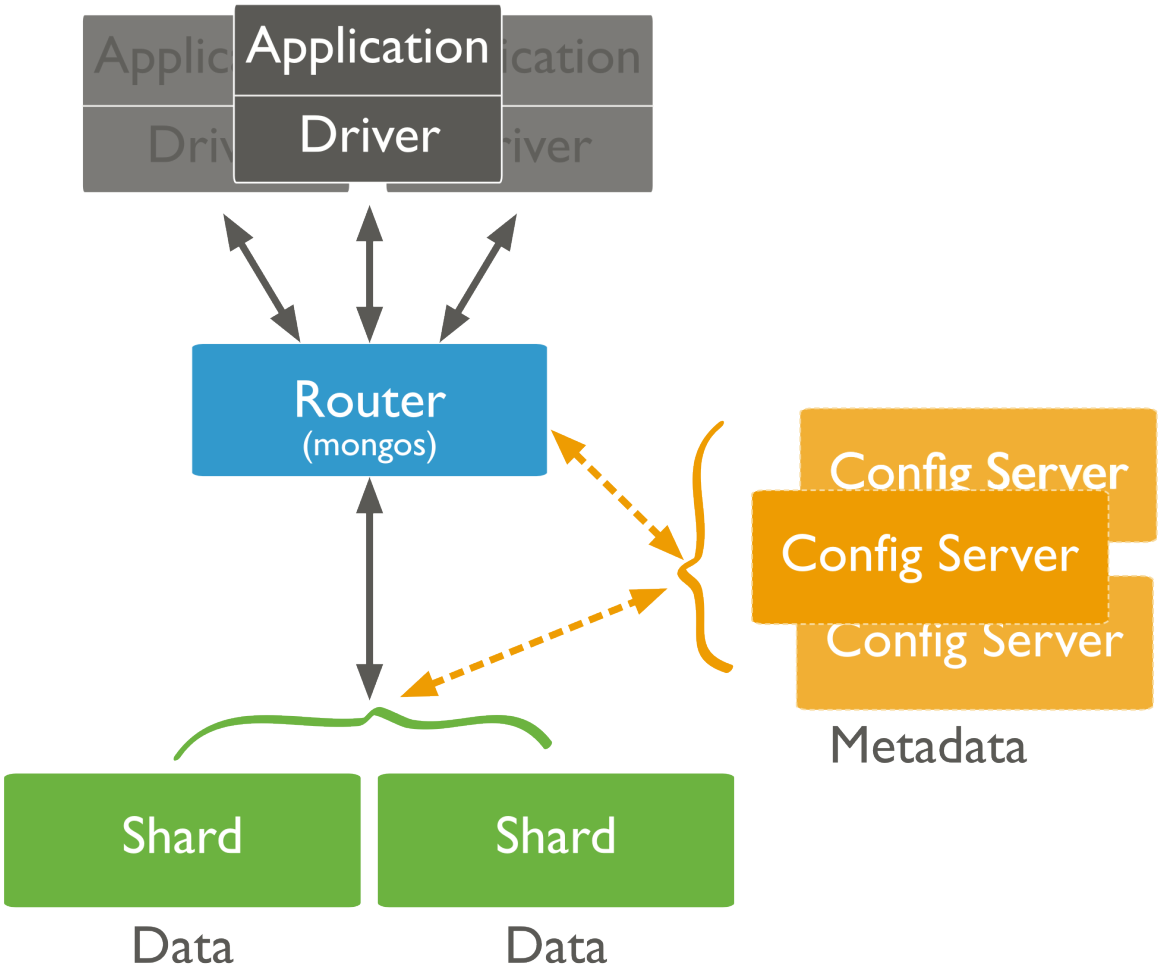
On this page

- Primary Shard (page 742)
- Shard Status (page 742)

A shard is a *replica set* or a single `mongod` that contains a subset of the data for the sharded cluster. Together, the cluster's shards hold the entire data set for the cluster.

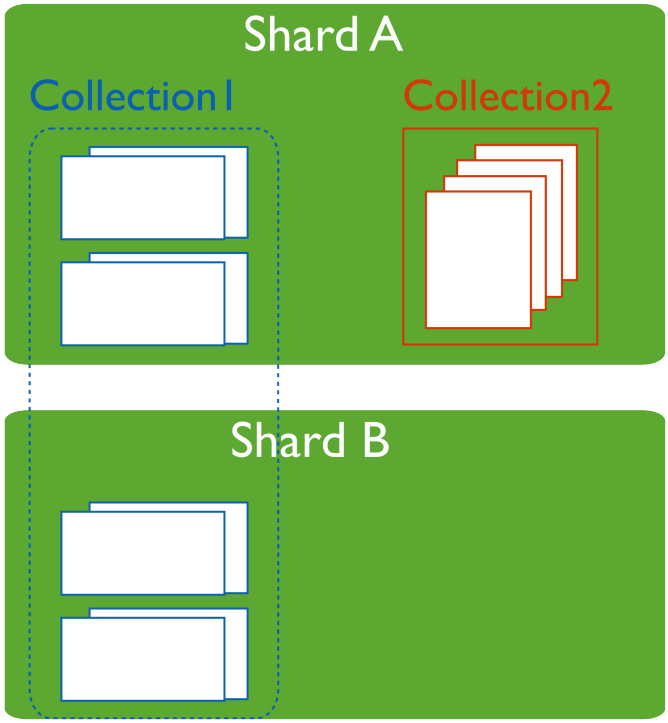
Typically each shard is a replica set. The replica set provides redundancy and high availability for the data in each shard.

Important: MongoDB shards data on a *per collection* basis. You *must* access all data in a sharded cluster via the `mongos` instances. If you connect directly to a shard, you will see only its fraction of the cluster's data. There is no particular order to the data set on a specific shard. MongoDB does not guarantee that any two contiguous chunks will reside on a single shard.


Application
Driver
Router
(mongos)
Config Server
Config Server
Config Server
Metadata
Shard
Shard
Data
Data

Primary Shard

Every database has a "primary" [9] shard that holds all the un-sharded collections in that database.



To change the primary shard for a database, use the `movePrimary` command. The process of migrating the primary shard may take significant time to complete, and you should not access the collections until it completes.

When you deploy a new *sharded cluster* with shards that were previously used as replica sets, all existing databases continue to reside on their original shard. Databases created subsequently may reside on any shard in the cluster.

Shard Status

Use the `sh.status()` method in the `mongo` shell to see an overview of the cluster. This reports includes which shard is primary for the database and the *chunk* distribution across the shards. See `sh.status()` method for more details.

Config Servers

On this page

- Replica Set Config Servers (page 743)
- Read and Write Operations on Config Servers (page 743)
- Config Server Availability (page 743)

[9] The term "primary" shard has nothing to do with the term *primary* in the context of *replica sets*.

Config servers store the *metadata* (page 764) for a sharded cluster.

Warning: If the config servers become inaccessible, the cluster is not accessible. If you cannot recover the data on a config server, the cluster will be inoperable.

MongoDB also uses the config servers to manage distributed locks.

Replica Set Config Servers

Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). Using a replica set for the config servers improves consistency across the config servers, since MongoDB can take advantage of the standard replica set read and write protocols for the config data. In addition, using a replica set for config servers allows a sharded cluster to have more than 3 config servers since a replica set can have up to 50 members. To deploy config servers as a relica set, the config servers must run the *WiredTiger storage engine* (page 595).

The following restrictions apply to a replica set configuration when used for config servers:

- Must have zero *arbiters* (page 635).
- Must have no *delayed members* (page 634).
- Must build indexes (i.e. no member should have `buildIndexes` setting set to false).

Earlier versions of MongoDB required *exactly three* mirrored `mongod` instances to act as the config servers. MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

Each sharded cluster must have its own config servers. Do not use the same config servers for different sharded clusters.

Read and Write Operations on Config Servers

Config servers store the cluster's metadata in the *config database* (page 823). The `mongos` instances cache this data and use it to route reads and writes to shards.

MongoDB only writes data to the config servers when the metadata changes, such as

- after a *chunk migration* (page 759), or
- after a *chunk split* (page 762).

When writing to the replica set config servers, MongoDB uses a *write concern* (page 180) of `"majority"`.

MongoDB reads data from the config server in the following cases:

- A new `mongos` starts for the first time, or an existing `mongos` restarts.
- After change in the cluster metadata, such as after a chunk migration.

When reading from the replica set config servers, MongoDB uses a *Read Concern* (page 181) level of `"majority"` (page 182).

Config Server Availability

If the config server replica set loses its primary and cannot elect a primary, the cluster's metadata becomes *read only*. You can still read and write data from the shards, but no chunk migration or chunk splits will occur until the replica set can elect a primary. If all config databases become unavailable, the cluster can become inoperable.

The `mongos` instances cache the metadata from the config servers. As such, if all config server members become unavailable, you can still use the cluster if you do not restart the `mongos` instances until after the config servers are accessible again. If you restart the `mongos` instances before the config servers are available, the `mongos` will be unable to route reads and writes.

Clusters become inoperable without the cluster metadata. To ensure that the config servers remain available and intact, backups of config servers are critical. The data on the config server is small compared to the data stored in a cluster, and the config server has a relatively low activity load.

See *A Config Server Replica Set Member Become Unavailable* (page 751) for more information.

13.2.2 Sharded Cluster Architectures

The following documents introduce deployment patterns for sharded clusters.

Sharded Cluster Requirements **(page 744)** Discusses the requirements for sharded clusters in MongoDB.

Production Cluster Architecture **(page 745)** Outlines the components required to deploy a redundant and highly available sharded cluster.

Sharded Cluster Test Architecture **(page 745)** Sharded clusters for testing and development can include fewer components.

Sharded Cluster Requirements

On this page

- Data Quantity Requirements (page 745)

While sharding is a powerful and compelling feature, sharded clusters have significant infrastructure requirements and increases the overall complexity of a deployment. As a result, only deploy sharded clusters when indicated by application and operational requirements

Sharding is the *only* solution for some classes of deployments. Use *sharded clusters* if:

- your data set approaches or exceeds the storage capacity of a single MongoDB instance.
- the size of your system's active *working set will soon* exceed the capacity of your system's *maximum* RAM.
- a single MongoDB instance cannot meet the demands of your write operations, and all other approaches have not reduced contention.

If these attributes are not present in your system, sharding will only add complexity to your system without adding much benefit.

Important: It takes time and resources to deploy sharding. If your system has *already* reached or exceeded its capacity, it will be difficult to deploy sharding without impacting your application.

As a result, if you think you will need to partition your database in the future, **do not** wait until your system is over capacity to enable sharding.

When designing your data model, take into consideration your sharding needs.

Data Quantity Requirements

Your cluster should manage a large quantity of data if sharding is to have an effect. The default *chunk* size is 64 megabytes. And the *balancer* (page 758) will not begin moving data across shards until the imbalance of chunks among the shards exceeds the *migration threshold* (page 759). In practical terms, unless your cluster has many hundreds of megabytes of data, your data will remain on a single shard.

In some situations, you may need to shard a small collection of data. But most of the time, sharding a small collection is not worth the added complexity and overhead unless you need additional write capacity. If you have a small data set, a properly configured single MongoDB instance or a replica set will usually be enough for your persistence layer needs.

Chunk size is `user configurable`. For most deployments, the default value is of 64 megabytes is ideal. See *Chunk Size* (page 762) for more information.

Production Cluster Architecture

In a production cluster, you must ensure that data is redundant and that your systems are highly available. To that end, a production cluster must have the following components:

- **Config Servers** Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

  A single *sharded cluster* must have exclusive use of its *config servers* (page 742). If you have multiple sharded clusters, each cluster must have its own replica set config servers.

- **Two or More Replica Sets As Shards** These replica sets are the *shards*. For information on replica sets, see *Replication* (page 623).

- **One or More Query Routers (`mongos`)** The `mongos` instances are the routers for the cluster. Typically, deployments have one `mongos` instance on each application server.

  You may also deploy a group of `mongos` instances and use a proxy/load balancer between the application and the `mongos`. In these deployments, you *must* configure the load balancer for *client affinity* so that every connection from a single client reaches the same `mongos`.

  Because cursors and other resources are specific to an single `mongos` instance, each client must interact with only one `mongos` instance.

See also:

Deploy a Sharded Cluster (page 765)

Sharded Cluster Test Architecture

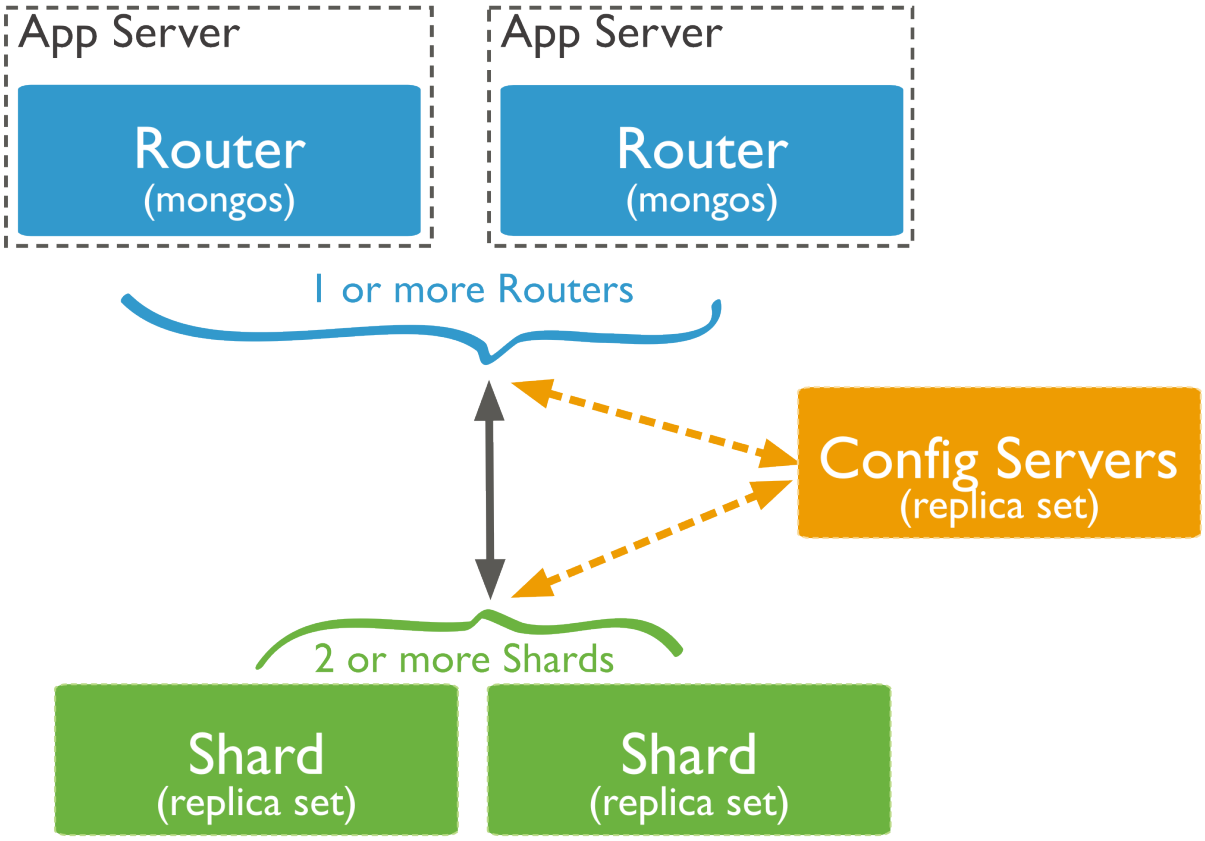
> **Warning:** Use the test cluster architecture for testing and development only.

For testing and development, you can deploy a sharded cluster with a minimum number of components. These **non-production** clusters have the following components:

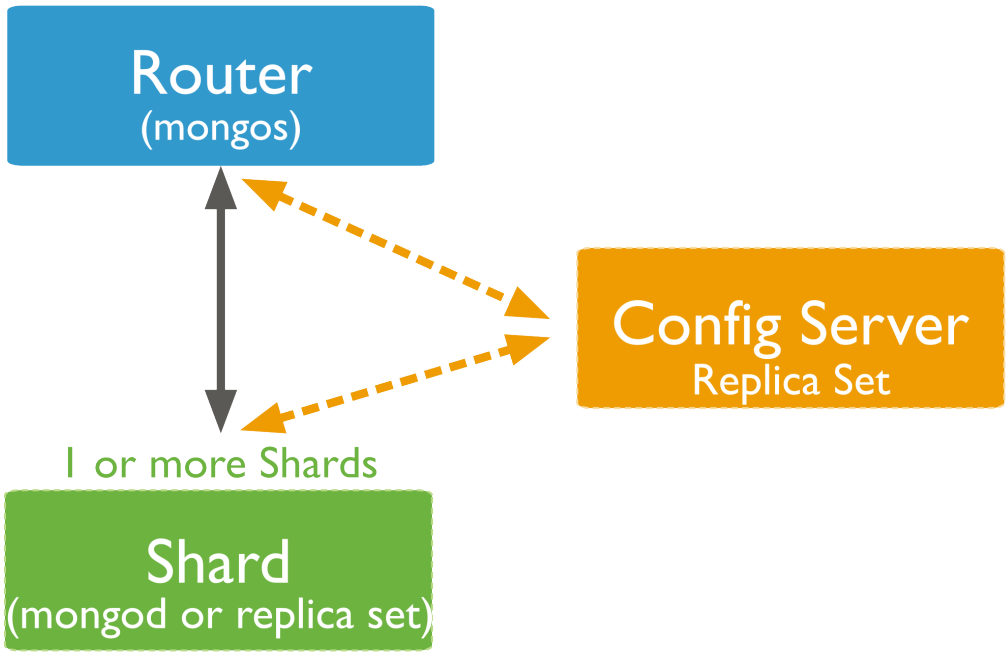
- A replica set *config server* (page 742) with one member.

  Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

- At least one shard. Shards are either *replica sets* or a standalone `mongod` instances.


App Server
Router
(mongos)
App Server
Router
(mongos)
1 or more Routers
Config Servers
(replica set)
2 or more Shards
Shard
(replica set)
Shard
(replica set)

- One `mongos` instance.



See

Production Cluster Architecture (page 745)

13.2.3 Sharded Cluster Behavior

These documents address the distribution of data and queries to a sharded cluster as well as specific security and availability considerations for sharded clusters.

Shard Keys **(page 747)** MongoDB uses the shard key to divide a collection's data across the cluster's shards.

Sharded Cluster High Availability **(page 750)** Sharded clusters provide ways to address some availability concerns.

Sharded Cluster Query Routing **(page 752)** The cluster's routers, or `mongos` instances, send reads and writes to the relevant shard or shards.

Tag Aware Sharding **(page 756)** Tags associate specific ranges of *shard key* values with specific shards for use in managing deployment patterns.

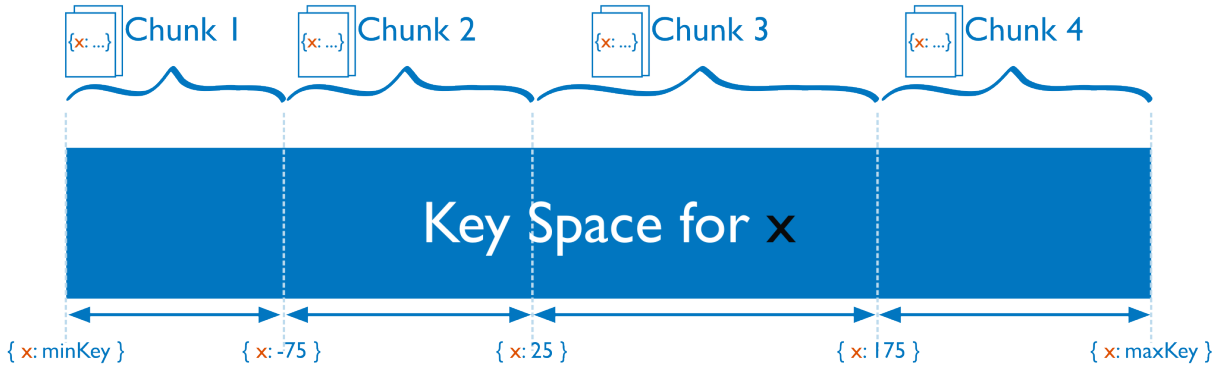
Shard Keys

On this page

- Considerations (page 748)
- Hashed Shard Keys (page 748)
- Impacts of Shard Keys on Cluster Operations (page 749)
- Additional Information (page 750)

The shard key determines the distribution of the collection's *documents* among the cluster's *shards*. The shard key is either an indexed *field* or an indexed compound field that exists in every document in the collection.

MongoDB partitions data in the collection using ranges of shard key values. Each range, or *chunk*, defines a non-overlapping range of shard key values. MongoDB distributes the chunks, and their documents, among the shards in the cluster.



When a chunk grows beyond the *chunk size* (page 762), MongoDB attempts to *split* the chunk into smaller chunks, always based on ranges in the shard key.

Considerations

Shard keys are immutable and cannot be changed after insertion. See the *system limits for sharded cluster* for more information.

The index on the shard key **cannot** be a *multikey index* (page 525).

Hashed Shard Keys

New in version 2.4.

Hashed shard keys use a *hashed index* (page 564) of a single field as the *shard key* to partition data across your sharded cluster.

The field you choose as your hashed shard key should have a good cardinality, or large number of different values. Hashed keys work well with fields that increase monotonically like *ObjectId* values or timestamps.

If you shard an empty collection using a hashed shard key, MongoDB will automatically create and migrate chunks so that each shard has two chunks. You can control how many chunks MongoDB will create with the `numInitialChunks` parameter to `shardCollection` or by manually creating chunks on the empty collection using the `split` command.

To shard a collection using a hashed shard key, see *Shard a Collection Using a Hashed Shard Key* (page 773).

Tip

MongoDB automatically computes the hashes when resolving queries using hashed indexes. Applications do **not** need to compute hashes.

Impacts of Shard Keys on Cluster Operations

The shard key affects write and query performance by determining how the MongoDB partitions data in the cluster and how effectively the `mongos` instances can direct operations to the cluster. Consider the following operational impacts of shard key selection:

Write Scaling Some possible shard keys will allow your application to take advantage of the increased write capacity that the cluster can provide, while others do not. Consider the following example where you shard by the values of the default *_id* field, which is *ObjectId*.

MongoDB generates `ObjectId` values upon document creation to produce a unique identifier for the object. However, the most significant bits of data in this value represent a time stamp, which means that they increment in a regular and predictable pattern. Even though this value has *high cardinality* (page 772), when using this, *any date, or other monotonically increasing number* as the shard key, all insert operations will be storing data into a single chunk, and therefore, a single shard. As a result, the write capacity of this shard will define the effective write capacity of the cluster.

A shard key that increases monotonically will not hinder performance if you have a very low insert rate, or if most of your write operations are `update()` operations distributed through your entire data set. Generally, choose shard keys that have *both* high cardinality and will distribute write operations across the *entire cluster*.

Typically, a computed shard key that has some amount of "randomness," such as ones that include a cryptographic hash (i.e. MD5 or SHA1) of other content in the document, will allow the cluster to scale write operations. However, random shard keys do not typically provide *query isolation* (page 749), which is another important characteristic of shard keys.

New in version 2.4: MongoDB makes it possible to shard a collection on a hashed index. This can greatly improve write scaling. See *Shard a Collection Using a Hashed Shard Key* (page 773).

Querying The `mongos` provides an interface for applications to interact with sharded clusters that hides the complexity of *data partitioning*. A `mongos` receives queries from applications, and uses metadata from the *config server* (page 742), to route queries to the `mongod` instances with the appropriate data. While the `mongos` succeeds in making all querying operational in sharded environments, the *shard key* you select can have a profound affect on query performance.

See also:

The *Sharded Cluster Query Routing* (page 752) and *config server* (page 742) sections for a more general overview of querying in sharded environments.

Query Isolation Generally, the fastest queries in a sharded environment are those that `mongos` will route to a single shard, using the *shard key* and the cluster meta data from the *config server* (page 742). For queries that don't include the shard key, `mongos` must query all shards, wait for their responses and then return the result to the application. These "scatter/gather" queries can be long running operations.

If your query includes the first component of a compound shard key [10], the `mongos` can route the query directly to a single shard, or a small number of shards, which provides better performance. Even if you query values of the shard key that reside in different chunks, the `mongos` will route queries directly to specific shards.

To select a shard key for a collection:

- determine the most commonly included fields in queries for a given application
- find which of these operations are most performance dependent.

[10] In many ways, you can think of the shard key a cluster-wide index. However, be aware that sharded systems cannot enforce cluster-wide unique indexes *unless* the unique field is in the shard key. Consider the *Indexes* (page 515) page for more information on indexes and compound indexes.

If this field has low cardinality (i.e not sufficiently selective) you should add a second field to the shard key making a compound shard key. The data may become more splittable with a compound shard key.

See

Sharded Cluster Query Routing (page 752) for more information on query operations in the context of sharded clusters.

Sorting In sharded systems, the `mongos` performs a merge-sort of all sorted query results from the shards. See *Sharded Cluster Query Routing* (page 752) and *Use Indexes to Sort Query Results* (page 587) for more information.

Indivisible Chunks An insufficiently granular shard key can result in chunks that are "unsplittable". See *Create a Shard Key that is Easily Divisible* (page 771) for more information.

Additional Information

- *Considerations for Selecting Shard Keys* (page 771)
- *Shard a Collection Using a Hashed Shard Key* (page 773).

Sharded Cluster High Availability

On this page

- Application Servers or `mongos` Instances Become Unavailable (page 750)
- A Single `mongod` Becomes Unavailable in a Shard (page 750)
- All Members of a Shard Become Unavailable (page 751)
- A Config Server Replica Set Member Become Unavailable (page 751)
- Renaming Mirrored Config Servers and Cluster Availability (page 751)
- Shard Keys and Cluster Availability (page 751)

A *production* (page 745) *cluster* has no single point of failure. This section introduces the availability concerns for MongoDB deployments in general and highlights potential failure scenarios and available resolutions.

Application Servers or `mongos` Instances Become Unavailable

If each application server has its own `mongos` instance, other application servers can continue to access the database. Furthermore, `mongos` instances do not maintain persistent state, and they can restart and become unavailable without losing any state or data. When a `mongos` instance starts, it retrieves a copy of the *config database* and can begin routing queries.

A Single `mongod` Becomes Unavailable in a Shard

Replica sets (page 623) provide high availability for shards. If the unavailable `mongod` is a *primary*, then the replica set will *elect* (page 644) a new primary. If the unavailable `mongod` is a *secondary*, and it disconnects the primary and secondary will continue to hold all data. In a three member replica set, even if a single member of the set experiences catastrophic failure, two other members have full copies of the data. [11]

[11] If an unavailable secondary becomes available while it still has current oplog entries, it can catch up to the latest state of the set using the normal *replication process*; otherwise, it must perform an *initial sync*.

Always investigate availability interruptions and failures. If a system is unrecoverable, replace it and create a new member of the replica set as soon as possible to replace the lost redundancy.

All Members of a Shard Become Unavailable

If all members of a replica set shard are unavailable, all data held in that shard is unavailable. However, the data on all other shards will remain available, and it is possible to read and write data to the other shards. However, your application must be able to deal with partial results, and you should investigate the cause of the interruption and attempt to recover the shard as soon as possible.

A Config Server Replica Set Member Become Unavailable

Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

Replica sets (page 623) provide high availability for the config servers. If an unavailable config server is a *primary*, then the replica set will *elect* (page 644) a new primary.

If the replica set config server loses its primary and cannot elect a primary, the cluster's metadata becomes *read only*. You can still read and write data from the shards, but no *chunk migration* (page 758) or *chunk splits* (page 808) will occur until a primary is available. If all config databases become unavailable, the cluster can become inoperable.

Note: All config servers must be running and available when you first initiate a *sharded cluster*.

Renaming Mirrored Config Servers and Cluster Availability

If the sharded cluster is using *mirrored* config servers instead of a replica set and the name or address that a sharded cluster uses to connect to a config server changes, you must restart **every** `mongod` and `mongos` instance in the sharded cluster. Avoid downtime by using CNAMEs to identify config servers within the MongoDB deployment.

To avoid downtime when renaming config servers, use DNS names unrelated to physical or virtual hostnames to refer to your *config servers* (page 742).

Generally, refer to each config server using the DNS alias (e.g. a CNAME record). When specifying the config server connection string to `mongos`, use these names. These records make it possible to change the IP address or rename config servers without changing the connection string and without having to restart the entire cluster.

Shard Keys and Cluster Availability

The most important consideration when choosing a *shard key* are:

- to ensure that MongoDB will be able to distribute data evenly among shards, and
- to scale writes across the cluster, and
- to ensure that `mongos` can isolate most queries to a specific `mongod`.

Furthermore:

- Each shard should be a *replica set*, if a specific `mongod` instance fails, the replica set members will elect another to be *primary* and continue operation. However, if an entire shard is unreachable or fails for some reason, that data will be unavailable.

- If the shard key allows the `mongos` to isolate most operations to a single shard, then the failure of a single shard will only render *some* data unavailable.
- If your shard key distributes data required for every operation throughout the cluster, then the failure of the entire shard will render the entire cluster unavailable.

In essence, this concern for reliability simply underscores the importance of choosing a shard key that isolates query operations to a single shard.

Sharded Cluster Query Routing

On this page

- Routing Process (page 752)
- Detect Connections to `mongos` Instances (page 753)
- Broadcast Operations and Targeted Operations (page 753)
- Sharded and Non-Sharded Data (page 756)

MongoDB `mongos` instances route queries and write operations to *shards* in a sharded cluster. `mongos` provide the only interface to a sharded cluster from the perspective of applications. Applications never connect or communicate directly with the shards.

The `mongos` tracks what data is on which shard by caching the metadata from the *config servers* (page 742). The `mongos` uses the metadata to route operations from applications and clients to the `mongod` instances. A `mongos` has no *persistent* state and consumes minimal system resources.

The most common practice is to run `mongos` instances on the same systems as your application servers, but you can maintain `mongos` instances on the shards or on other dedicated resources.

Changed in version 3.2: For *aggregation operations* (page 199) that run on multiple shards, if the operations do not require running on the database's primary shard, these operations can route the results to any shard to merge the results and avoid overloading the primary shard for that database.

Routing Process

A `mongos` instance uses the following processes to route queries and return results.

How `mongos` Determines which Shards Receive a Query A `mongos` instance routes a query to a *cluster* by:

1. Determining the list of *shards* that must receive the query.
2. Establishing a cursor on all targeted shards.

In some cases, when the *shard key* or a prefix of the shard key is a part of the query, the `mongos` can route the query to a subset of the shards. Otherwise, the `mongos` must direct the query to *all* shards that hold documents for that collection.

Example

Given the following shard key:

```
{ zipcode: 1, u_id: 1, c_date: 1 }
```

Depending on the distribution of chunks in the cluster, the `mongos` may be able to target the query at a subset of shards, if the query contains the following fields:

```
{ zipcode: 1 }
{ zipcode: 1, u_id: 1 }
{ zipcode: 1, u_id: 1, c_date: 1 }
```

How mongos Handles Query Modifiers If the result of the query is not sorted, the mongos instance opens a result cursor that "round robins" results from all cursors on the shards.

If the query specifies sorted results using the sort() cursor method, the mongos instance passes the $orderby option to the shards. The primary shard for the database receives and performs a merge sort for all results before returning the data to the client via the mongos.

If the query limits the size of the result set using the limit() cursor method, the mongos instance passes that limit to the shards and then re-applies the limit to the result before returning the result to the client.

If the query specifies a number of records to *skip* using the skip() cursor method, the mongos *cannot* pass the skip to the shards, but rather retrieves unskipped results from the shards and skips the appropriate number of documents when assembling the complete result. However, when used in conjunction with a limit(), the mongos will pass the *limit* plus the value of the skip() to the shards to improve the efficiency of these operations.

Detect Connections to mongos Instances

To detect if the MongoDB instance that your client is connected to is mongos, use the isMaster command. When a client connects to a mongos, isMaster returns a document with a msg field that holds the string isdbgrid. For example:

```
{
   "ismaster" : true,
   "msg" : "isdbgrid",
   "maxBsonObjectSize" : 16777216,
   "ok" : 1
}
```

If the application is instead connected to a mongod, the returned document does not include the isdbgrid string.

Broadcast Operations and Targeted Operations

In general, operations in a sharded environment are either:

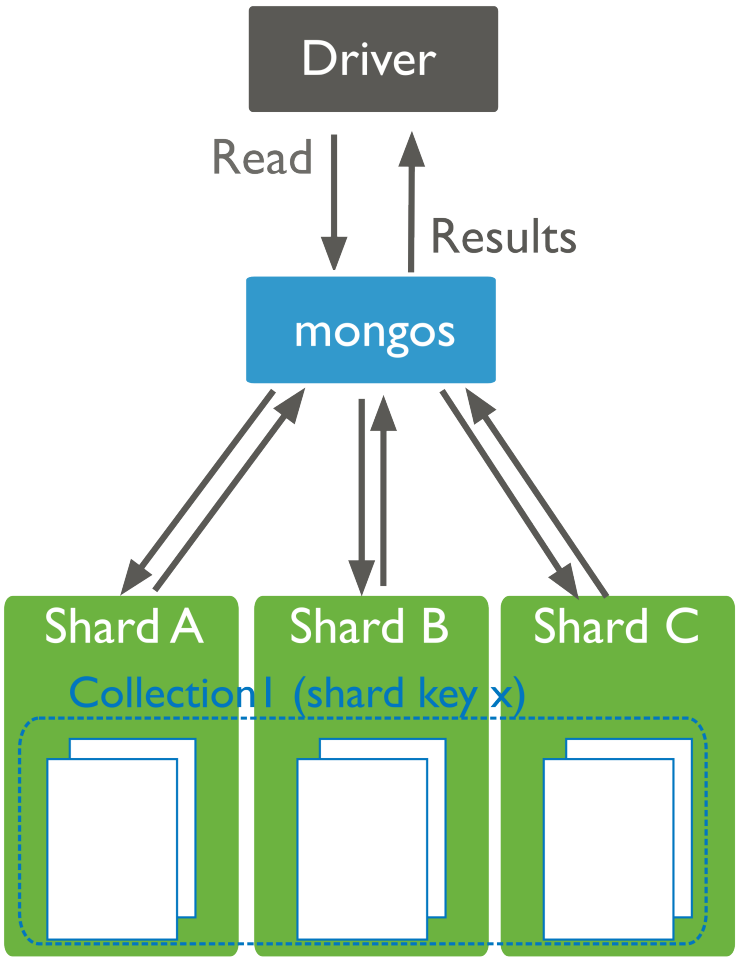
- Broadcast to all shards in the cluster that hold documents in a collection
- Targeted at a single shard or a limited group of shards, based on the shard key

For best performance, use targeted operations whenever possible. While some operations must broadcast to all shards, you can ensure MongoDB uses targeted operations whenever possible by always including the shard key.

Broadcast Operations mongos instances broadcast queries to all shards for the collection **unless** the mongos can determine which shard or subset of shards stores this data.

Multi-update operations are always broadcast operations.

The remove() operation is always a broadcast operation, unless the operation specifies the shard key in full.


Driver
Read
Results
mongos
Shard A
Shard B
Shard C
Collection 1 (shard key x)

Targeted Operations All `insert()` operations target to one shard.

All single `update()` (including *upsert* operations) and `remove()` operations must target to one shard.

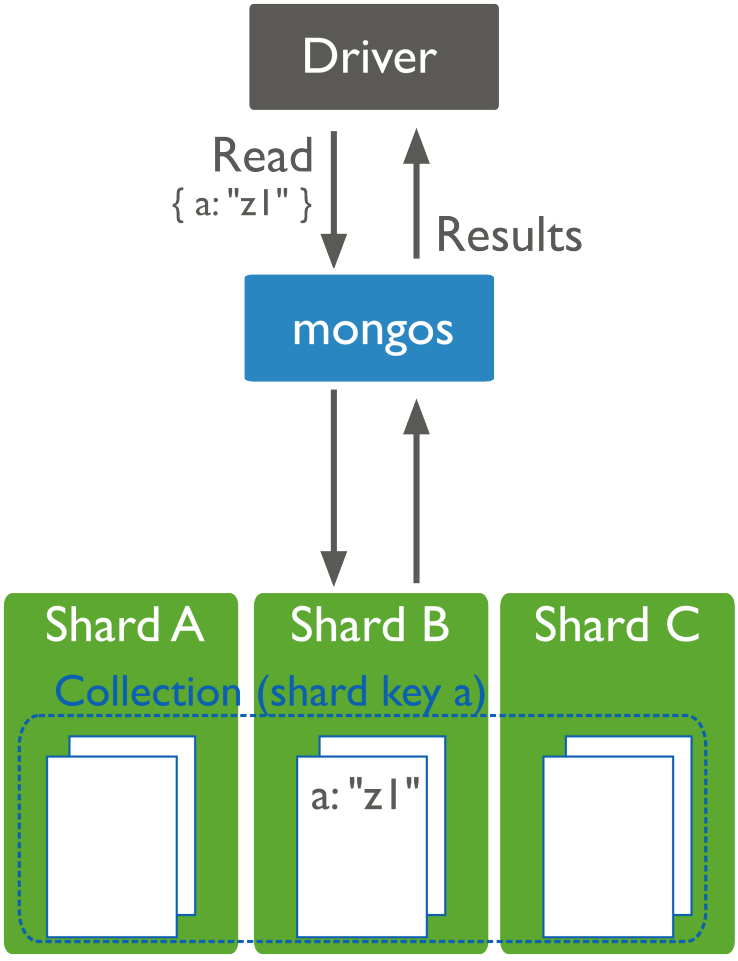
Important: All `update()` and `remove()` operations for a sharded collection that specify the `justOne` or `multi: false` option must include the *shard key or* the `_id` field in the query specification. `update()` and `remove()` operations specifying `justOne` or `multi: false` in a sharded collection without the *shard key or* the `_id` field return an error.

For queries that include the shard key or portion of the shard key, `mongos` can target the query at a specific shard or set of shards. This is the case only if the portion of the shard key included in the query is a *prefix* of the shard key. For example, if the shard key is:

```
{ a: 1, b: 1, c: 1 }
```

The `mongos` program *can* route queries that include the full shard key or either of the following shard key prefixes at a specific shard or set of shards:

```
{ a: 1 }
{ a: 1, b: 1 }
```

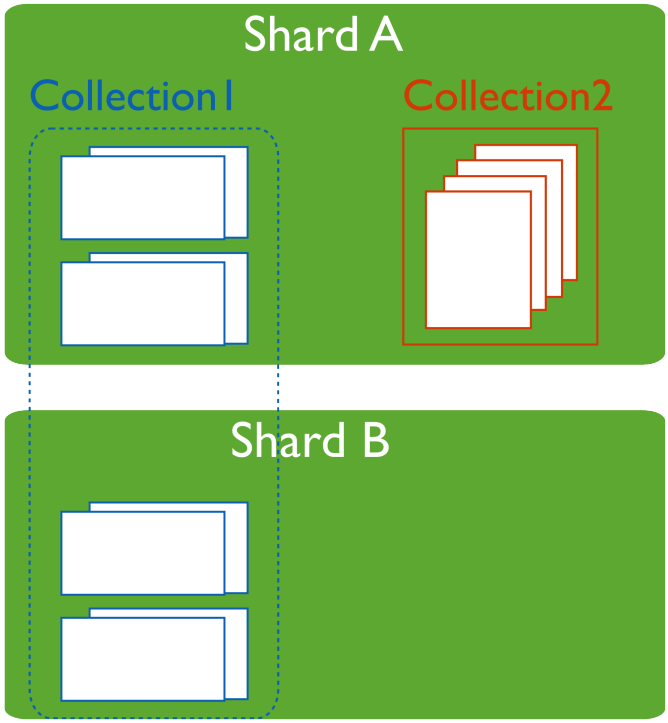


Depending on the distribution of data in the cluster and the selectivity of the query, `mongos` may still have to contact

multiple shards [12] to fulfill these queries.

Sharded and Non-Sharded Data

Sharding operates on the collection level. You can shard multiple collections within a database or have multiple databases with sharding enabled. [13] However, in production deployments, some databases and collections will use sharding, while other databases and collections will only reside on a single shard.



Regardless of the data architecture of your *sharded cluster*, ensure that all queries and operations use the *mongos* router to access the data cluster. Use the `mongos` even for operations that do not impact the sharded data.

Tag Aware Sharding

On this page

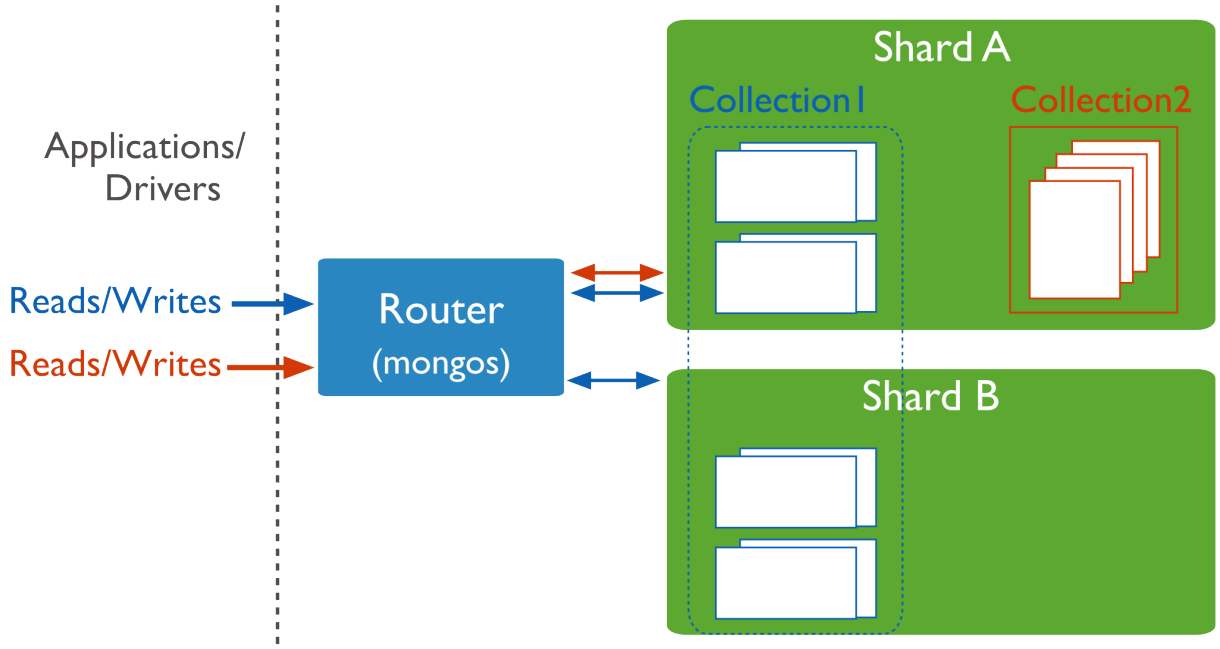
- Considerations (page 757)
- Behavior and Operations (page 757)
- Additional Resource (page 758)

MongoDB supports tagging a range of *shard key* values to associate that range with a shard or group of shards. Those shards receive all inserts within the tagged range.

The balancer obeys tagged range associations, which enables the following deployment patterns:

[12] `mongos` will route some queries, even some that include the shard key, to all shards, if needed.

[13] As you configure sharding, you will use the `enableSharding` command to enable sharding for a database. This simply makes it possible to use the `shardCollection` command on a collection within that database.



- isolate a specific subset of data on a specific set of shards.
- ensure that the most relevant data reside on shards that are geographically closest to the application servers.

This document describes the behavior, operation, and use of tag aware sharding in MongoDB deployments.

Considerations

- *Shard key* range tags are distinct from *replica set member tags* (page 653).
- *Hash-based sharding* only supports tag-aware sharding on an entire collection.
- Shard ranges are always inclusive of the lower value and exclusive of the upper boundary.

Behavior and Operations

The balancer migrates chunks of documents in a sharded collection to the shards associated with a tag that has a *shard key* range with an *upper* bound *greater* than the chunk's *lower* bound.

During balancing rounds, if the balancer detects that any chunks violate configured tags, the balancer migrates those chunks to shards associated with those tags.

After configuring a tag with a shard key range and associating it with a shard or shards, the cluster may take some time to balance the data among the shards. This depends on the division of chunks and the current distribution of data in the cluster.

Once configured, the balancer respects tag ranges during future *balancing rounds* (page 758).

See also:

Manage Shard Tags (page 816)

Additional Resource

- Whitepaper: MongoDB Multi-Data Center Deployments[14]
- Webinar: Multi-Data Center Deployment[15]

13.2.4 Sharding Mechanics

The following documents describe sharded cluster processes.

Sharded Collection Balancing **(page 758)** Balancing distributes a sharded collection's data cluster to all of the shards.

Chunk Migration Across Shards **(page 759)** MongoDB migrates chunks to shards as part of the balancing process.

Chunk Splits in a Sharded Cluster **(page 762)** When a chunk grows beyond the configured size, MongoDB splits the chunk in half.

Shard Key Indexes **(page 763)** Sharded collections must keep an index that starts with the shard key.

Sharded Cluster Metadata **(page 764)** The cluster maintains internal metadata that reflects the location of data within the cluster.

Sharded Collection Balancing

On this page

- Cluster Balancer (page 758)
- Migration Thresholds (page 759)
- Shard Size (page 759)

Balancing is the process MongoDB uses to distribute data of a sharded collection evenly across a *sharded cluster*. When a *shard* has too many of a sharded collection's *chunks* compared to other shards, MongoDB automatically balances the chunks across the shards. The balancing procedure for *sharded clusters* is entirely transparent to the user and application layer.

Cluster Balancer

The *balancer* process is responsible for redistributing the chunks of a sharded collection evenly among the shards for every sharded collection. By default, the balancer process is always enabled.

Any `mongos` instance in the cluster can start a balancing round. When a balancer process is active, the responsible `mongos` acquires a "lock" by modifying a document in the `lock` collection in the *Config Database* (page 823).

Changed in version 3.2: With replica set config servers, clock skew does not affect distributed lock management. If you are using *mirrored* config servers, large differences in timekeeping can lead to failed distributed locks. With *mirrored* config servers, minimize clock skew by running the network time protocol (NTP) `ntpd` on your servers.

To address uneven chunk distribution for a sharded collection, the balancer *migrates chunks* (page 759) from shards with more chunks to shards with a fewer number of chunks. The balancer migrates the chunks, one at a time, until there is an even distribution of chunks for the collection across the shards. For details about chunk migration, see *Chunk Migration Procedure* (page 760).

[14] http://www.mongodb.com/lp/white-paper/multi-dc?jmp=docs
[15] https://www.mongodb.com/presentations/webinar-multi-data-center-deployment?jmp=docs

Changed in version 2.6: Chunk migrations can have an impact on disk space. Starting in MongoDB 2.6, the source shard automatically archives the migrated documents by default. For details, see *moveChunk directory* (page 761).

Chunk migrations carry some overhead in terms of bandwidth and workload, both of which can impact database performance. The *balancer* attempts to minimize the impact by:

- Moving only one chunk at a time. See also *Chunk Migration Queuing* (page 761).
- Starting a balancing round **only** when the difference in the number of chunks between the shard with the greatest number of chunks for a sharded collection and the shard with the lowest number of chunks for that collection reaches the *migration threshold* (page 759).

You may disable the balancer temporarily for maintenance. See *Disable the Balancer* (page 802) for details.

You can also limit the window during which the balancer runs to prevent it from impacting production traffic. See *Schedule the Balancing Window* (page 801) for details.

Note: The specification of the balancing window is relative to the local time zone of all individual `mongos` instances in the cluster.

See also:

Manage Sharded Cluster Balancer (page 800).

Migration Thresholds

To minimize the impact of balancing on the cluster, the *balancer* will not begin balancing until the distribution of chunks for a sharded collection has reached certain thresholds. The thresholds apply to the difference in number of *chunks* between the shard with the most chunks for the collection and the shard with the fewest chunks for that collection. The balancer has the following thresholds:

Number of Chunks	Migration Threshold
Fewer than 20	2
20-79	4
80 and greater	8

Once a balancing round starts, the balancer will not stop until, for the collection, the difference between the number of chunks on any two shards for that collection is *less than two* or a chunk migration fails.

Shard Size

By default, MongoDB will attempt to fill all available disk space with data on every shard as the data set grows. To ensure that the cluster always has the capacity to handle data growth, monitor disk usage as well as other performance metrics.

When adding a shard, you may set a "maximum size" for that shard. This prevents the *balancer* from migrating chunks to the shard when the value of `mem.mapped` exceeds the "maximum size". Use the `maxSize` parameter of the `addShard` command to set the "maximum size" for the shard.

See also:

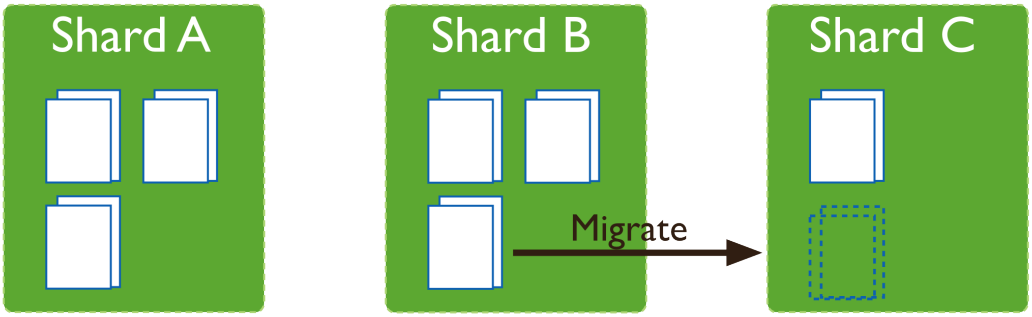
Change the Maximum Storage Size for a Given Shard (page 798) and *Monitoring for MongoDB* (page 285).

Chunk Migration Across Shards

On this page

- Chunk Migration (page 760)
- `moveChunk` directory (page 761)
- Jumbo Chunks (page 762)

Chunk migration moves the chunks of a sharded collection from one shard to another and is part of the *balancer* (page 758) process.



Chunk Migration

MongoDB migrates chunks in a *sharded cluster* to distribute the chunks of a sharded collection evenly among shards. Migrations may be either:

- Manual. Only use manual migration in limited cases, such as to distribute data during bulk inserts. See *Migrating Chunks Manually* (page 809) for more details.
- Automatic. The *balancer* (page 758) process automatically migrates chunks when there is an uneven distribution of a sharded collection's chunks across the shards. See *Migration Thresholds* (page 759) for more details.

Chunk Migration Procedure All chunk migrations use the following procedure:

1. The balancer process sends the `moveChunk` command to the source shard.
2. The source starts the move with an internal `moveChunk` command. During the migration process, operations to the chunk route to the source shard. The source shard is responsible for incoming write operations for the chunk.
3. The destination shard builds any indexes required by the source that do not exist on the destination.
4. The destination shard begins requesting documents in the chunk and starts receiving copies of the data.
5. After receiving the final document in the chunk, the destination shard starts a synchronization process to ensure that it has the changes to the migrated documents that occurred during the migration.
6. When fully synchronized, the destination shard connects to the *config database* and updates the cluster metadata with the new location for the chunk.
7. After the destination shard completes the update of the metadata, and once there are no open cursors on the chunk, the source shard deletes its copy of the documents.

Note: If the balancer needs to perform additional chunk migrations from the source shard, the balancer can

start the next chunk migration without waiting for the current migration process to finish this deletion step. See *Chunk Migration Queuing* (page 761).

Changed in version 2.6: The source shard automatically archives the migrated documents by default. For more information, see *moveChunk directory* (page 761).

The migration process ensures consistency and maximizes the availability of chunks during balancing.

Chunk Migration Queuing To migrate multiple chunks from a shard, the balancer migrates the chunks one at a time. However, the balancer does not wait for the current migration's delete phase to complete before starting the next chunk migration. See *Chunk Migration* (page 760) for the chunk migration process and the delete phase.

This queuing behavior allows shards to unload chunks more quickly in cases of heavily imbalanced cluster, such as when performing initial data loads without pre-splitting and when adding new shards.

This behavior also affect the `moveChunk` command, and migration scripts that use the `moveChunk` command may proceed more quickly.

In some cases, the delete phases may persist longer. If multiple delete phases are queued but not yet complete, a crash of the replica set's primary can orphan data from multiple migrations.

The `_waitForDelete`, available as a setting for the balancer as well as the `moveChunk` command, can alter the behavior so that the delete phase of the current migration blocks the start of the next chunk migration. The `_waitForDelete` is generally for internal testing purposes. For more information, see *Wait for Delete* (page 800).

Chunk Migration and Replication Changed in version 3.0: The default value `secondaryThrottle` became `true` for all chunk migrations.

The new `writeConcern` field in the balancer configuration document allows you to specify a *write concern* (page 179) semantics with the `_secondaryThrottle` option.

By default, each document operation during chunk migration propagates to at least one secondary before the balancer proceeds with the next document, which is equivalent to a write concern of `{ w: 2 }`. You can set the `writeConcern` option on the balancer configuration to set different write concern semantics.

To override this behavior and allow the balancer to continue without waiting for replication to a secondary, set the `_secondaryThrottle` parameter to `false`. See *Change Replication Behavior for Chunk Migration* (page 799) to update the `_secondaryThrottle` parameter for the balancer.

For the `moveChunk` command, the `secondaryThrottle` parameter is independent of the `_secondaryThrottle` parameter for the balancer.

Independent of the `secondaryThrottle` setting, certain phases of the chunk migration have the following replication policy:

- MongoDB briefly pauses all application writes to the source shard before updating the config servers with the new location for the chunk, and resumes the application writes after the update. The chunk move requires all writes to be acknowledged by majority of the members of the replica set both before and after committing the chunk move to config servers.
- When an outgoing chunk migration finishes and cleanup occurs, all writes must be replicated to a majority of servers before further cleanup (from other outgoing migrations) or new incoming migrations can proceed.

`moveChunk` directory

Starting in MongoDB 2.6, `sharding.archiveMovedChunks` is enabled by default. With `sharding.archiveMovedChunks` enabled, the source shard archives the documents in the migrated chunks in a directory named after the collection namespace under the `moveChunk` directory in the `storage.dbPath`.

Jumbo Chunks

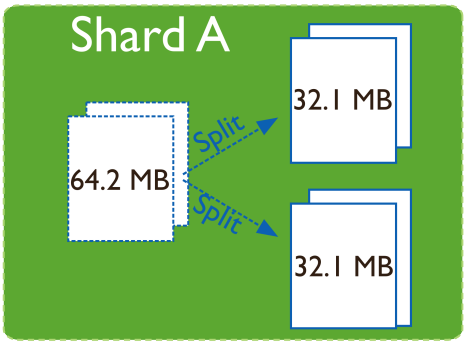
During chunk migration, if the chunk exceeds the *specified chunk size* (page 762) or if the number of documents in the chunk exceeds `Maximum Number of Documents Per Chunk to Migrate`, MongoDB does not migrate the chunk. Instead, MongoDB attempts to *split* (page 762) the chunk. If the split is unsuccessful, MongoDB labels the chunk as *jumbo* to avoid repeated attempts to migrate the chunk.

Chunk Splits in a Sharded Cluster

On this page

- Chunk Size (page 762)
- Limitations (page 763)
- Indivisible Chunks (page 763)

As chunks grow beyond the *specified chunk size* (page 762) a `mongos` instance will attempt to split the chunk in half. Splits may lead to an uneven distribution of the chunks for a collection across the shards. In such cases, the `mongos` instances will initiate a round of migrations to redistribute chunks across shards. See *Sharded Collection Balancing* (page 758) for more details on balancing chunks across shards.



Chunk Size

The default *chunk* size in MongoDB is 64 megabytes. You can *increase or reduce the chunk size* (page 813), mindful of its effect on the cluster's efficiency.

1. Small chunks lead to a more even distribution of data at the expense of more frequent migrations. This creates expense at the query routing (`mongos`) layer.
2. Large chunks lead to fewer migrations. This is more efficient both from the networking perspective *and* in terms of internal overhead at the query routing layer. But, these efficiencies come at the expense of a potentially more uneven distribution of data.
3. Chunk size affects the `Maximum Number of Documents Per Chunk to Migrate`.

For many deployments, it makes sense to avoid frequent and potentially spurious migrations at the expense of a slightly less evenly distributed data set.

Limitations

Changing the chunk size affects when chunks split but there are some limitations to its effects.

- Automatic splitting only occurs during inserts or updates. If you lower the chunk size, it may take time for all chunks to split to the new size.
- Splits cannot be "undone". If you increase the chunk size, existing chunks must grow through inserts or updates until they reach the new size.

Note: Chunk ranges are inclusive of the lower boundary and exclusive of the upper boundary.

Indivisible Chunks

In some cases, chunks can grow beyond the *specified chunk size* (page 762) but cannot undergo a split; e.g. if a chunk represents a single shard key value. See *Considerations for Selecting Shard Keys* (page 771) for considerations for selecting a shard key.

Shard Key Indexes

On this page

- Example (page 763)

All sharded collections **must** have an index that starts with the *shard key*; i.e. the index can be an index on the shard key or a *compound index* where the shard key is a prefix of the index.

If you shard a collection without any documents and *without* such an index, the `shardCollection` command will create the index on the shard key. If the collection already has documents, you must create the index before using `shardCollection`.

Important: The index on the shard key **cannot** be a *multikey index* (page 525).

Example

A sharded collection named `people` has for its shard key the field `zipcode`. It currently has the index `{ zipcode: 1 }`. You can replace this index with a compound index `{ zipcode: 1, username: 1 }`, as follows:

1. Create an index on `{ zipcode: 1, username: 1 }`:

   ```
   db.people.createIndex( { zipcode: 1, username: 1 } );
   ```

2. When MongoDB finishes building the index, you can safely drop the existing index on `{ zipcode: 1 }`:

   ```
   db.people.dropIndex( { zipcode: 1 } );
   ```

Since the index on the shard key cannot be a multikey index, the index `{ zipcode: 1, username: 1 }` can only replace the index `{ zipcode: 1 }` if there are no array values for the `username` field.

If you drop the last valid index for the shard key, recover by recreating an index on just the shard key.

For restrictions on shard key indexes, see *limits-shard-keys*.

Sharded Cluster Metadata

Config servers (page 742) store the metadata for a sharded cluster. The metadata reflects state and organization of the sharded data sets and system. The metadata includes the list of chunks on every shard and the ranges that define the chunks. The `mongos` instances cache this data and use it to route read and write operations to shards.

Config servers store the metadata in the *Config Database* (page 823).

Important: Always back up the `config` database before doing any maintenance on the config server.

To access the `config` database, issue the following command from the `mongo` shell:

```
use config
```

In general, you should *never* edit the content of the `config` database directly. The `config` database contains the following collections:

- `changelog` (page 824)
- `chunks` (page 826)
- `collections` (page 826)
- `databases` (page 826)
- `lockpings` (page 827)
- `locks` (page 827)
- `mongos` (page 827)
- `settings` (page 828)
- `shards` (page 828)
- `version` (page 829)

For more information on these collections and their role in sharded clusters, see *Config Database* (page 823). See *Read and Write Operations on Config Servers* (page 743) for more information about reads and updates to the metadata.

13.3 Sharded Cluster Tutorials

The following tutorials provide instructions for administering *sharded clusters*. For a higher-level overview, see *Sharding* (page 733).

Sharded Cluster Deployment Tutorials **(page 765)** Instructions for deploying sharded clusters, adding shards, selecting shard keys, and the initial configuration of sharded clusters.

> ***Deploy a Sharded Cluster*** **(page 765)** Set up a sharded cluster by creating the needed data directories, starting the required MongoDB instances, and configuring the cluster settings.
>
> ***Considerations for Selecting Shard Keys*** **(page 771)** Choose the field that MongoDB uses to parse a collection's documents for distribution over the cluster's shards. Each shard holds documents with values within a certain range.
>
> ***Shard a Collection Using a Hashed Shard Key*** **(page 773)** Shard a collection based on hashes of a field's values in order to ensure even distribution over the collection's shards.
>
> ***Add Shards to a Cluster*** **(page 773)** Add a shard to add capacity to a sharded cluster.
>
> Continue reading from *Sharded Cluster Deployment Tutorials* (page 765) for additional tutorials.

Sharded Cluster Maintenance Tutorials **(page 789)** Procedures and tasks for common operations on active sharded clusters.

View Cluster Configuration **(page 789)** View status information about the cluster's databases, shards, and chunks.

Remove Shards from an Existing Sharded Cluster **(page 805)** Migrate a single shard's data and remove the shard.

Manage Shard Tags **(page 816)** Use tags to associate specific ranges of shard key values with specific shards.

Continue reading from *Sharded Cluster Maintenance Tutorials* (page 789) for additional tutorials.

Sharded Cluster Data Management **(page 807)** Practices that address common issues in managing large sharded data sets.

Troubleshoot Sharded Clusters **(page 821)** Presents solutions to common issues and concerns relevant to the administration and use of sharded clusters. Refer to *FAQ: MongoDB Diagnostics* (page 856) for general diagnostic information.

13.3.1 Sharded Cluster Deployment Tutorials

The following tutorials provide information on deploying sharded clusters.

Deploy a Sharded Cluster **(page 765)** Set up a sharded cluster by creating the needed data directories, starting the required MongoDB instances, and configuring the cluster settings.

Considerations for Selecting Shard Keys **(page 771)** Choose the field that MongoDB uses to parse a collection's documents for distribution over the cluster's shards. Each shard holds documents with values within a certain range.

Shard a Collection Using a Hashed Shard Key **(page 773)** Shard a collection based on hashes of a field's values in order to ensure even distribution over the collection's shards.

Add Shards to a Cluster **(page 773)** Add a shard to add capacity to a sharded cluster.

Convert a Replica Set to a Sharded Cluster **(page 775)** Convert a replica set to a sharded cluster in which each shard is its own replica set.

Upgrade Config Servers to Replica Set **(page 780)** Convert three mirrored config servers to a replica set. The procedure does not require any downtime. Requires MongoDB version 3.2.4 or greater.

Upgrade Config Servers to Replica Set (Downtime) **(page 784)** Convert three mirrored config servers to a replica set. The procedure requires downtime.

Convert Sharded Cluster to Replica Set **(page 788)** Replace your sharded cluster with a single replica set.

Deploy a Sharded Cluster

On this page

- Considerations (page 766)
- Deploy the Config Server Replica Set (page 766)
- Start the `mongos` Instances (page 767)
- Add Shards to the Cluster (page 767)
- Enable Sharding for a Database (page 768)
- Shard a Collection (page 768)
- Using 3 Mirrored Config Servers (Deprecated) (page 769)

Changed in version 3.2.

Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

The following tutorial deploys a new sharded cluster for MongoDB 3.2. To deploy a sharded cluster for earlier versions of MongoDB, refer to the corresponding version of the MongoDB Manual.

Considerations

Host Identifier

> **Warning:** Sharding and "localhost" Addresses
> If you use either "localhost" or `127.0.0.1` as the hostname portion of any host identifier, for example as the `host` argument to `addShard` or the value to the *`--configdb`* run time option, then you must use "localhost" or `127.0.0.1` for *all* host settings for any MongoDB instances in the cluster. If you mix localhost addresses and remote host address, MongoDB will error.

Connectivity All members of a sharded cluster must be able to connect to *all* other members of a sharded cluster, including all shards and all config servers. Ensure that the network and security systems, including all interfaces and firewalls, allow these connections.

Deploy the Config Server Replica Set

Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

The following restrictions apply to a replica set configuration when used for config servers:

- Must have zero *arbiters* (page 635).
- Must have no *delayed members* (page 634).
- Must build indexes (i.e. no member should have `buildIndexes` setting set to false).

The config servers store the sharded cluster's metadata. The following steps deploy a three member replica set for the config servers.

1. Start all the config servers with both the *`--configsvr`* and *`--replSet <name>`* options:

   ```
   mongod --configsvr --replSet configReplSet --port <port> --dbpath <path>
   ```

   Or if using a `configuration file`, include the `sharding.clusterRole` and `replication.replSetName` setting:

   ```
   sharding:
      clusterRole: configsvr
   replication:
      replSetName: configReplSet
   net:
      port: <port>
   storage:
      dbpath: <path>
   ```

   For additional options, see `https://docs.mongodb.org/manual/reference/program/mongod` or `https://docs.mongodb.org/manual/reference/configuration-options`.

2. Connect a `mongo` shell to one of the config servers and run `rs.initiate()` to initiate the replica set.

```
rs.initiate( {
   _id: "configReplSet",
   configsvr: true,
   members: [
      { _id: 0, host: "<host1>:<port1>" },
      { _id: 1, host: "<host2>:<port2>" },
      { _id: 2, host: "<host3>:<port3>" }
   ]
} )
```

To use the deprecated mirrored config server deployment topology, see *Start 3 Mirrored Config Servers (Deprecated)* (page 769).

Start the `mongos` Instances

The `mongos` instances are lightweight and do not require data directories. You can run a `mongos` instance on a system that runs other cluster components, such as on an application server or a server running a `mongod` process. By default, a `mongos` instance runs on port `27017`.

When you start the `mongos` instance, specify the config servers, using either the `sharding.configDB` setting in the configuration file or the *`--configdb`* command line option.

Note: All config servers must be running and available when you first initiate a *sharded cluster*.

---

1. Start one or more `mongos` instances. For *`--configdb`*, or `sharding.configDB`, specify the config server replica set name followed by a slash `https://docs.mongodb.org/manual/` and at least one of the config server hostnames and ports:

   ```
   mongos --configdb configReplSet/<cfgsvr1:port1>,<cfgsvr2:port2>,<cfgsvr3:port3>
   ```

If using the deprecated mirrored config server deployment topology, see *Start the mongos Instances (Deprecated)* (page 770).

Add Shards to the Cluster

A *shard* can be a standalone `mongod` or a *replica set*. In a production environment, each shard should be a replica set. Use the procedure in *Deploy a Replica Set* (page 667) to deploy replica sets for each shard.

1. From a `mongo` shell, connect to the `mongos` instance. Issue a command using the following syntax:

   ```
   mongo --host <hostname of machine running mongos> --port <port mongos listens on>
   ```

   For example, if a `mongos` is accessible at `mongos0.example.net` on port `27017`, issue the following command:

   ```
   mongo --host mongos0.example.net --port 27017
   ```

2. Add each shard to the cluster using the `sh.addShard()` method, as shown in the examples below. Issue `sh.addShard()` separately for each shard. If the shard is a replica set, specify the name of the replica set and specify a member of the set. In production deployments, all shards should be replica sets.

   Optional

You can instead use the `addShard` database command, which lets you specify a name and maximum size for the shard. If you do not specify these, MongoDB automatically assigns a name and maximum size. To use the database command, see `addShard`.

The following are examples of adding a shard with `sh.addShard()`:

- To add a shard for a replica set named `rs1` with a member running on port `27017` on `mongodb0.example.net`, issue the following command:

  ```
  sh.addShard( "rs1/mongodb0.example.net:27017" )
  ```

- To add a shard for a standalone `mongod` on port `27017` of `mongodb0.example.net`, issue the following command:

  ```
  sh.addShard( "mongodb0.example.net:27017" )
  ```

Note: It might take some time for *chunks* to migrate to the new shard.

Enable Sharding for a Database

Before you can shard a collection, you must enable sharding for the collection's database. Enabling sharding for a database does not redistribute data but make it possible to shard the collections in that database.

Once you enable sharding for a database, MongoDB assigns a *primary shard* for that database where MongoDB stores all data before sharding begins.

1. From a `mongo` shell, connect to the `mongos` instance. Issue a command using the following syntax:

   ```
   mongo --host <hostname of machine running mongos> --port <port mongos listens on>
   ```

2. Issue the `sh.enableSharding()` method, specifying the name of the database for which to enable sharding. Use the following syntax:

   ```
   sh.enableSharding("<database>")
   ```

Optionally, you can enable sharding for a database using the `enableSharding` command, which uses the following syntax:

```
db.runCommand( { enableSharding: <database> } )
```

Shard a Collection

You shard on a per-collection basis.

1. Determine what you will use for the *shard key*. Your selection of the shard key affects the efficiency of sharding. See the selection considerations listed in the *Considerations for Selecting Shard Key* (page 771).
2. If the collection already contains data you must create an index on the *shard key* using `createIndex()`. If the collection is empty then MongoDB will create the index as part of the `sh.shardCollection()` step.
3. Shard a collection by issuing the `sh.shardCollection()` method in the `mongo` shell. The method uses the following syntax:

   ```
   sh.shardCollection("<database>.<collection>", shard-key-pattern)
   ```

Replace the <database>.<collection> string with the full namespace of your database, which consists of the name of your database, a dot (e.g. .), and the full name of the collection. The shard-key-pattern represents your shard key, which you specify in the same form as you would an index key pattern.

Example

The following sequence of commands shards four collections:

```
sh.shardCollection("records.people", { "zipcode": 1, "name": 1 } )
sh.shardCollection("people.addresses", { "state": 1, "_id": 1 } )
sh.shardCollection("assets.chairs", { "type": 1, "_id": 1 } )
sh.shardCollection("events.alerts", { "_id": "hashed" } )
```

In order, these operations shard:

(a) The people collection in the records database using the shard key { "zipcode": 1, "name": 1 }.

This shard key distributes documents by the value of the zipcode field. If a number of documents have the same value for this field, then that *chunk* will be *splittable* (page 772) by the values of the name field.

(b) The addresses collection in the people database using the shard key { "state": 1, "_id": 1 }.

This shard key distributes documents by the value of the state field. If a number of documents have the same value for this field, then that *chunk* will be *splittable* (page 772) by the values of the _id field.

(c) The chairs collection in the assets database using the shard key { "type": 1, "_id": 1 }.

This shard key distributes documents by the value of the type field. If a number of documents have the same value for this field, then that *chunk* will be *splittable* (page 772) by the values of the _id field.

(d) The alerts collection in the events database using the shard key { "_id": "hashed" }.

This shard key distributes documents by a hash of the value of the _id field. MongoDB computes the hash of the _id field for the *hashed index* (page 564), which should provide an even distribution of documents across a cluster.

Using 3 Mirrored Config Servers (Deprecated)

Start 3 Mirrored Config Servers (Deprecated) Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored mongod instances for config servers.

In production deployments, if using mirrored config servers, you must deploy exactly three config server instances, each running on different servers to assure good uptime and data safety. In test environments, you can run all three instances on a single server.

Important: All members of a sharded cluster must be able to connect to *all* other members of a sharded cluster, including all shards and all config servers. Ensure that the network and security systems including all interfaces and firewalls, allow these connections.

---

1. Create data directories for each of the three config server instances. By default, a config server stores its data files in the */data/configdb* directory. You can choose a different location. To create a data directory, issue a command similar to the following:

```
mkdir /data/configdb
```

2. Start the three config server instances. Start each by issuing a command using the following syntax:

   ```
   mongod --configsvr --dbpath <path> --port <port>
   ```

   The default port for config servers is `27019`. You can specify a different port. The following example starts a config server using the default port and default data directory:

   ```
   mongod --configsvr --dbpath /data/configdb --port 27019
   ```

   For additional command options, see `https://docs.mongodb.org/manual/reference/program/mongod` or `https://docs.mongodb.org/manual/reference/configuration-options`.

   Note: All config servers must be running and available when you first initiate a *sharded cluster*.

Start the `mongos` Instances (Deprecated) Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

If using 3 mirrored config servers, when you start the `mongos` instance, specify the hostnames of the three config servers, either in the configuration file or as command line parameters.

Tip

To avoid downtime, give each config server a logical DNS name (unrelated to the server's physical or virtual hostname). Without logical DNS names, moving or renaming a config server requires shutting down every `mongod` and `mongos` instance in the sharded cluster.

To start a `mongos` instance, issue a command using the following syntax:

```
mongos --configdb <config server hostnames>
```

For example, to start a `mongos` that connects to config server instance running on the following hosts and on the default ports:

- `cfg0.example.net`
- `cfg1.example.net`
- `cfg2.example.net`

You would issue the following command:

```
mongos --configdb cfg0.example.net:27019,cfg1.example.net:27019,cfg2.example.net:27019
```

Each `mongos` in a sharded cluster must use the same `configDB` string, with identical host names listed in identical order.

If you start a `mongos` instance with a string that *does not* exactly match the string used by the other `mongos` instances in the cluster, the `mongos` instance returns a *Config Database String Error* (page 821) error and refuses to start.

To add shards, enable sharding and shard a collection, see *Add Shards to the Cluster* (page 767), *Enable Sharding for a Database* (page 768), and *Shard a Collection* (page 768).

Considerations for Selecting Shard Keys

Choosing a Shard Key

For many collections there may be no single, naturally occurring key that possesses all the qualities of a good shard key. The following strategies may help construct a useful shard key from existing data:

1. Compute a more ideal shard key in your application layer, and store this in all of your documents, potentially in the `_id` field.
2. Use a compound shard key that uses two or three values from all documents that provide the right mix of cardinality with scalable write operations and query isolation.
3. Determine that the impact of using a less than ideal shard key is insignificant in your use case, given:
   - limited write volume,
   - expected data size, or
   - application query patterns.
4. Use a *hashed shard key*. Choose a field that has high cardinality and create a *hashed index* (page 564) on that field. MongoDB uses these hashed index values as shard key values, which ensures an even distribution of documents across the shards.

Tip

MongoDB automatically computes the hashes when resolving queries using hashed indexes. Applications do **not** need to compute hashes.

Considerations for Selecting Shard Key

Choosing the correct shard key can have a great impact on the performance, capability, and functioning of your database and cluster. Appropriate shard key choice depends on the schema of your data and the way that your applications query and write data.

Create a Shard Key that is Easily Divisible An easily divisible shard key makes it easy for MongoDB to distribute content among the shards. Shard keys that have a limited number of possible values can result in chunks that are "unsplittable".

For instance, if a chunk represents a single shard key value, then MongoDB cannot split the chunk even when the chunk exceeds the size at which *splits* (page 762) occur.

See also:

Cardinality (page 772)

Create a Shard Key that has High Degree of Randomness A shard key with high degree of randomness prevents any single shard from becoming a bottleneck and will distribute write operations among the cluster.

See also:

Write Scaling (page 749)

Create a Shard Key that Targets a Single Shard A shard key that targets a single shard makes it possible for the **mongos** program to return most query operations directly from a single *specific* **mongod** instance. Your shard key should be the primary field used by your queries. Fields with a high degree of "randomness" make it difficult to target operations to specific shards.

See also:

Query Isolation (page 749)

Shard Using a Compound Shard Key The challenge when selecting a shard key is that there is not always an obvious choice. Often, an existing field in your collection may not be the optimal key. In those situations, computing a special purpose shard key into an additional field or using a compound shard key may help produce one that is more ideal.

Cardinality Cardinality in the context of MongoDB, refers to the ability of the system to *partition* data into *chunks*. For example, consider a collection of data such as an "address book" that stores address records:

- Consider the use of a `state` field as a shard key:

  The state key's value holds the US state for a given address document. This field has a *low cardinality* as all documents that have the same value in the `state` field *must* reside on the same shard, even if a particular state's chunk exceeds the maximum chunk size.

  Since there are a limited number of possible values for the `state` field, MongoDB may distribute data unevenly among a small number of fixed chunks. This may have a number of effects:

  - If MongoDB cannot split a chunk because all of its documents have the same shard key, migrations involving these un-splittable chunks will take longer than other migrations, and it will be more difficult for your data to stay balanced.
  - If you have a fixed maximum number of chunks, you will never be able to use more than that number of shards for this collection.

- Consider the use of a `zipcode` field as a shard key:

  While this field has a large number of possible values, and thus has potentially higher cardinality, it's possible that a large number of users could have the same value for the shard key, which would make this chunk of users un-splittable.

  In these cases, cardinality depends on the data. If your address book stores records for a geographically distributed contact list (e.g. "Dry cleaning businesses in America,") then a value like zipcode would be sufficient. However, if your address book is more geographically concentrated (e.g "ice cream stores in Boston Massachusetts,") then you may have a much lower cardinality.

- Consider the use of a `phone-number` field as a shard key:

  Phone number has a *high cardinality,* because users will generally have a unique value for this field, MongoDB will be able to split as many chunks as needed.

While "high cardinality," is necessary for ensuring an even distribution of data, having a high cardinality does not guarantee sufficient *query isolation* (page 749) or appropriate *write scaling* (page 749).

If you choose a shard key with low cardinality, some chunks may grow too large for MongoDB to migrate. See *Jumbo Chunks* (page 762) for more information.

Shard Key Selection Strategy

When selecting a shard key, it is difficult to balance the qualities of an ideal shard key, which sometimes dictate opposing strategies. For instance, it's difficult to produce a key that has both a high degree randomness for even data

distribution and a shard key that allows your application to target specific shards. For some workloads, it's more important to have an even data distribution, and for others targeted queries are essential.

Therefore, the selection of a shard key is about balancing both your data and the performance characteristics caused by different possible data distributions and system workloads.

Shard a Collection Using a Hashed Shard Key

On this page

- Shard the Collection (page 773)
- Specify the Initial Number of Chunks (page 773)

New in version 2.4.

Hashed shard keys (page 748) use a *hashed index* (page 564) of a field as the *shard key* to partition data across your sharded cluster.

For suggestions on choosing the right field as your hashed shard key, see *Hashed Shard Keys* (page 748). For limitations on hashed indexes, see *Hashed Indexes* (page 564).

Note: If chunk migrations are in progress while creating a hashed shard key collection, the initial chunk distribution may be uneven until the balancer automatically balances the collection.

Shard the Collection

To shard a collection using a hashed shard key, use an operation in the `mongo` that resembles the following:

```
sh.shardCollection( "records.active", { a: "hashed" } )
```

This operation shards the `active` collection in the `records` database, using a hash of the `a` field as the shard key.

Specify the Initial Number of Chunks

If you shard an empty collection using a hashed shard key, MongoDB automatically creates and migrates empty chunks so that each shard has two chunks. To control how many chunks MongoDB creates when sharding the collection, use `shardCollection` with the `numInitialChunks` parameter.

Important: MongoDB 2.4 adds support for hashed shard keys. After sharding a collection with a hashed shard key, you must use the MongoDB 2.4 or higher `mongos` and `mongod` instances in your sharded cluster.

Warning: MongoDB `hashed` indexes truncate floating point numbers to 64-bit integers before hashing. For example, a `hashed` index would store the same value for a field that held a value of `2.3`, `2.2`, and `2.9`. To prevent collisions, do not use a `hashed` index for floating point numbers that cannot be reliably converted to 64-bit integers (and then back to floating point). MongoDB `hashed` indexes do not support floating point values larger than 2^{53}.

Add Shards to a Cluster

On this page

- Considerations (page 774)
- Add a Shard to a Cluster (page 774)

You add shards to a *sharded cluster* after you create the cluster or any time that you need to add capacity to the cluster. If you have not created a sharded cluster, see *Deploy a Sharded Cluster* (page 765).

In production environments, all shards should be *replica sets*.

Considerations

Balancing When you add a shard to a sharded cluster, you affect the balance of *chunks* among the shards of a cluster for all existing sharded collections. The balancer will begin migrating chunks so that the cluster will achieve balance. See *Sharded Collection Balancing* (page 758) for more information.

Changed in version 2.6: Chunk migrations can have an impact on disk space. Starting in MongoDB 2.6, the source shard automatically archives the migrated documents by default. For details, see *moveChunk directory* (page 761).

Capacity Planning When adding a shard to a cluster, always ensure that the cluster has enough capacity to support the migration required for balancing the cluster without affecting legitimate production traffic.

Add a Shard to a Cluster

You interact with a sharded cluster by connecting to a `mongos` instance.

1. From a `mongo` shell, connect to the `mongos` instance. For example, if a `mongos` is accessible at `mongos0.example.net` on port `27017`, issue the following command:

   ```
   mongo --host mongos0.example.net --port 27017
   ```

2. Add a shard to the cluster using the `sh.addShard()` method, as shown in the examples below. Issue `sh.addShard()` separately for each shard. If the shard is a replica set, specify the name of the replica set and specify a member of the set. In production deployments, all shards should be replica sets.

   Optional

   You can instead use the `addShard` database command, which lets you specify a name and maximum size for the shard. If you do not specify these, MongoDB automatically assigns a name and maximum size. To use the database command, see `addShard`.

   The following are examples of adding a shard with `sh.addShard()`:

   - To add a shard for a replica set named `rs1` with a member running on port `27017` on `mongodb0.example.net`, issue the following command:

     ```
     sh.addShard( "rs1/mongodb0.example.net:27017" )
     ```

   - To add a shard for a standalone `mongod` on port `27017` of `mongodb0.example.net`, issue the following command:

     ```
     sh.addShard( "mongodb0.example.net:27017" )
     ```

Note: It might take some time for *chunks* to migrate to the new shard.

Convert a Replica Set to a Sharded Cluster

On this page

- Overview (page 775)
- Prerequisites (page 775)
- Procedures (page 775)

Overview

This tutorial converts a single three-member replica set to a sharded cluster with two shards. Each shard is an independent three-member replica set. This tutorial is specific to MongoDB 3.2. For other versions of MongoDB, refer to the corresponding version of the MongoDB Manual.

The procedure is as follows:

1. Create the initial three-member replica set and insert data into a collection. See *Set Up Initial Replica Set* (page 775).
2. Start the config servers and a `mongos`. See *Deploy Config Server Replica Set and mongos* (page 776).
3. Add the initial replica set as a shard. See *Add Initial Replica Set as a Shard* (page 777).
4. Create a second shard and add to the cluster. See *Add Second Shard* (page 777).
5. Shard the desired collection. See *Shard a Collection* (page 778).

Prerequisites

This tutorial uses a total of ten servers: one server for the `mongos` and three servers each for the first *replica set*, the second replica set, and the *config server replica set* (page 742).

Each server must have a resolvable domain, hostname, or IP address within your system.

The tutorial uses the default data directories (e.g. `/data/db` and `/data/configdb`). Create the appropriate directories with appropriate permissions. To use different paths, see `https://docs.mongodb.org/manual/reference/configuration-options`.

The tutorial uses the `default ports` (e.g. `27017` and `27019`). To use different ports, see `https://docs.mongodb.org/manual/reference/configuration-options`.

Procedures

Set Up Initial Replica Set This procedure creates the initial three-member replica set `rs0`. The replica set members are on the following hosts: `mongodb0.example.net`, `mongodb1.example.net`, and `mongodb2.example.net`.

Step 1: Start each member of the replica set with the appropriate options. For each member, start a `mongod`, specifying the replica set name through the `replSet` option. Include any other parameters specific to your deployment. For replication-specific parameters, see *cli-mongod-replica-set*.

```
mongod --replSet "rs0"
```

Repeat this step for the other two members of the `rs0` replica set.

Step 2: Connect a `mongo` shell to a replica set member. Connect a `mongo` shell to *one* member of the replica set (e.g. `mongodb0.example.net`)

```
mongo mongodb0.example.net
```

Step 3: Initiate the replica set. From the `mongo` shell, run `rs.initiate()` to initiate a replica set that consists of the current member.

```
rs.initiate()
```

Step 4: Add the remaining members to the replica set.

```
rs.add("mongodb1.example.net")
rs.add("mongodb2.example.net")
```

Step 5: Create and populate a new collection. The following step adds one million documents to the collection `test_collection` and can take several minutes depending on your system.

Issue the following operations on the primary of the replica set:

```
use test
var bulk = db.test_collection.initializeUnorderedBulkOp();
people = ["Marc", "Bill", "George", "Eliot", "Matt", "Trey", "Tracy", "Greg", "Steve", "Kristina", "
for(var i=0; i<1000000; i++){
   user_id = i;
   name = people[Math.floor(Math.random()*people.length)];
   number = Math.floor(Math.random()*10001);
   bulk.insert( { "user_id":user_id, "name":name, "number":number });
}
bulk.execute();
```

For more information on deploying a replica set, see *Deploy a Replica Set* (page 667).

Deploy Config Server Replica Set and `mongos` Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

This procedure deploys the three-member replica set for the *config servers* (page 742) and the `mongos`.

- The config servers use the following hosts: `mongodb7.example.net`, `mongodb8.example.net`, and `mongodb9.example.net`.
- The `mongos` uses `mongodb6.example.net`.

Step 1: Deploy the config servers as a three-member replica set. Start a config server on `mongodb7.example.net`, `mongodb8.example.net`, and `mongodb9.example.net`. Specify the same replica set name. The config servers use the default data directory `/data/configdb` and the default port `27019`.

```
mongod --configsvr --replSet configReplSet
```

To modify the default settings or to include additional options specific to your deployment, see `https://docs.mongodb.org/manual/reference/program/mongod` or `https://docs.mongodb.org/manual/reference/configuration-options`.

Connect a `mongo` shell to one of the config servers and run `rs.initiate()` to initiate the replica set.

```
rs.initiate( {
   _id: "configReplSet",
   configsvr: true,
   members: [
      { _id: 0, host: "mongodb07.example.net:27019" },
      { _id: 1, host: "mongodb08.example.net:27019" },
      { _id: 2, host: "mongodb09.example.net:27019" }
   ]
} )
```

Step 2: Start a `mongos` instance. On `mongodb6.example.net`, start the `mongos` specifying the config server replica set name followed by a slash `https://docs.mongodb.org/manual/` and at least one of the config server hostnames and ports.

This tutorial specifies a small *`--chunkSize`* of 1 MB to test sharding with the `test_collection` created earlier.

Note: In production environments, do **not** use a small chunkSize size.

```
mongos --configdb configReplSet/mongodb07.example.net:27019,mongodb08.example.net:27019,mongodb09.exa
```

Add Initial Replica Set as a Shard The following procedure adds the initial replica set `rs0` as a shard.

Step 1: Connect a `mongo` shell to the `mongos`.

```
mongo mongodb6.example.net:27017/admin
```

Step 2: Add the shard. Add a shard to the cluster with the `sh.addShard` method:

```
sh.addShard( "rs0/mongodb0.example.net:27017,mongodb1.example.net:27017,mongodb2.example.net:27017" )
```

Add Second Shard The following procedure deploys a new replica set `rs1` for the second shard and adds it to the cluster. The replica set members are on the following hosts: `mongodb3.example.net`, `mongodb4.example.net`, and `mongodb5.example.net`.

Step 1: Start each member of the replica set with the appropriate options. For each member, start a `mongod`, specifying the replica set name through the `replSet` option. Include any other parameters specific to your deployment. For replication-specific parameters, see *cli-mongod-replica-set*.

```
mongod --replSet "rs1"
```

Repeat this step for the other two members of the `rs1` replica set.

Step 2: Connect a mongo shell to a replica set member. Connect a `mongo` shell to *one* member of the replica set (e.g. `mongodb3.example.net`)

```
mongo mongodb3.example.net
```

Step 3: Initiate the replica set. From the `mongo` shell, run `rs.initiate()` to initiate a replica set that consists of the current member.

```
rs.initiate()
```

Step 4: Add the remaining members to the replica set. Add the remaining members with the `rs.add()` method.

```
rs.add("mongodb4.example.net")
rs.add("mongodb5.example.net")
```

Step 5: Connect a mongo shell to the mongos.

```
mongo mongodb6.example.net:27017/admin
```

Step 6: Add the shard. In a `mongo` shell connected to the `mongos`, add the shard to the cluster with the `sh.addShard()` method:

```
sh.addShard( "rs1/mongodb3.example.net:27017,mongodb4.example.net:27017,mongodb5.example.net:27017" )
```

Shard a Collection

Step 1: Connect a mongo shell to the mongos.

```
mongo mongodb6.example.net:27017/admin
```

Step 2: Enable sharding for a database. Before you can shard a collection, you must first enable sharding for the collection's database. Enabling sharding for a database does not redistribute data but makes it possible to shard the collections in that database.

The following operation enables sharding on the `test` database:

```
sh.enableSharding( "test" )
```

The operation returns the status of the operation:

```
{ "ok" : 1 }
```

Step 3: Determine the shard key. For the collection to shard, determine the shard key. The *shard key* (page 747) determines how MongoDB distributes the documents between shards. Good shard keys:

- have values that are evenly distributed among all documents,
- group documents that are often accessed at the same time into contiguous chunks, and
- allow for effective distribution of activity among shards.

Once you shard a collection with the specified shard key, you **cannot** change the shard key. For more information on shard keys, see *Shard Keys* (page 747) and *Considerations for Selecting Shard Keys* (page 771).

This procedure will use the `number` field as the shard key for `test_collection`.

Step 4: Create an index on the shard key. Before sharding a non-empty collection, create an *index on the shard key* (page 763).

```
use test
db.test_collection.createIndex( { number : 1 } )
```

Step 5: Shard the collection. In the `test` database, shard the `test_collection`, specifying `number` as the shard key.

```
use test
sh.shardCollection( "test.test_collection", { "number" : 1 } )
```

The method returns the status of the operation:

```
{ "collectionsharded" : "test.test_collection", "ok" : 1 }
```

The *balancer* (page 758) will redistribute chunks of documents when it next runs. As clients insert additional documents into this collection, the `mongos` will route the documents between the shards.

Step 6: Confirm the shard is balancing. To confirm balancing activity, run `db.stats()` or `db.printShardingStatus()` in the `test` database.

```
use test
db.stats()
db.printShardingStatus()
```

Example output of the `db.stats()`:

```
{
  "raw" : {
      "rs0/mongodb0.example.net:27017,mongodb1.example.net:27017,mongodb2.example.net:27017" : {
         "db" : "test",
         "collections" : 3,
         "objects" : 989316,
         "avgObjSize" : 111.99974123535857,
         "dataSize" : 110803136,
         "storageSize" : 174751744,
         "numExtents" : 14,
         "indexes" : 2,
         "indexSize" : 57370992,
         "fileSize" : 469762048,
         "ok" : 1
      },
      "rs1/mongodb3.example.net:27017,mongodb4.example.net:27017,mongodb5.example.net:27017" : {
         "db" : "test",
         "collections" : 3,
         "objects" : 14697,
         "avgObjSize" : 111.98258147921345,
         "dataSize" : 1645808,
         "storageSize" : 2809856,
         "numExtents" : 7,
         "indexes" : 2,
         "indexSize" : 1169168,
         "fileSize" : 67108864,
         "ok" : 1
      }
  },
  "objects" : 1004013,
```

```
  "avgObjSize" : 111,
  "dataSize" : 112448944,
  "storageSize" : 177561600,
  "numExtents" : 21,
  "indexes" : 4,
  "indexSize" : 58540160,
  "fileSize" : 536870912,
  "extentFreeList" : {
      "num" : 0,
      "totalSize" : 0
  },
  "ok" : 1
}
```

Example output of the `db.printShardingStatus()`:

```
--- Sharding Status ---
sharding version: {
   "_id" : 1,
   "minCompatibleVersion" : 5,
   "currentVersion" : 6,
   "clusterId" : ObjectId("5446970c04ad5132c271597c")
}
shards:
   {  "_id" : "rs0",  "host" : "rs0/mongodb0.example.net:27017,mongodb1.example.net:27017,mongodb2.ex
   {  "_id" : "rs1",  "host" : "rs1/mongodb3.example.net:27017,mongodb4.example.net:27017,mongodb5.ex
active mongoses:
   "3.2.0" : 2
balancer:
   Currently enabled:  yes
   Currently running:  no
Failed balancer rounds in last 5 attempts:  0
Migration Results for the last 24 hours:
   1 : Success
databases:
   {  "_id" : "test", "primary" : "rs0", "partitioned" : true }
      test.test_collection
            shard key: { "number" : 1 }
            unique: false
            balancing: true
            chunks:
               rs1    5
               rs0    186
            too many chunks to print, use verbose if you want to force print
```

Run these commands for a second time to demonstrate that *chunks* are migrating from `rs0` to `rs1`.

Upgrade Config Servers to Replica Set

On this page

- Prerequisites (page 781)
- Procedure (page 781)

Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). Using a replica set for the config servers improves consistency across the config servers, since MongoDB can take advantage

of the standard replica set read and write protocols for the config data. In addition, using a replica set for config servers allows a sharded cluster to have more than 3 config servers since a replica set can have up to 50 members. To deploy config servers as a relica set, the config servers must run the *WiredTiger storage engine* (page 595).

The following procedure upgrades three mirrored config servers to a *config server replica set* (page 743) without downtime. To use this procedure, all the sharded cluster binaries must be at least version 3.2.4.

During this procedure there will be a period of time where the config servers will be read-only. During this period, certain catalog operations will fail if attempted. Operations that will not be available include adding and dropping shards, creating and dropping databases, creating and dropping sharded collections, and migrating chunks (both manually and via the balancer process). Normal read and write operations to existing collections will not be affected.

See also:

Upgrade Config Servers to Replica Set (Downtime) (page 784)

Prerequisites

- All binaries in the sharded clusters must be at least version **3.2.4**. See *Upgrade a Sharded Cluster to 3.2* (page 902) for instructions to upgrade the sharded cluster.
- The existing config servers must be in sync.

Procedure

Note: The procedure refers to the first config server, second config server, and the third config server as listed in the `configDB` setting of the `mongos`. This means, that for the following example:

```
mongos --configdb confServer1:port1,confServer2:port2,confServer3:port3
```

- The first config server refers to `confServer1`.
- The second config server refers to `confServer2`.
- The third config server refers to `confServer3`.

---

1. **Disable the balancer** as described in *Disable the Balancer* (page 802).
2. Connect a `mongo` shell to the *first* config server listed in the `configDB` setting of the `mongos` and run `rs.initiate()` to initiate the single member replica set.

   ```
   rs.initiate( {
      _id: "csReplSet",
      configsvr: true,
      version: 1,
      members: [ { _id: 0, host: "<host>:<port>" } ]
   } )
   ```

   - `_id` (page 718) corresponds to the replica set name for the config servers.
   - `configsvr` (page 718) must be set to `true`.
   - `version` (page 718) must be set to `1`.
   - `members` (page 719) array contains a document that specifies:
     - `members._id` (page 719) which is a numeric identifier for the member.
     - `members.host` (page 719) which is a string corresponding to the config server's hostname and port.

3. Restart this config server as a single member replica set with:

   - the *--replSet* option set to the replica set name specified during the `rs.initiate()`,
   - the *--configsvrMode* option set to the legacy config server mode Sync Cluster Connection Config (`sccc`),
   - the *--configsvr* option,
   - the *--storageEngine* option set to the storage engine used by this config server. For this upgrade procedure, the existing config server can be using either MMAPv1 or WiredTiger, and
   - the *--port* option set to the same port as before restart, and
   - the *--dbpath* option set to the same path as before restart.

   Include additional options as specific to your deployment.

   Important: The config server must use the same port as before. [16]

   ```
   mongod --configsvr --replSet csReplSet --configsvrMode=sccc --storageEngine <storageEngine> --pc
   ```

   Or if using a `configuration file`, specify the:

   - `sharding.clusterRole`,
   - `sharding.configsvrMode`,
   - `replication.replSetName`,
   - `storage.dbPath`,
   - `storage.engine`, and
   - `net.port`.

   ```
   sharding:
      clusterRole: configsvr
      configsvrMode: sccc
   replication:
      replSetName: csReplSet
   net:
      port: <port>
   storage:
      dbPath: <path>
      engine: <storageEngine>
   ```

4. Start the new `mongod` instances to add to the replica set. These instances must use the *WiredTiger* (page 595) storage engine. Starting in 3.2, the default storage engine is WiredTiger for new `mongod` instances with new data paths.

   Important:

   - Do not add existing config servers to the replica set.
   - Use new dbpaths for the new instances.

   The number of new `mongod` instances to add depends on the config server currently in the single-member replica set:

[16] If before the restart, your config server did not explicitly specify the `--configsvr` option or the `--port` option, the restart with the `--configsvr` will result in a change of port.

To ensure that the port used by the config server does not change, include the *--port* option or `net.port` set to the same port as before the restart.

- If the config server is using MMAPv1, start 3 new `mongod` instances.
- If the config server is using WiredTiger, start 2 new `mongod` instances.

Note: The example in this procedure assumes that the existing config servers use MMAPv1.

For each new `mongod` instance to add, include the `--configsvr` and the `--replSet` options:

```
mongod --configsvr --replSet csReplSet --port <port> --dbpath <path>
```

Or if using a `configuration file`:

```
sharding:
   clusterRole: configsvr
replication:
   replSetName: csReplSet
net:
   port: <port>
storage:
   dbPath: <path>
```

5. Using the `mongo` shell connected to the replica set config server, add the new `mongod` instances as *non-voting* (page 646), *priority 0* (page 631) members:

   ```
   rs.add( { host: <host:port>, priority: 0, votes: 0 } )
   ```

6. Once all the new members have been added as *non-voting* (page 646), *priority 0* (page 631) members, ensure that the new nodes have completed the *initial sync* (page 658) and have reached `SECONDARY` (page 726) state. To check the state of the replica set members, run `rs.status()` in the `mongo` shell:

   ```
   rs.status()
   ```

7. Shut down one of the other non-replica set config servers; i.e. either the second and third config server listed in the `configDB` setting of the `mongos`. At this point the config servers will go read-only, meaning certain operations - such as creating and dropping databases and sharded collections - will not be available.

8. Reconfigure the replica set to allow all members to vote and have default priority of `1`.

   ```
   var cfg = rs.conf();

   cfg.members[0].priority = 1;
   cfg.members[1].priority = 1;
   cfg.members[2].priority = 1;
   cfg.members[3].priority = 1;
   cfg.members[0].votes = 1;
   cfg.members[1].votes = 1;
   cfg.members[2].votes = 1;
   cfg.members[3].votes = 1;

   rs.reconfig(cfg);
   ```

9. Step down the first config server, i.e. the server started with `--configsvrMode=sccc`.

   ```
   rs.stepDown(600)
   ```

10. Shut down the first config server.

11. Restart the first config server in config server replica set (`CSRS`) mode; i.e. restart **without** the `--configsvrMode=sccc` option:

```
mongod --configsvr --replSet csReplSet --storageEngine <storageEngine> --port <port> --dbpath <p
```

Or if using a `configuration file`, omit the `sharding.configsvrMode` setting:

```
sharding:
   clusterRole: configsvr
replication:
   replSetName: csReplSet
net:
   port: <port>
storage:
   dbPath: <path>
   engine: <storageEngine>
```

If the first config server uses the MMAPv1 storage engine, the member will transition to `"REMOVED"` state.

At this point the config server data will return to being writeable and all catalog operations - including creating and dropping databases and sharded collections - will once again be possible.

12. Restart `mongos` instances with updated `--configdb` or `sharding.configDB` setting.

    For the updated `--configdb` or `sharding.configDB` setting, specify the replica set name for the config servers and the members in the replica set.

    ```
    mongos --configdb csReplSet/<rsconfigsver1:port1>,<rsconfigsver2:port2>,<rsconfigsver3:port3>
    ```

13. Verify that the restarted `mongos` instances are aware of the protocol change. Connect a `mongo` shell to a `mongos` instance and check the `mongos` collection in the `config` database:

    ```
    use config
    db.mongos.find()
    ```

    The `ping` value for the `mongos` instances should indicate some time after the restart.

14. If the first config server uses the MMAPv1 storage engine, remove the member from the replica set. Connect a `mongo` shell to the current primary and use `rs.remove()`:

    Important: Only if the config server uses the MMAPv1 storage engine.

    ```
    rs.remove("<hostname>:<port>")
    ```

15. Shut down the remaining non-replica set config server.

16. **Re-enable the balancer** as described in *Enable the Balancer* (page 803).

Upgrade Config Servers to Replica Set (Downtime)

On this page

- Prerequisites (page 785)
- Procedure (page 785)

Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). Using a replica set for the config servers improves consistency across the config servers, since MongoDB can take advantage of the standard replica set read and write protocols for the config data. In addition, using a replica set for config servers allows a sharded cluster to have more than 3 config servers since a replica set can have up

to 50 members. To deploy config servers as a relica set, the config servers must run the *WiredTiger storage engine* (page 595).

The following procedure upgrades three mirrored config servers to a *config server replica set* (page 743).

Prerequisites

- All binaries in the sharded clusters must be at least version 3.2. See *Upgrade a Sharded Cluster to 3.2* (page 902) for instructions to upgrade the sharded cluster.
- The existing config servers must be in sync.

Procedure

Important: The procedure outlined in this tutorial requires downtime. If all the sharded cluster binaries are at least version 3.2.4, you can also convert the config servers to replica set without downtime. For details, see *Upgrade Config Servers to Replica Set* (page 780).

---

1. **Disable the balancer** as described in *Disable the Balancer* (page 802).
2. Connect a `mongo` shell to the *first* config server listed in the `configDB` setting of the `mongos` and run `rs.initiate()` to initiate the single member replica set.

   ```
   rs.initiate( {
      _id: "csReplSet",
      version: 1,
      configsvr: true,
      members: [ { _id: 0, host: "<host>:<port>" } ]
   } )
   ```

   - `_id` (page 718) corresponds to the replica set name for the config servers.
   - `version` (page 718) set to 1, corresponding to the initial version of the replica set configuration.
   - `configsvr` (page 718) must be set be `true`.
   - `members` (page 719) array contains a document that specifies:
     - `members._id` (page 719) which is a numeric identifier for the member.
     - `members.host` (page 719) which is a string corresponding to the config server's hostname and port.
3. Restart this config server as a single member replica set with:
   - the *--replSet* option set to the replica set name specified during the `rs.initiate()`,
   - the *--configsvrMode* option set to the legacy config server mode Sync Cluster Connection Config (`sccc`),
   - the *--configsvr* option, and
   - the *--storageEngine* option set to the storage engine used by this config server. For this upgrade procedure, the existing config server can be using either MMAPv1 or WiredTiger.

   Include additional options as specific to your deployment.

   ```
   mongod --configsvr --replSet csReplSet --configsvrMode=sccc --storageEngine <storageEngine> --pc
   ```

   Or if using a `configuration file`, specify the `replication.replSetName:`, `sharding.clusterRole`, `sharding.configsvrMode` and `net.port`.

```
sharding:
   clusterRole: configsvr
   configsvrMode: sccc
replication:
   replSetName: csReplSet
net:
   port: <port>
storage:
   dbPath: <path>
   engine: <storageEngine>
```

4. Start the new `mongod` instances to add to the replica set. These instances must use the *WiredTiger* (page 595) storage engine. Starting in 3.2, the default storage engine is WiredTiger for new `mongod` instances with new data paths.

   Important:

   - Do not add existing config servers to the replica set.
   - Use new dbpaths for the new instances.

   The number of new `mongod` instances to add depends on the config server currently in the single-member replica set:

   - If the config server is using MMAPv1, start 3 new `mongod` instances.
   - If the config server is using WiredTiger, start 2 new `mongod` instances.

   Note: The example in this procedure assumes that the existing config servers use MMAPv1.

   For each new `mongod` instance to add, include the `--configsvr` and the `--replSet` options:

   ```
   mongod --configsvr --replSet csReplSet --port <port> --dbpath <path>
   ```

   Or if using a `configuration file`:

   ```
   sharding:
      clusterRole: configsvr
   replication:
      replSetName: csReplSet
   net:
      port: <port>
   storage:
      dbPath: <path>
   ```

5. Using the `mongo` shell connected to the replica set config server, add the new `mongod` instances as *non-voting* (page 646), *priority 0* (page 631) members:

   ```
   rs.add( { host: <host:port>, priority: 0, votes: 0 } )
   ```

6. Once all the new members have been added as *non-voting* (page 646), *priority 0* (page 631) members, ensure that the new nodes have completed the *initial sync* (page 658) and have reached `SECONDARY` (page 726) state. To check the state of the replica set members, run `rs.status()` in the `mongo` shell:

   ```
   rs.status()
   ```

7. Shut down one of the other non-replica set config servers; i.e. either the second and third config server listed in the `configDB` setting of the `mongos`.

8. Reconfigure the replica set to allow all members to vote and have default priority of `1`.

```
var cfg = rs.conf();

cfg.members[0].priority = 1;
cfg.members[1].priority = 1;
cfg.members[2].priority = 1;
cfg.members[3].priority = 1;
cfg.members[0].votes = 1;
cfg.members[1].votes = 1;
cfg.members[2].votes = 1;
cfg.members[3].votes = 1;

rs.reconfig(cfg);
```

9. Step down the first config server, i.e. the server started with `--configsvrMode=sccc`.

   ```
   rs.stepDown()
   ```

10. Shut down the following members of the sharded cluster:

    - The `mongos` instances.
    - The shards.
    - The remaining non-replica set config servers.

11. Shut down the first config server.

    If the first config server uses the MMAPv1 storage engine, remove the member from the replica set. Connect a `mongo` shell to the current primary and use `rs.remove()`:

    Important: If the first config server uses the WiredTiger storage engine, do not remove.

    ```
    rs.remove("<hostname>:<port>")
    ```

12. If the first config server uses *WiredTiger* (page 595), restart the first config server in config server replica set (`CSRS`) mode; i.e. restart **without** the `--configsvrMode=sccc` option:

    Important: If the first config server uses the MMAPv1 storage engine, do not restart.

    ```
    mongod --configsvr --replSet csReplSet --storageEngine wiredTiger --port <port> --dbpath <path>
    ```

    Or if using a `configuration file`, omit the `sharding.configsvrMode` setting:

    ```
    sharding:
       clusterRole: configsvr
    replication:
       replSetName: csReplSet
    net:
       port: <port>
    storage:
       dbPath: <path>
       engine: <storageEngine>
    ```

13. Restart the shards.

14. Restart `mongos` instances with updated *`--configdb`* or `configDB` setting.

    For the updated *`--configdb`* or `configDB` setting, specify the replica set name for the config servers and the members in the replica set.

```
mongos --configdb csReplSet/<rsconfigsver1:port1>,<rsconfigsver2:port2>,<rsconfigsver3:port3>
```

15. **Re-enable the balancer** as described in *Enable the Balancer* (page 803).

Convert Sharded Cluster to Replica Set

On this page

- Convert a Cluster with a Single Shard into a Replica Set (page 788)
- Convert a Sharded Cluster into a Replica Set (page 788)

This tutorial describes the process for converting a *sharded cluster* to a non-sharded *replica set*. To convert a replica set into a sharded cluster *Convert a Replica Set to a Sharded Cluster* (page 775). See the *Sharding* (page 733) documentation for more information on sharded clusters.

Convert a Cluster with a Single Shard into a Replica Set

In the case of a *sharded cluster* with only one shard, that shard contains the full data set. Use the following procedure to convert that cluster into a non-sharded *replica set*:

1. Reconfigure the application to connect to the primary member of the replica set hosting the single shard that system will be the new replica set.
2. Optionally remove the `--shardsrv` option, if your `mongod` started with this option.

   Tip

   Changing the `--shardsrv` option will change the port that `mongod` listens for incoming connections on.

The single-shard cluster is now a non-sharded *replica set* that will accept read and write operations on the data set.

You may now decommission the remaining sharding infrastructure.

Convert a Sharded Cluster into a Replica Set

Use the following procedure to transition from a *sharded cluster* with more than one shard to an entirely new *replica set*.

1. With the *sharded cluster* running, *deploy a new replica set* (page 667) in addition to your sharded cluster. The replica set must have sufficient capacity to hold all of the data files from all of the current shards combined. Do not configure the application to connect to the new replica set until the data transfer is complete.
2. Stop all writes to the *sharded cluster*. You may reconfigure your application or stop all `mongos` instances. If you stop all `mongos` instances, the applications will not be able to read from the database. If you stop all `mongos` instances, start a temporary `mongos` instance on that applications cannot access for the data migration procedure.
3. Use *mongodump and mongorestore* (page 349) to migrate the data from the `mongos` instance to the new *replica set*.

   Note: Not all collections on all databases are necessarily sharded. Do not solely migrate the sharded collections. Ensure that all databases and all collections migrate correctly.

4. Reconfigure the application to use the non-sharded *replica set* instead of the `mongos` instance.

The application will now use the un-sharded *replica set* for reads and writes. You may now decommission the remaining unused sharded cluster infrastructure.

13.3.2 Sharded Cluster Maintenance Tutorials

The following tutorials provide information in maintaining sharded clusters.

View Cluster Configuration **(page 789)** View status information about the cluster's databases, shards, and chunks.

Replace a Config Server **(page 791)** Replace a config server in a config server replica set.

Migrate Config Servers with the Same Hostname **(page 792)** For a sharded cluster with three mirrored config servers, migrate a config server to a new system while keeping the same hostname. This procedure requires changing the DNS entry to point to the new system.

Migrate Config Servers with Different Hostnames **(page 793)** For a sharded cluster with three mirrored config servers, migrate a config server to a new system that uses a new hostname. If possible, avoid changing the hostname and instead use the *Migrate Config Servers with the Same Hostname* (page 792) procedure.

Migrate a Sharded Cluster to Different Hardware **(page 794)** Migrate a sharded cluster to a different hardware system, for example, when moving a pre-production environment to production.

Backup Cluster Metadata **(page 797)** Create a backup of a sharded cluster's metadata while keeping the cluster operational.

Configure Behavior of Balancer Process in Sharded Clusters **(page 798)** Manage the balancer's behavior by scheduling a balancing window, changing size settings, or requiring replication before migration.

Manage Sharded Cluster Balancer **(page 800)** View balancer status and manage balancer behavior.

Remove Shards from an Existing Sharded Cluster **(page 805)** Migrate a single shard's data and remove the shard.

View Cluster Configuration

On this page

- List Databases with Sharding Enabled (page 789)
- List Shards (page 790)
- View Cluster Details (page 790)

List Databases with Sharding Enabled

To list the databases that have sharding enabled, query the `databases` collection in the *Config Database* (page 823). A database has sharding enabled if the value of the `partitioned` field is `true`. Connect to a `mongos` instance with a `mongo` shell, and run the following operation to get a full list of databases with sharding enabled:

```
use config
db.databases.find( { "partitioned": true } )
```

Example

You can use the following sequence of commands when to return a list of all databases in the cluster:

```
use config
db.databases.find()
```

If this returns the following result set:

```
{ "_id" : "admin", "partitioned" : false, "primary" : "config" }
{ "_id" : "animals", "partitioned" : true, "primary" : "m0.example.net:30001" }
{ "_id" : "farms", "partitioned" : false, "primary" : "m1.example2.net:27017" }
```

Then sharding is only enabled for the `animals` database.

List Shards

To list the current set of configured shards, use the `listShards` command, as follows:

```
use admin
db.runCommand( { listShards : 1 } )
```

View Cluster Details

To view cluster details, issue `db.printShardingStatus()` or `sh.status()`. Both methods return the same output.

Example

In the following example output from `sh.status()`

- `sharding version` displays the version number of the shard metadata.
- `shards` displays a list of the `mongod` instances used as shards in the cluster.
- `databases` displays all databases in the cluster, including database that do not have sharding enabled.
- The `chunks` information for the `foo` database displays how many chunks are on each shard and displays the range of each chunk.

```
--- Sharding Status ---
  sharding version: { "_id" : 1, "version" : 3 }
  shards:
    {  "_id" : "shard0000",  "host" : "m0.example.net:30001" }
    {  "_id" : "shard0001",  "host" : "m3.example2.net:50000" }
  databases:
    {  "_id" : "admin",  "partitioned" : false,  "primary" : "config" }
    {  "_id" : "contacts",  "partitioned" : true,  "primary" : "shard0000" }
        foo.contacts
            shard key: { "zip" : 1 }
            chunks:
                shard0001    2
                shard0002    3
                shard0000    2
            { "zip" : { "$minKey" : 1 } } -->> { "zip" : "56000" } on : shard0001 { "t" : 2, "i" : 0
            { "zip" : 56000 } -->> { "zip" : "56800" } on : shard0002 { "t" : 3, "i" : 4 }
            { "zip" : 56800 } -->> { "zip" : "57088" } on : shard0002 { "t" : 4, "i" : 2 }
            { "zip" : 57088 } -->> { "zip" : "57500" } on : shard0002 { "t" : 4, "i" : 3 }
            { "zip" : 57500 } -->> { "zip" : "58140" } on : shard0001 { "t" : 4, "i" : 0 }
            { "zip" : 58140 } -->> { "zip" : "59000" } on : shard0000 { "t" : 4, "i" : 1 }
```

```
        { "zip" : 59000 } -->> { "zip" : { "$maxKey" : 1 } } on : shard0000 { "t" : 3, "i" : 3 }
{  "_id" : "test",  "partitioned" : false,  "primary" : "shard0000" }
```

Replace a Config Server

On this page

- Overview (page 791)
- Considerations (page 791)
- Procedure (page 791)

Changed in version 3.2: Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

For replacing config servers deployed as three mirrored `mongod` instances, see *Migrate Config Servers with the Same Hostname* (page 792) and *Migrate Config Servers with Different Hostnames* (page 793).

Overview

If the config server replica set becomes read only, i.e. does not have a primary, the sharded cluster cannot support operations that change the cluster metadata, such as chunk splits and migrations. Although no chunks can be split or migrated, applications will be able to write data to the sharded cluster.

If one of the config servers is unavailable or inoperable, repair or replace it as soon as possible. The following procedure replaces a member of a *config server replica set* (page 742) with a new member.

The tutorial is specific to MongoDB 3.2. For earlier versions of MongoDB, refer to the corresponding version of the MongoDB Manual.

Considerations

The following restrictions apply to a replica set configuration when used for config servers:

- Must have zero *arbiters* (page 635).
- Must have no *delayed members* (page 634).
- Must build indexes (i.e. no member should have `buildIndexes` setting set to false).

Procedure

Step 1: Start the replacement config server. Start a `mongod` instance, specifying both the `--configsvr` and `--replSet` options.

```
mongod --configsvr --replSet <replicaSetName>
```

Step 2: Add the new config server to the replica set. Connect a `mongo` shell to the primary of the config server replica set and use `rs.add()` to add the new member.

```
rs.add("<hostnameNew>:<portNew>")
```

The initial sync process copies all the data from one member of the config server replica set to the new member without restarting.

`mongos` instances automatically recognize the change in the config server replica set members without restarting.

Step 3: Shut down the member to replace. If replacing the primary member, step down the primary first before shutting down.

Step 4: Remove the member to replace from the config server replica set. Upon completion of initial sync of the replacement config server, from a `mongo` shell connected to the primary, use `rs.remove()` to remove the old member.

```
rs.remove("<hostnameOld>:<portOld>")
```

`mongos` instances automatically recognize the change in the config server replica set members without restarting.

Step 5: If necessary, update `mongos` configuration or DNS entry. With replica set config servers, the `mongos` instances specify in the *`--configdb`* or `sharding.configDB` setting the config server replica set name and at least one of the replica set members.

As such, if the `mongos` instance does not specify the removed replica set member in the *`--configdb`* or `sharding.configDB` setting, no further action is necessary.

If, however, a `mongos` instance specified the removed member in the `--configdb` or `configDB` setting, either:

- Update the setting for the next time you restart the `mongos`, or
- Modify the DNS entry that points to the system that provided the old config server, so that the *same* hostname points to the new config server.

Migrate Config Servers with the Same Hostname

Important: This procedure applies to migrating config servers when using three mirrored `mongod` instances as config servers.

Starting in MongoDB 3.2, config servers can be deployed as *replica set* (page 623). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

For replacing config servers deployed as members of a replica set, see *Replace a Config Server* (page 791).

For a *sharded cluster* (page 739) that uses 3 mirrored config servers, use the following procedure migrates a *config server* (page 742) to a new system that uses *the same* hostname.

To migrate all three mirrored config servers, perform this procedure for each config server separately and migrate the config servers in reverse order from how they are listed in the `mongos` instances' `configDB` string. Start with the last config server listed in the `configDB` string.

1. Shut down the config server.

   This renders all config data for the sharded cluster "read only."

2. Change the DNS entry that points to the system that provided the old config server, so that the *same* hostname points to the new system. How you do this depends on how you organize your DNS and hostname resolution services.

3. Copy the contents of `dbPath` from the old config server to the new config server.

   For example, to copy the contents of `dbPath` to a machine named `mongodb.config2.example.net`, you might issue a command similar to the following:

   ```
   rsync -az /data/configdb/ mongodb.config2.example.net:/data/configdb
   ```

4. Start the config server instance on the new system. The default invocation is:

   ```
   mongod --configsvr
   ```

When you start the third config server, your cluster will become writable and it will be able to create new splits and migrate chunks as needed.

Migrate Config Servers with Different Hostnames

On this page

- Overview (page 793)
- Considerations (page 793)
- Procedure (page 794)

Important: This procedure applies to migrating config servers when using three mirrored `mongod` instances as config servers.

Changed in version 3.2: Starting in MongoDB 3.2, config servers can be deployed as *replica set* (page 623). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

For replacing config servers deployed as members of a replica set, see *Replace a Config Server* (page 791).

Overview

For a *sharded cluster* (page 739) that uses three mirrored config servers, all three config servers must be available in order to support operations that result in cluster metadata changes, e.g. chunk splits and migrations. If one of the config servers is unavailable or inoperable, you must replace it as soon as possible.

For a *sharded cluster* (page 739) that uses three mirrored config servers, this procedure migrates a *config server* (page 742) to a new server that uses a different hostname. Use this procedure only if the config server *will not* be accessible via the same hostname. If possible, avoid changing the hostname so that you can instead use the procedure to *migrate a config server and use the same hostname* (page 792).

Considerations

With three mirrored config servers, changing a *config server's* (page 742) hostname **requires downtime** and requires restarting every process in the sharded cluster.

While migrating config servers, always make sure that all `mongos` instances have three config servers specified in the `configDB` setting at all times. Also ensure that you specify the config servers in the same order for each `mongos` instance's `configDB` setting.

Procedure

Important: This procedure applies to migrating config servers when using three mirrored `mongod` instances as config servers. For replacing config servers deployed as members of a replica set, see *Replace a Config Server* (page 791).

1. Disable the cluster balancer process temporarily. See *Disable the Balancer* (page 802) for more information.
2. Shut down the config server to migrate.

   This renders all config data for the sharded cluster "read only."
3. Copy the contents of `dbPath` from the old config server to the new config server. For example, to copy the contents of `dbPath` to a machine named `mongodb.config2.example.net`, use a command that resembles the following:

   ```
   rsync -az /data/configdb mongodb.config2.example.net:/data/configdb
   ```
4. Start the config server instance on the new system. The default invocation is:

   ```
   mongod --configsvr
   ```
5. Shut down all existing MongoDB processes. This includes:
   - the `mongod` instances for the shards.
   - the `mongod` instances for the existing *config databases* (page 823).
   - the `mongos` instances.
6. Restart all shard `mongod` instances.
7. Restart the `mongod` instances for the two existing non-migrated config servers.
8. Update the `configDB` setting for each `mongos` instances.
9. Restart the `mongos` instances.
10. Re-enable the balancer to allow the cluster to resume normal balancing operations. See the *Disable the Balancer* (page 802) section for more information on managing the balancer process.

Migrate a Sharded Cluster to Different Hardware

On this page

- Disable the Balancer (page 795)
- Migrate Each Config Server Separately (page 795)
- Restart the `mongos` Instances (page 796)
- Migrate the Shards (page 796)
- Re-Enable the Balancer (page 797)

The tutorial is specific to MongoDB 3.2. For earlier versions of MongoDB, refer to the corresponding version of the MongoDB Manual.

Changed in version 3.2.

Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

This procedure moves the components of the *sharded cluster* to a new hardware system without downtime for reads and writes.

Important: While the migration is in progress, do not attempt to change to the *cluster metadata* (page 764). Do not use any operation that modifies the cluster metadata *in any way*. For example, do not create or drop databases, create or drop collections, or use any sharding commands.

If your cluster includes a shard backed by a *standalone* `mongod` instance, consider *converting the standalone to a replica set* (page 678) to simplify migration and to let you keep the cluster online during future maintenance. Migrating a shard as standalone is a multi-step process that may require downtime.

Disable the Balancer

Disable the balancer to stop *chunk migration* (page 759) and do not perform any metadata write operations until the process finishes. If a migration is in progress, the balancer will complete the in-progress migration before stopping.

To disable the balancer, connect to one of the cluster's `mongos` instances and issue the following method:

```
sh.stopBalancer()
```

To check the balancer state, issue the `sh.getBalancerState()` method.

For more information, see *Disable the Balancer* (page 802).

Migrate Each Config Server Separately

Changed in version 3.2.

Starting in MongoDB 3.2, config servers for sharded clusters can be deployed as a *replica set* (page 623). The replica set config servers must run the *WiredTiger storage engine* (page 595). MongoDB 3.2 deprecates the use of three mirrored `mongod` instances for config servers.

The following restrictions apply to a replica set configuration when used for config servers:

- Must have zero *arbiters* (page 635).
- Must have no *delayed members* (page 634).
- Must build indexes (i.e. no member should have `buildIndexes` setting set to false).

For each member of the config server replica set:

Important: Replace the secondary members before replacing the primary.

Step 1: Start the replacement config server. Start a `mongod` instance, specifying both the `--configsvr` and `--replSet` options.

```
mongod --configsvr --replSet <replicaSetName>
```

Step 2: Add the new config server to the replica set. Connect a `mongo` shell to the primary of the config server replica set and use `rs.add()` to add the new member.

```
rs.add("<hostnameNew>:<portNew>")
```

The initial sync process copies all the data from one member of the config server replica set to the new member without restarting.

`mongos` instances automatically recognize the change in the config server replica set members without restarting.

Step 3: Shut down the member to replace. If replacing the primary member, step down the primary first before shutting down.

Restart the `mongos` Instances

Changed in version 3.2: With replica set config servers, the `mongos` instances specify in the `--configdb` or `sharding.configDB` setting the config server replica set name and at least one of the replica set members. The `mongos` instances for the sharded cluster must specify the same config server replica set name but can specify different members of the replica set.

If a `mongos` instance specifies a migrated replica set member in the `--configdb` or `sharding.configDB` setting, update the config server setting for the next time you restart the `mongos` instance.

For more information, see *Start the mongos Instances* (page 767).

Migrate the Shards

Migrate the shards one at a time. For each shard, follow the appropriate procedure in this section.

Migrate a Replica Set Shard To migrate a sharded cluster, migrate each member separately. First migrate the non-primary members, and then migrate the *primary* last.

If the replica set has two voting members, add an *arbiter* (page 635) to the replica set to ensure the set keeps a majority of its votes available during the migration. You can remove the arbiter after completing the migration.

Migrate a Member of a Replica Set Shard

1. Shut down the `mongod` process. To ensure a clean shutdown, use the `shutdown` command.
2. Move the data directory (i.e., the `dbPath`) to the new machine.
3. Restart the `mongod` process at the new location.
4. Connect to the replica set's current primary.
5. If the hostname of the member has changed, use `rs.reconfig()` to update the *replica set configuration document* (page 717) with the new hostname.

   For example, the following sequence of commands updates the hostname for the instance at position `2` in the `members` array:

   ```
   cfg = rs.conf()
   cfg.members[2].host = "pocatello.example.net:27017"
   rs.reconfig(cfg)
   ```

   For more information on updating the configuration document, see *replica-set-reconfiguration-usage*.
6. To confirm the new configuration, issue `rs.conf()`.
7. Wait for the member to recover. To check the member's state, issue `rs.status()`.

Migrate the Primary in a Replica Set Shard While migrating the replica set's primary, the set must elect a new primary. This failover process which renders the replica set unavailable to perform reads or accept writes for the duration of the election, which typically completes quickly. If possible, plan the migration during a maintenance window.

1. Step down the primary to allow the normal *failover* (page 644) process. To step down the primary, connect to the primary and issue the either the `replSetStepDown` command or the `rs.stepDown()` method. The following example shows the `rs.stepDown()` method:

   ```
   rs.stepDown()
   ```

2. Once the primary has stepped down and another member has become `PRIMARY` (page 725) state. To migrate the stepped-down primary, follow the *Migrate a Member of a Replica Set Shard* (page 796) procedure

   You can check the output of `rs.status()` to confirm the change in status.

Migrate a Standalone Shard The ideal procedure for migrating a standalone shard is to *convert the standalone to a replica set* (page 678) and then use the procedure for *migrating a replica set shard* (page 796). In production clusters, all shards should be replica sets, which provides continued availability during maintenance windows.

Migrating a shard as standalone is a multi-step process during which part of the shard may be unavailable. If the shard is the *primary shard* for a database,the process includes the `movePrimary` command. While the `movePrimary` runs, you should stop modifying data in that database. To migrate the standalone shard, use the *Remove Shards from an Existing Sharded Cluster* (page 805) procedure.

Re-Enable the Balancer

To complete the migration, re-enable the balancer to resume *chunk migrations* (page 759).

Connect to one of the cluster's `mongos` instances and pass `true` to the `sh.setBalancerState()` method:

```
sh.setBalancerState(true)
```

To check the balancer state, issue the `sh.getBalancerState()` method.

For more information, see *Enable the Balancer* (page 803).

Backup Cluster Metadata

This procedure shuts down the `mongod` instance of a *config server* (page 742) in order to create a backup of a *sharded cluster's* (page 733) metadata. The cluster's config servers store all of the cluster's metadata, most importantly the mapping from *chunks* to *shards*.

When you perform this procedure, the cluster remains operational [17].

1. Disable the cluster balancer process temporarily. See *Disable the Balancer* (page 802) for more information.
2. Shut down one of the config databases.
3. Create a full copy of the data files (i.e. the path specified by the `dbPath` option for the config instance.)
4. Restart the original configuration server.
5. Re-enable the balancer to allow the cluster to resume normal balancing operations. See the *Disable the Balancer* (page 802) section for more information on managing the balancer process.

[17] While one of the three config servers is unavailable, the cluster cannot split any chunks nor can it migrate chunks between shards. Your application will be able to write data to the cluster. See *Config Servers* (page 742) for more information.

See also:

MongoDB Backup Methods (page 282).

Configure Behavior of Balancer Process in Sharded Clusters

On this page

- Schedule a Window of Time for Balancing to Occur (page 798)
- Configure Default Chunk Size (page 798)
- Change the Maximum Storage Size for a Given Shard (page 798)
- Change Replication Behavior for Chunk Migration (page 799)

The balancer is a process that runs on *one* of the `mongos` instances in a cluster and ensures that *chunks* are evenly distributed throughout a sharded cluster. In most deployments, the default balancer configuration is sufficient for normal operation. However, administrators might need to modify balancer behavior depending on application or operational requirements. If you encounter a situation where you need to modify the behavior of the balancer, use the procedures described in this document.

For conceptual information about the balancer, see *Sharded Collection Balancing* (page 758) and *Cluster Balancer* (page 758).

Schedule a Window of Time for Balancing to Occur

You can schedule a window of time during which the balancer can migrate chunks, as described in the following procedures:

- *Schedule the Balancing Window* (page 801)
- *Remove a Balancing Window Schedule* (page 802).

The `mongos` instances use their own local timezones when respecting balancer window.

Configure Default Chunk Size

The default chunk size for a sharded cluster is 64 megabytes. In most situations, the default size is appropriate for splitting and migrating chunks. For information on how chunk size affects deployments, see details, see *Chunk Size* (page 762).

Changing the default chunk size affects chunks that are processes during migrations and auto-splits but does not retroactively affect all chunks.

To configure default chunk size, see *Modify Chunk Size in a Sharded Cluster* (page 813).

Change the Maximum Storage Size for a Given Shard

The `maxSize` field in the `shards` (page 828) collection in the *config database* (page 823) sets the maximum size for a shard, allowing you to control whether the balancer will migrate chunks to a shard. If `mem.mapped` size [18] is above a shard's `maxSize`, the balancer will not move chunks to the shard. Also, the balancer will not move chunks off an overloaded shard. This must happen manually. The `maxSize` value only affects the balancer's selection of destination shards.

[18] This value includes the mapped size of all data files including the‘‘local‘‘ and `admin` databases. Account for this when setting `maxSize`.

By default, maxSize is not specified, allowing shards to consume the total amount of available space on their machines if necessary.

You can set maxSize both when adding a shard and once a shard is running.

To set maxSize when adding a shard, set the addShard command's maxSize parameter to the maximum size in megabytes. For example, the following command run in the mongo shell adds a shard with a maximum size of 125 megabytes:

```
db.runCommand( { addshard : "example.net:34008", maxSize : 125 } )
```

To set maxSize on an existing shard, insert or update the maxSize field in the shards (page 828) collection in the *config database* (page 823). Set the maxSize in megabytes.

Example

Assume you have the following shard without a maxSize field:

```
{ "_id" : "shard0000", "host" : "example.net:34001" }
```

Run the following sequence of commands in the mongo shell to insert a maxSize of 125 megabytes:

```
use config
db.shards.update( { _id : "shard0000" }, { $set : { maxSize : 125 } } )
```

To later increase the maxSize setting to 250 megabytes, run the following:

```
use config
db.shards.update( { _id : "shard0000" }, { $set : { maxSize : 250 } } )
```

Change Replication Behavior for Chunk Migration

Secondary Throttle Changed in version 3.0.0: The balancer configuration document added configurable writeConcern to control the semantics of the _secondaryThrottle option.

The _secondaryThrottle parameter of the balancer and the moveChunk command affects the replication behavior during *chunk migration* (page 761). By default, _secondaryThrottle is true, which means each document move during chunk migration propagates to at least one secondary before the balancer proceeds with the next document: this is equivalent to a write concern of { w: 2 }.

You can also configure the writeConcern for the _secondaryThrottle operation, to configure how migrations will wait for replication to complete. For more information on the replication behavior during various steps of chunk migration, see *Chunk Migration and Replication* (page 761).

To change the balancer's _secondaryThrottle and writeConcern values, connect to a mongos instance and directly update the _secondaryThrottle value in the settings (page 828) collection of the *config database* (page 823). For example, from a mongo shell connected to a mongos, issue the following command:

```
use config
db.settings.update(
   { "_id" : "balancer" },
   { $set : { "_secondaryThrottle" : false ,
              "writeConcern": { "w": "majority" } } },
   { upsert : true }
)
```

The effects of changing the _secondaryThrottle and writeConcern value may not be immediate. To ensure an immediate effect, stop and restart the balancer to enable the selected value of _secondaryThrottle. See *Manage Sharded Cluster Balancer* (page 800) for details.

Wait for Delete The `_waitForDelete` setting of the balancer and the `moveChunk` command affects how the balancer migrates multiple chunks from a shard. By default, the balancer does not wait for the on-going migration's delete phase to complete before starting the next chunk migration. To have the delete phase **block** the start of the next chunk migration, you can set the `_waitForDelete` to true.

For details on chunk migration, see *Chunk Migration* (page 760). For details on the chunk migration queuing behavior, see *Chunk Migration Queuing* (page 761).

The `_waitForDelete` is generally for internal testing purposes. To change the balancer's `_waitForDelete` value:

1. Connect to a `mongos` instance.
2. Update the `_waitForDelete` value in the `settings` (page 828) collection of the *config database* (page 823). For example:

```
use config
db.settings.update(
   { "_id" : "balancer" },
   { $set : { "_waitForDelete" : true } },
   { upsert : true }
)
```

Once set to `true`, to revert to the default behavior:

1. Connect to a `mongos` instance.
2. Update or unset the `_waitForDelete` field in the `settings` (page 828) collection of the *config database* (page 823):

```
use config
db.settings.update(
   { "_id" : "balancer", "_waitForDelete": true },
   { $unset : { "_waitForDelete" : "" } }
)
```

Manage Sharded Cluster Balancer

On this page

- Check the Balancer State (page 801)
- Check the Balancer Lock (page 801)
- Schedule the Balancing Window (page 801)
- Remove a Balancing Window Schedule (page 802)
- Disable the Balancer (page 802)
- Enable the Balancer (page 803)
- Disable Balancing During Backups (page 803)
- Disable Balancing on a Collection (page 804)
- Enable Balancing on a Collection (page 804)
- Confirm Balancing is Enabled or Disabled (page 804)

This page describes common administrative procedures related to balancing. For an introduction to balancing, see *Sharded Collection Balancing* (page 758). For lower level information on balancing, see *Cluster Balancer* (page 758).

See also:

Configure Behavior of Balancer Process in Sharded Clusters (page 798)

Check the Balancer State

`sh.getBalancerState()` checks if the balancer is enabled (i.e. that the balancer is permitted to run). `sh.getBalancerState()` does not check if the balancer is actively balancing chunks.

To see if the balancer is enabled in your *sharded cluster*, issue the following command, which returns a boolean:

```
sh.getBalancerState()
```

New in version 3.0.0: You can also see if the balancer is enabled using `sh.status()`. The `currently-enabled` field indicates whether the balancer is enabled, while the `currently-running` field indicates if the balancer is currently running.

Check the Balancer Lock

To see if the balancer process is active in your *cluster*, do the following:

1. Connect to any `mongos` in the cluster using the `mongo` shell.
2. Issue the following command to switch to the *Config Database* (page 823):

   ```
   use config
   ```

3. Use the following query to return the balancer lock:

   ```
   db.locks.find( { _id : "balancer" } ).pretty()
   ```

   When this command returns, you will see output like the following:

   ```
   {   "_id" : "balancer",
   "process" : "mongos0.example.net:1292810611:1804289383",
     "state" : 2,
        "ts" : ObjectId("4d0f872630c42d1978be8a2e"),
     "when" : "Mon Dec 20 2010 11:41:10 GMT-0500 (EST)",
      "who" : "mongos0.example.net:1292810611:1804289383:Balancer:846930886",
      "why" : "doing balance round" }
   ```

   This output confirms that:

   - The balancer originates from the `mongos` running on the system with the hostname `mongos0.example.net`.
   - The value in the `state` field indicates that a `mongos` has the lock. For version 2.0 and later, the value of an active lock is `2`; for earlier versions the value is `1`.

Schedule the Balancing Window

In some situations, particularly when your data set grows slowly and a migration can impact performance, it is useful to ensure that the balancer is active only at certain times. The following procedure specifies the `activeWindow`, which is the timeframe during which the *balancer* will be able to migrate chunks:

Step 1: Connect to `mongos` using the `mongo` shell. You can connect to any `mongos` in the cluster.

Step 2: Switch to the *Config Database*. Issue the following command to switch to the config database.

```
use config
```

Step 3: Ensure that the balancer is not `stopped`. The balancer will not activate in the `stopped` state. To ensure that the balancer is not `stopped`, use `sh.setBalancerState()`, as in the following:

```
sh.setBalancerState( true )
```

The balancer will not start if you are outside of the `activeWindow` timeframe.

Step 4: Modify the balancer's window. Set the `activeWindow` using `update()`, as in the following:

```
db.settings.update(
   { _id: "balancer" },
   { $set: { activeWindow : { start : "<start-time>", stop : "<stop-time>" } } },
   { upsert: true }
)
```

Replace `<start-time>` and `<end-time>` with time values using two digit hour and minute values (i.e. `HH:MM`) that specify the beginning and end boundaries of the balancing window.

- For `HH` values, use hour values ranging from `00` - `23`.
- For `MM` value, use minute values ranging from `00` - `59`.

MongoDB evaluates the start and stop times relative to the time zone of each individual `mongos` instance in the sharded cluster. If your `mongos` instances are physically located in different time zones, set the time zone on each server to UTC+-00:00 so that the balancer window is uniformly interpreted.

Note: The balancer window must be sufficient to *complete* the migration of all data inserted during the day.

As data insert rates can change based on activity and usage patterns, it is important to ensure that the balancing window you select will be sufficient to support the needs of your deployment.

Do not use the `sh.startBalancer()` method when you have set an `activeWindow`.

Remove a Balancing Window Schedule

If you have *set the balancing window* (page 801) and wish to remove the schedule so that the balancer is always running, use `$unset` to clear the `activeWindow`, as in the following:

```
use config
db.settings.update({ _id : "balancer" }, { $unset : { activeWindow : true } })
```

Disable the Balancer

By default, the balancer may run at any time and only moves chunks as needed. To disable the balancer for a short period of time and prevent all migration, use the following procedure:

1. Connect to any `mongos` in the cluster using the `mongo` shell.
2. Issue the following operation to disable the balancer:

   ```
   sh.stopBalancer()
   ```

If a migration is in progress, the system will complete the in-progress migration before stopping.

3. To verify that the balancer will not start, issue the following command, which returns `false` if the balancer is disabled:

   ```
   sh.getBalancerState()
   ```

   Optionally, to verify no migrations are in progress after disabling, issue the following operation in the `mongo` shell:

   ```
   use config
   while( sh.isBalancerRunning() ) {
             print("waiting...");
             sleep(1000);
   }
   ```

Note: To disable the balancer from a driver that does not have the `sh.stopBalancer()` or `sh.setBalancerState()` helpers, issue the following command from the `config` database:

```
db.settings.update( { _id: "balancer" }, { $set : { stopped: true } } ,  { upsert: true } )
```

Enable the Balancer

Use this procedure if you have disabled the balancer and are ready to re-enable it:

1. Connect to any `mongos` in the cluster using the `mongo` shell.
2. Issue one of the following operations to enable the balancer:

   From the `mongo` shell, issue:

   ```
   sh.setBalancerState(true)
   ```

   From a driver that does not have the `sh.startBalancer()` helper, issue the following from the `config` database:

   ```
   db.settings.update( { _id: "balancer" }, { $set : { stopped: false } } , { upsert: true } )
   ```

Disable Balancing During Backups

If MongoDB migrates a *chunk* during a *backup* (page 282), you can end with an inconsistent snapshot of your *sharded cluster*. Never run a backup while the balancer is active. To ensure that the balancer is inactive during your backup operation:

- Set the *balancing window* (page 801) so that the balancer is inactive during the backup. Ensure that the backup can complete while you have the balancer disabled.
- *manually disable the balancer* (page 802) for the duration of the backup procedure.

If you turn the balancer off while it is in the middle of a balancing round, the shut down is not instantaneous. The balancer completes the chunk move in-progress and then ceases all further balancing rounds.

Before starting a backup operation, confirm that the balancer is not active. You can use the following command to determine if the balancer is active:

```
!sh.getBalancerState() && !sh.isBalancerRunning()
```

When the backup procedure is complete you can reactivate the balancer process.

Disable Balancing on a Collection

You can disable balancing for a specific collection with the `sh.disableBalancing()` method. You may want to disable the balancer for a specific collection to support maintenance operations or atypical workloads, for example, during data ingestions or data exports.

When you disable balancing on a collection, MongoDB will not interrupt in progress migrations.

To disable balancing on a collection, connect to a `mongos` with the `mongo` shell and call the `sh.disableBalancing()` method.

For example:

```
sh.disableBalancing("students.grades")
```

The `sh.disableBalancing()` method accepts as its parameter the full *namespace* of the collection.

Enable Balancing on a Collection

You can enable balancing for a specific collection with the `sh.enableBalancing()` method.

When you enable balancing for a collection, MongoDB will not *immediately* begin balancing data. However, if the data in your sharded collection is not balanced, MongoDB will be able to begin distributing the data more evenly.

To enable balancing on a collection, connect to a `mongos` with the `mongo` shell and call the `sh.enableBalancing()` method.

For example:

```
sh.enableBalancing("students.grades")
```

The `sh.enableBalancing()` method accepts as its parameter the full *namespace* of the collection.

Confirm Balancing is Enabled or Disabled

To confirm whether balancing for a collection is enabled or disabled, query the `collections` collection in the `config` database for the collection *namespace* and check the `noBalance` field. For example:

```
db.getSiblingDB("config").collections.findOne({_id : "students.grades"}).noBalance;
```

This operation will return a null error, `true`, `false`, or no output:

- A null error indicates the collection namespace is incorrect.
- If the result is `true`, balancing is disabled.
- If the result is `false`, balancing is enabled currently but has been disabled in the past for the collection. Balancing of this collection will begin the next time the balancer runs.
- If the operation returns no output, balancing is enabled currently and has never been disabled in the past for this collection. Balancing of this collection will begin the next time the balancer runs.

New in version 3.0.0: You can also see if the balancer is enabled using `sh.status()`. The `currently-enabled` field indicates if the balancer is enabled.

Remove Shards from an Existing Sharded Cluster

On this page

- Ensure the Balancer Process is Enabled (page 805)
- Determine the Name of the Shard to Remove (page 805)
- Remove Chunks from the Shard (page 805)
- Check the Status of the Migration (page 806)
- Move Unsharded Data (page 806)
- Finalize the Migration (page 807)

To remove a *shard* you must ensure the shard's data is migrated to the remaining shards in the cluster. This procedure describes how to safely migrate data and how to remove a shard.

This procedure describes how to safely remove a *single* shard. *Do not* use this procedure to migrate an entire cluster to new hardware. To migrate an entire shard to new hardware, migrate individual shards as if they were independent replica sets.

To remove a shard, first connect to one of the cluster's `mongos` instances using `mongo` shell. Then use the sequence of tasks in this document to remove a shard from the cluster.

Ensure the Balancer Process is Enabled

To successfully migrate data from a shard, the *balancer* process **must** be enabled. Check the balancer state using the `sh.getBalancerState()` helper in the `mongo` shell. For more information, see the section on *balancer operations* (page 802).

Determine the Name of the Shard to Remove

To determine the name of the shard, connect to a `mongos` instance with the `mongo` shell and either:

- Use the `listShards` command, as in the following:

  ```
  db.adminCommand( { listShards: 1 } )
  ```

- Run either the `sh.status()` or the `db.printShardingStatus()` method.

The `shards._id` field lists the name of each shard.

Remove Chunks from the Shard

From the `admin` database, run the `removeShard` command. This begins "draining" chunks from the shard you are removing to other shards in the cluster. For example, for a shard named `mongodb0`, run:

```
use admin
db.runCommand( { removeShard: "mongodb0" } )
```

This operation returns immediately, with the following response:

```
{
    "msg" : "draining started successfully",
    "state" : "started",
    "shard" : "mongodb0",
```

```
    "ok" : 1
}
```

Depending on your network capacity and the amount of data, this operation can take from a few minutes to several days to complete.

Check the Status of the Migration

To check the progress of the migration at any stage in the process, run `removeShard` from the `admin` database again. For example, for a shard named `mongodb0`, run:

```
use admin
db.runCommand( { removeShard: "mongodb0" } )
```

The command returns output similar to the following:

```
{
     "msg" : "draining ongoing",
    "state" : "ongoing",
    "remaining" : {
        "chunks" : 42,
        "dbs" : 1
    },
    "ok" : 1
}
```

In the output, the `remaining` document displays the remaining number of chunks that MongoDB must migrate to other shards and the number of MongoDB databases that have "primary" status on this shard.

Continue checking the status of the *removeShard* command until the number of chunks remaining is `0`. Always run the command on the `admin` database. If you are on a database other than `admin`, you can use `sh._adminCommand` to run the command on `admin`.

Move Unsharded Data

If the shard is the *primary shard* for one or more databases in the cluster, then the shard will have unsharded data. If the shard is not the primary shard for any databases, skip to the next task, *Finalize the Migration* (page 807).

In a cluster, a database with unsharded collections stores those collections only on a single shard. That shard becomes the primary shard for that database. (Different databases in a cluster can have different primary shards.)

Warning: Do not perform this procedure until you have finished draining the shard.

1. To determine if the shard you are removing is the primary shard for any of the cluster's databases, issue one of the following methods:

   - `sh.status()`
   - `db.printShardingStatus()`

   In the resulting document, the `databases` field lists each database and its primary shard. For example, the following `database` field shows that the `products` database uses `mongodb0` as the primary shard:

   ```
   {  "_id" : "products",  "partitioned" : true,  "primary" : "mongodb0" }
   ```

2. To move a database to another shard, use the `movePrimary` command. For example, to migrate all remaining unsharded data from `mongodb0` to `mongodb1`, issue the following command:

```
db.runCommand( { movePrimary: "products", to: "mongodb1" })
```

This command does not return until MongoDB completes moving all data, which may take a long time. The response from this command will resemble the following:

```
{ "primary" : "mongodb1", "ok" : 1 }
```

Finalize the Migration

To clean up all metadata information and finalize the removal, run `removeShard` again. For example, for a shard named `mongodb0`, run:

```
use admin
db.runCommand( { removeShard: "mongodb0" } )
```

A success message appears at completion:

```
{
    "msg" : "removeshard completed successfully",
    "state" : "completed",
    "shard" : "mongodb0",
    "ok" : 1
}
```

Once the value of the `state` field is "completed", you may safely stop the processes comprising the `mongodb0` shard.

See also:

Backup and Restore Sharded Clusters (page 355)

13.3.3 Sharded Cluster Data Management

The following documents provide information in managing data in sharded clusters.

Create Chunks in a Sharded Cluster **(page 808)** Create chunks, or *pre-split* empty collection to ensure an even distribution of chunks during data ingestion.

Split Chunks in a Sharded Cluster **(page 808)** Manually create chunks in a sharded collection.

Migrate Chunks in a Sharded Cluster **(page 809)** Manually migrate chunks without using the automatic balance process.

Merge Chunks in a Sharded Cluster **(page 810)** Use the `mergeChunks` to manually combine chunk ranges.

Modify Chunk Size in a Sharded Cluster **(page 813)** Modify the default chunk size in a sharded collection

Clear jumbo Flag **(page 814)** Clear *jumbo* flag from a shard.

Manage Shard Tags **(page 816)** Use tags to associate specific ranges of shard key values with specific shards.

Enforce Unique Keys for Sharded Collections **(page 818)** Ensure that a field is always unique in all collections in a sharded cluster.

Shard GridFS Data Store **(page 820)** Choose whether to shard GridFS data in a sharded collection.

Create Chunks in a Sharded Cluster

Pre-splitting the chunk ranges in an empty sharded collection allows clients to insert data into an already partitioned collection. In most situations a *sharded cluster* will create and distribute chunks automatically without user intervention. However, in a limited number of cases, MongoDB cannot create enough chunks or distribute data fast enough to support required throughput. For example:

- If you want to partition an existing data collection that resides on a single shard.
- If you want to ingest a large volume of data into a cluster that isn't balanced, or where the ingestion of data will lead to data imbalance. For example, monotonically increasing or decreasing shard keys insert all data into a single chunk.

These operations are resource intensive for several reasons:

- Chunk migration requires copying all the data in the chunk from one shard to another.
- MongoDB can migrate only a single chunk at a time.
- MongoDB creates splits only after an insert operation.

Warning: Only pre-split an empty collection. If a collection already has data, MongoDB automatically splits the collection's data when you enable sharding for the collection. Subsequent attempts to manually create splits can lead to unpredictable chunk ranges and sizes as well as inefficient or ineffective balancing behavior.

To create chunks manually, use the following procedure:

1. Split empty chunks in your collection by manually performing the `split` command on chunks.

   Example

   To create chunks for documents in the `myapp.users` collection using the `email` field as the *shard key*, use the following operation in the `mongo` shell:

   ```
   for ( var x=97; x<97+26; x++ ){
     for( var y=97; y<97+26; y+=6 ) {
       var prefix = String.fromCharCode(x) + String.fromCharCode(y);
       db.runCommand( { split : "myapp.users" , middle : { email : prefix } } );
     }
   }
   ```

   This assumes a collection size of 100 million documents.

   For information on the balancer and automatic distribution of chunks across shards, see *Cluster Balancer* (page 758) and *Chunk Migration* (page 760). For information on manually migrating chunks, see *Migrate Chunks in a Sharded Cluster* (page 809).

Split Chunks in a Sharded Cluster

Normally, MongoDB splits a *chunk* after an insert if the chunk exceeds the maximum *chunk size* (page 762). However, you may want to split chunks manually if:

- you have a large amount of data in your cluster and very few *chunks*, as is the case after deploying a cluster using existing data.
- you expect to add a large amount of data that would initially reside in a single chunk or shard. For example, you plan to insert a large amount of data with *shard key* values between `300` and `400`, *but* all values of your shard keys are between `250` and `500` are in a single chunk.

Note: New in version 2.6: MongoDB provides the `mergeChunks` command to combine contiguous chunk ranges into a single chunk. See *Merge Chunks in a Sharded Cluster* (page 810) for more information.

The *balancer* may migrate recently split chunks to a new shard immediately if `mongos` predicts future insertions will benefit from the move. The balancer does not distinguish between chunks split manually and those split automatically by the system.

Warning: Be careful when splitting data in a sharded collection to create new chunks. When you shard a collection that has existing data, MongoDB automatically creates chunks to evenly distribute the collection. To split data effectively in a sharded cluster you must consider the number of documents in a chunk and the average document size to create a uniform chunk size. When chunks have irregular sizes, shards may have an equal number of chunks but have very different data sizes. Avoid creating splits that lead to a collection with differently sized chunks.

Use `sh.status()` to determine the current chunk ranges across the cluster.

To split chunks manually, use the `split` command with either fields `middle` or `find`. The `mongo` shell provides the helper methods `sh.splitFind()` and `sh.splitAt()`.

`splitFind()` splits the chunk that contains the *first* document returned that matches this query into two equally sized chunks. You must specify the full namespace (i.e. "`<database>.<collection>`") of the sharded collection to `splitFind()`. The query in `splitFind()` does not need to use the shard key, though it nearly always makes sense to do so.

Example

The following command splits the chunk that contains the value of `63109` for the `zipcode` field in the `people` collection of the `records` database:

```
sh.splitFind( "records.people", { "zipcode": "63109" } )
```

Use `splitAt()` to split a chunk in two, using the queried document as the lower bound in the new chunk:

Example

The following command splits the chunk that contains the value of `63109` for the `zipcode` field in the `people` collection of the `records` database.

```
sh.splitAt( "records.people", { "zipcode": "63109" } )
```

Note: `splitAt()` does not necessarily split the chunk into two equally sized chunks. The split occurs at the location of the document matching the query, regardless of where that document is in the chunk.

Migrate Chunks in a Sharded Cluster

In most circumstances, you should let the automatic *balancer* migrate *chunks* between *shards*. However, you may want to migrate chunks manually in a few cases:

- When *pre-splitting* an empty collection, migrate chunks manually to distribute them evenly across the shards. Use pre-splitting in limited situations to support bulk data ingestion.
- If the balancer in an active cluster cannot distribute chunks within the *balancing window* (page 801), then you will have to migrate chunks manually.

To manually migrate chunks, use the `moveChunk` command. For more information on how the automatic balancer moves chunks between shards, see *Cluster Balancer* (page 758) and *Chunk Migration* (page 760).

Example

Migrate a single chunk

The following example assumes that the field `username` is the *shard key* for a collection named `users` in the `myapp` database, and that the value `smith` exists within the *chunk* to migrate. Migrate the chunk using the following command in the `mongo` shell.

```
db.adminCommand( { moveChunk : "myapp.users",
                   find : {username : "smith"},
                   to : "mongodb-shard3.example.net" } )
```

This command moves the chunk that includes the shard key value "smith" to the *shard* named `mongodb-shard3.example.net`. The command will block until the migration is complete.

Tip

To return a list of shards, use the `listShards` command.

Example

Evenly migrate chunks

To evenly migrate chunks for the `myapp.users` collection, put each prefix chunk on the next shard from the other and run the following commands in the mongo shell:

```
var shServer = [ "sh0.example.net", "sh1.example.net", "sh2.example.net", "sh3.example.net", "sh4.exa
for ( var x=97; x<97+26; x++ ){
  for( var y=97; y<97+26; y+=6 ) {
    var prefix = String.fromCharCode(x) + String.fromCharCode(y);
    db.adminCommand({moveChunk : "myapp.users", find : {email : prefix}, to : shServer[(y-97)/6]})
  }
}
```

See *Create Chunks in a Sharded Cluster* (page 808) for an introduction to pre-splitting.

The `moveChunk` command has the: `_secondaryThrottle` parameter. When set to `true`, MongoDB ensures that changes to shards as part of chunk migrations replicate to *secondaries* throughout the migration operation. For more information, see *Change Replication Behavior for Chunk Migration* (page 799).

Changed in version 2.4: In 2.4, `_secondaryThrottle` is `true` by default.

Warning: The `moveChunk` command may produce the following error message:

```
The collection's metadata lock is already taken.
```

This occurs when clients have too many open *cursors* that access the migrating chunk. You may either wait until the cursors complete their operations or close the cursors manually.

Merge Chunks in a Sharded Cluster

On this page

- Overview (page 811)
- Procedure (page 811)

Overview

The `mergeChunks` command allows you to collapse empty chunks into neighboring chunks on the same shard. A *chunk* is empty if it has no documents associated with its shard key range.

Important: Empty *chunks* can make the *balancer* assess the cluster as properly balanced when it is not.

Empty chunks can occur under various circumstances, including:

- If a *pre-split* (page 808) creates too many chunks, the distribution of data to chunks may be uneven.
- If you delete many documents from a sharded collection, some chunks may no longer contain data.

This tutorial explains how to identify chunks available to merge, and how to merge those chunks with neighboring chunks.

Procedure

Note: Examples in this procedure use a `users` *collection* in the `test` *database*, using the `username` filed as a *shard key*.

Identify Chunk Ranges In the `mongo` shell, identify the *chunk* ranges with the following operation:

```
sh.status()
```

The output of the `sh.status()` will resemble the following:

```
--- Sharding Status ---
sharding version: {
     "_id" : 1,
     "version" : 4,
     "minCompatibleVersion" : 4,
     "currentVersion" : 5,
     "clusterId" : ObjectId("5260032c901f6712dcd8f400")
}
shards:
     {  "_id" : "shard0000",  "host" : "localhost:30000" }
     {  "_id" : "shard0001",  "host" : "localhost:30001" }
  databases:
     {  "_id" : "admin",  "partitioned" : false,  "primary" : "config" }
     {  "_id" : "test",  "partitioned" : true,  "primary" : "shard0001" }
            test.users
                    shard key: { "username" : 1 }
                    chunks:
                            shard0000       7
                            shard0001       7
                    { "username" : { "$minKey" : 1 } } -->> { "username" : "user16643" } on : shard0
                    { "username" : "user16643" } -->> { "username" : "user2329" } on : shard0000 Tim
```

```
                    { "username" : "user2329" } -->> { "username" : "user29937" } on : shard0000 Tir
                    { "username" : "user29937" } -->> { "username" : "user36583" } on : shard0000 Ti
                    { "username" : "user36583" } -->> { "username" : "user43229" } on : shard0000 Ti
                    { "username" : "user43229" } -->> { "username" : "user49877" } on : shard0000 Ti
                    { "username" : "user49877" } -->> { "username" : "user56522" } on : shard0000 Ti
                    { "username" : "user56522" } -->> { "username" : "user63169" } on : shard0001 Ti
                    { "username" : "user63169" } -->> { "username" : "user69816" } on : shard0001 Ti
                    { "username" : "user69816" } -->> { "username" : "user76462" } on : shard0001 Ti
                    { "username" : "user76462" } -->> { "username" : "user83108" } on : shard0001 Ti
                    { "username" : "user83108" } -->> { "username" : "user89756" } on : shard0001 Ti
                    { "username" : "user89756" } -->> { "username" : "user96401" } on : shard0001 Ti
                    { "username" : "user96401" } -->> { "username" : { "$maxKey" : 1 } } on : shard0
```

The chunk ranges appear after the chunk counts for each sharded collection, as in the following excerpts:

Chunk counts:

```
    chunks:
            shard0000       7
            shard0001       7
```

Chunk range:

```
    { "username" : "user36583" } -->> { "username" : "user43229" } on : shard0000 Timestamp(6, 0)
```

Verify a Chunk is Empty The `mergeChunks` command requires at least one empty input chunk. To check the size of a chunk, use the `dataSize` command in the sharded collection's database. For example, the following checks the amount of data in the chunk for the `users` collection in the `test` database:

Important: You must use the `use <db>` helper to switch to the database containing the sharded collection before running the `dataSize` command.

```
use test
db.runCommand({
   "dataSize": "test.users",
   "keyPattern": { username: 1 },
   "min": { "username": "user36583" },
   "max": { "username": "user43229" }
})
```

If the input chunk to `dataSize` is empty, `dataSize` produces output similar to:

```
{ "size" : 0, "numObjects" : 0, "millis" : 0, "ok" : 1 }
```

Merge Chunks Merge two contiguous *chunks* on the same *shard*, where at least one of the contains no data, with an operation that resembles the following:

```
db.runCommand( { mergeChunks: "test.users",
                 bounds: [ { "username": "user68982" },
                           { "username": "user95197" } ]
               } )
```

On success, `mergeChunks` produces the following output:

```
{ "ok" : 1 }
```

On any failure condition, `mergeChunks` returns a document where the value of the `ok` field is `0`.

View Merged Chunks Ranges After merging all empty chunks, confirm the new chunk, as follows:

```
sh.status()
```

The output of `sh.status()` should resemble:

```
--- Sharding Status ---
sharding version: {
     "_id" : 1,
     "version" : 4,
     "minCompatibleVersion" : 4,
     "currentVersion" : 5,
     "clusterId" : ObjectId("5260032c901f6712dcd8f400")
}
shards:
     {  "_id" : "shard0000",  "host" : "localhost:30000" }
     {  "_id" : "shard0001",  "host" : "localhost:30001" }
  databases:
     {  "_id" : "admin",  "partitioned" : false,  "primary" : "config" }
     {  "_id" : "test",  "partitioned" : true,  "primary" : "shard0001" }
             test.users
                     shard key: { "username" : 1 }
                     chunks:
                             shard0000       2
                             shard0001       2
                     { "username" : { "$minKey" : 1 } } -->> { "username" : "user16643" } on : shard0
                     { "username" : "user16643" } -->> { "username" : "user56522" } on : shard0000 T
                     { "username" : "user56522" } -->> { "username" : "user96401" } on : shard0001 T
                     { "username" : "user96401" } -->> { "username" : { "$maxKey" : 1 } } on : shard0
```

Modify Chunk Size in a Sharded Cluster

When the first `mongos` connects to a set of *config servers*, it initializes the sharded cluster with a default chunk size of 64 megabytes. This default chunk size works well for most deployments; however, if you notice that automatic migrations have more I/O than your hardware can handle, you may want to reduce the chunk size. For automatic splits and migrations, a small chunk size leads to more rapid and frequent migrations. The allowed range of the chunk size is between 1 and 1024 megabytes, inclusive.

To modify the chunk size, use the following procedure:

1. Connect to any `mongos` in the cluster using the `mongo` shell.
2. Issue the following command to switch to the *Config Database* (page 823):

   ```
   use config
   ```

3. Issue the following `save()` operation to store the global chunk size configuration value:

   ```
   db.settings.save( { _id:"chunksize", value: <sizeInMB> } )
   ```

Note: The `chunkSize` and *`--chunkSize`* options, passed at startup to the `mongos`, **do not** affect the chunk size after you have initialized the cluster.

To avoid confusion, *always* set the chunk size using the above procedure instead of the startup options.

Modifying the chunk size has several limitations:

- Automatic splitting only occurs on insert or update.

- If you lower the chunk size, it may take time for all chunks to split to the new size.
- Splits cannot be undone.
- If you increase the chunk size, existing chunks grow only through insertion or updates until they reach the new size.
- The allowed range of the chunk size is between 1 and 1024 megabytes, inclusive.

Clear `jumbo` Flag

On this page

- Procedures (page 814)

If MongoDB cannot split a chunk that exceeds the *specified chunk size* (page 762) or contains a number of documents that exceeds the `max`, MongoDB labels the chunk as *jumbo* (page 762).

If the chunk size no longer hits the limits, MongoDB clears the `jumbo` flag for the chunk when the `mongos` reloads or rewrites the chunk metadata.

In cases where you need to clear the flag manually, the following procedures outline the steps to manually clear the `jumbo` flag.

Procedures

Divisible Chunks The preferred way to clear the `jumbo` flag from a chunk is to attempt to split the chunk. If the chunk is divisible, MongoDB removes the flag upon successful split of the chunk.

Step 1: Connect to `mongos`. Connect a `mongo` shell to a `mongos`.

Step 2: Find the `jumbo` Chunk. Run `sh.status(true)` to find the chunk labeled `jumbo`.

```
sh.status(true)
```

For example, the following output from sh.status(true) shows that chunk with shard key range `{ "x" : 2 } -->> { "x" : 4 }` is `jumbo`.

```
--- Sharding Status ---
  sharding version: {
     ...
  }
  shards:
     ...
  databases:
     ...
        test.foo
           shard key: { "x" : 1 }
        chunks:
             shard-b  2
             shard-a  2
        { "x" : { "$minKey" : 1 } } -->> { "x" : 1 } on : shard-b Timestamp(2, 0)
        { "x" : 1 } -->> { "x" : 2 } on : shard-a Timestamp(3, 1)
        { "x" : 2 } -->> { "x" : 4 } on : shard-a Timestamp(2, 2) jumbo
        { "x" : 4 } -->> { "x" : { "$maxKey" : 1 } } on : shard-b Timestamp(3, 0)
```

Step 3: Split the `jumbo` Chunk. Use either `sh.splitAt()` or `sh.splitFind()` to split the `jumbo` chunk.

```
sh.splitAt( "test.foo", { x: 3 })
```

MongoDB removes the `jumbo` flag upon successful split of the chunk.

Indivisible Chunks In some instances, MongoDB cannot split the no-longer `jumbo` chunk, such as a chunk with a range of single shard key value, and the preferred method to clear the flag is not applicable. In such cases, you can clear the flag using the following steps.

Important: Only use this method if the *preferred method* (page 814) is *not* applicable.

Before modifying the *config database* (page 823), *always* back up the config database.

If you clear the `jumbo` flag for a chunk that still exceeds the chunk size and/or the document number limit, MongoDB will re-label the chunk as `jumbo` when MongoDB tries to move the chunk.

Step 1: Stop the balancer. Disable the cluster balancer process temporarily, following the steps outlined in *Disable the Balancer* (page 802).

Step 2: Create a backup of `config` database. Use `mongodump` against a config server to create a backup of the `config` database. For example:

```
mongodump --db config --port <config server port> --out <output file>
```

Step 3: Connect to `mongos`. Connect a `mongo` shell to a `mongos`.

Step 4: Find the `jumbo` Chunk. Run `sh.status(true)` to find the chunk labeled `jumbo`.

```
sh.status(true)
```

For example, the following output from sh.status(true) shows that chunk with shard key range `{ "x" : 2 } -->> { "x" : 3 }` is `jumbo`.

```
--- Sharding Status ---
  sharding version: {
     ...
  }
  shards:
     ...
  databases:
     ...
        test.foo
           shard key: { "x" : 1 }
        chunks:
             shard-b  2
             shard-a  2
        { "x" : { "$minKey" : 1 } } -->> { "x" : 1 } on : shard-b Timestamp(2, 0)
        { "x" : 1 } -->> { "x" : 2 } on : shard-a Timestamp(3, 1)
        { "x" : 2 } -->> { "x" : 3 } on : shard-a Timestamp(2, 2) jumbo
        { "x" : 3 } -->> { "x" : { "$maxKey" : 1 } } on : shard-b Timestamp(3, 0)
```

Step 5: Update `chunks` collection. In the `chunks` collection of the `config` database, unset the `jumbo` flag for the chunk. For example,

```
db.getSiblingDB("config").chunks.update(
   { ns: "test.foo", min: { x: 2 }, jumbo: true },
   { $unset: { jumbo: "" } }
)
```

Step 6: Restart the balancer. Restart the balancer, following the steps in *Enable the Balancer* (page 803).

Step 7: Optional. Clear current cluster meta information. To ensure that `mongos` instances update their cluster information cache, run `flushRouterConfig` in the `admin` database.

```
db.adminCommand({ flushRouterConfig: 1 } )
```

Manage Shard Tags

On this page

- Tag a Shard (page 816)
- Tag a Shard Key Range (page 816)
- Remove a Tag From a Shard Key Range (page 817)
- View Existing Shard Tags (page 817)
- Additional Resource (page 817)

In a sharded cluster, you can use tags to associate specific ranges of a *shard key* with a specific *shard* or subset of shards.

Tag a Shard

Associate tags with a particular shard using the `sh.addShardTag()` method when connected to a `mongos` instance. A single shard may have multiple tags, and multiple shards may also have the same tag.

Example

The following example adds the tag `NYC` to two shards, and the tags `SFO` and `NRT` to a third shard:

```
sh.addShardTag("shard0000", "NYC")
sh.addShardTag("shard0001", "NYC")
sh.addShardTag("shard0002", "SFO")
sh.addShardTag("shard0002", "NRT")
```

You may remove tags from a particular shard using the `sh.removeShardTag()` method when connected to a `mongos` instance, as in the following example, which removes the `NRT` tag from a shard:

```
sh.removeShardTag("shard0002", "NRT")
```

Tag a Shard Key Range

To assign a tag to a range of shard keys use the `sh.addTagRange()` method when connected to a `mongos` instance. Any given shard key range may only have *one* assigned tag. You cannot overlap defined ranges, or tag the same range

more than once.

Example

Given a collection named `users` in the `records` database, sharded by the `zipcode` field. The following operations assign:

- two ranges of zip codes in Manhattan and Brooklyn the `NYC` tag
- one range of zip codes in San Francisco the `SFO` tag

```
sh.addTagRange("records.users", { zipcode: "10001" }, { zipcode: "10281" }, "NYC")
sh.addTagRange("records.users", { zipcode: "11201" }, { zipcode: "11240" }, "NYC")
sh.addTagRange("records.users", { zipcode: "94102" }, { zipcode: "94135" }, "SFO")
```

Note: Shard ranges are always inclusive of the lower value and exclusive of the upper boundary.

Remove a Tag From a Shard Key Range

The `mongod` does not provide a helper for removing a tag range. You may delete tag assignment from a shard key range by removing the corresponding document from the `tags` (page 828) collection of the `config` database.

Each document in the `tags` (page 828) holds the *namespace* of the sharded collection and a minimum shard key value.

Example

The following example removes the `NYC` tag assignment for the range of zip codes within Manhattan:

```
use config
db.tags.remove({ _id: { ns: "records.users", min: { zipcode: "10001" }}, tag: "NYC" })
```

View Existing Shard Tags

The output from `sh.status()` lists tags associated with a shard, if any, for each shard. A shard's tags exist in the shard's document in the `shards` (page 828) collection of the `config` database. To return all shards with a specific tag, use a sequence of operations that resemble the following, which will return only those shards tagged with `NYC`:

```
use config
db.shards.find({ tags: "NYC" })
```

You can find tag ranges for all *namespaces* in the `tags` (page 828) collection of the `config` database. The output of `sh.status()` displays all tag ranges. To return all shard key ranges tagged with `NYC`, use the following sequence of operations:

```
use config
db.tags.find({ tags: "NYC" })
```

Additional Resource

- Whitepaper: MongoDB Multi-Data Center Deployments[19]

[19] http://www.mongodb.com/lp/white-paper/multi-dc?jmp=docs

- Webinar: Multi-Data Center Deployment[20]

Enforce Unique Keys for Sharded Collections

On this page

- Overview (page 818)
- Procedures (page 818)

Overview

The `unique` constraint on indexes ensures that only one document can have a value for a field in a *collection*. For *sharded collections these unique indexes cannot enforce uniqueness* because insert and indexing operations are local to each shard.

MongoDB does not support creating new unique indexes in sharded collections and will not allow you to shard collections with unique indexes on fields other than the `_id` field.

If you need to ensure that a field is always unique in a sharded collection, there are three options:

1. Enforce uniqueness of the *shard key* (page 747).

   MongoDB *can* enforce uniqueness for the *shard key*. For compound shard keys, MongoDB will enforce uniqueness on the *entire* key combination, and not for a specific component of the shard key.

   You cannot specify a unique constraint on a *hashed index* (page 564).

2. Use a secondary collection to enforce uniqueness.

   Create a minimal collection that only contains the unique field and a reference to a document in the main collection. If you always insert into a secondary collection *before* inserting to the main collection, MongoDB will produce an error if you attempt to use a duplicate key.

   If you have a small data set, you may not need to shard this collection and you can create multiple unique indexes. Otherwise you can shard on a single unique key.

3. Use guaranteed unique identifiers.

   Universally unique identifiers (i.e. UUID) like the `ObjectId` are guaranteed to be unique.

Procedures

Unique Constraints on the Shard Key

Process To shard a collection using the `unique` constraint, specify the `shardCollection` command in the following form:

```
db.runCommand( { shardCollection : "test.users" , key : { email : 1 } , unique : true } );
```

Remember that the `_id` field index is always unique. By default, MongoDB inserts an `ObjectId` into the `_id` field. However, you can manually insert your own value into the `_id` field and use this as the shard key. To use the `_id` field as the shard key, use the following operation:

[20] https://www.mongodb.com/presentations/webinar-multi-data-center-deployment?jmp=docs

```
db.runCommand( { shardCollection : "test.users" } )
```

Limitations

- You can only enforce uniqueness on one single field in the collection using this method.
- If you use a compound shard key, you can only enforce uniqueness on the *combination* of component keys in the shard key.

In most cases, the best shard keys are compound keys that include elements that permit *write scaling* (page 749) and *query isolation* (page 749), as well as *high cardinality* (page 772). These ideal shard keys are not often the same keys that require uniqueness and enforcing unique values in these collections requires a different approach.

Unique Constraints on Arbitrary Fields If you cannot use a unique field as the shard key or if you need to enforce uniqueness over multiple fields, you must create another *collection* to act as a "proxy collection". This collection must contain both a reference to the original document (i.e. its `ObjectId`) and the unique key.

If you must shard this "proxy" collection, then shard on the unique key using the *above procedure* (page 818); otherwise, you can simply create multiple unique indexes on the collection.

Process Consider the following for the "proxy collection:"

```
{
  "_id" : ObjectId("...")
  "email" ": "..."
}
```

The `_id` field holds the `ObjectId` of the *document* it reflects, and the `email` field is the field on which you want to ensure uniqueness.

To shard this collection, use the following operation using the `email` field as the *shard key*:

```
db.runCommand( { shardCollection : "records.proxy" ,
                 key : { email : 1 } ,
                 unique : true } );
```

If you do not need to shard the proxy collection, use the following command to create a unique index on the `email` field:

```
db.proxy.createIndex( { "email" : 1 }, { unique : true } )
```

You may create multiple unique indexes on this collection if you do not plan to shard the `proxy` collection.

To insert documents, use the following procedure in the *JavaScript shell*:

```
db = db.getSiblingDB('records');

var primary_id = ObjectId();

db.proxy.insert({
   "_id" : primary_id
   "email" : "example@example.net"
})

// if: the above operation returns successfully,
// then continue:

db.information.insert({
```

```
   "_id" : primary_id
   "email": "example@example.net"
   // additional information...
})
```

You must insert a document into the `proxy` collection first. If this operation succeeds, the `email` field is unique, and you may continue by inserting the actual document into the `information` collection.

See

The full documentation of: `createIndex()` and `shardCollection`.

Considerations

- Your application must catch errors when inserting documents into the "proxy" collection and must enforce consistency between the two collections.
- If the proxy collection requires sharding, you must shard on the single field on which you want to enforce uniqueness.
- To enforce uniqueness on more than one field using sharded proxy collections, you must have *one* proxy collection for *every* field for which to enforce uniqueness. If you create multiple unique indexes on a single proxy collection, you will *not* be able to shard proxy collections.

Use Guaranteed Unique Identifier The best way to ensure a field has unique values is to generate universally unique identifiers (UUID,) such as MongoDB's `ObjectId` values.

This approach is particularly useful for the''_id'' field, which *must* be unique: for collections where you are *not* sharding by the `_id` field the application is responsible for ensuring that the `_id` field is unique.

Shard GridFS Data Store

On this page

- `files` Collection (page 820)
- `chunks` Collection (page 820)

When sharding a *GridFS* store, consider the following:

`files` Collection

Most deployments will not need to shard the `files` collection. The `files` collection is typically small, and only contains metadata. None of the required keys for GridFS lend themselves to an even distribution in a sharded situation. If you *must* shard the `files` collection, use the `_id` field possibly in combination with an application field.

Leaving `files` unsharded means that all the file metadata documents live on one shard. For production GridFS stores you *must* store the `files` collection on a replica set.

`chunks` Collection

To shard the `chunks` collection by `{ files_id : 1 , n : 1 }`, issue commands similar to the following:

```
db.fs.chunks.createIndex( { files_id : 1 , n : 1 } )

db.runCommand( { shardCollection : "test.fs.chunks" , key : { files_id : 1 , n : 1 } } )
```

You may also want to shard using just the `file_id` field, as in the following operation:

```
db.runCommand( { shardCollection : "test.fs.chunks" , key : {  files_id : 1 } } )
```

Important: `{ files_id : 1 , n : 1 }` and `{ files_id : 1 }` are the **only** supported shard keys for the `chunks` collection of a GridFS store.

The default `files_id` value is an *ObjectId*, as a result the values of `files_id` are always ascending, and applications will insert all new GridFS data to a single chunk and shard. If your write load is too high for a single server to handle, consider a different shard key or use a different value for `_id` in the `files` collection.

13.3.4 Troubleshoot Sharded Clusters

On this page

- Config Database String Error (page 821)
- Cursor Fails Because of Stale Config Data (page 821)
- Avoid Downtime when Moving Config Servers (page 821)

This section describes common strategies for troubleshooting *sharded cluster* deployments.

Config Database String Error

Changed in version 3.2: Starting in MongoDB 3.2, config servers are deployed as replica sets by default. The `mongos` instances for the sharded cluster must specify the same config server replica set name but can specify hostname and port of different members of the replica set.

If using the deprecated topology of three mirrored `mongod` instances for config servers, `mongos` instances in a sharded cluster must specify identical `configDB` string.

Cursor Fails Because of Stale Config Data

A query returns the following warning when one or more of the `mongos` instances has not yet updated its cache of the cluster's metadata from the *config database*:

```
could not initialize cursor across all shards because : stale config detected
```

This warning *should* not propagate back to your application. The warning will repeat until all the `mongos` instances refresh their caches. To force an instance to refresh its cache, run the `flushRouterConfig` command.

Avoid Downtime when Moving Config Servers

Use CNAMEs to identify your config servers to the cluster so that you can rename and renumber your config servers without downtime.

13.4 Sharding Reference

On this page

- Sharding Methods in the `mongo` Shell (page 822)
- Sharding Database Commands (page 823)
- Reference Documentation (page 823)

13.4.1 Sharding Methods in the `mongo` Shell

Name	Description
`sh._adminCommand()`	Runs a *database command* against the admin database, like `db.runCommand()`, but can confirm that it is issued against a `mongos`.
`sh.getBalancerLockDetails()`	Reports on the active balancer lock, if it exists.
`sh._checkFullName()`	Tests a namespace to determine if its well formed.
`sh._checkMongos()`	Tests to see if the `mongo` shell is connected to a `mongos` instance.
`sh._lastMigration()`	Reports on the last *chunk* migration.
`sh.addShard()`	Adds a *shard* to a sharded cluster.
`sh.addShardTag()`	Associates a shard with a tag, to support *tag aware sharding* (page 756).
`sh.addTagRange()`	Associates range of shard keys with a shard tag, to support *tag aware sharding* (page 756).
`sh.removeTagRange()`	Removes an association between a range shard keys and a shard tag. Use to manage *tag aware sharding* (page 756).
`sh.disableBalancing()`	Disable balancing on a single collection in a sharded database. Does not affect balancing of other collections in a sharded cluster.
`sh.enableBalancing()`	Activates the sharded collection balancer process if previously disabled using `sh.disableBalancing()`.
`sh.enableSharding()`	Enables sharding on a specific database.
`sh.getBalancerHost()`	Returns the name of a `mongos` that's responsible for the balancer process.
`sh.getBalancerState()`	Returns a boolean to report if the *balancer* is currently enabled.
`sh.help()`	Returns help text for the `sh` methods.
`sh.isBalancerRunning()`	Returns a boolean to report if the balancer process is currently migrating chunks.
`sh.moveChunk()`	Migrates a *chunk* in a *sharded cluster*.
`sh.removeShardTag()`	Removes the association between a shard and a shard tag.
`sh.setBalancerState()`	Enables or disables the *balancer* which migrates *chunks* between *shards*.
`sh.shardCollection()`	Enables sharding for a collection.
`sh.splitAt()`	Divides an existing *chunk* into two chunks using a specific value of the *shard key* as the dividing point.
`sh.splitFind()`	Divides an existing *chunk* that contains a document matching a query into two approximately equal chunks.
`sh.startBalancer()`	Enables the *balancer* and waits for balancing to start.
`sh.status()`	Reports on the status of a *sharded cluster*, as `db.printShardingStatus()`.
`sh.stopBalancer()`	Disables the *balancer* and waits for any in progress balancing rounds to complete.
`sh.waitForBalancer()`	Internal. Waits for the balancer state to change.
`sh.waitForBalancerOff()`	Internal. Waits until the balancer stops running.
`sh.waitForDLock()`	Internal. Waits for a specified distributed *sharded cluster* lock.
`sh.waitForPingChange()`	Internal. Waits for a change in ping state from one of the `mongos` in the sharded cluster.

13.4.2 Sharding Database Commands

The following database commands support *sharded clusters*.

Name	Description
`flushRouterConfig`	Forces an update to the cluster metadata cached by a `mongos`.
`addShard`	Adds a *shard* to a *sharded cluster*.
`cleanupOrphaned`	Removes orphaned data with shard key values outside of the ranges of the chunks owned by a shard.
`checkShardingIndex`	Internal command that validates index on shard key.
`enableSharding`	Enables sharding on a specific database.
`listShards`	Returns a list of configured shards.
`removeShard`	Starts the process of removing a shard from a sharded cluster.
`getShardMap`	Internal command that reports on the state of a sharded cluster.
`getShardVersion`	Internal command that returns the *config server* version.
`mergeChunks`	Provides the ability to combine chunks on a single shard.
`setShardVersion`	Internal command to sets the *config server* version.
`shardCollection`	Enables the sharding functionality for a collection, allowing the collection to be sharded.
`shardingState`	Reports whether the `mongod` is a member of a sharded cluster.
`unsetSharding`	Internal command that affects connections between instances in a MongoDB deployment.
`split`	Creates a new *chunk*.
`splitChunk`	Internal command to split chunk. Instead use the methods `sh.splitFind()` and `sh.splitAt()`.
`splitVector`	Internal command that determines split points.
`medianKey`	Deprecated internal command. See `splitVector`.
`moveChunk`	Internal command that migrates chunks between shards.
`movePrimary`	Reassigns the *primary shard* when removing a shard from a sharded cluster.
`isdbgrid`	Verifies that a process is a `mongos`.

13.4.3 Reference Documentation

Config Database **(page 823)** Complete documentation of the content of the `local` database that MongoDB uses to store sharded cluster metadata.

Config Database

On this page

- Collections (page 824)

The `config` database supports *sharded cluster* operation. See the *Sharding* (page 733) section of this manual for full documentation of sharded clusters.

Important: The schema of the `config` database is *internal* and may change between releases of MongoDB. The `config` database is not a dependable API, and users should not write data to the `config` database in the course of normal operation or maintenance.

To access the `config` database, connect to a `mongos` instance in a sharded cluster, and use the following helper:

```
use config
```

You can return a list of the collections, with the following helper:

```
show collections
```

The config database is for internal use only, and during normal operations you should never manually insert or store data in it. However, if you need to verify the write availability of a config server, you can insert a document into a test collection (after making sure that no collection of that name already exists):

> **Warning:** Modification of the `config` database on a functioning system may lead to instability or inconsistent data sets. If you must modify the `config` database, use `mongodump` to create a full backup of the `config` database.

```
db.testConfigServerWriteAvail.insert( { a : 1 } )
```

If the operation succeeds, the config server is available to process writes.

Future releases of the server may include different collections in the config database, so be careful when selecting a name for your test collection.

Collections

config

config.**changelog**

Internal MongoDB Metadata

The `config` (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The `changelog` (page 824) collection stores a document for each change to the metadata of a sharded collection.

Example

The following example displays a single record of a chunk split from a `changelog` (page 824) collection:

```
{
 "_id" : "<hostname>-<timestamp>-<increment>",
 "server" : "<hostname><:port>",
 "clientAddr" : "127.0.0.1:63381",
 "time" : ISODate("2012-12-11T14:09:21.039Z"),
 "what" : "split",
 "ns" : "<database>.<collection>",
 "details" : {
    "before" : {
       "min" : {
          "<database>" : { $minKey : 1 }
       },
       "max" : {
          "<database>" : { $maxKey : 1 }
       },
       "lastmod" : Timestamp(1000, 0),
       "lastmodEpoch" : ObjectId("000000000000000000000000")
```

```
    },
    "left" : {
       "min" : {
          "<database>" : { $minKey : 1 }
       },
       "max" : {
          "<database>" : "<value>"
       },
       "lastmod" : Timestamp(1000, 1),
       "lastmodEpoch" : ObjectId(<...>)
    },
    "right" : {
       "min" : {
          "<database>" : "<value>"
       },
       "max" : {
          "<database>" : { $maxKey : 1 }
       },
       "lastmod" : Timestamp(1000, 2),
       "lastmodEpoch" : ObjectId("<...>")
    }
 }
}
```

Each document in the changelog (page 824) collection contains the following fields:

config.changelog.**_id**
: The value of changelog._id is: <hostname>-<timestamp>-<increment>.

config.changelog.**server**
: The hostname of the server that holds this data.

config.changelog.**clientAddr**
: A string that holds the address of the client, a mongos instance that initiates this change.

config.changelog.**time**
: A *ISODate* timestamp that reflects when the change occurred.

config.changelog.**what**
: Reflects the type of change recorded. Possible values are:

  - dropCollection
  - dropCollection.start
  - dropDatabase
  - dropDatabase.start
  - moveChunk.start
  - moveChunk.commit
  - split
  - multi-split

config.changelog.**ns**
: Namespace where the change occurred.

config.changelog.**details**
: A *document* that contains additional details regarding the change. The structure of the details (page 825) document depends on the type of change.

config.**chunks**

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The chunks (page 826) collection stores a document for each chunk in the cluster. Consider the following example of a document for a chunk named records.pets-animal_\"cat\":

```
{
   "_id" : "mydb.foo-a_\"cat\"",
   "lastmod" : Timestamp(1000, 3),
   "lastmodEpoch" : ObjectId("5078407bd58b175c5c225fdc"),
   "ns" : "mydb.foo",
   "min" : {
         "animal" : "cat"
   },
   "max" : {
         "animal" : "dog"
   },
   "shard" : "shard0004"
}
```

These documents store the range of values for the shard key that describe the chunk in the min and max fields. Additionally the shard field identifies the shard in the cluster that "owns" the chunk.

config.**collections**

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The collections (page 826) collection stores a document for each sharded collection in the cluster. Given a collection named pets in the records database, a document in the collections (page 826) collection would resemble the following:

```
{
   "_id" : "records.pets",
   "lastmod" : ISODate("1970-01-16T15:00:58.107Z"),
   "dropped" : false,
   "key" : {
         "a" : 1
   },
   "unique" : false,
   "lastmodEpoch" : ObjectId("5078407bd58b175c5c225fdc")
}
```

config.**databases**

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The databases (page 826) collection stores a document for each database in the cluster, and tracks if the database has sharding enabled. databases (page 826) represents each database in a distinct document. When a databases have sharding enabled, the primary field holds the name of the *primary shard*.

```
{ "_id" : "admin", "partitioned" : false, "primary" : "config" }
{ "_id" : "mydb", "partitioned" : true, "primary" : "shard0000" }
```

config.**lockpings**

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The lockpings (page 827) collection keeps track of the active components in the sharded cluster. Given a cluster with a mongos running on example.com:30000, the document in the lockpings (page 827) collection would resemble:

```
{ "_id" : "example.com:30000:1350047994:16807", "ping" : ISODate("2012-10-12T18:32:54.892Z") }
```

config.**locks**

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The locks (page 827) collection stores a distributed lock. This ensures that only one mongos instance can perform administrative tasks on the cluster at once. The mongos acting as *balancer* takes a lock by inserting a document resembling the following into the locks collection.

```
{
    "_id" : "balancer",
    "process" : "example.net:40000:1350402818:16807",
    "state" : 2,
    "ts" : ObjectId("507daeedf40e1879df62e5f3"),
    "when" : ISODate("2012-10-16T19:01:01.593Z"),
    "who" : "example.net:40000:1350402818:16807:Balancer:282475249",
    "why" : "doing balance round"
}
```

If a mongos holds the balancer lock, the state field has a value of 2, which means that balancer is active. The when field indicates when the balancer began the current operation.

config.**mongos**

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The mongos (page 827) collection stores a document for each mongos instance affiliated with the cluster. mongos instances send pings to all members of the cluster every 30 seconds so the cluster can verify that the mongos is active. The ping field shows the time of the last ping, while the up field reports the uptime of the mongos as of the last ping. The cluster maintains this collection for reporting purposes.

The following document shows the status of the mongos running on example.com:30000.

```
{ "_id" : "example.com:30000", "ping" : ISODate("2012-10-12T17:08:13.538Z"), "up" : 13699, "wait
```

config.settings

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The settings (page 828) collection holds the following sharding configuration settings:

- Chunk size. To change chunk size, see *Modify Chunk Size in a Sharded Cluster* (page 813).
- Balancer status. To change status, see *Disable the Balancer* (page 802).

The following is an example settings collection:

```
{ "_id" : "chunksize", "value" : 64 }
{ "_id" : "balancer", "stopped" : false }
```

config.shards

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The shards (page 828) collection represents each shard in the cluster in a separate document. If the shard is a replica set, the host field displays the name of the replica set, then a slash, then the hostname, as in the following example:

```
{ "_id" : "shard0000", "host" : "shard1/localhost:30000" }
```

If the shard has *tags* (page 756) assigned, this document has a tags field, that holds an array of the tags, as in the following example:

```
{ "_id" : "shard0001", "host" : "localhost:30001", "tags": [ "NYC" ] }
```

config.tags

Internal MongoDB Metadata

The config (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The tags (page 828) collection holds documents for each tagged shard key range in the cluster. The documents in the tags (page 828) collection resemble the following:

```
{
    "_id" : { "ns" : "records.users", "min" : { "zipcode" : "10001" } },
    "ns" : "records.users",
    "min" : { "zipcode" : "10001" },
    "max" : { "zipcode" : "10281" },
    "tag" : "NYC"
}
```

config.**version**

Internal MongoDB Metadata

The `config` (page 824) database is internal: applications and administrators should not modify or depend upon its content in the course of normal operation.

The `version` (page 829) collection holds the current metadata version number. This collection contains only one document:

To access the `version` (page 829) collection you must use the `db.getCollection()` method. For example, to display the collection's document:

```
mongos> db.getCollection("version").find()
{ "_id" : 1, "version" : 3 }
```

CHAPTER 14

Frequently Asked Questions

14.1 FAQ: MongoDB Fundamentals

On this page

- What platforms does MongoDB support? (page 831)
- How does a collection differ from a table? (page 831)
- How do I create a database and a collection? (page 831)
- How do I define or alter the collection schema? (page 832)
- Does MongoDB support SQL? (page 832)
- Does MongoDB support transactions? (page 832)
- Does MongoDB handle caching? (page 832)
- How does MongoDB address SQL or Query injection? (page 833)

This document answers some common questions about MongoDB.

14.1.1 What platforms does MongoDB support?

For the list of supported platforms, see *Supported Platforms* (page 296).

14.1.2 How does a collection differ from a table?

Instead of tables, a MongoDB database stores its data in *collections*. A collection holds one or more *BSON documents* (page 8). Documents are analogous to records or rows in a relational database table. Each document has *one or more fields* (page 8); fields are similar to the columns in a relational database table.

See also:

SQL to MongoDB Mapping Chart (page 183), *Introduction to MongoDB* (page 3)

14.1.3 How do I create a database and a collection?

If a database does not exist, MongoDB creates the database when you first store data for that database.

If a collection does not exist, MongoDB creates the collection when you first store data for that collection. [1]

[1] You can also create a collection explicitly using `db.createCollection` if you want to specify specific options, such as maximum size or document validation rules.

As such, you can switch to a non-existent database (`use <dbname>`) and perform the following operation:

```
use myNewDB

db.myNewCollection1.insert( { x: 1 } )
db.myNewCollection2.createIndex( { a: 1 } )
```

The `insert` operation creates both the database `myNewDB` and the collection `myNewCollection1` if they do not already exist.

The `createIndex` operation, which occurs after the `myNewDB` has been created, creates the index and the collection `myNewCollection2` if the collection does not exist. If `myNewDb` did not exist, the `createIndex` operation would have also created the `myNewDB`.

14.1.4 How do I define or alter the collection schema?

You do not need to specify a schema for a collection in MongoDB. Although it is common for the documents in a collection to have a largely homogeneous structure, it is not a requirement; i.e. documents in a single collection do not need to have the same set of fields. The data type for a field can differ across documents in a collection as well.

To change the structure of the documents in a collection, update the documents to the new structure. For instance, add new fields, remove existing ones, or update the value of a field to a new type.

Changed in version 3.2: Starting in MongoDB 3.2, however, you can enforce *document validation rules* (page 250) for a collection during update and insert operations.

Some collection properties, such as specifying a maximum size, can be specified during the explicit creation of a collection and be modified. See `db.createCollection` and `collMod`. If you are not specifying these properties, you do not need to explicitly create the collection since MongoDB creates new collections when you first store data for the collections.

14.1.5 Does MongoDB support SQL?

No. However, MongoDB does support a rich query language of its own. For examples on using MongoDB's query language, see *MongoDB CRUD Tutorials* (page 136)

See also:

SQL to MongoDB Mapping Chart (page 183)

14.1.6 Does MongoDB support transactions?

MongoDB does not support multi-document transactions. However, MongoDB does provide atomic operations on a single document.

For more details on MongoDB's isolation guarantees and behavior under concurrency, see *FAQ: Concurrency* (page 835).

14.1.7 Does MongoDB handle caching?

Yes. MongoDB keeps most recently used data in RAM. If you have created indexes for your queries and your working data set fits in RAM, MongoDB serves all queries from memory.

MongoDB does not cache the query results in order to return the cached results for identical queries.

For more information on MongoDB and memory use, see *WiredTiger and Memory Use* (page 597) and *MMAPv1 and Memory Use* (page 605).

14.1.8 How does MongoDB address SQL or Query injection?

BSON

As a client program assembles a query in MongoDB, it builds a BSON object, not a string. Thus traditional SQL injection attacks are not a problem. More details and some nuances are covered below.

MongoDB represents queries as *BSON* objects. Typically `client libraries` provide a convenient, injection free, process to build these objects. Consider the following C++ example:

```
BSONObj my_query = BSON( "name" << a_name );
auto_ptr<DBClientCursor> cursor = c.query("tutorial.persons", my_query);
```

Here, `my_query` then will have a value such as `{ name : "Joe" }`. If `my_query` contained special characters, for example `,`, `:`, and `{`, the query simply wouldn't match any documents. For example, users cannot hijack a query and convert it to a delete.

JavaScript

Note: You can disable all server-side execution of JavaScript, by passing the *`--noscripting`* option on the command line or setting `security.javascriptEnabled` in a configuration file.

All of the following MongoDB operations permit you to run arbitrary JavaScript expressions directly on the server:

- `$where`
- `mapReduce`
- `group`

You must exercise care in these cases to prevent users from submitting malicious JavaScript.

Fortunately, you can express most queries in MongoDB without JavaScript and for queries that require JavaScript, you can mix JavaScript and non-JavaScript in a single query. Place all the user-supplied fields directly in a *BSON* field and pass JavaScript code to the `$where` field.

If you need to pass user-supplied values in a `$where` clause, you may escape these values with the `CodeWScope` mechanism. When you set user-submitted values as variables in the scope document, you can avoid evaluating them on the database server.

14.2 FAQ: Indexes

On this page

- How do I create an index? (page 834)
- How does an index build affect database performance? (page 834)
- How do I see what indexes exist on a collection? (page 834)
- How can I see if a query uses an index? (page 834)
- How do I determine which fields to index? (page 834)
- How can I see the size of an index? (page 835)
- How do write operations affect indexes? (page 835)

This document addresses some common questions regarding MongoDB *indexes* (page 515). For more information on indexes, see *Indexes* (page 515).

14.2.1 How do I create an index?

To create an index on a collection, use the `db.collection.createIndex()` method. Creating an index is an administrative operation. In general, applications should not call `db.collection.createIndex()` on a regular basis.

Note: Index builds can impact performance; see *How does an index build affect database performance?* (page 834). Administrators should consider the performance implications before building indexes.

14.2.2 How does an index build affect database performance?

When building an index on a collection, the database that holds the collection is unavailable for read or write operations until the index build completes. If you need to build a large index, consider building the index in the *background* (page 577). See *Index Build* (page 576) and *Build Indexes on Replica Sets* (page 579).

To return information on currently running index creation operations, see *currentOp-index-creation*. To kill a running index creation operation, see `db.killOp()`. The partially built index will be deleted.

14.2.3 How do I see what indexes exist on a collection?

To list a collection's indexes, use the `db.collection.getIndexes()` method.

14.2.4 How can I see if a query uses an index?

To inspect how MongoDB processes a query, use the `explain()` method.

14.2.5 How do I determine which fields to index?

A number of factors determine which fields to index, including *selectivity* (page 590), the support for multiple *query shapes*, and *size of the index* (page 589). For more information, see *Operational Considerations for Indexes* (page 256) and *Indexing Strategies* (page 586).

14.2.6 How can I see the size of an index?

The `db.collection.stats()` includes an `indexSizes` document which provides size information for each index on the collection.

Depending on its size, an index may not fit into RAM. An index fits into RAM when your server has enough RAM available for both the index and the rest of the *working set*. When an index is too large to fit into RAM, MongoDB must read the index from disk, which is a much slower operation than reading from RAM.

In certain cases, an index does not need to fit *entirely* into RAM. For details, see *Indexes that Hold Only Recent Values in RAM* (page 590).

14.2.7 How do write operations affect indexes?

Write operations may require updates to indexes:

- If a write operation modifies an indexed field, MongoDB updates all indexes that have the modified field as a key.
- When running with the *MMAPv1* (page 603) storage engine, if an update to a document causes the document to grow past its allocated record size, MongoDB moves the document to a new record and updates all indexes that refer to the document, regardless of the field modified.

Therefore, if your application is write-heavy, indexes might affect performance.

14.3 FAQ: Concurrency

On this page

- What type of locking does MongoDB use? (page 836)
- How granular are locks in MongoDB? (page 836)
- How do I see the status of locks on my `mongod` instances? (page 837)
- Does a read or write operation ever yield the lock? (page 837)
- Which operations lock the database? (page 837)
- Which administrative commands lock the database? (page 838)
- Does a MongoDB operation ever lock more than one database? (page 839)
- How does sharding affect concurrency? (page 839)
- How does concurrency affect a replica set primary? (page 839)
- How does concurrency affect secondaries? (page 839)
- Does MongoDB support transactions? (page 839)
- What isolation guarantees does MongoDB provide? (page 840)
- Can reads see changes that have not been committed to disk? (page 840)

Changed in version 3.0.

MongoDB allows multiple clients to read and write the same data. In order to ensure consistency, it uses locking and other *concurrency control* measures to prevent multiple clients from modifying the same piece of data simultaneously. Together, these mechanisms guarantee that all writes to a single document occur either in full or not at all and that clients never see an inconsistent view of the data.

14.3.1 What type of locking does MongoDB use?

MongoDB uses multi-granularity locking [2] that allows operations to lock at the global, database or collection level, and allows for individual storage engines to implement their own concurrency control below the collection level (e.g., at the document-level in WiredTiger).

MongoDB uses reader-writer locks that allow concurrent readers shared access to a resource, such as a database or collection, but in MMAPv1, give exclusive access to a single write operation.

In addition to a shared (S) locking mode for reads and an exclusive (X) locking mode for write operations, intent shared (IS) and intent exclusive (IX) modes indicate an intent to read or write a resource using a finer granularity lock. When locking at a certain granularity all higher levels are locked using an *intent lock*.

For example, when locking a collection for writing (using mode X), both the corresponding database lock and the global lock must be locked in intent exclusive (IX) mode. A single database can simultaneously be locked in IS and IX mode, but an exclusive (X) lock cannot coexist with any other modes, and a shared (S) lock can only coexists with intent shared (IS) locks.

Locks are fair, with reads and writes being queued in order. However, to optimize throughput, when one request is granted, all other compatible requests will be granted at the same time, potentially releasing them before a conflicting request. For example, consider a case in which an X lock was just released, and in which the conflict queue contains the following items:

IS $\rightarrow$ IS $\rightarrow$ X $\rightarrow$ X $\rightarrow$ S $\rightarrow$ IS

In strict first-in, first-out (FIFO) ordering, only the first two IS modes would be granted. Instead MongoDB will actually grant all IS and S modes, and once they all drain, it will grant X, even if new IS or S requests have been queued in the meantime. As a grant will always move all other requests ahead in the queue, no starvation of any request is possible.

14.3.2 How granular are locks in MongoDB?

Changed in version 3.0.

For WiredTiger

Beginning with version 3.0, MongoDB ships with the *WiredTiger* (page 595) storage engine.

For most read and write operations, WiredTiger uses optimistic concurrency control. WiredTiger uses only intent locks at the global, database and collection levels. When the storage engine detects conflicts between two operations, one will incur a write conflict causing MongoDB to transparently retry that operation.

Some global operations, typically short lived operations involving multiple databases, still require a global "instance-wide" lock. Some other operations, such as dropping a collection, still require an exclusive database lock.

For MMAPv1

The MMAPv1 storage engine uses collection-level locking as of the 3.0 release series, an improvement on earlier versions in which the database lock was the finest-grain lock. Third-party storage engines may either use collection-level locking or implement their own finer-grained concurrency control.

For example, if you have six collections in a database using the MMAPv1 storage engine and an operation takes a collection-level write lock, the other five collections are still available for read and write operations. An exclusive database lock makes all six collections unavailable for the duration of the operation holding the lock.

[2] See the Wikipedia page on Multiple granularity locking (http://en.wikipedia.org/wiki/Multiple_granularity_locking) for more information.

14.3.3 How do I see the status of locks on my mongod instances?

For reporting on lock utilization information on locks, use any of the following methods:

- db.serverStatus(),
- db.currentOp(),
- mongotop,
- mongostat, and/or
- the MongoDB Cloud Manager[3] or Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[4]

Specifically, the locks document in the output of serverStatus, or the locks field in the current operation reporting provides insight into the type of locks and amount of lock contention in your mongod instance.

To terminate an operation, use db.killOp().

14.3.4 Does a read or write operation ever yield the lock?

In some situations, read and write operations can yield their locks.

Long running read and write operations, such as queries, updates, and deletes, yield under many conditions. MongoDB operations can also yield locks between individual document modifications in write operations that affect multiple documents like update() with the multi parameter.

MongoDB's *MMAPv1* (page 603) storage engine uses heuristics based on its access pattern to predict whether data is likely in physical memory before performing a read. If MongoDB *predicts* that the data is not in physical memory, an operation will yield its lock while MongoDB loads the data into memory. Once data is available in memory, the operation will reacquire the lock to complete the operation.

For storage engines supporting document level *concurrency control*, such as *WiredTiger* (page 595), yielding is not necessary when accessing storage as the *intent locks*, held at the global, database and collection level, do not block other readers and writers.

Changed in version 2.6: MongoDB does not yield locks when scanning an index even if it predicts that the index is not in memory.

14.3.5 Which operations lock the database?

The following table lists common database operations and the types of locks they use.

[3] https://cloud.mongodb.com/?jmp=docs
[4] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs

Operation	Lock Type
Issue a query	Read lock
Get more data from a *cursor*	Read lock
Insert data	Write lock
Remove data	Write lock
Update data	Write lock
Map-reduce	Read lock and write lock, unless operations are specified as non-atomic. Portions of map-reduce jobs can run concurrently.
Create an index	Building an index in the foreground, which is the default, locks the database for extended periods of time.
`db.eval()` Deprecated since version 3.0.	Write lock. The `db.eval()` method takes a global write lock while evaluating the JavaScript function. To avoid taking this global write lock, you can use the `eval` command with `nolock: true`.
`eval` Deprecated since version 3.0.	Write lock. By default, `eval` command takes a global write lock while evaluating the JavaScript function. If used with `nolock: true`, the `eval` command does *not* take a global write lock while evaluating the JavaScript function. However, the logic within the JavaScript function may take write locks for write operations.
`aggregate()`	Read lock

14.3.6 Which administrative commands lock the database?

Certain administrative commands can exclusively lock the database for extended periods of time. In some deployments, for large databases, you may consider taking the `mongod` instance offline so that clients are not affected. For example, if a `mongod` is part of a *replica set*, take the `mongod` offline and let other members of the set service load while maintenance is in progress.

The following administrative operations require an exclusive (i.e. write) lock on the database for extended periods:

- `db.collection.createIndex()`, when issued *without* setting `background` to `true`,
- `reIndex`,
- `compact`,
- `db.repairDatabase()`,
- `db.createCollection()`, when creating a very large (i.e. many gigabytes) capped collection,
- `db.collection.validate()`, and
- `db.copyDatabase()`. This operation may lock all databases. See *Does a MongoDB operation ever lock more than one database?* (page 839).

The following administrative commands lock the database but only hold the lock for a very short time:

- `db.collection.dropIndex()`,
- `db.getLastError()`,
- `db.isMaster()`,
- `rs.status()` (i.e. `replSetGetStatus`),
- `db.serverStatus()`,
- `db.auth()`, and
- `db.addUser()`.

14.3.7 Does a MongoDB operation ever lock more than one database?

The following MongoDB operations lock multiple databases:

- `db.copyDatabase()` must lock the entire `mongod` instance at once.
- `db.repairDatabase()` obtains a global write lock and will block other operations until it finishes.
- *Journaling*, which is an internal operation, locks all databases for short intervals. All databases share a single journal.
- *User authentication* (page 393) requires a read lock on the `admin` database for deployments using *2.6 user credentials* (page 497). For deployments using the 2.4 schema for user credentials, authentication locks the `admin` database as well as the database the user is accessing.
- All writes to a replica set's *primary* lock both the database receiving the writes and then the `local` database for a short time. The lock for the `local` database allows the `mongod` to write to the primary's *oplog* and accounts for a small portion of the total time of the operation.

14.3.8 How does sharding affect concurrency?

Sharding improves concurrency by distributing collections over multiple `mongod` instances, allowing shard servers (i.e. `mongos` processes) to perform any number of operations concurrently to the various downstream `mongod` instances.

In a sharded cluster, locks apply to each individual shard, not to the whole cluster; i.e. each `mongod` instance is independent of the others in the shard cluster and uses its own *locks* (page 836). The operations on one `mongod` instance do not block the operations on any others.

14.3.9 How does concurrency affect a replica set primary?

With replica sets, when MongoDB writes to a collection on the *primary*, MongoDB also writes to the primary's *oplog*, which is a special collection in the `local` database. Therefore, MongoDB must lock both the collection's database and the `local` database. The `mongod` must lock both databases at the same time to keep the database consistent and ensure that write operations, even with replication, are "all-or-nothing" operations.

When writing to a *replica set*, the lock's scope applies to the *primary*.

14.3.10 How does concurrency affect secondaries?

In *replication*, MongoDB does not apply writes serially to *secondaries*. Secondaries collect oplog entries in batches and then apply those batches in parallel. Secondaries do not allow reads while applying the write operations, and apply write operations in the order that they appear in the oplog.

14.3.11 Does MongoDB support transactions?

MongoDB does not support multi-document transactions.

However, MongoDB does provide atomic operations on a single document. Often these document-level atomic operations are sufficient to solve problems that would require ACID transactions in a relational database.

For example, in MongoDB, you can embed related data in nested arrays or nested documents within a single document and update the entire document in a single atomic operation. Relational databases might represent the same kind of data with multiple tables and rows, which would require transaction support to update the data atomically.

See also:

Atomicity and Transactions (page 125)

14.3.12 What isolation guarantees does MongoDB provide?

MongoDB provides the following guarantees in the presence of concurrent read and write operations. These guarantees hold on systems configured with either the MMAPv1 or WiredTiger storage engines.

1. Write operations are atomic with respect to a single document; i.e. if a write is updating multiple fields in the document, a reader will never see the document with only some of the fields updated.

   With a standalone `mongod` instance, a set of read and write operations to a single document is serializable. With a replica set, a set of read and write operations to a single document is serializable *only* in the absence of a rollback.

2. Correctness with respect to query predicates, e.g. `db.collection.find()` will only return documents that match and `db.collection.update()` will only write to matching documents.

3. Correctness with respect to sort. For read operations that request a sort order (e.g. `db.collection.find()` or `db.collection.aggregate()`), the sort order will not be violated due to concurrent writes.

Although MongoDB provides these strong guarantees for single-document operations, read and write operations may access an arbitrary number of documents during execution. Multi-document operations do *not* occur transactionally and are not isolated from concurrent writes. This means that the following behaviors are expected under the normal operation of the system, for both the MMAPv1 and WiredTiger storage engines:

1. Non-point-in-time read operations. Suppose a read operation begins at time t_1 and starts reading documents. A write operation then commits an update to one of the documents at some later time t_2. The reader may see the updated version of the document, and therefore does not see a point-in-time snapshot of the data.

2. Non-serializable operations. Suppose a read operation reads a document d_1 at time t_1 and a write operation updates d_1 at some later time t_3. This introduces a read-write dependency such that, if the operations were to be serialized, the read operation must precede the write operation. But also suppose that the write operation updates document d_2 at time t_2 and the read operation subsequently reads d_2 at some later time t_4. This introduces a write-read dependency which would instead require the read operation to come *after* the write operation in a serializable schedule. There is a dependency cycle which makes serializability impossible.

3. Reads may miss matching documents that are updated during the course of the read operation.

See also:

Atomicity and Transactions (page 125)

14.3.13 Can reads see changes that have not been committed to disk?

Changed in version 3.2: MongoDB 3.2 introduces *readConcern* (page 889) option. Clients using `majority` `readConcern` cannot see the results of writes before they are made *durable*.

Readers, using `"local"` (page 182) `readConcern` can see the results of writes before they are made *durable*, regardless of write concern level or journaling configuration. As a result, applications may observe the following behaviors:

1. MongoDB will allow a concurrent reader to see the result of the write operation before the write is acknowledged to the client application. For details on when writes are acknowledged for different write concern levels, see *Write Concern* (page 179).

2. Reads can see data which may subsequently be rolled back in cases such as replica set failover or power loss. It does *not* mean that read operations can see documents in a partially written or otherwise inconsistent state.

Other systems refer to these semantics as *read uncommitted*.

Changed in version 3.2.

14.4 FAQ: Sharding with MongoDB

On this page

- Is sharding appropriate for a new deployment? (page 841)
- How does sharding work with replication? (page 842)
- Can I change the shard key after sharding a collection? (page 842)
- What happens to unsharded collections in sharded databases? (page 842)
- How does MongoDB distribute data across shards? (page 842)
- What happens if a client updates a document in a chunk during a migration? (page 842)
- What happens to queries if a shard is inaccessible or slow? (page 842)
- How does MongoDB distribute queries among shards? (page 843)
- How does MongoDB sort queries in sharded environments? (page 843)
- How does MongoDB ensure unique `_id` field values when using a shard key *other* than `_id`? (page 843)
- I've enabled sharding and added a second shard, but all the data is still on one server. Why? (page 843)
- Is it safe to remove old files in the `moveChunk` directory? (page 843)
- How does `mongos` use connections? (page 844)
- Why does `mongos` hold connections open? (page 844)
- Where does MongoDB report on connections used by `mongos`? (page 844)
- What does `writebacklisten` in the log mean? (page 844)
- How should administrators deal with failed migrations? (page 844)
- What is the process for moving, renaming, or changing the number of config servers? (page 844)
- When do the `mongos` servers detect config server changes? (page 845)
- Is it possible to quickly update `mongos` servers after updating a replica set configuration? (page 845)
- What does the `maxConns` setting on `mongos` do? (page 845)
- How do indexes impact queries in sharded systems? (page 845)
- Can shard keys be randomly generated? (page 845)
- Can shard keys have a non-uniform distribution of values? (page 845)
- Can you shard on the `_id` field? (page 846)
- What do `moveChunk commit failed` errors mean? (page 846)
- How does draining a shard affect the balancing of uneven chunk distribution? (page 846)

This document answers common questions about horizontal scaling using MongoDB's *sharding*.

If you don't find the answer you're looking for, check the *complete list of FAQs* (page 831) or post your question to the MongoDB User Mailing List[5].

14.4.1 Is sharding appropriate for a new deployment?

Sometimes.

If your data set fits on a single server, you should begin with an unsharded deployment.

Converting an unsharded database to a *sharded cluster* is easy and seamless, so there is *little advantage* in configuring sharding while your data set is small.

Still, all production deployments should use *replica sets* to provide high availability and disaster recovery.

[5] https://groups.google.com/forum/?fromgroups#!forum/mongodb-user

14.4.2 How does sharding work with replication?

To use replication with sharding, deploy each *shard* as a *replica set*.

14.4.3 Can I change the shard key after sharding a collection?

No.

There is no automatic support in MongoDB for changing a shard key after sharding a collection. This reality underscores the importance of choosing a good *shard key* (page 747). If you *must* change a shard key after sharding a collection, the best option is to:

- dump all data from MongoDB into an external format.
- drop the original sharded collection.
- configure sharding using a more ideal shard key.
- *pre-split* (page 808) the shard key range to ensure initial even distribution.
- restore the dumped data into MongoDB.

See `shardCollection`, `sh.shardCollection()`, the *Shard Key* (page 747), *Deploy a Sharded Cluster* (page 765), and SERVER-4000[6] for more information.

14.4.4 What happens to unsharded collections in sharded databases?

In the current implementation, all databases in a *sharded cluster* have a "primary *shard*." All unsharded collection within that database will reside on the same shard.

14.4.5 How does MongoDB distribute data across shards?

Sharding must be specifically enabled on a collection. After enabling sharding on the collection, MongoDB will assign various ranges of collection data to the different shards in the cluster. The cluster automatically corrects imbalances between shards by migrating ranges of data from one shard to another.

14.4.6 What happens if a client updates a document in a chunk during a migration?

The `mongos` routes the operation to the "old" shard, where it will succeed immediately. Then the *shard* `mongod` instances will replicate the modification to the "new" shard before the *sharded cluster* updates that chunk's "ownership," which effectively finalizes the migration process.

14.4.7 What happens to queries if a shard is inaccessible or slow?

If a *shard* is inaccessible or unavailable, queries will return with an error.

However, a client may set the `partial` query bit, which will then return results from all available shards, regardless of whether a given shard is unavailable.

If a shard is responding slowly, `mongos` will merely wait for the shard to return results.

[6] https://jira.mongodb.org/browse/SERVER-4000

14.4.8 How does MongoDB distribute queries among shards?

The exact method for distributing queries to *shards* in a *cluster* depends on the nature of the query and the configuration of the sharded cluster. Consider a sharded collection, using the *shard key* `user_id`, that has `last_login` and `email` attributes:

- For a query that selects one or more values for the `user_id` key:

  `mongos` determines which shard or shards contains the relevant data, based on the cluster metadata, and directs a query to the required shard or shards, and returns those results to the client.

- For a query that selects `user_id` and also performs a sort:

  `mongos` can make a straightforward translation of this operation into a number of queries against the relevant shards, ordered by `user_id`. When the sorted queries return from all shards, the `mongos` merges the sorted results and returns the complete result to the client.

- For queries that select on `last_login`:

  These queries must run on all shards: `mongos` must parallelize the query over the shards and perform a merge-sort on the `email` of the documents found.

14.4.9 How does MongoDB sort queries in sharded environments?

If you call the `cursor.sort()` method on a query in a sharded environment, the `mongod` for each shard will sort its results, and the `mongos` merges each shard's results before returning them to the client.

14.4.10 How does MongoDB ensure unique `_id` field values when using a shard key *other* than `_id`?

If you do not use `_id` as the shard key, then your application/client layer must be responsible for keeping the `_id` field unique. It is problematic for collections to have duplicate `_id` values.

If you're not sharding your collection by the `_id` field, then you should be sure to store a globally unique identifier in that field. The default *BSON ObjectId* (page 14) works well in this case.

14.4.11 I've enabled sharding and added a second shard, but all the data is still on one server. Why?

First, ensure that you've declared a *shard key* for your collection. Until you have configured the shard key, MongoDB will not create *chunks*, and *sharding* will not occur.

Next, keep in mind that the default chunk size is 64 MB. As a result, in most situations, the collection needs to have at least 64 MB of data before a migration will occur.

Additionally, the system which balances chunks among the servers attempts to avoid superfluous migrations. Depending on the number of shards, your shard key, and the amount of data, systems often require at least 10 chunks of data to trigger migrations.

You can run `db.printShardingStatus()` to see all the chunks present in your cluster.

14.4.12 Is it safe to remove old files in the `moveChunk` directory?

Yes. `mongod` creates these files as backups during normal *shard* balancing operations. If some error occurs during a *migration* (page 759), these files may be helpful in recovering documents affected during the migration.

Once the migration has completed successfully and there is no need to recover documents from these files, you may safely delete these files. Or, if you have an existing backup of the database that you can use for recovery, you may also delete these files after migration.

To determine if all migrations are complete, run `sh.isBalancerRunning()` while connected to a `mongos` instance.

14.4.13 How does `mongos` use connections?

Each client maintains a connection to a `mongos` instance. Each `mongos` instance maintains a pool of connections to the members of a replica set supporting the sharded cluster. Clients use connections between `mongos` and `mongod` instances one at a time. Requests are not multiplexed or pipelined. When client requests complete, the `mongos` returns the connection to the pool.

See the *System Resource Utilization* (page 373) section of the *UNIX ulimit Settings* (page 372) document.

14.4.14 Why does `mongos` hold connections open?

`mongos` uses a set of connection pools to communicate with each *shard*. These pools do not shrink when the number of clients decreases.

This can lead to an unused `mongos` with a large number of open connections. If the `mongos` is no longer in use, it is safe to restart the process to close existing connections.

14.4.15 Where does MongoDB report on connections used by `mongos`?

Connect to the `mongos` with the `mongo` shell, and run the following command:

```
db._adminCommand("connPoolStats");
```

14.4.16 What does `writebacklisten` in the log mean?

The writeback listener is a process that opens a long poll to relay writes back from a `mongod` or `mongos` after migrations to make sure they have not gone to the wrong server. The writeback listener sends writes back to the correct server if necessary.

These messages are a key part of the sharding infrastructure and should not cause concern.

14.4.17 How should administrators deal with failed migrations?

Failed migrations require no administrative intervention. Chunk migrations always preserve a consistent state. If a migration fails to complete for some reason, the *cluster* retries the operation. When the migration completes successfully, the data resides only on the new shard.

14.4.18 What is the process for moving, renaming, or changing the number of config servers?

See *Sharded Cluster Tutorials* (page 764) for information on migrating and replacing config servers.

14.4.19 When do the mongos servers detect config server changes?

mongos instances maintain a cache of the *config database* that holds the metadata for the *sharded cluster*. This metadata includes the mapping of *chunks* to *shards*.

mongos updates its cache lazily by issuing a request to a shard and discovering that its metadata is out of date. There is no way to control this behavior from the client, but you can run the flushRouterConfig command against any mongos to force it to refresh its cache.

14.4.20 Is it possible to quickly update mongos servers after updating a replica set configuration?

The mongos instances will detect these changes without intervention over time. However, if you want to force the mongos to reload its configuration, run the flushRouterConfig command against to each mongos directly.

14.4.21 What does the maxConns setting on mongos do?

The maxIncomingConnections option limits the number of connections accepted by mongos.

This is particularly useful for a mongos if you have a client that creates multiple connections and allows them to timeout rather than closing them.

In this case, set maxIncomingConnections to a value slightly higher than the maximum number of connections that the client creates, or the maximum size of the connection pool.

This setting prevents the mongos from causing connection spikes on the individual *shards*. Spikes like these may disrupt the operation and memory allocation of the *sharded cluster*.

14.4.22 How do indexes impact queries in sharded systems?

If the query does not include the *shard key*, the mongos must send the query to all shards as a "scatter/gather" operation. Each shard will, in turn, use *either* the shard key index or another more efficient index to fulfill the query.

If the query includes multiple sub-expressions that reference the fields indexed by the shard key *and* the secondary index, the mongos can route the queries to a specific shard and the shard will use the index that will allow it to fulfill most efficiently. See this presentation[7] for more information.

14.4.23 Can shard keys be randomly generated?

Shard keys can be random. Random keys ensure optimal distribution of data across the cluster.

Sharded clusters, attempt to route queries to *specific* shards when queries include the shard key as a parameter, because these directed queries are more efficient. In many cases, random keys can make it difficult to direct queries to specific shards.

14.4.24 Can shard keys have a non-uniform distribution of values?

Yes. There is no requirement that documents be evenly distributed by the shard key.

However, documents that have the same shard key *must* reside in the same *chunk* and therefore on the same server. If your sharded data set has too many documents with the exact same shard key you will not be able to distribute *those* documents across your sharded cluster.

[7] http://www.slideshare.net/mongodb/how-queries-work-with-sharding

14.4.25 Can you shard on the `_id` field?

You can use any field for the shard key. The `_id` field is a common shard key.

Be aware that `ObjectId()` values, which are the default value of the `_id` field, increment as a timestamp. As a result, when used as a shard key, all new documents inserted into the collection will initially belong to the same chunk on a single shard. Although the system will eventually divide this chunk and migrate its contents to distribute data more evenly, at any moment the cluster can only direct insert operations at a single shard. This can limit the throughput of inserts. If most of your write operations are updates, this limitation should not impact your performance. However, if you have a high insert volume, this may be a limitation.

To address this issue, MongoDB 2.4 provides *hashed shard keys* (page 748).

14.4.26 What do `moveChunk commit failed` errors mean?

At the end of a *chunk migration* (page 760), the *shard* must connect to the *config database* to update the chunk's record in the cluster metadata. If the *shard* fails to connect to the *config database*, MongoDB reports the following error:

```
ERROR: moveChunk commit failed: version is at <n>|<nn> instead of
<N>|<NN>" and "ERROR: TERMINATING"
```

When this happens, the *primary* member of the shard's replica set then terminates to protect data consistency. If a *secondary* member can access the config database, data on the shard becomes accessible again after an election.

The user will need to resolve the chunk migration failure independently. If you encounter this issue, contact the MongoDB User Group[8] or MongoDB Support[9] to address this issue.

14.4.27 How does draining a shard affect the balancing of uneven chunk distribution?

The sharded cluster balancing process controls both migrating chunks from decommissioned shards (i.e. draining) and normal cluster balancing activities. Consider the following behaviors for different versions of MongoDB in situations where you remove a shard in a cluster with an uneven chunk distribution:

- After MongoDB 2.2, the balancer first removes the chunks from the draining shard and then balances the remaining uneven chunk distribution.
- Before MongoDB 2.2, the balancer handles the uneven chunk distribution and *then* removes the chunks from the draining shard.

14.5 FAQ: Replication and Replica Sets

[8] http://groups.google.com/group/mongodb-user
[9] https://www.mongodb.org/about/support

On this page

- What kinds of replication does MongoDB support? (page 847)
- What does the term "primary" mean? (page 847)
- What does the term "secondary" mean? (page 847)
- How long does replica set failover take? (page 847)
- Does replication work over the Internet and WAN connections? (page 848)
- Can MongoDB replicate over a "noisy" connection? (page 848)
- Why use journaling if replication already provides data redundancy? (page 848)
- How many arbiters do replica sets need? (page 848)
- What information do arbiters exchange with the rest of the replica set? (page 849)
- Which members of a replica set vote in elections? (page 849)
- Do hidden members vote in replica set elections? (page 849)
- Is it normal for replica set members to use different amounts of disk space? (page 850)
- Can I rename a replica set? (page 850)

This document answers common questions about database replication in MongoDB.

If you don't find the answer you're looking for, check the *complete list of FAQs* (page 831) or post your question to the MongoDB User Mailing List[10].

14.5.1 What kinds of replication does MongoDB support?

MongoDB supports *replica sets* (page 623).

Changed in version 3.0.0: In MongoDB 3.0.0, replica sets can have up to *50 nodes* (page 942). Previous versions limited the maximum number of replica set members to 12.

MongoDB also supports master-slave replication; however, replica sets are the recommended replication topology. However, if your deployment requires more than 50 nodes, you must use master/slave replication.

14.5.2 What does the term "primary" mean?

Primary is a *replica set* (page 623) member that can accept writes. Only the primary can accept write operations. [11]

14.5.3 What does the term "secondary" mean?

Secondary nodes are the read-only nodes in *replica sets* (page 623).

14.5.4 How long does replica set failover take?

It varies, but a replica set will generally select a new primary within a minute.

For instance, it may take 10-30 seconds for the members of a *replica set* to declare a *primary* inaccessible (see `electionTimeoutMillis` (page 722)). One of the remaining secondaries holds an *election* to elect itself as a new primary. During the election, the cluster is unavailable for writes.

[10] https://groups.google.com/forum/?fromgroups#!forum/mongodb-user

[11] In *some circumstances* (page 729), two nodes in a replica set may *transiently* believe that they are the primary, but at most, one of them will be able to complete writes with `{ w: "majority" }` (page 180) write concern. The node that can complete `{ w: "majority" }` (page 180) writes is the current primary, and the other node is a former primary that has not yet recognized its demotion, typically due to a *network partition*. When this occurs, clients that connect to the former primary may observe stale data despite having requested read preference `primary` (page 728), and new writes to the former primary will eventually roll back.

The election itself may take another 10-30 seconds.

Changed in version 3.2: Starting in MongoDB 3.2, with the *replication election enhancements* (page 888), MongoDB reduces replica set failover time. See *replication election enhancements* (page 888) for details.

14.5.5 Does replication work over the Internet and WAN connections?

Yes.

For example, a deployment may maintain a *primary* and *secondary* in an East-coast data center along with a *secondary* member for disaster recovery in a West-coast data center.

See also:

Deploy a Geographically Redundant Replica Set (page 672)

14.5.6 Can MongoDB replicate over a "noisy" connection?

Yes, but not without connection failures and the obvious latency.

Members of the set will attempt to reconnect to the other members of the set in response to networking flaps. This does not require administrator intervention. However, if the network connections among the nodes in the replica set are very slow, it might not be possible for the members of the node to keep up with the replication.

If the TCP connection between the secondaries and the *primary* instance breaks, a *replica set* will automatically elect one of the *secondary* members of the set as primary.

14.5.7 Why use journaling if replication already provides data redundancy?

Journaling facilitates faster crash recovery. Prior to journaling, crashes often required `database repairs` or full data resync. Both were slow, and the first was unreliable.

Journaling is particularly useful for protection against power failures, especially if your replica set resides in a single data center or power circuit.

When a *replica set* runs with journaling, `mongod` instances can safely restart without any administrator intervention.

Note: Journaling requires some resource overhead for write operations. Journaling has no effect on read performance, however.

Journaling is enabled by default on all 64-bit builds of MongoDB v2.0 and greater.

14.5.8 How many arbiters do replica sets need?

Some configurations do not require any *arbiter* instances. Arbiters vote in *elections* for *primary* but do not replicate the data like *secondary* members.

Replica sets require a majority of the remaining nodes present to elect a primary. Arbiters allow you to construct this majority without the overhead of adding replicating nodes to the system.

There are many possible replica set *architectures* (page 636).

A replica set with an odd number of voting nodes does not need an arbiter.

A common configuration consists of two replicating nodes that include a *primary* and a *secondary*, as well as an *arbiter* for the third node. This configuration makes it possible for the set to elect a primary in the event of failure, without requiring three replicating nodes.

You may also consider adding an arbiter to a set if it has an equal number of nodes in two facilities and network partitions between the facilities are possible. In these cases, the arbiter will break the tie between the two facilities and allow the set to elect a new primary.

See also:

Replica Set Deployment Architectures (page 636)

14.5.9 What information do arbiters exchange with the rest of the replica set?

Arbiters never receive the contents of a collection but do exchange the following data with the rest of the replica set:

- Credentials used to authenticate the arbiter with the replica set. All MongoDB processes within a replica set use keyfiles. These exchanges are encrypted.
- Replica set configuration data and voting data. This information is not encrypted. Only credential exchanges are encrypted.

If your MongoDB deployment uses TLS/SSL, then all communications between arbiters and the other members of the replica set are secure. See the documentation for *Configure mongod and mongos for TLS/SSL* (page 451) for more information. Run all arbiters on secure networks, as with all MongoDB components.

See

The overview of *Arbiter Members of Replica Sets* (page **??**).

14.5.10 Which members of a replica set vote in elections?

All members of a replica set, unless the value of `votes` (page 721) is equal to `0`, vote in elections. This includes all *delayed* (page 634), *hidden* (page 633) and *secondary-only* (page 631) members. *Arbiters* (page **??**) always vote in elections and always have `1` vote.

Additionally, the `state` of the voting members also determine whether the member can vote. Only voting members in the following states are eligible to vote:

- `PRIMARY`
- `SECONDARY`
- `RECOVERING`
- `ARBITER`
- `ROLLBACK`

See also:

Replica Set Elections (page 644)

14.5.11 Do hidden members vote in replica set elections?

Hidden members (page 633) of *replica sets do* vote in elections. To exclude a member from voting in an *election*, change the value of the member's `members[n].votes` (page 721) configuration to `0`.

See also:

Replica Set Elections (page 644)

14.5.12 Is it normal for replica set members to use different amounts of disk space?

Yes.

Factors including: different oplog sizes, different levels of storage fragmentation, and MongoDB's data file preallocation can lead to some variation in storage utilization between nodes. Storage use disparities will be most pronounced when you add members at different times.

14.5.13 Can I rename a replica set?

As of MongoDB 2.6 there are no tools or functions designed specifically to rename a replica set.

You can use the backup and restore procedure described in the *Restore a Replica Set from MongoDB Backups* (page 348) tutorial to create a new replica set with the desired name. Downtime may be necessary in order to ensure parity between the original replica set and the new one.

14.6 FAQ: MongoDB Storage

On this page

- Storage Engine Fundamentals (page 850)
- Can you mix storage engines in a replica set? (page 850)
- WiredTiger Storage Engine (page 851)
- MMAPv1 Storage Engine (page 852)
- Can I manually pad documents to prevent moves during updates? (page 855)
- Data Storage Diagnostics (page 855)

This document addresses common questions regarding MongoDB's storage system.

14.6.1 Storage Engine Fundamentals

What is a storage engine?

A storage engine is the part of a database that is responsible for managing how data is stored, both in memory and on disk. Many databases support multiple storage engines, where different engines perform better for specific workloads. For example, one storage engine might offer better performance for read-heavy workloads, and another might support a higher-throughput for write operations.

See also:

Storage Engines (page 595)

14.6.2 Can you mix storage engines in a replica set?

Yes. You can have a replica set members that use different storage engines.

When designing these multi-storage engine deployments consider the following:

- the oplog on each member may need to be sized differently to account for differences in throughput between different storage engines.
- recovery from backups may become more complex if your backup captures data files from MongoDB: you may need to maintain backups for each storage engine.

14.6.3 WiredTiger Storage Engine

Can I upgrade an existing deployment to a WiredTiger?

Yes. See:

- *Change Standalone to WiredTiger* (page 597)
- *Change Replica Set to WiredTiger* (page 598)
- *Change Sharded Cluster to WiredTiger* (page 599)

How much compression does WiredTiger provide?

The ratio of compressed data to uncompressed data depends on your data and the compression library used. By default, collection data in WiredTiger use *Snappy block compression*; *zlib* compression is also available. Index data use *prefix compression* by default.

To what size should I set the WiredTiger cache?

With WiredTiger, MongoDB utilizes both the WiredTiger cache and the filesystem cache.

Changed in version 3.2: Starting in MongoDB 3.2, the WiredTiger cache, by default, will use the larger of either:

- 60% of RAM minus 1 GB, or
- 1 GB.

For systems with up to 10 GB of RAM, the new default setting is less than or equal to the 3.0 default setting (For MongoDB 3.0, the WiredTiger cache uses either 1 GB or half of the installed physical RAM, whichever is larger).

For systems with more than 10 GB of RAM, the new default setting is greater than the 3.0 setting.

Via the filesystem cache, MongoDB automatically uses all free memory that is not used by the WiredTiger cache or by other processes. Data in the filesystem cache is compressed.

To adjust the size of the WiredTiger cache, see `storage.wiredTiger.engineConfig.cacheSizeGB` and `--wiredTigerCacheSizeGB`. Avoid increasing the WiredTiger cache size above its default value.

Note: The `storage.wiredTiger.engineConfig.cacheSizeGB` only limits the size of the WiredTiger cache, not the total amount of memory used by `mongod`. The WiredTiger cache is only one component of the RAM used by MongoDB. MongoDB also automatically uses all free memory on the machine via the filesystem cache (data in the filesystem cache is compressed).

In addition, the operating system will use any free RAM to buffer filesystem blocks.

To accommodate the additional consumers of RAM, you may have to decrease WiredTiger cache size.

The default WiredTiger cache size value assumes that there is a single `mongod` instance per machine. If a single machine contains multiple MongoDB instances, then you should decrease the setting to accommodate the other `mongod` instances.

If you run `mongod` in a container (e.g. `lxc`, `cgroups`, Docker, etc.) that does *not* have access to all of the RAM available in a system, you must set `storage.wiredTiger.engineConfig.cacheSizeGB` to a value less than the amount of RAM available in the container. The exact amount depends on the other processes running in the container.

To view statistics on the cache and eviction rate, see the `wiredTiger.cache` field returned from the `serverStatus` command.

How frequently does WiredTiger write to disk?

MongoDB configures WiredTiger to create checkpoints (i.e. write the snapshot data to disk) at intervals of 60 seconds or 2 gigabytes of journal data.

For journal data, MongoDB writes to disk according to the following intervals or condition:

- New in version 3.2: Every 50 milliseconds.
- MongoDB sets checkpoints to occur in WiredTiger on user data at an interval of 60 seconds or when 2 GB of journal data has been written, whichever occurs first.
- If the write operation includes a write concern of `j: true` (page 181), WiredTiger forces a sync of the WiredTiger journal files.
- Because MongoDB uses a journal file size limit of 100 MB, WiredTiger creates a new journal file approximately every 100 MB of data. When WiredTiger creates a new journal file, WiredTiger syncs the previous journal file.

14.6.4 MMAPv1 Storage Engine

What are memory mapped files?

A memory-mapped file is a file with data that the operating system places in memory by way of the `mmap()` system call. `mmap()` thus *maps* the file to a region of virtual memory. Memory-mapped files are the critical piece of the MMAPv1 storage engine in MongoDB. By using memory mapped files, MongoDB can treat the contents of its data files as if they were in memory. This provides MongoDB with an extremely fast and simple method for accessing and manipulating data.

How do memory mapped files work?

MongoDB uses memory mapped files for managing and interacting with all data.

Memory mapping assigns files to a block of virtual memory with a direct byte-for-byte correlation. MongoDB memory maps data files to memory as it accesses documents. Unaccessed data is *not* mapped to memory.

Once mapped, the relationship between file and memory allows MongoDB to interact with the data in the file as if it were memory.

How frequently does MMAPv1 write to disk?

In the default configuration for the *MMAPv1 storage engine* (page 603), MongoDB writes to the data files on disk every 60 seconds and writes to the *journal* files roughly every 100 milliseconds.

To change the interval for writing to the data files, use the `storage.syncPeriodSecs` setting. For the journal files, see `storage.journal.commitIntervalMs` setting.

These values represent the *maximum* amount of time between the completion of a write operation and when MongoDB writes to the data files or to the journal files. In many cases MongoDB and the operating system flush data to disk more frequently, so that the above values represents a theoretical maximum.

Why are the files in my data directory larger than the data in my database?

The data files in your data directory, which is the `/data/db` directory in default configurations, might be larger than the data set inserted into the database. Consider the following possible causes:

Preallocated data files

MongoDB preallocates its data files to avoid filesystem fragmentation, and because of this, the size of these files do not necessarily reflect the size of your data.

The `storage.mmapv1.smallFiles` option will reduce the size of these files, which may be useful if you have many small databases on disk.

The `oplog`

If this `mongod` is a member of a replica set, the data directory includes the *oplog.rs* file, which is a preallocated *capped collection* in the `local` database.

The default allocation is approximately 5% of disk space on 64-bit installations. In most cases, you should not need to resize the oplog. See *Oplog Sizing* (page 657) for more information.

The `journal`

The data directory contains the journal files, which store write operations on disk before MongoDB applies them to databases. See *Journaling* (page 606).

Empty records

MongoDB maintains lists of empty records in data files as it deletes documents and collections. MongoDB can reuse this space, but will not, by default, return this space to the operating system.

To allow MongoDB to more effectively reuse the space, you can de-fragment your data. To de-fragment, use the `compact` command. The `compact` requires up to 2 gigabytes of extra disk space to run. Do not use `compact` if you are critically low on disk space. For more information on its behavior and other considerations, see `compact`.

`compact` only removes fragmentation from MongoDB data files within a collection and does not return any disk space to the operating system. To return disk space to the operating system, see *How do I reclaim disk space?* (page 853).

How do I reclaim disk space?

The following provides some options to consider when reclaiming disk space.

Note: You do not need to reclaim disk space for MongoDB to reuse freed space. See *Empty records* (page 853) for information on reuse of freed space.

repairDatabase

You can use `repairDatabase` on a database to rebuilds the database, de-fragmenting the associated storage in the process.

`repairDatabase` requires free disk space equal to the size of your current data set plus 2 gigabytes. If the volume that holds dbpath lacks sufficient space, you can mount a separate volume and use that for the repair. For additional information and considerations, see `repairDatabase`.

> **Warning:** Do not use `repairDatabase` if you are critically low on disk space.
> `repairDatabase` will block all other operations and may take a long time to complete.

You can only run `repairDatabase` on a standalone `mongod` instance.

You can also run the `repairDatabase` operation for all databases on the server by restarting your `mongod` standalone instance with the *`--repair`* and *`--repairpath`* options. All databases on the server will be unavailable during this operation.

Resync the Member of the Replica Set

For a secondary member of a replica set, you can perform a *resync of the member* (page 699) by: stopping the secondary member to resync, deleting all data and subdirectories from the member's data directory, and restarting.

For details, see *Resync a Member of a Replica Set* (page 699).

What is the working set?

Working set represents the total body of data that the application uses in the course of normal operation. Often this is a subset of the total data size, but the specific size of the working set depends on actual moment-to-moment use of the database.

If you run a query that requires MongoDB to scan every document in a collection, the working set will expand to include every document. Depending on physical memory size, this may cause documents in the working set to "page out," or to be removed from physical memory by the operating system. The next time MongoDB needs to access these documents, MongoDB may incur a hard page fault.

For best performance, the majority of your *active* set should fit in RAM.

What are page faults?

With the MMAPv1 storage engine, page faults can occur as MongoDB reads from or writes data to parts of its data files that are not currently located in physical memory. In contrast, operating system page faults happen when physical memory is exhausted and pages of physical memory are swapped to disk.

If there is free memory, then the operating system can find the page on disk and load it to memory directly. However, if there is no free memory, the operating system must:

- find a page in memory that is stale or no longer needed, and write the page to disk.
- read the requested page from disk and load it into memory.

This process, on an active system, can take a long time, particularly in comparison to reading a page that is already in memory.

See *Page Faults* (page 311) for more information.

What is the difference between soft and hard page faults?

Page faults occur when MongoDB, with the MMAP storage engine, needs access to data that isn't currently in active memory. A "hard" page fault refers to situations when MongoDB must access a disk to access the data. A "soft" page fault, by contrast, merely moves memory pages from one list to another, such as from an operating system file cache.

See *Page Faults* (page 311) for more information.

14.6.5 Can I manually pad documents to prevent moves during updates?

Changed in version 3.0.0.

With the *MMAPv1 storage engine* (page 603), an update can cause a document to move on disk if the document grows in size. To *minimize* document movements, MongoDB uses *padding*.

You should not have to pad manually because by default, MongoDB uses *Power of 2 Sized Allocations* (page 604) to add *padding automatically* (page 604). The *Power of 2 Sized Allocations* (page 604) ensures that MongoDB allocates document space in sizes that are powers of 2, which helps ensure that MongoDB can efficiently reuse free space created by document deletion or relocation as well as reduce the occurrences of reallocations in many cases.

However, *if you must* pad a document manually, you can add a temporary field to the document and then `$unset` the field, as in the following example.

Warning: Do not manually pad documents in a capped collection. Applying manual padding to a document in a capped collection can break replication. Also, the padding is not preserved if you re-sync the MongoDB instance.

```
var myTempPadding = [ "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa",
                      "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa",
                      "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa",
                      "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa"];

db.myCollection.insert( { _id: 5, paddingField: myTempPadding } );

db.myCollection.update( { _id: 5 },
                        { $unset: { paddingField: "" } }
                      )

db.myCollection.update( { _id: 5 },
                        { $set: { realField: "Some text that I might have needed padding for" } }
                      )
```

See also:

Record Allocation Strategies (page 604)

14.6.6 Data Storage Diagnostics

How can I check the size of a collection?

To view the statistics for a collection, including the data size, use the `db.collection.stats()` method from the `mongo` shell. The following example issues `db.collection.stats()` for the `orders` collection:

```
db.orders.stats();
```

MongoDB also provides the following methods to return specific sizes for the collection:

- `db.collection.dataSize()` to return data size in bytes for the collection.

- `db.collection.storageSize()` to return allocation size in bytes, including unused space.
- `db.collection.totalSize()` to return the data size plus the index size in bytes.
- `db.collection.totalIndexSize()` to return the index size in bytes.

The following script prints the statistics for each database:

```
db._adminCommand("listDatabases").databases.forEach(function (d) {
   mdb = db.getSiblingDB(d.name);
   printjson(mdb.stats());
})
```

The following script prints the statistics for each collection in each database:

```
db._adminCommand("listDatabases").databases.forEach(function (d) {
   mdb = db.getSiblingDB(d.name);
   mdb.getCollectionNames().forEach(function(c) {
      s = mdb[c].stats();
      printjson(s);
   })
})
```

How can I check the size of indexes for a collection?

To view the size of the data allocated for an index, use the `db.collection.stats()` method and check the `indexSizes` field in the returned document.

How can I get information on the storage use of a database?

The `db.stats()` method in the `mongo` shell returns the current state of the "active" database. For the description of the returned fields, see *dbStats Output*.

14.7 FAQ: MongoDB Diagnostics

On this page

- Where can I find information about a `mongod` process that stopped running unexpectedly? (page 857)
- Does TCP `keepalive` time affect MongoDB Deployments? (page 857)
- Why does MongoDB log so many "Connection Accepted" events? (page 858)
- What tools are available for monitoring MongoDB? (page 858)
- Memory Diagnostics for the MMAPv1 Storage Engine (page 859)
- Memory Diagnostics for the WiredTiger Storage Engine (page 860)
- Sharded Cluster Diagnostics (page 862)

This document provides answers to common diagnostic questions and issues.

If you don't find the answer you're looking for, check the *complete list of FAQs* (page 831) or post your question to the MongoDB User Mailing List[12].

[12] https://groups.google.com/forum/?fromgroups#!forum/mongodb-user

14.7.1 Where can I find information about a mongod process that stopped running unexpectedly?

If mongod shuts down unexpectedly on a UNIX or UNIX-based platform, and if mongod fails to log a shutdown or error message, then check your system logs for messages pertaining to MongoDB. For example, for logs located in /var/log/messages, use the following commands:

```
sudo grep mongod /var/log/messages
sudo grep score /var/log/messages
```

14.7.2 Does TCP keepalive time affect MongoDB Deployments?

If you experience socket errors between clients and servers or between members of a sharded cluster or replica set that do not have other reasonable causes, check the TCP keepalive value (e.g. on Linux systems store, the tcp_keepalive_time value). A common keepalive period is 7200 seconds (2 hours); however, different distributions and OS X may have different settings.

For MongoDB, you will have better results with shorter keepalive periods, on the order of 120 seconds (two minutes).

If your MongoDB deployment experiences keepalive-related issues, you must alter the keep alive value on *all* machines hosting MongoDB processes. This includes all machines hosting mongos or mongod servers and all machines hosting client processes that connect to MongoDB.

Note: For non-Linux systems, values greater than or equal to 600 seconds (10 minutes) will be ignored by mongod and mongos. For Linux, values greater than 300 seconds (5 minutes) will be overridden on the mongod and mongos sockets with a maximum of 300 seconds.

On Linux systems:

- To view the keep alive setting, you can use one of the following commands:

  ```
  sysctl net.ipv4.tcp_keepalive_time
  ```

  Or:

  ```
  cat /proc/sys/net/ipv4/tcp_keepalive_time
  ```

  The value is measured in seconds.

- To change the tcp_keepalive_time value, you can use one of the following command:

  ```
  sudo sysctl -w net.ipv4.tcp_keepalive_time=<value>
  ```

  Or:

  ```
  echo <value> | sudo tee /proc/sys/net/ipv4/tcp_keepalive_time
  ```

  These operations do not persist across system reboots. To persist the setting, add the following line to /etc/sysctl.conf:

  ```
  net.ipv4.tcp_keepalive_time = <value>
  ```

  On Linux, mongod and mongos processes limit the keepalive to a maximum of 300 seconds (5 minutes) on their own sockets by overriding keepalive values greater than 5 minutes.

For OS X systems:

- To view the keep alive setting, issue the following command:

```
sysctl net.inet.tcp.keepinit
```

- To change the `net.inet.tcp.keepinit` value, you can use the following command:

  ```
  sysctl -w net.inet.tcp.keepinit=<value>
  ```

  The above method for setting the TCP keepalive is not persistent; you will need to reset the value each time you reboot or restart a system. See your operating system's documentation for instructions on setting the TCP keepalive value persistently.

For Windows systems:

- To view the keep alive setting, issue the following command:

  ```
  reg query HKLM\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters /v KeepAliveTime
  ```

  The registry value is not present by default. The system default, used if the value is absent, is 7200000 *milliseconds* or `0x6ddd00` in hexadecimal.

- To change the `KeepAliveTime` value, use the following command in an Administrator *Command Prompt*, where `<value>` is expressed in hexadecimal (e.g. `0x0124c0` is 120000):

  ```
  reg add HKLM\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\ /v KeepAliveTime /d <value>
  ```

  Windows users should consider the Windows Server Technet Article on KeepAliveTime[13] for more information on setting keep alive for MongoDB deployments on Windows systems.

You will need to restart `mongod` and `mongos` servers for new system-wide keepalive settings to take effect.

14.7.3 Why does MongoDB log so many "Connection Accepted" events?

If you see a very large number connection and re-connection messages in your MongoDB log, then clients are frequently connecting and disconnecting to the MongoDB server. This is normal behavior for applications that do not use request pooling, such as CGI. Consider using FastCGI, an Apache Module, or some other kind of persistent application server to decrease the connection overhead.

If these connections do not impact your performance you can use the run-time `quiet` option or the command-line option `--quiet` to suppress these messages from the log.

14.7.4 What tools are available for monitoring MongoDB?

The MongoDB Cloud Manager[14] and Ops Manager, an on-premise solution available in MongoDB Enterprise Advanced[15] include monitoring functionality, which collects data from running MongoDB deployments and provides visualization and alerts based on that data.

For more information, see also the MongoDB Cloud Manager documentation[16] and Ops Manager documentation[17].

A full list of third-party tools is available as part of the *Monitoring for MongoDB* (page 285) documentation.

[13] https://technet.microsoft.com/en-us/library/cc957549.aspx
[14] https://cloud.mongodb.com/?jmp=docs
[15] https://www.mongodb.com/products/mongodb-enterprise-advanced?jmp=docs
[16] https://docs.cloud.mongodb.com/
[17] https://docs.opsmanager.mongodb.com/current/application

14.7.5 Memory Diagnostics for the MMAPv1 Storage Engine

Do I need to configure swap space?

Always configure systems to have swap space. Without swap, your system may not be reliant in some situations with extreme memory constraints, memory leaks, or multiple programs using the same memory. Think of the swap space as something like a steam release valve that allows the system to release extra pressure without affecting the overall functioning of the system.

Nevertheless, systems running MongoDB *do not* need swap for routine operation. Database files are *memory-mapped* (page 852) and should constitute most of your MongoDB memory use. Therefore, it is unlikely that `mongod` will ever use any swap space in normal operation. The operating system will release memory from the memory mapped files without needing swap and MongoDB can write data to the data files without needing the swap system.

What is a "working set"?

The *working set* is the portion of your data that clients access most often.

Must my working set size fit RAM?

Your working set should stay in memory to achieve good performance. Otherwise many random disk IO's will occur, and unless you are using SSD, this can be quite slow.

One area to watch specifically in managing the size of your working set is index access patterns. If you are inserting into indexes at random locations (as would happen with id's that are randomly generated by hashes), you will continually be updating the whole index. If instead you are able to create your id's in approximately ascending order (for example, day concatenated with a random id), all the updates will occur at the right side of the b-tree and the working set size for index pages will be much smaller.

It is fine if databases and thus virtual size are much larger than RAM.

How do I calculate how much RAM I need for my application?

The amount of RAM you need depends on several factors, including but not limited to:

- The relationship between *database storage* (page 850) and working set.
- The operating system's cache strategy for LRU (Least Recently Used)
- The impact of *journaling* (page 606)
- The number or rate of page faults and other MongoDB Cloud Manager gauges to detect when you need more RAM
- Each database connection thread will need up to 1 MB of RAM.

MongoDB defers to the operating system when loading data into memory from disk. It simply *memory maps* (page 852) all its data files and relies on the operating system to cache data. The OS typically evicts the least-recently-used data from RAM when it runs low on memory. For example if clients access indexes more frequently than documents, then indexes will more likely stay in RAM, but it depends on your particular usage.

To calculate how much RAM you need, you must calculate your working set size, or the portion of your data that clients use most often. This depends on your access patterns, what indexes you have, and the size of your documents. Because MongoDB uses a thread per connection model, each database connection also will need up to 1 MB of RAM, whether active or idle.

If page faults are infrequent, your working set fits in RAM. If fault rates rise higher than that, you risk performance degradation. This is less critical with SSD drives than with spinning disks.

How do I read memory statistics in the UNIX `top` command

Because `mongod` uses *memory-mapped files* (page 852), the memory statistics in `top` require interpretation in a special way. On a large database, `VSIZE` (virtual bytes) tends to be the size of the entire database. If the `mongod` doesn't have other processes running, `RSIZE` (resident bytes) is the total memory of the machine, as this counts file system cache contents.

For Linux systems, use the `vmstat` command to help determine how the system uses memory. On OS X systems use `vm_stat`.

14.7.6 Memory Diagnostics for the WiredTiger Storage Engine

Must my working set size fit RAM?

No.

If the cache does not have enough space to load additional data, WiredTiger evicts pages from the cache to free up space.

Note: The `storage.wiredTiger.engineConfig.cacheSizeGB` only limits the size of the WiredTiger cache, not the total amount of memory used by `mongod`. The WiredTiger cache is only one component of the RAM used by MongoDB. MongoDB also automatically uses all free memory on the machine via the filesystem cache (data in the filesystem cache is compressed).

In addition, the operating system will use any free RAM to buffer filesystem blocks.

To accommodate the additional consumers of RAM, you may have to decrease WiredTiger cache size.

The default WiredTiger cache size value assumes that there is a single `mongod` instance per machine. If a single machine contains multiple MongoDB instances, then you should decrease the setting to accommodate the other `mongod` instances.

If you run `mongod` in a container (e.g. `lxc`, `cgroups`, Docker, etc.) that does *not* have access to all of the RAM available in a system, you must set `storage.wiredTiger.engineConfig.cacheSizeGB` to a value less than the amount of RAM available in the container. The exact amount depends on the other processes running in the container.

To see statistics on the cache and eviction, use the `serverStatus` command. The `wiredTiger.cache` field holds the information on the cache and eviction.

```
...
"wiredTiger" : {
   ...
   "cache" : {
      "tracked dirty bytes in the cache" : <num>,
      "bytes currently in the cache" : <num>,
      "maximum bytes configured" : <num>,
      "bytes read into cache" :<num>,
      "bytes written from cache" : <num>,
      "pages evicted by application threads" : <num>,
      "checkpoint blocked page eviction" : <num>,
      "unmodified pages evicted" : <num>,
      "page split during eviction deepened the tree" : <num>,
      "modified pages evicted" : <num>,
```

```
    "pages selected for eviction unable to be evicted" : <num>,
    "pages evicted because they exceeded the in-memory maximum" : <num>,,
    "pages evicted because they had chains of deleted items" : <num>,
    "failed eviction of pages that exceeded the in-memory maximum" : <num>,
    "hazard pointer blocked page eviction" : <num>,
    "internal pages evicted" : <num>,
    "maximum page size at eviction" : <num>,
    "eviction server candidate queue empty when topping up" : <num>,
    "eviction server candidate queue not empty when topping up" : <num>,
    "eviction server evicting pages" : <num>,
    "eviction server populating queue, but not evicting pages" : <num>,
    "eviction server unable to reach eviction goal" : <num>,
    "pages split during eviction" : <num>,
    "pages walked for eviction" : <num>,
    "eviction worker thread evicting pages" : <num>,
    "in-memory page splits" : <num>,
    "percentage overhead" : <num>,
    "tracked dirty pages in the cache" : <num>,
    "pages currently held in the cache" : <num>,
    "pages read into cache" : <num>,
    "pages written from cache" : <num>,
},
...
```

For an explanation of some key cache and eviction statistics, such as `wiredTiger.cache.bytes currently in the cache` and `wiredTiger.cache.tracked dirty bytes in the cache`, see `wiredTiger.cache`.

To adjust the size of the WiredTiger cache, see `storage.wiredTiger.engineConfig.cacheSizeGB` and *`--wiredTigerCacheSizeGB`*. Avoid increasing the WiredTiger cache size above its default value.

How do I calculate how much RAM I need for my application?

With WiredTiger, MongoDB utilizes both the WiredTiger cache and the filesystem cache.

Changed in version 3.2: Starting in MongoDB 3.2, the WiredTiger cache, by default, will use the larger of either:

- 60% of RAM minus 1 GB, or
- 1 GB.

For systems with up to 10 GB of RAM, the new default setting is less than or equal to the 3.0 default setting (For MongoDB 3.0, the WiredTiger cache uses either 1 GB or half of the installed physical RAM, whichever is larger).

For systems with more than 10 GB of RAM, the new default setting is greater than the 3.0 setting.

Via the filesystem cache, MongoDB automatically uses all free memory that is not used by the WiredTiger cache or by other processes. Data in the filesystem cache is compressed.

To adjust the size of the WiredTiger cache, see `storage.wiredTiger.engineConfig.cacheSizeGB` and *`--wiredTigerCacheSizeGB`*. Avoid increasing the WiredTiger cache size above its default value.

Note: The `storage.wiredTiger.engineConfig.cacheSizeGB` only limits the size of the WiredTiger cache, not the total amount of memory used by `mongod`. The WiredTiger cache is only one component of the RAM used by MongoDB. MongoDB also automatically uses all free memory on the machine via the filesystem cache (data in the filesystem cache is compressed).

In addition, the operating system will use any free RAM to buffer filesystem blocks.

To accommodate the additional consumers of RAM, you may have to decrease WiredTiger cache size.

The default WiredTiger cache size value assumes that there is a single `mongod` instance per machine. If a single machine contains multiple MongoDB instances, then you should decrease the setting to accommodate the other `mongod` instances.

If you run `mongod` in a container (e.g. `lxc`, `cgroups`, Docker, etc.) that does *not* have access to all of the RAM available in a system, you must set `storage.wiredTiger.engineConfig.cacheSizeGB` to a value less than the amount of RAM available in the container. The exact amount depends on the other processes running in the container.

To view statistics on the cache and eviction rate, see the `wiredTiger.cache` field returned from the `serverStatus` command.

14.7.7 Sharded Cluster Diagnostics

The two most important factors in maintaining a successful sharded cluster are:

- *choosing an appropriate shard key* (page 747) and
- *sufficient capacity to support current and future operations* (page 744).

You can prevent most issues encountered with sharding by ensuring that you choose the best possible *shard key* for your deployment and ensure that you are always adding additional capacity to your cluster well before the current resources become saturated. Continue reading for specific issues you may encounter in a production environment.

In a new sharded cluster, why does all data remains on one shard?

Your cluster must have sufficient data for sharding to make sense. Sharding works by migrating chunks between the shards until each shard has roughly the same number of chunks.

The default chunk size is 64 megabytes. MongoDB will not begin migrations until the imbalance of chunks in the cluster exceeds the *migration threshold* (page 759). While the default chunk size is configurable with the `chunkSize` setting, these behaviors help prevent unnecessary chunk migrations, which can degrade the performance of your cluster as a whole.

If you have just deployed a sharded cluster, make sure that you have enough data to make sharding effective. If you do not have sufficient data to create more than eight 64 megabyte chunks, then all data will remain on one shard. Either lower the *chunk size* (page 762) setting, or add more data to the cluster.

As a related problem, the system will split chunks only on inserts or updates, which means that if you configure sharding and do not continue to issue insert and update operations, the database will not create any chunks. You can either wait until your application inserts data *or* *split chunks manually* (page 808).

Finally, if your shard key has a low *cardinality* (page 772), MongoDB may not be able to create sufficient splits among the data.

Why would one shard receive a disproportion amount of traffic in a sharded cluster?

In some situations, a single shard or a subset of the cluster will receive a disproportionate portion of the traffic and workload. In almost all cases this is the result of a shard key that does not effectively allow *write scaling* (page 749).

It's also possible that you have "hot chunks." In this case, you may be able to solve the problem by splitting and then migrating parts of these chunks.

In the worst case, you may have to consider re-sharding your data and *choosing a different shard key* (page 771) to correct this pattern.

What can prevent a sharded cluster from balancing?

If you have just deployed your sharded cluster, you may want to consider the *troubleshooting suggestions for a new cluster where data remains on a single shard* (page 862).

If the cluster was initially balanced, but later developed an uneven distribution of data, consider the following possible causes:

- You have deleted or removed a significant amount of data from the cluster. If you have added additional data, it may have a different distribution with regards to its shard key.
- Your *shard key* has low *cardinality* (page 772) and MongoDB cannot split the chunks any further.
- Your data set is growing faster than the balancer can distribute data around the cluster. This is uncommon and typically is the result of:
  - a *balancing window* (page 801) that is too short, given the rate of data growth.
  - an uneven distribution of *write operations* (page 749) that requires more data migration. You may have to choose a different shard key to resolve this issue.
  - poor network connectivity between shards, which may lead to chunk migrations that take too long to complete. Investigate your network configuration and interconnections between shards.

Why do chunk migrations affect sharded cluster performance?

If migrations impact your cluster or application's performance, consider the following options, depending on the nature of the impact:

1. If migrations only interrupt your clusters sporadically, you can limit the *balancing window* (page 801) to prevent balancing activity during peak hours. Ensure that there is enough time remaining to keep the data from becoming out of balance again.
2. If the balancer is always migrating chunks to the detriment of overall cluster performance:
   - You may want to attempt *decreasing the chunk size* (page 813) to limit the size of the migration.
   - Your cluster may be over capacity, and you may want to attempt to *add one or two shards* (page 773) to the cluster to distribute load.

It's also possible that your shard key causes your application to direct all writes to a single shard. This kind of activity pattern can require the balancer to migrate most data soon after writing it. Consider redeploying your cluster with a shard key that provides better *write scaling* (page 749).

CHAPTER 15

Release Notes

Always install the latest, stable version of MongoDB. See *MongoDB Version Numbers* (page 1070) for more information.

See the following release notes for an account of the changes in major versions. Release notes also include instructions for upgrade.

15.1 Current Stable Release

(*3.2-series*)

15.1.1 Release Notes for MongoDB 3.2

On this page

- Minor Releases (page 866)
- WiredTiger as Default (page 888)
- Replication Election Enhancements (page 888)
- Sharded Cluster Enhancements (page 889)
- `readConcern` (page 889)
- Partial Indexes (page 889)
- Document Validation (page 890)
- Aggregation Framework Enhancements (page 890)
- MongoDB Tools Enhancements (page 893)
- Encrypted Storage Engine (page 893)
- Text Search Enhancements (page 893)
- New Storage Engines (page 894)
- General Enhancements (page 895)
- Changes Affecting Compatibility (page 898)
- Upgrade Process (page 901)
- Known Issues in 3.2.1 (page 910)
- Known Issues in 3.2.0 (page 910)
- Download (page 911)
- Additional Resources (page 911)

Dec 8, 2015

MongoDB 3.2 is now available. Key features include WiredTiger as the default storage engine, replication election enhancements, config servers as replica sets, `readConcern`, and document validations.

OpsManager 2.0 is also available. See the Ops Manager documentation[1] and the Ops Manager release notes[2] for more information.

Minor Releases

3.2 Changelog

On this page

- 3.2.6 Changelog (page 866)
- 3.2.5 Changelog (page 868)
- 3.2.4 Changelog (page 872)
- 3.2.3 Changelog (page 877)
- 3.2.1 Changelog (page 883)

3.2.6 Changelog

Security

- SERVER-23184[3] Reduce listCollections privileges
- SERVER-23394[4] AuthorizationManager may deadlock while building role graph if profiling is enabled
- SERVER-23591[5] Avoid using rawMongoProgramOutput() in js_protection.js and js_protection_roundtrip.js
- SERVER-23838[6] Remove startup warnings for no access control and bind_ip

Sharding

- SERVER-23544[7] Race condition can allow using a SyncClusterConnection to talk to config servers even after swapping CatalogManager to CSRS mode
- SERVER-23586[8] Increase timeouts in csrs upgrade tests to reduce flakiness
- SERVER-23589[9] Run csrs upgrade tests serially to avoid overloading the test machine
- SERVER-23704[10] shard_keycount.js does not invoke anonymous function
- SERVER-23784[11] Don't use 30 second network timeout on commands sent to shards through the ShardRegistry
- SERVER-23796[12] Incorrect warning when using mongos with keyfile: Access control is not enabled for the database

[1] http://docs.opsmanager.mongodb.com/current/
[2] http://docs.opsmanager.mongodb.com/current/release-notes/application/
[3] https://jira.mongodb.org/browse/SERVER-23184
[4] https://jira.mongodb.org/browse/SERVER-23394
[5] https://jira.mongodb.org/browse/SERVER-23591
[6] https://jira.mongodb.org/browse/SERVER-23838
[7] https://jira.mongodb.org/browse/SERVER-23544
[8] https://jira.mongodb.org/browse/SERVER-23586
[9] https://jira.mongodb.org/browse/SERVER-23589
[10] https://jira.mongodb.org/browse/SERVER-23704
[11] https://jira.mongodb.org/browse/SERVER-23784
[12] https://jira.mongodb.org/browse/SERVER-23796

- SERVER-23858[13] server22767.js in noPassthrough suite is failing due to checking for wrong error code

Replication

- SERVER-23775[14] oplog default size must be differently calculated for inMemory storage engine
- SERVER-23828[15] replsets_priority1.js needs to wait for repl after elections

Query

- SERVER-7005[16] Documents containing keys with embedded null characters can be created
- SERVER-23807[17] Updates should always throw WriteConflictException on unindexing

JavaScript SERVER-23571[18] Make debug builds of SpiderMonkey distinct from –dbg in scons

Storage

- SERVER-18844[19] Reacquire the snapshot after commit/abort
- SERVER-21414[20] Add information to server status to tell if data is persisted to disk
- SERVER-22970[21] Compound background Index contains mismatched index keys and documents
- SERVER-23766[22] Remove beta startup warning for inMemory storage engine

WiredTiger

- SERVER-23504[23] Coverity analysis defect 98177: Resource leak
- SERVER-23526[24] Replication relies on storage engines reporting a non-zero size for correctness
- SERVER-23588[25] mongod with WiredTiger won't start on Windows when built with –dbg=on –opt=off
- SERVER-23682[26] WiredTiger changes for MongoDB 3.2.6

Operations

- SERVER-22043[27] count helper doesn't apply read preference
- SERVER-23044[28] Fall back to system CA certs in the shell if CA file isn't provided

[13] https://jira.mongodb.org/browse/SERVER-23858
[14] https://jira.mongodb.org/browse/SERVER-23775
[15] https://jira.mongodb.org/browse/SERVER-23828
[16] https://jira.mongodb.org/browse/SERVER-7005
[17] https://jira.mongodb.org/browse/SERVER-23807
[18] https://jira.mongodb.org/browse/SERVER-23571
[19] https://jira.mongodb.org/browse/SERVER-18844
[20] https://jira.mongodb.org/browse/SERVER-21414
[21] https://jira.mongodb.org/browse/SERVER-22970
[22] https://jira.mongodb.org/browse/SERVER-23766
[23] https://jira.mongodb.org/browse/SERVER-23504
[24] https://jira.mongodb.org/browse/SERVER-23526
[25] https://jira.mongodb.org/browse/SERVER-23588
[26] https://jira.mongodb.org/browse/SERVER-23682
[27] https://jira.mongodb.org/browse/SERVER-22043
[28] https://jira.mongodb.org/browse/SERVER-23044

Build and Packaging

- SERVER-23719[29] Control build verbosity via a VERBOSE variable rather than the –mute flag
- SERVER-23804[30] Reduce num_jobs_available on ppc64le rhel builder

Internals

- SERVER-23217[31] Hang in network_interface_asio_test
- SERVER-23474[32] set a more reasonable –dialTimeout in runMongoTool
- SERVER-23523[33] shell scripts in evergreen.yml should always exit on error
- SERVER-23566[34] Update distros (AMI) for Evergreen performance projects
- SERVER-23642[35] system_perf.yml refactoring
- SERVER-23652[36] Add automatic generation of timeseries.py to system_perf.yml
- SERVER-23655[37] Invalidate CollectionInfoCache when starting an index build
- SERVER-23762[38] ValidateAdaptor::validate() should return non-OK status if it fails.
- SERVER-23788[39] Disable sharding_csrs_upgrade on ppc64le

3.2.5 Changelog

Security SERVER-22708[40] Add exposure startup warnings

Sharding

- SERVER-17468[41] actionlog should not log every single balancer round
- SERVER-21994[42] cleanup_orphaned_basic.js
- SERVER-22081[43] Enable CSRS continuous stepdown workload in evergreen
- SERVER-22151[44] Blacklist lagged_config_secondary.js and similar tests from the sharding_csrs_continuous_config_stepdown_WT suite
- SERVER-22511[45] Blacklist sharding_rs1.js from CSRS continuous config primary step down suite because of config db writes without retry
- SERVER-22725[46] prevent concurrent exitCleanly execution in mongos

[29] https://jira.mongodb.org/browse/SERVER-23719
[30] https://jira.mongodb.org/browse/SERVER-23804
[31] https://jira.mongodb.org/browse/SERVER-23217
[32] https://jira.mongodb.org/browse/SERVER-23474
[33] https://jira.mongodb.org/browse/SERVER-23523
[34] https://jira.mongodb.org/browse/SERVER-23566
[35] https://jira.mongodb.org/browse/SERVER-23642
[36] https://jira.mongodb.org/browse/SERVER-23652
[37] https://jira.mongodb.org/browse/SERVER-23655
[38] https://jira.mongodb.org/browse/SERVER-23762
[39] https://jira.mongodb.org/browse/SERVER-23788
[40] https://jira.mongodb.org/browse/SERVER-22708
[41] https://jira.mongodb.org/browse/SERVER-17468
[42] https://jira.mongodb.org/browse/SERVER-21994
[43] https://jira.mongodb.org/browse/SERVER-22081
[44] https://jira.mongodb.org/browse/SERVER-22151
[45] https://jira.mongodb.org/browse/SERVER-22511
[46] https://jira.mongodb.org/browse/SERVER-22725

- SERVER-22767[47] mongos segfault when invoking .explain() on certain operations.
- SERVER-22794[48] Add retry to continuous config primary step-down thread when primary steps down and closes all connections
- SERVER-22918[49] SyncClusterConnection::_connect can leak DBClientConnections on failure
- SERVER-22937[50] Retry catalog operations whenever possible
- SERVER-23030[51] Increase number of iterations of aggregations performed in csrs_upgrade_during_agg.js
- SERVER-23036[52] ShardRegistry accesses _configServerCS without locking mutex
- SERVER-23283[53] RangeDeleter does not log cursor ids correctly in deleteNow()

Replication

- SERVER-21863[54] map/reduce permits documents larger than 16MB to be inserted
- SERVER-21975[55] test_command.js failed in replset
- SERVER-22130[56] Reset applier lastAppliedOptime after rollback
- SERVER-22504[57] Do not blindly add self to heartbeat member data array in the TopologyCoordinator
- SERVER-22845[58] Do not busy loop on bgsync errors
- SERVER-22873[59] disallow_adding_initialized_node2.js should handle heartbeat message set by liveness timeout
- SERVER-22929[60] remove rollback4.js
- SERVER-22933[61] Update last opTime to latest after applyOps no-op
- SERVER-22934[62] add applyOps command opTime testing
- SERVER-23003[63] Recovery problems after network partition.
- SERVER-23086[64] avoid rollbacks in replsetprio1.js
- SERVER-23274[65] Aggregate with out, then stepdown, out collection dropped.

Query

- SERVER-18468[66] Include query planning details on query log lines

[47] https://jira.mongodb.org/browse/SERVER-22767
[48] https://jira.mongodb.org/browse/SERVER-22794
[49] https://jira.mongodb.org/browse/SERVER-22918
[50] https://jira.mongodb.org/browse/SERVER-22937
[51] https://jira.mongodb.org/browse/SERVER-23030
[52] https://jira.mongodb.org/browse/SERVER-23036
[53] https://jira.mongodb.org/browse/SERVER-23283
[54] https://jira.mongodb.org/browse/SERVER-21863
[55] https://jira.mongodb.org/browse/SERVER-21975
[56] https://jira.mongodb.org/browse/SERVER-22130
[57] https://jira.mongodb.org/browse/SERVER-22504
[58] https://jira.mongodb.org/browse/SERVER-22845
[59] https://jira.mongodb.org/browse/SERVER-22873
[60] https://jira.mongodb.org/browse/SERVER-22929
[61] https://jira.mongodb.org/browse/SERVER-22933
[62] https://jira.mongodb.org/browse/SERVER-22934
[63] https://jira.mongodb.org/browse/SERVER-23003
[64] https://jira.mongodb.org/browse/SERVER-23086
[65] https://jira.mongodb.org/browse/SERVER-23274
[66] https://jira.mongodb.org/browse/SERVER-18468

- SERVER-19936[67] Performance pass on unicode-aware text processing logic (text index v3)
- SERVER-22945[68] Rewrite update_yield1.js to not depend heavily on timing

Write Operations SERVER-22947[69] fix pessimizing move in batch_executor

Aggregation SERVER-23097[70] Segfault on drop of source collection during MapReduce

Storage SERVER-21681[71] In-memory storage engine not reporting index size

WiredTiger

- SERVER-22117[72] WiredTiger journal files not deleted/ Way too many journal files
- SERVER-22791[73] Invariant failure when creating WT collection with crafted configString
- SERVER-22831[74] Low query rate with heavy cache pressure and an idle collection
- SERVER-22964[75] IX GlobalLock being held while wating for wt cache eviction

Operations

- SERVER-22493[76] MongoRunner uses non-thread-safe function to find path to mongod
- SERVER-23109[77] Typo in src/mongo/shell/collection.js

Build and Packaging

- SERVER-17563[78] GPerfTools does not build on PPC64 (Power8) platform
- SERVER-21834[79] scons detection for ppc64le needs to change
- SERVER-22090[80] ssl_cert_password.js must be disabled on ppc64
- SERVER-22110[81] ppc64le builds should detect target arch as ppc64le, not ppc64
- SERVER-22111[82] packager.py needs support for ppc64le
- SERVER-22197[83] Only select tcmalloc as the default allocator on x86 platforms
- SERVER-22313[84] ppc64le: go compiled programs should statically link with libgo

[67] https://jira.mongodb.org/browse/SERVER-19936
[68] https://jira.mongodb.org/browse/SERVER-22945
[69] https://jira.mongodb.org/browse/SERVER-22947
[70] https://jira.mongodb.org/browse/SERVER-23097
[71] https://jira.mongodb.org/browse/SERVER-21681
[72] https://jira.mongodb.org/browse/SERVER-22117
[73] https://jira.mongodb.org/browse/SERVER-22791
[74] https://jira.mongodb.org/browse/SERVER-22831
[75] https://jira.mongodb.org/browse/SERVER-22964
[76] https://jira.mongodb.org/browse/SERVER-22493
[77] https://jira.mongodb.org/browse/SERVER-23109
[78] https://jira.mongodb.org/browse/SERVER-17563
[79] https://jira.mongodb.org/browse/SERVER-21834
[80] https://jira.mongodb.org/browse/SERVER-22090
[81] https://jira.mongodb.org/browse/SERVER-22110
[82] https://jira.mongodb.org/browse/SERVER-22111
[83] https://jira.mongodb.org/browse/SERVER-22197
[84] https://jira.mongodb.org/browse/SERVER-22313

- SERVER-22986[85] Linking against SSL3 OpenSSL symbols fails on Archlinux
- SERVER-23088[86] boost 1.56 libstdc++ version detection is broken when compiling with clang
- SERVER-23254[87] eslint.py returns 1 on successful patch lint instead of 0

Internals

- SERVER-21529[88] Sign of log severity values may get lost, depending on compiler
- SERVER-21836[89] Generate spidermonkey config for ppc64
- SERVER-21843[90] secure_allocator_test fails on ppc64
- SERVER-21844[91] processinfo_test fails on ppc64
- SERVER-21845[92] ppc64 debug build fails
- SERVER-21850[93] mmapv1 fails to start on ppc64
- SERVER-22019[94] Checksum::gen should use signed char
- SERVER-22067[95] artifacts.tgz should contain stripped tests before uploading
- SERVER-22338[96] Integrate JavaScript linting into scons "lint" target
- SERVER-22339[97] Integrate JavaScript formatting into scons "lint" target
- SERVER-22391[98] Mongos 3.2.1 crashes after this error, failure _checkedOutPool.empty() src/mongo/executor/connection_pool.cpp 570
- SERVER-22468[99] Format JS code with approved style in jstests/
- SERVER-22469[100] Format JS code with approved style in src/mongo/shell & src/mongo/scripting
- SERVER-22470[101] Format JS code with approved style in Enterprise repo
- SERVER-22842[102] Support JavaScript style for clang-format
- SERVER-22856[103] Cleanup pkill logic in etc/evergreen.yml
- SERVER-22864[104] Fix minor javascript errors found by eslint in v3.2 branch
- SERVER-22871[105] splitChunk needs to check for a failed index scan

[85] https://jira.mongodb.org/browse/SERVER-22986
[86] https://jira.mongodb.org/browse/SERVER-23088
[87] https://jira.mongodb.org/browse/SERVER-23254
[88] https://jira.mongodb.org/browse/SERVER-21529
[89] https://jira.mongodb.org/browse/SERVER-21836
[90] https://jira.mongodb.org/browse/SERVER-21843
[91] https://jira.mongodb.org/browse/SERVER-21844
[92] https://jira.mongodb.org/browse/SERVER-21845
[93] https://jira.mongodb.org/browse/SERVER-21850
[94] https://jira.mongodb.org/browse/SERVER-22019
[95] https://jira.mongodb.org/browse/SERVER-22067
[96] https://jira.mongodb.org/browse/SERVER-22338
[97] https://jira.mongodb.org/browse/SERVER-22339
[98] https://jira.mongodb.org/browse/SERVER-22391
[99] https://jira.mongodb.org/browse/SERVER-22468
[100] https://jira.mongodb.org/browse/SERVER-22469
[101] https://jira.mongodb.org/browse/SERVER-22470
[102] https://jira.mongodb.org/browse/SERVER-22842
[103] https://jira.mongodb.org/browse/SERVER-22856
[104] https://jira.mongodb.org/browse/SERVER-22864
[105] https://jira.mongodb.org/browse/SERVER-22871

- SERVER-22894[106] Don't use curl when downloading jstestfuzz
- SERVER-22950[107] mongos shutdown is non-deterministic when the shutdown command is executed
- SERVER-22967[108] race in destruction of user cache invalidation thread can cause use-after-free in MongoS shutdown
- SERVER-22968[109] Add JS support in .clang-format in Enterprise repo
- SERVER-23006[110] hang_analyzer should use GDB in the mongodb tool chain
- SERVER-23007[111] hang_analyzer support on Solaris
- SERVER-23016[112] Update .eslintrc.yml for newer versions of ESLint
- SERVER-23018[113] Clean up JS linting errors in JS tests
- SERVER-23019[114] Add .eslintrc.yml file to enterprise repo
- SERVER-23020[115] Disable clang-formatting for template string in jstests/noPassthrough/update_yield1.js
- SERVER-23023[116] Disable clang-format for function values in jstests/tool/csvexport1.js
- SERVER-23066[117] killOp should accept negative opid
- SERVER-23067[118] Final round of JS linting and formatting
- SERVER-23190[119] Add build variant for rhel 7.1 on POWER to 3.2 branch
- SERVER-23288[120] Update 3.2 branch mongo-perf to check against 3.0.9 baseline
- TOOLS-1058[121] goconvey fails with gccgo
- TOOLS-1064[122] mongoimport alwasy returns 0 imported documents when compiled with gccgo

3.2.4 Changelog

Security SERVER-22237[123] Built-in role that allows full control over data, but not security or topology

Sharding

- SERVER-21758[124] Test behavior when 'nearest' config server has severe replication lag

[106] https://jira.mongodb.org/browse/SERVER-22894
[107] https://jira.mongodb.org/browse/SERVER-22950
[108] https://jira.mongodb.org/browse/SERVER-22967
[109] https://jira.mongodb.org/browse/SERVER-22968
[110] https://jira.mongodb.org/browse/SERVER-23006
[111] https://jira.mongodb.org/browse/SERVER-23007
[112] https://jira.mongodb.org/browse/SERVER-23016
[113] https://jira.mongodb.org/browse/SERVER-23018
[114] https://jira.mongodb.org/browse/SERVER-23019
[115] https://jira.mongodb.org/browse/SERVER-23020
[116] https://jira.mongodb.org/browse/SERVER-23023
[117] https://jira.mongodb.org/browse/SERVER-23066
[118] https://jira.mongodb.org/browse/SERVER-23067
[119] https://jira.mongodb.org/browse/SERVER-23190
[120] https://jira.mongodb.org/browse/SERVER-23288
[121] https://jira.mongodb.org/browse/TOOLS-1058
[122] https://jira.mongodb.org/browse/TOOLS-1064
[123] https://jira.mongodb.org/browse/SERVER-22237
[124] https://jira.mongodb.org/browse/SERVER-21758

- SERVER-22184[125] Operations that fail against a secondary in a sharded cluster may have their error message swallowed
- SERVER-22239[126] wait for replication after duplicate key error from insert operations
- SERVER-22297[127] Add targeted jstests for csrs upgrade during common operations
- SERVER-22299[128] Add a jstest that runs moveChunk directly against a mongod that is not yet sharding aware, providing an SCCC connection string for the config servers
- SERVER-22524[129] Only interrupt mapReduce on catalog manager swap if it is outputting to a sharded collection
- SERVER-22543[130] multi_write_target.js should not rely on the order of shard ids
- SERVER-22547[131] add support for config server ReplSetTest options to ShardingTest
- SERVER-22553[132] mongos_shard_failure_tolerance.js should not rely on order of shard ids
- SERVER-22569[133] Initialization of eooElement static local variable isn't thread safe with MSVC 2013
- SERVER-22584[134] Make sure IncompatibleCatalogManager errors fully propagate wherever thrown
- SERVER-22585[135] CatalogManagerLegacy needs retry logic on config server reads
- SERVER-22590[136] applyChunkOpsDeprecated retries and throws an error on preCondition no longer matching because the original write worked
- SERVER-22592[137] Remove duplicate check for 'enabled' from ShardingState::_refreshMetadata
- SERVER-22627[138] ShardRegistry should mark hosts which failed due to OperationTimeout as faulty
- SERVER-22783[139] CSRS catalog manager writes should retry on WriteConcernFailed error
- SERVER-22789[140] CSRS catalog manager writes should use writeConcern majority
- SERVER-22797[141] Calls to ShardRegistry::reload needs to be serialized
- SERVER-22822[142] Prevent mongod step down during moveChunk in balance_repl.js and sharding_rs2.js
- SERVER-22849[143] Shard registry should update config last visible opTime even on command errors
- SERVER-22859[144] SCCC config server reads need to specify slaveOk
- SERVER-22862[145] Deadlock between ReplicaSetMonitor updating the connection string for a shard and reloading the ShardRegistry

[125] https://jira.mongodb.org/browse/SERVER-22184
[126] https://jira.mongodb.org/browse/SERVER-22239
[127] https://jira.mongodb.org/browse/SERVER-22297
[128] https://jira.mongodb.org/browse/SERVER-22299
[129] https://jira.mongodb.org/browse/SERVER-22524
[130] https://jira.mongodb.org/browse/SERVER-22543
[131] https://jira.mongodb.org/browse/SERVER-22547
[132] https://jira.mongodb.org/browse/SERVER-22553
[133] https://jira.mongodb.org/browse/SERVER-22569
[134] https://jira.mongodb.org/browse/SERVER-22584
[135] https://jira.mongodb.org/browse/SERVER-22585
[136] https://jira.mongodb.org/browse/SERVER-22590
[137] https://jira.mongodb.org/browse/SERVER-22592
[138] https://jira.mongodb.org/browse/SERVER-22627
[139] https://jira.mongodb.org/browse/SERVER-22783
[140] https://jira.mongodb.org/browse/SERVER-22789
[141] https://jira.mongodb.org/browse/SERVER-22797
[142] https://jira.mongodb.org/browse/SERVER-22822
[143] https://jira.mongodb.org/browse/SERVER-22849
[144] https://jira.mongodb.org/browse/SERVER-22859
[145] https://jira.mongodb.org/browse/SERVER-22862

- SERVER-22863[146] sharding read_after_optime.js test timeout should exceed ping interval
- SERVER-22878[147] rewrite checks in csrs_upgrade_mongod_using_movechunk.js after new assert.contains implementation
- SERVER-22880[148] add requires_persistence tag to csrs_upgrade_mongod_using_movechunk.js

Replication

- SERVER-21698[149] Add error-checking for isMaster() return values in jstests/libs/election_timing_test.js
- SERVER-21972[150] improve naming of ReplicationCoordinator and TopologyCoordinator unittests
- SERVER-22269[151] ReadConcern: majority does not reflect journaled state on PRIMARY
- SERVER-22276[152] implement "j" flag in write concern apply to secondary as well as primary
- SERVER-22277[153] test "j" flag in write concern apply to secondary as well as primary
- SERVER-22287[154] Merging replica sets with replication protocol version 1 may result in two primaries
- SERVER-22426[155] priority_takeover_one_node_higher_priority.js should call ReplSetTest.awaitReplication() before stepping primary down
- SERVER-22428[156] Log read-after-optime timeouts
- SERVER-22495[157] Running without journaling doesn't set all OpTimes (lastDurableOpTime)
- SERVER-22521[158] default timeout for ReplSetTest.initiate() from 60 seconds to 120 seconds should be longer to accommodate slow hosts
- SERVER-22595[159] Reactivate rollback4.js
- SERVER-22598[160] ensure all default write concern options use sync unset
- SERVER-22617[161] SnapshotThread hits invariant due to reading oplog entries out of order
- SERVER-22683[162] enableMajorityReadConcern option cannot be disabled if specified
- SERVER-22728[163] if journaling is disabled, update durableOpTime when appliedOpTime updates
- SERVER-22731[164] give correct error message when running initiate on a non-replset mongod

[146] https://jira.mongodb.org/browse/SERVER-22863
[147] https://jira.mongodb.org/browse/SERVER-22878
[148] https://jira.mongodb.org/browse/SERVER-22880
[149] https://jira.mongodb.org/browse/SERVER-21698
[150] https://jira.mongodb.org/browse/SERVER-21972
[151] https://jira.mongodb.org/browse/SERVER-22269
[152] https://jira.mongodb.org/browse/SERVER-22276
[153] https://jira.mongodb.org/browse/SERVER-22277
[154] https://jira.mongodb.org/browse/SERVER-22287
[155] https://jira.mongodb.org/browse/SERVER-22426
[156] https://jira.mongodb.org/browse/SERVER-22428
[157] https://jira.mongodb.org/browse/SERVER-22495
[158] https://jira.mongodb.org/browse/SERVER-22521
[159] https://jira.mongodb.org/browse/SERVER-22595
[160] https://jira.mongodb.org/browse/SERVER-22598
[161] https://jira.mongodb.org/browse/SERVER-22617
[162] https://jira.mongodb.org/browse/SERVER-22683
[163] https://jira.mongodb.org/browse/SERVER-22728
[164] https://jira.mongodb.org/browse/SERVER-22731

Query

- SERVER-22344[165] certain cursor options can trigger an invariant failure in GetMoreCmd
- SERVER-22425[166] execStats in system.profile reports winning plan and rejected plans
- SERVER-22532[167] $type with invalid integer type code fails with unhelpful message and leaks memory
- SERVER-22626[168] fix $type unit tests on experimental decimal build
- SERVER-22793[169] Unbounded memory usage by long-running query using projection

Aggregation SERVER-22537[170] segfault running aggregation query

JavaScript

- SERVER-9131[171] Ensure documents with code elements do not conflict with internal JS functions
- SERVER-22587[172] Upgrade to spidermonkey 38.6.1esr

Storage

- SERVER-21419[173] The ephemeralForTest storage engine should support the fsync command
- SERVER-21924[174] Add log message for inMemory and ephemeralForTest storage engine
- SERVER-22534[175] Change ephemeral storage to update durable OpTime

WiredTiger

- SERVER-22437[176] Coverity analysis defect 77704: Redundant test
- SERVER-22438[177] Coverity analysis defect 77705: Dereference before null check
- SERVER-22570[178] WiredTiger changes for MongoDB 3.2.4
- SERVER-22691[179] Incorrect initialization order in WiredTigerKVEngine
- SERVER-22898[180] High fragmentation on WiredTiger databases under write workloads

Operations SERVER-22440[181] Shell incorrectly issues first query in legacy read mode

[165] https://jira.mongodb.org/browse/SERVER-22344
[166] https://jira.mongodb.org/browse/SERVER-22425
[167] https://jira.mongodb.org/browse/SERVER-22532
[168] https://jira.mongodb.org/browse/SERVER-22626
[169] https://jira.mongodb.org/browse/SERVER-22793
[170] https://jira.mongodb.org/browse/SERVER-22537
[171] https://jira.mongodb.org/browse/SERVER-9131
[172] https://jira.mongodb.org/browse/SERVER-22587
[173] https://jira.mongodb.org/browse/SERVER-21419
[174] https://jira.mongodb.org/browse/SERVER-21924
[175] https://jira.mongodb.org/browse/SERVER-22534
[176] https://jira.mongodb.org/browse/SERVER-22437
[177] https://jira.mongodb.org/browse/SERVER-22438
[178] https://jira.mongodb.org/browse/SERVER-22570
[179] https://jira.mongodb.org/browse/SERVER-22691
[180] https://jira.mongodb.org/browse/SERVER-22898
[181] https://jira.mongodb.org/browse/SERVER-22440

Build and Packaging

- SERVER-20930[182] RPM package overwrites /etc/sysconfig/mongod
- SERVER-22003[183] inMemory windows build variant should be run less often

Tools TOOLS-1043[184] mongorestore –noIndexRestore inhibits empty collection creation

Internals

- SERVER-14501[185] De-inline ReplSettings class
- SERVER-21881[186] dbhash checking in FSM framework doesn't handle TTL deletes
- SERVER-22101[187] Generate minidumps when the hang analyzer is triggered on Windows
- SERVER-22231[188] Add additional test suites to run resmoke.py validation hook
- SERVER-22292[189] Use more reliable mechanism in the mongo shell to wait for process to terminate on windows
- SERVER-22314[190] Fix the detection of Python processes in the hang analyzer script
- SERVER-22317[191] Make checkReplDBHash hook work with dbhash quirks on 3.2
- SERVER-22332[192] Move the repl_write_threads_start_param.js JS test out of the jsCore suite
- SERVER-22340[193] Fix JS lint errors in src/mongo/shell & src/mongo/scripting with ESLint –fix
- SERVER-22341[194] Fix JS lint errors in jstests/ with ESLint –fix
- SERVER-22342[195] Fix JS lint errors in the enterprise repo with ESLint –fix
- SERVER-22479[196] upgrade_downgrade_mongod.js incorrectly checks if the TTL monitor has run
- SERVER-22513[197] Don't redirect jstestfuzz self-test's stderr.
- SERVER-22539[198] Add an ESLint configuration file
- SERVER-22546[199] Enable more ESLint rules
- SERVER-22559[200] Add –retry option to curl command in evergreen.yml
- SERVER-22597[201] Fix minor javascript errors found by eslint
- SERVER-22636[202] Disable jstestfuzz's self-tests on Evergreen

[182] https://jira.mongodb.org/browse/SERVER-20930
[183] https://jira.mongodb.org/browse/SERVER-22003
[184] https://jira.mongodb.org/browse/TOOLS-1043
[185] https://jira.mongodb.org/browse/SERVER-14501
[186] https://jira.mongodb.org/browse/SERVER-21881
[187] https://jira.mongodb.org/browse/SERVER-22101
[188] https://jira.mongodb.org/browse/SERVER-22231
[189] https://jira.mongodb.org/browse/SERVER-22292
[190] https://jira.mongodb.org/browse/SERVER-22314
[191] https://jira.mongodb.org/browse/SERVER-22317
[192] https://jira.mongodb.org/browse/SERVER-22332
[193] https://jira.mongodb.org/browse/SERVER-22340
[194] https://jira.mongodb.org/browse/SERVER-22341
[195] https://jira.mongodb.org/browse/SERVER-22342
[196] https://jira.mongodb.org/browse/SERVER-22479
[197] https://jira.mongodb.org/browse/SERVER-22513
[198] https://jira.mongodb.org/browse/SERVER-22539
[199] https://jira.mongodb.org/browse/SERVER-22546
[200] https://jira.mongodb.org/browse/SERVER-22559
[201] https://jira.mongodb.org/browse/SERVER-22597
[202] https://jira.mongodb.org/browse/SERVER-22636

- SERVER-22641[203] Disable clang-format for template strings in JS code
- SERVER-22732[204] assert.contains() has unreachable code after return
- SERVER-22746[205] don't run CheckReplDBHash on 3.2 on the config database
- SERVER-22776[206] CheckReplDBHash hook should dump the oplog upon failure
- SERVER-22806[207] CheckReplDBHash hook should fsync before doing await_repl
- SERVER-22846[208] Add applyOps command to readConcern passthrough
- SERVER-22850[209] Clean up additional javascript issues found by eslint

3.2.3 Changelog

Sharding

- SERVER-18671[210] SecondaryPreferred can end up using unversioned connections
- SERVER-20030[211] ForwardingCatalogManager::shutdown races with _replaceCatalogManager
- SERVER-20036[212] Add interruption points to operations that hold distributed locks for a long time
- SERVER-20037[213] Transfer responsibility for the release of distributed locks to new catalog manager
- SERVER-20290[214] Recipient shard for migration can continue on retrieving data even after donor shard aborts
- SERVER-20418[215] Make sure mongod and mongos always start the distlock pinger when running in SCCC mode
- SERVER-20422[216] setShardVersion configdb string mismatch during config rs upgrade
- SERVER-20580[217] Failure in csrs_upgrade_during_migrate.js
- SERVER-20694[218] user-initiated finds against the config servers can fail with "need to swap catalog manager" error
- SERVER-21382[219] Sharding migration transfers all document deletions
- SERVER-21789[220] mongos replica set monitor should choose primary based on (rs config version, electionId)
- SERVER-21896[221] Chunk metadata will not get refreshed after shard is removed
- SERVER-21906[222] Race in ShardRegistry::reload and config.shard update can cause shard not found error

[203] https://jira.mongodb.org/browse/SERVER-22641
[204] https://jira.mongodb.org/browse/SERVER-22732
[205] https://jira.mongodb.org/browse/SERVER-22746
[206] https://jira.mongodb.org/browse/SERVER-22776
[207] https://jira.mongodb.org/browse/SERVER-22806
[208] https://jira.mongodb.org/browse/SERVER-22846
[209] https://jira.mongodb.org/browse/SERVER-22850
[210] https://jira.mongodb.org/browse/SERVER-18671
[211] https://jira.mongodb.org/browse/SERVER-20030
[212] https://jira.mongodb.org/browse/SERVER-20036
[213] https://jira.mongodb.org/browse/SERVER-20037
[214] https://jira.mongodb.org/browse/SERVER-20290
[215] https://jira.mongodb.org/browse/SERVER-20418
[216] https://jira.mongodb.org/browse/SERVER-20422
[217] https://jira.mongodb.org/browse/SERVER-20580
[218] https://jira.mongodb.org/browse/SERVER-20694
[219] https://jira.mongodb.org/browse/SERVER-21382
[220] https://jira.mongodb.org/browse/SERVER-21789
[221] https://jira.mongodb.org/browse/SERVER-21896
[222] https://jira.mongodb.org/browse/SERVER-21906

- SERVER-21956[223] applyOps does not correctly propagate operation cancellation exceptions
- SERVER-21994[224] cleanup_orphaned_basic.js
- SERVER-21995[225] Queries against sharded collections fail after upgrade to CSRS due to caching of config server string in setShardVersion
- SERVER-22010[226] min_optime_recovery.js failure in the sharding continuous config stepdown suite
- SERVER-22016[227] Fatal assertion 28723 trying to rollback applyOps on a CSRS config server
- SERVER-22027[228] AsyncResultMerger should not retry killed operations
- SERVER-22079[229] Make sharding_rs1.js more compact
- SERVER-22112[230] Circular call dependency between CatalogManager and CatalogCache
- SERVER-22113[231] Remove unused sharding-specific getLocsInRange code in dbhelpers
- SERVER-22114[232] Mongos can accumulate multiple copies of ChunkManager when a shard restarts
- SERVER-22169[233] Deadlock during CatalogManager swap from SCCC -> CSRS
- SERVER-22232[234] Increase stability of csrs_upgrade_during_migrate.js test
- SERVER-22247[235] Parsing old config.collection documents fails because of missing 'lastmodEpoch' field
- SERVER-22249[236] stats.js - Not starting chunk migration because another migration is already in progress
- SERVER-22270[237] applyOps to config rs does not wait for majority
- SERVER-22303[238] Wait longer for initial sync to finish in csrs_upgrade_during_migrate.js

Replication

- SERVER-21583[239] ApplyOps background index creation may deadlock
- SERVER-21678[240] fromMigrate flag never set for deletes in oplog
- SERVER-21744[241] Clients may fail to discover new primaries when clock skew between nodes is greater than electionTimeout
- SERVER-21958[242] Eliminate unused flags from Cloner methods
- SERVER-21988[243] Rollback does not wait for applier to finish before starting

[223] https://jira.mongodb.org/browse/SERVER-21956
[224] https://jira.mongodb.org/browse/SERVER-21994
[225] https://jira.mongodb.org/browse/SERVER-21995
[226] https://jira.mongodb.org/browse/SERVER-22010
[227] https://jira.mongodb.org/browse/SERVER-22016
[228] https://jira.mongodb.org/browse/SERVER-22027
[229] https://jira.mongodb.org/browse/SERVER-22079
[230] https://jira.mongodb.org/browse/SERVER-22112
[231] https://jira.mongodb.org/browse/SERVER-22113
[232] https://jira.mongodb.org/browse/SERVER-22114
[233] https://jira.mongodb.org/browse/SERVER-22169
[234] https://jira.mongodb.org/browse/SERVER-22232
[235] https://jira.mongodb.org/browse/SERVER-22247
[236] https://jira.mongodb.org/browse/SERVER-22249
[237] https://jira.mongodb.org/browse/SERVER-22270
[238] https://jira.mongodb.org/browse/SERVER-22303
[239] https://jira.mongodb.org/browse/SERVER-21583
[240] https://jira.mongodb.org/browse/SERVER-21678
[241] https://jira.mongodb.org/browse/SERVER-21744
[242] https://jira.mongodb.org/browse/SERVER-21958
[243] https://jira.mongodb.org/browse/SERVER-21988

- SERVER-22109[244] Invariant failure when running applyOps to create an index with a bad ns field
- SERVER-22152[245] priority_takeover_two_nodes_equal_priority.js fails if default priority node gets elected at beginning of test
- SERVER-22190[246] electionTime field not set in heartbeat response from primary under protocol version 1
- SERVER-22335[247] Do not prepare getmore when un-needed in bgsync fetcher
- SERVER-22362[248] election_timing.js waits for wrong node to become primary
- SERVER-22420[249] priority_takeover_two_nodes_equal_priority.js fails if existing primary's step down period expires
- SERVER-22456[250] The oplog find query timeout is too low

Query

- SERVER-17011[251] Cursor can return objects out of order if updated during query ("legacy" readMode only)
- SERVER-18115[252] The planner can add an unnecessary in-memory sort stage for .min()/.max() queries
- SERVER-20083[253] Add log statement at default log level for when an index filter is set or cleared successfully
- SERVER-21776[254] Move per-operation log lines for queries out of the QUERY log component
- SERVER-21869[255] Avoid wrapping of spherical queries in geo_full.js
- SERVER-22002[256] Do not retry findAndModify operations on MMAPv1
- SERVER-22100[257] memory pressure from find/getMore buffer preallocation causes concurrency suite slowness on Windows DEBUG
- SERVER-22448[258] Query planner does not filter 2dsphere Index Version 3 correctly

Write Operations

- SERVER-11983[259] Update on document without _id, in capped collection without _id index, creates an _id field
- SERVER-21647[260] $rename changes field ordering

Aggregation

- SERVER-21887[261] $sample takes disproportionately long time on newly created collection

[244] https://jira.mongodb.org/browse/SERVER-22109
[245] https://jira.mongodb.org/browse/SERVER-22152
[246] https://jira.mongodb.org/browse/SERVER-22190
[247] https://jira.mongodb.org/browse/SERVER-22335
[248] https://jira.mongodb.org/browse/SERVER-22362
[249] https://jira.mongodb.org/browse/SERVER-22420
[250] https://jira.mongodb.org/browse/SERVER-22456
[251] https://jira.mongodb.org/browse/SERVER-17011
[252] https://jira.mongodb.org/browse/SERVER-18115
[253] https://jira.mongodb.org/browse/SERVER-20083
[254] https://jira.mongodb.org/browse/SERVER-21776
[255] https://jira.mongodb.org/browse/SERVER-21869
[256] https://jira.mongodb.org/browse/SERVER-22002
[257] https://jira.mongodb.org/browse/SERVER-22100
[258] https://jira.mongodb.org/browse/SERVER-22448
[259] https://jira.mongodb.org/browse/SERVER-11983
[260] https://jira.mongodb.org/browse/SERVER-21647
[261] https://jira.mongodb.org/browse/SERVER-21887

- SERVER-22048[262] Index access stats should be recorded for $match & mapReduce

JavaScript SERVER-21528[263] Clean up core/capped6.js

Storage

- SERVER-21388[264] Invariant Failure in CappedRecordStoreV1::cappedTruncateAfter
- SERVER-22011[265] Direct writes to the local database can cause deadlock involving the WiredTiger write throttle
- SERVER-22058[266] 'not all control paths return a value' warning in non-MMAP V1 implementations of '::writingPtr'
- SERVER-22167[267] Failed to insert document larger than 256k
- SERVER-22199[268] Collection drop command during checkpoint causes complete stall until end of checkpoint

WiredTiger

- SERVER-21833[269] Compact does not release space to the system with WiredTiger
- SERVER-21944[270] WiredTiger changes for 3.2.3
- SERVER-22064[271] Coverity analysis defect 77699: Unchecked return value
- SERVER-22279[272] SubplanStage fails to register its MultiPlanStage

MMAP

- SERVER-21997[273] kill_cursors.js deadlocks
- SERVER-22261[274] MMAPv1 LSNFile may be updated ahead of what is synced to data files

Operations

- SERVER-20358[275] Usernames can contain NULL characters
- SERVER-22007[276] List all commands crashes server
- SERVER-22075[277] election_timing.js election timed out

[262] https://jira.mongodb.org/browse/SERVER-22048
[263] https://jira.mongodb.org/browse/SERVER-21528
[264] https://jira.mongodb.org/browse/SERVER-21388
[265] https://jira.mongodb.org/browse/SERVER-22011
[266] https://jira.mongodb.org/browse/SERVER-22058
[267] https://jira.mongodb.org/browse/SERVER-22167
[268] https://jira.mongodb.org/browse/SERVER-22199
[269] https://jira.mongodb.org/browse/SERVER-21833
[270] https://jira.mongodb.org/browse/SERVER-21944
[271] https://jira.mongodb.org/browse/SERVER-22064
[272] https://jira.mongodb.org/browse/SERVER-22279
[273] https://jira.mongodb.org/browse/SERVER-21997
[274] https://jira.mongodb.org/browse/SERVER-22261
[275] https://jira.mongodb.org/browse/SERVER-20358
[276] https://jira.mongodb.org/browse/SERVER-22007
[277] https://jira.mongodb.org/browse/SERVER-22075

Build and Packaging

- SERVER-21905[278] Can't compile Mongo 3.2
- SERVER-22042[279] If ssl libraries not present, configure fails with a misleading error about boost
- SERVER-22350[280] Package generation failure doesn't fail compile tasks

Tools TOOLS-1039[281] mongoexport chokes on data with quotes

Internals

- SERVER-12108[282] setup_multiversion_mongodb.py script should support downloading windows binaries
- SERVER-20409[283] Negative scaling with more than 10K connections
- SERVER-21035[284] Delete the disabled fsm_all_sharded.js test runner
- SERVER-21050[285] Add a failover workload to cause CSRS config server primary failovers
- SERVER-21309[286] Remove Install step from jstestfuzz in evergreen
- SERVER-21421[287] Update concurrency suite's ThreadManager constructor to provide default executionMode
- SERVER-21499[288] Enable fsm_all_simultaneous.js (FSM parallel mode)
- SERVER-21565[289] resmoke.py can not start replica sets with more than 7 nodes
- SERVER-21597[290] Fix connPoolStats command to work with many TaskExecutor-NetworkInterface pairs
- SERVER-21747[291] CheckReplDBHash should not print error message when the system collections differ in the presence of other errors
- SERVER-21801[292] CheckReplDBHash testing hook should check document type (resmoke.py)
- SERVER-21875[293] AttributeError in hang_analyzer.py when sending SIGKILL on Windows
- SERVER-21892[294] Include thread ID in concurrency suite error report
- SERVER-21894[295] Remove unused 'hashed' resmoke.py tags from JS tests
- SERVER-21902[296] Use multiple shard nodes in the jstestfuzz_sharded suite
- SERVER-21916[297] Add missing tasks/suites to ASan Evergreen variant

[278] https://jira.mongodb.org/browse/SERVER-21905
[279] https://jira.mongodb.org/browse/SERVER-22042
[280] https://jira.mongodb.org/browse/SERVER-22350
[281] https://jira.mongodb.org/browse/TOOLS-1039
[282] https://jira.mongodb.org/browse/SERVER-12108
[283] https://jira.mongodb.org/browse/SERVER-20409
[284] https://jira.mongodb.org/browse/SERVER-21035
[285] https://jira.mongodb.org/browse/SERVER-21050
[286] https://jira.mongodb.org/browse/SERVER-21309
[287] https://jira.mongodb.org/browse/SERVER-21421
[288] https://jira.mongodb.org/browse/SERVER-21499
[289] https://jira.mongodb.org/browse/SERVER-21565
[290] https://jira.mongodb.org/browse/SERVER-21597
[291] https://jira.mongodb.org/browse/SERVER-21747
[292] https://jira.mongodb.org/browse/SERVER-21801
[293] https://jira.mongodb.org/browse/SERVER-21875
[294] https://jira.mongodb.org/browse/SERVER-21892
[295] https://jira.mongodb.org/browse/SERVER-21894
[296] https://jira.mongodb.org/browse/SERVER-21902
[297] https://jira.mongodb.org/browse/SERVER-21916

- SERVER-21917[298] Add the httpinterface test suite to the Enterprise RHEL 6.2 variant
- SERVER-21934[299] Add extra information to OSX stack traces to facilitate addr2line translation
- SERVER-21940[300] Workload connection cache in FSM suite is not nulled out properly
- SERVER-21949[301] Add validation testing hook to resmoke.py
- SERVER-21952[302] jstestfuzz tasks should not run with –continueOnFailure
- SERVER-21959[303] Do not truncate stack traces in log messages
- SERVER-21960[304] Include symbol name in stacktrace json when available
- SERVER-21964[305] Remove startPort option from ReplSetTest options in jstests/replsets/auth1.js
- SERVER-21978[306] move_primary_basic.js should always set a fixed primary shard
- SERVER-21990[307] Deprecation warning from resmoke.py - replicaset.py insert is deprecated
- SERVER-22028[308] hang_analyzer should fail when run against unsupported lldb
- SERVER-22034[309] Server presents clusterFile certificate for incoming connections
- SERVER-22054[310] Authentication failure reports incorrect IP address
- SERVER-22055[311] Cleanup unused legacy client functionality from the server code
- SERVER-22059[312] Add the authSchemaUpgrade command to the readConcern passthrough
- SERVER-22066[313] range_deleter_test:ImmediateDelete is flaky
- SERVER-22083[314] Delete the disabled fsm_all_master_slave.js test runner
- SERVER-22098[315] Split FSM sharded tests for SCCC into a separate suite
- SERVER-22099[316] Remove unreliable check in cleanup_orphaned_basic.js
- SERVER-22120[317] No data found after force sync in no_chaining.js
- SERVER-22121[318] Add resmoke.py validation testing hook to test suites
- SERVER-22142[319] resmoke.py's FlushThread attempts to reference imported members during Python interpreter shutdown
- SERVER-22154[320] csrs_upgrade.js, csrs_upgrade_during_migrate.js should be blacklisted on in-mem

[298] https://jira.mongodb.org/browse/SERVER-21917
[299] https://jira.mongodb.org/browse/SERVER-21934
[300] https://jira.mongodb.org/browse/SERVER-21940
[301] https://jira.mongodb.org/browse/SERVER-21949
[302] https://jira.mongodb.org/browse/SERVER-21952
[303] https://jira.mongodb.org/browse/SERVER-21959
[304] https://jira.mongodb.org/browse/SERVER-21960
[305] https://jira.mongodb.org/browse/SERVER-21964
[306] https://jira.mongodb.org/browse/SERVER-21978
[307] https://jira.mongodb.org/browse/SERVER-21990
[308] https://jira.mongodb.org/browse/SERVER-22028
[309] https://jira.mongodb.org/browse/SERVER-22034
[310] https://jira.mongodb.org/browse/SERVER-22054
[311] https://jira.mongodb.org/browse/SERVER-22055
[312] https://jira.mongodb.org/browse/SERVER-22059
[313] https://jira.mongodb.org/browse/SERVER-22066
[314] https://jira.mongodb.org/browse/SERVER-22083
[315] https://jira.mongodb.org/browse/SERVER-22098
[316] https://jira.mongodb.org/browse/SERVER-22099
[317] https://jira.mongodb.org/browse/SERVER-22120
[318] https://jira.mongodb.org/browse/SERVER-22121
[319] https://jira.mongodb.org/browse/SERVER-22142
[320] https://jira.mongodb.org/browse/SERVER-22154

- SERVER-22165[321] Deadlock in resmoke.py between logger pipe and timer thread
- SERVER-22171[322] The lint task is running on 3 Evergreen variants
- SERVER-22219[323] Use the subprocess32 package on POSIX systems in resmoke.py if it's available
- SERVER-22324[324] Update findAndModify FSM workloads to handle not matching anything
- TOOLS-1028[325] expose qr/qw and ar/aw fields in mongostat JSON output mode.

3.2.1 Changelog

Security

- SERVER-21724[326] Backup role can't read system.profile
- SERVER-21824[327] Disable kmip.js test in ESE suite; re-enable once fixed
- SERVER-21890[328] Create a flag to allow server realm to be specified explicitly on Windows

Sharding

- SERVER-20824[329] Test for sharding state recovery
- SERVER-21076[330] Write tests to ensure that operations using DBDirectClient handle shard versioning properly
- SERVER-21132[331] Add more basic tests for moveChunk
- SERVER-21133[332] Add more basic test for mergeChunk
- SERVER-21134[333] Add more basic tests for shardCollection
- SERVER-21135[334] Add more basic tests for sharded implicit database creation
- SERVER-21136[335] Add more basic tests for enableSharding
- SERVER-21137[336] Add more basic tests for movePrimary
- SERVER-21138[337] Add more basic tests for dropDatabase
- SERVER-21139[338] Add more basic tests for drop collection
- SERVER-21366[339] Long-running transactions in MigrateStatus::apply
- SERVER-21586[340] Investigate v3.0 mongos and v3.2 cluster compatibility issues in jstests/sharding

[321] https://jira.mongodb.org/browse/SERVER-22165
[322] https://jira.mongodb.org/browse/SERVER-22171
[323] https://jira.mongodb.org/browse/SERVER-22219
[324] https://jira.mongodb.org/browse/SERVER-22324
[325] https://jira.mongodb.org/browse/TOOLS-1028
[326] https://jira.mongodb.org/browse/SERVER-21724
[327] https://jira.mongodb.org/browse/SERVER-21824
[328] https://jira.mongodb.org/browse/SERVER-21890
[329] https://jira.mongodb.org/browse/SERVER-20824
[330] https://jira.mongodb.org/browse/SERVER-21076
[331] https://jira.mongodb.org/browse/SERVER-21132
[332] https://jira.mongodb.org/browse/SERVER-21133
[333] https://jira.mongodb.org/browse/SERVER-21134
[334] https://jira.mongodb.org/browse/SERVER-21135
[335] https://jira.mongodb.org/browse/SERVER-21136
[336] https://jira.mongodb.org/browse/SERVER-21137
[337] https://jira.mongodb.org/browse/SERVER-21138
[338] https://jira.mongodb.org/browse/SERVER-21139
[339] https://jira.mongodb.org/browse/SERVER-21366
[340] https://jira.mongodb.org/browse/SERVER-21586

- SERVER-21704[341] JS Test single_node_config_server_smoke has race condition
- SERVER-21706[342] Certain parameters to mapReduce trigger segmentation fault in a sharded cluster
- SERVER-21786[343] Fix code coverage gaps in s/query directory exposed by code coverage tool
- SERVER-21848[344] bulk write operations on config/admin triggers invariant failure

Replication

- SERVER-21248[345] jstests for fast-failover correctness
- SERVER-21667[346] do not set lastop on clients used by replication on secondaries
- SERVER-21795[347] Do not reschedule more than one liveness timeout callback at a time
- SERVER-21847[348] log range of operations read from sync source during replication
- SERVER-21868[349] Shutdown may not be handled correctly on secondary nodes
- SERVER-21930[350] Restart oplog query if oplog entries are not monotonically increasing

Query

- SERVER-21600[351] Increase test coverage for killCursors command and OP_KILLCURSORS
- SERVER-21602[352] Reduce execution time of cursor_timeout.js
- SERVER-21637[353] Add mixed version tests for find/getMore commands
- SERVER-21638[354] Audit and improve logging in new find/getMore commands code
- SERVER-21750[355] getMore command does not set "nreturned" operation counter

Storage

- SERVER-21384[356] Expand testing for in memory storage engines
- SERVER-21545[357] collMod and invalid parameter triggers fassert on dropCollection on mmapv1
- SERVER-21885[358] capped_truncate.js cannot be run with –repeat
- SERVER-21920[359] Use enhanced WiredTiger next_random cursors for oplog stones

[341] https://jira.mongodb.org/browse/SERVER-21704
[342] https://jira.mongodb.org/browse/SERVER-21706
[343] https://jira.mongodb.org/browse/SERVER-21786
[344] https://jira.mongodb.org/browse/SERVER-21848
[345] https://jira.mongodb.org/browse/SERVER-21248
[346] https://jira.mongodb.org/browse/SERVER-21667
[347] https://jira.mongodb.org/browse/SERVER-21795
[348] https://jira.mongodb.org/browse/SERVER-21847
[349] https://jira.mongodb.org/browse/SERVER-21868
[350] https://jira.mongodb.org/browse/SERVER-21930
[351] https://jira.mongodb.org/browse/SERVER-21600
[352] https://jira.mongodb.org/browse/SERVER-21602
[353] https://jira.mongodb.org/browse/SERVER-21637
[354] https://jira.mongodb.org/browse/SERVER-21638
[355] https://jira.mongodb.org/browse/SERVER-21750
[356] https://jira.mongodb.org/browse/SERVER-21384
[357] https://jira.mongodb.org/browse/SERVER-21545
[358] https://jira.mongodb.org/browse/SERVER-21885
[359] https://jira.mongodb.org/browse/SERVER-21920

WiredTiger

- SERVER-21792[360] 75% performance regression in insert workload under Windows between 3.0.7 and 3.2 with WiredTiger
- SERVER-21872[361] WiredTiger changes for 3.2.1

Operations SERVER-21870[362] Missing space in error message

Build and Packaging

- SERVER-13370[363] Generate Enterprise RPM's for Amazon Linux
- SERVER-21781[364] Nightly packages are in the wrong repo directories
- SERVER-21796[365] fix startup_log.js test to handle git describe versioning
- SERVER-21864[366] streamline artifact signing procedure to support coherent release process

Tools

- TOOLS-954[367] Add bypassDocumentValidation option to mongorestore and mongoimport
- TOOLS-982[368] Missing "from" text in mongorestore status message

Internals

- SERVER-21164[369] Change assert to throw in rslib.js's wait loop
- SERVER-21214[370] Dump config server data when the sharded concurrency suites fail
- SERVER-21426[371] Add writeConcern support to benchRun
- SERVER-21450[372] Modify MongoRunner to add enableMajorityReadConcern flag based on jsTestOptions
- SERVER-21500[373] Include the name of the FSM workload in the WorkloadFailure description
- SERVER-21516[374] Remove dbStats command from readConcern testing override
- SERVER-21665[375] Suppress tar output in jstestfuzz tasks
- SERVER-21714[376] Increase replSetTest.initiate() timeout for FSM tests
- SERVER-21719[377] Add initiateTimeout rsOption for ShardingTest

[360] https://jira.mongodb.org/browse/SERVER-21792
[361] https://jira.mongodb.org/browse/SERVER-21872
[362] https://jira.mongodb.org/browse/SERVER-21870
[363] https://jira.mongodb.org/browse/SERVER-13370
[364] https://jira.mongodb.org/browse/SERVER-21781
[365] https://jira.mongodb.org/browse/SERVER-21796
[366] https://jira.mongodb.org/browse/SERVER-21864
[367] https://jira.mongodb.org/browse/TOOLS-954
[368] https://jira.mongodb.org/browse/TOOLS-982
[369] https://jira.mongodb.org/browse/SERVER-21164
[370] https://jira.mongodb.org/browse/SERVER-21214
[371] https://jira.mongodb.org/browse/SERVER-21426
[372] https://jira.mongodb.org/browse/SERVER-21450
[373] https://jira.mongodb.org/browse/SERVER-21500
[374] https://jira.mongodb.org/browse/SERVER-21516
[375] https://jira.mongodb.org/browse/SERVER-21665
[376] https://jira.mongodb.org/browse/SERVER-21714
[377] https://jira.mongodb.org/browse/SERVER-21719

- SERVER-21725[378] Enable the analysis script move
- SERVER-21737[379] remove deprecated release process configuration from master branch evergreen configuration
- SERVER-21752[380] slow2_wt fails by exhausting host machine's memory
- SERVER-21768[381] Remove the 'numCollections' field from dbHash's response
- SERVER-21772[382] findAndModify not captured by Profiler
- SERVER-21793[383] create v3.2 branch and update evergreen configuration
- SERVER-21849[384] Fix timestamp compare in min_optime_recovery.js
- SERVER-21852[385] kill_cursors.js fails in small_oplog* configurations
- SERVER-21871[386] Do not run min_optime_recovery.js on ephemeralForTest storageEngine
- SERVER-21901[387] CheckReplDBHash checks the wrong node when dumping docs from missing collections
- SERVER-21923[388] ReplSetTest.awaitSecondaryNodes does not propagate supplied timeout
- TOOLS-944[389] write concern mongos tests are flaky
- TOOLS-1002[390] oplog_rollover test is flaky

3.2.6 – Upcoming

- First production release of the *in-memory storage engine* (page 605).
- Fixed issues with background index build may result in extra index key entries that do not correspond to indexed documents: SERVER-22970[391]
- Fixed issues with `mongo` shell method `count()` where the method ignored read preference: SERVER-22043[392]
- All issues closed in 3.2.6[393]

3.2.5 – Apr 14, 2016

- Fixed issue with IX GlobalLock being held while waiting for WiredTiger cache eviction: SERVER-22964[394]
- Fixed issue that may cause low query rate with heavy cache pressure and an idle collection: SERVER-22831[395]

[378] https://jira.mongodb.org/browse/SERVER-21725
[379] https://jira.mongodb.org/browse/SERVER-21737
[380] https://jira.mongodb.org/browse/SERVER-21752
[381] https://jira.mongodb.org/browse/SERVER-21768
[382] https://jira.mongodb.org/browse/SERVER-21772
[383] https://jira.mongodb.org/browse/SERVER-21793
[384] https://jira.mongodb.org/browse/SERVER-21849
[385] https://jira.mongodb.org/browse/SERVER-21852
[386] https://jira.mongodb.org/browse/SERVER-21871
[387] https://jira.mongodb.org/browse/SERVER-21901
[388] https://jira.mongodb.org/browse/SERVER-21923
[389] https://jira.mongodb.org/browse/TOOLS-944
[390] https://jira.mongodb.org/browse/TOOLS-1002
[391] https://jira.mongodb.org/browse/SERVER-22970
[392] https://jira.mongodb.org/browse/SERVER-22043
[393] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.2.6%20AND%20resolution%20%3D%2
[394] https://jira.mongodb.org/browse/SERVER-22964
[395] https://jira.mongodb.org/browse/SERVER-22831

- Report index size stats for in-memory storage engine: SERVER-21681[396]
- All issues closed in 3.2.5[397]

3.2.4 – Mar 8, 2016

- Fixed issue with setting optime when running with journaling disabled: SERVER-22495[398], SERVER-22728[399]
- Have read concern `majority` reflect journaled state on the primary: SERVER-22269[400]
- Fixed issue where specifying `replication.enableMajorityReadConcern` implied `true` regardless of the actual boolean value: SERVER-22683[401]
- Fixed issue with the `count()` method in the `mongo` shell where the method did not apply the read preferences: SERVER-22043[402]
- All issues closed in 3.2.4[403]

3.2.3 – Feb 17, 2016

- Fixed issue with MMAPv1 journaling where the "last sequence number" file (`lsn` file) may be ahead of what is synced to the data files: SERVER-22261[404].
- Fixed issue where in some cases, insert operations fails to add the `_id` field to large documents: SERVER-22167[405].
- Increased timeout for querying oplog to 1 minute: SERVER-22456[406].
- All issues closed in 3.2.3[407]

3.2.2 – Feb 16, 2016

Replaced by MongoDB 3.2.3. Users wishing to run MongoDB 3.2 should skip 3.2.2 and upgrade directly to 3.2.3.

3.2.1 – Jan 12, 2016

- Fixed error where during a regular shutdown of a replica set, secondaries may mark certain replicated but yet to be applied operations as successfully applied: SERVER-21868[408].
- Improve insert workload performance with WiredTiger on Windows: SERVER-20262[409].
- Fixed long-running transactions during chunk moves: SERVER-21366[410]

[396] https://jira.mongodb.org/browse/SERVER-21681
[397] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.2.5%20AND%20resolution%20%3D%2
[398] https://jira.mongodb.org/browse/SERVER-22495
[399] https://jira.mongodb.org/browse/SERVER-22728
[400] https://jira.mongodb.org/browse/SERVER-22269
[401] https://jira.mongodb.org/browse/SERVER-22683
[402] https://jira.mongodb.org/browse/SERVER-22043
[403] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.2.4%20AND%20resolution%20%3D%2
[404] https://jira.mongodb.org/browse/SERVER-22261
[405] https://jira.mongodb.org/browse/SERVER-22167
[406] https://jira.mongodb.org/browse/SERVER-22456
[407] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.2.3%20AND%20resolution%20%3D%2
[408] https://jira.mongodb.org/browse/SERVER-21868
[409] https://jira.mongodb.org/browse/SERVER-20262
[410] https://jira.mongodb.org/browse/SERVER-21366

- All issues closed in 3.2.1[411]

WiredTiger as Default

Starting in 3.2, MongoDB uses the WiredTiger as the default storage engine.

To specify the MMAPv1 storage engine, you must specify the storage engine setting either:

- On the command line with the `--storageEngine` option:

```
mongod --storageEngine mmapv1
```

- Or in a `configuration file`, using the `storage.engine` setting:

```
storage:
   engine: mmapv1
```

Note: For existing deployments, if you do not specify the `--storageEngine` or the `storage.engine` setting, MongoDB 3.2 can automatically determine the storage engine used to create the data files in the `--dbpath` or `storage.dbPath`.

If specifying `--storageEngine` or `storage.engine`, `mongod` will not start if `dbPath` contains data files created by a storage engine other than the one specified.

See also:

Default Storage Engine Change (page 898)

WiredTiger Default Cache Size

Starting in MongoDB 3.2, the WiredTiger cache, by default, will use the larger of either:

- 60% of RAM minus 1 GB, or
- 1 GB.

For more information, see *WiredTiger and Memory Use* (page 597).

WiredTiger Journal Write Frequency

MongoDB 3.2 configures WiredTiger to write to the journal files at every 50 milliseconds. This is in addition to the existing joural write intervals and conditions. For more information, see *Journaling Process* (page 607).

Replication Election Enhancements

Starting in MongoDB 3.2, MongoDB reduces replica set failover time and accelerates the detection of multiple simultaneous primaries.

As part of this enhancement, MongoDB introduces a version 1 of the replication protocol. New replica sets will, by default, use `protocolVersion: 1` (page 718). Previous versions of MongoDB use version 0 of the protocol.

In addition, MongoDB introduces a new *replica set configuration* (page 717) option `electionTimeoutMillis` (page 722). `electionTimeoutMillis` (page 722) specifies the time limit in milliseconds for detecting when a replica set's primary is unreachable.

[411] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.2.1%20AND%20resolution%20%3D%2

electionTimeoutMillis (page 722) only applies if using the version 1 of the replication protocol (page 718).

Sharded Cluster Enhancements

MongoDB 3.2 deprecates the use of three mirrored mongod instances for config servers.

Instead, starting in 3.2, the *config servers* (page 742) for a sharded cluster can be deployed as a replica set. The replica set config servers must run the WiredTiger storage engine.

This change improves consistency across the config servers, since MongoDB can take advantage of the standard replica set read and write protocols for sharding config data. In addition, this allows a sharded cluster to have more than 3 config servers since a replica set can have up to 50 members.

For more information, see *Config Servers* (page 742). To deploy a **new** sharded cluster with replica set config servers, see *Deploy a Sharded Cluster* (page 765).

readConcern

MongoDB 3.2 introduces the readConcern query option for replica sets and replica set shards. For the *WiredTiger storage engine* (page 595), the readConcern option allows clients to choose a level of isolation for their reads. You can specify a readConcern of "majority" to read data that has been written to a majority of nodes and thus cannot be rolled back. By default, MongoDB uses a readConcern of "local" to return the most recent data available to the node at the time of the query, even if the data has not been persisted to a majority of nodes and may be rolled back. With the *MMAPv1 storage engine* (page 603), you can only specify a readConcern of "local".

readConcern requires MongoDB drivers updated for MongoDB 3.2.

Only replica sets using protocol version 1 (page 718) support "majority" (page 182) read concern. Replica sets running protocol version 0 do not support "majority" (page 182) read concern.

For details on readConcern, including operations that support the option, see *Read Concern* (page 181).

Partial Indexes

MongoDB 3.2 provides the option to create indexes that only index the documents in a collection that meet a specified filter expression. By indexing a subset of the documents in a collection, partial indexes have lower storage requirements and reduced performance costs for index creation and maintenance. You can specify a partialFilterExpression option for all MongoDB *index types* (page 516).

The partialFilterExpression option accepts a document that specifies the condition using:

- equality expressions (i.e. field: value or using the $eq operator),
- $exists: true expression,
- $gt, $gte, $lt, $lte expressions,
- $type expressions,
- $and operator at the top-level only

For details, see *Partial Indexes* (page 570).

Document Validation

Starting in 3.2, MongoDB provides the capability to validate documents during updates and insertions. Validation rules are specified on a per-collection basis.

To specify document validation on a new collection, use the new `validator` option in the `db.createCollection()` method. To add document validation to an existing collection, use the new `validator` option in the `collMod` command. For more information, see *Document Validation* (page 250).

To view the validation specifications for a collection, use the `db.getCollectionInfos()` method.

The following commands can bypass validation per operation using the new option `bypassDocumentValidation`:

- `applyOps` command
- `findAndModify` command and `db.collection.findAndModify()` method
- `mapReduce` command and `db.collection.mapReduce()` method
- `insert` command
- `update` command
- `$out` for the `aggregate` command and `db.collection.aggregate()` method

For deployments that have enabled access control, you must have `bypassDocumentValidation` (page 501) action. The built-in roles `dbAdmin` (page 487) and `restore` (page 492) provide this action.

Aggregation Framework Enhancements

MongoDB introduces:

- New stages, accumulators, and expressions.
- *Availability of accumulator expressions* (page 892) in `$project` stage.
- *Performance improvements* (page 892) on sharded clusters.

New Aggregation Stages

Stage	Description	Syntax
`$sample`	Randomly selects N documents from its input.	`{ $sample: { size: <positive integer> } }`
`$indexStats`	Returns statistics on index usage.	`{ $indexStats: { } }`
`$lookup`	Performs a left outer join with another collection.	`{ $lookup: { from: <collection to join>, localField: <fieldA>, foreignField: <fieldB>, as: <output array field> } }`

New Accumulators for `$group` Stage

Accumulator	Description	Syntax
`$stdDevSamp`	Calculates standard deviation.	`{ $stdDevSamp: <array> }`
`$stdDevPop`	Calculates population standard deviation.	`{ $stdDevPop: <array> }`

New Aggregation Arithmetic Operators

Operator	Description	Syntax
`$sqrt`	Calculates the square root.	`{ $sqrt: <number> }`
`$abs`	Returns the absolute value of a number.	`{ $abs: <number> }`
`$log`	Calculates the log of a number in the specified base.	`{ $log: [ <number>, <base> ] }`
`$log10`	Calculates the log base 10 of a number.	`{ $log10: <number> }`
`$ln`	Calculates the natural log of a number.	`{ $ln: <number> }`
`$pow`	Raises a number to the specified exponent.	`{ $pow: [ <number>, <exponent> ] }`
`$exp`	Raises *e* to the specified exponent.	`{ exp: <number> }`
`$trunc`	Truncates a number to its integer.	`{ $trunc: <number> }`
`$ceil`	Returns the smallest integer greater than or equal to the specified number.	`{ $ceil: <number> }`
`$floor`	Returns the largest integer less than or equal to the specified number.	`{ floor: <number> }`

New Aggregation Array Operators

Operator	Description	Syntax
`$slice`	Returns a subset of an array.	`{ $slice: [ <array>, <n> ] }` or `{ $slice: [ <array>, <position>, <n> ] }`
`$arrayElemAt`	Returns the element at the specified array index.	`{ $arrayElemAt: [ <array>, <idx> ] }`
`$concatArrays`	Concatenates arrays.	`{ $concatArrays: [ <array1>, <array2` `}`
`$isArray`	Determines if the operand is an array.	`{ $isArray: [ <expression> ] }`
`$filter`	Selects a subset of the array based on the condition.	`{ $filter: { input: <array>, as: <string>, cond: <expression> } }`

Accumulator Expression Availability

Starting in version 3.2, the following accumulator expressions, previously only available in the `$group` stage, are now also available in the `$project` stage:

- `$avg`
- `$min`
- `$max`
- `$sum`
- `$stdDevPop`
- `$stdDevSamp`

When used as part of the `$project` stage, these accumulator expressions can accept either:

- A single argument: `<accumulator> : <arg>`
- Multiple arguments: `<accumulator> : [ <arg1>, <arg2>, ... ]`

General Enhancements

- In MongoDB 3.2, `$project` stage supports using the square brackets `[]` to directly create new array fields. For an example, see *example-project-new-array-fields*.
- MongoDB 3.2 introduces the `minDistance` option for the `$geoNear` stage.
- `$unwind` stage no longer errors on non-array operand. If the operand does not resolve to an array but is not missing, null, or an empty array, `$unwind` treats the operand as a single element array.

  `$unwind` stage can:

  - include the array index of the array element in the output by specifying a new option `includeArrayIndex` in the stage specification.
  - output those documents where the array field is missing, null or an empty array by specifying a new option `preserveNullAndEmptyArrays` in the stage specification.

  To support these new features, `$unwind` can now take an alternate syntax. See `$unwind` for details.

Optimization

Indexes can *cover* (page 106) aggregation operations.

MongoDB improves the overall performance of the pipeline in large sharded clusters.

If the pipeline starts with an exact `$match` on a shard key, the entire pipeline runs on the matching shard only. Previously, the pipeline would have been split, and the work of merging it would have to be done on the primary shard.

For aggregation operations that run on multiple shards, if the operations do not require running on the database's primary shard, these operations can route the results to any shard to merge the results and avoid overloading the primary shard for that database. Aggregation operations that require running on the database's primary shard are the `$out` stage and `$lookup` stage.

Compatibility

For compatibility changes, see *Aggregation Compatibility Changes* (page 899).

MongoDB Tools Enhancements

- `mongodump` and `mongorestore` add support for archive files and standard output/input streams with a new `--archive` option. This enhancement allows for the streaming of the dump data over a network device via a pipe. For examples, see
  - *mongodump to an Archive File* and *mongodump an Archive to Standard Output*
  - *mongorestore-example-archive-file* and *mongorestore-example-archive-stdin*.
- `mongodump` and `mongorestore` add support for compressed data dumps with a new `--gzip` option. This enhancement reduces storage space for the dump files. For examples, see:
  - *Compress mongodump Output*
  - *mongorestore-example-gzip*.

Encrypted Storage Engine

Enterprise Feature

Available in MongoDB Enterprise only.

Important: Available for the WiredTiger Storage Engine only.

Encryption at rest, when used in conjunction with transport encryption and good security policies that protect relevant accounts, passwords, and encryption keys, can help ensure compliance with security and privacy standards, including HIPAA, PCI-DSS, and FERPA.

MongoDB Enterprise 3.2 introduces a native encryption option for the WiredTiger storage engine. This feature allows MongoDB to encrypt data files such that only parties with the decryption key can decode and read the data. For detail, see *Encrypted Storage Engine* (page 461).

Text Search Enhancements

`text` Index Version 3

MongoDB 3.2 introduces a version 3 of the *text index* (page 533). Key features of the new version of the index are:

- Improved *case insensitivity* (page 535).
- *Diacritic insensitivity* (page 536).
- Additional *delimiters for tokenization* (page 536).

Starting in MongoDB 3.2, version 3 is the default version for new *text* (page 533) indexes.

See also:

Text Index Version 3 Compatibility (page 899)

`$text` Operator Enhancements

The `$text` operator adds support for:

- *case sensitive text search* with the new `$caseSensitive` option, and

- *diacritic sensitive text search* with the new `$diacriticSensitive` option.

For more information and examples, see the `$text` operator reference sections *text-operator-case-sensitivity* and *text-operator-diacritic-sensitivity*.

Support for Additional Languages

Enterprise Feature

Available in MongoDB Enterprise only.

Starting in 3.2, MongoDB Enterprise provides support for the following languages: Arabic, Farsi (specifically Dari and Iranian Persian dialects), Urdu, Simplified Chinese, and Traditional Chinese.

For details, see *Text Search with Basis Technology Rosette Linguistics Platform* (page 244).

New Storage Engines

`inMemory` Storage Engine

Enterprise Feature

Available in MongoDB Enterprise only.

MongoDB Enterprise 3.2 provides an in-memory storage engine. Other than some metadata, the in-memory storage engine does not maintain any on-disk data. By avoiding disk I/O, the in-memory storage engine allows for more predictable latency of database operations.

> **Warning:** Currently in beta. Do not use in production.

To select this storage engine, specify

- `inMemory` for the *`--storageEngine`* option or the `storage.engine` setting.
- `--dbpath`. Although the in-memory storage engine does not write data to the filesystem, it maintains in the `--dbpath` small metadata files and diagnostic data as well temporary files for building large indexes.

The `inMemory` storage engine uses document-level locking. For more details, see *In-Memory Storage Engine* (page 605).

`ephemeralForTest` Storage Engine

MongoDB 3.2 provides a new for-test storage engine. Other than some metadata, the for-test storage engine does not maintain any on-disk data, removing the need to clean up between test runs. The for-test storage engine is unsupported.

> **Warning:** For test purposes only. Do not use in production.

To select this storage engine, specify

- `ephemeralForTest` for the *`--storageEngine`* option or the `storage.engine` setting.
- `--dbpath`. Although the for-test storage engine does not write data to the filesystem, it maintains small metadata files in the `--dbpath`.

The `ephemeralForTest` storage engine uses collection-level locking.

General Enhancements

Bit Test Query Operators

MongoDB 3.2 provides new query operators to test bit values:

- `$bitsAllSet`
- `$bitsAllClear`
- `$bitsAnySet`
- `$bitsAnyClear`

SpiderMonkey JavaScript Engine

MongoDB 3.2 uses SpiderMonkey as the JavaScript engine for the `mongo` shell and `mongod` server. SpiderMonkey provides support for additional platforms and has an improved memory management model.

This change affects all JavaScript behavior including the commands `mapReduce`, `group`, and the query operator `$where`; *however*, this change should be completely transparent to the user.

See also:

SpiderMonkey Compatibility Changes (page 899)

`mongo` Shell and CRUD API

To provide consistency with the MongoDB drivers' CRUD (Create/Read/Update/Delete) API, the `mongo` shell introduces additional CRUD methods that are consistent with the drivers' CRUD API:

New API	Description
`db.collection.bulkWrite()`	Equivalent to initializing `Bulk()` operations builder, using *Bulk methods* to add operations, and running `Bulk.execute()` to execute the operations. MongoDB 3.2 deprecates `Bulk()` and its associated `methods`.
`db.collection.deleteMany()`	Equivalent to `db.collection.remove()`.
`db.collection.deleteOne()`	Equivalent to `db.collection.remove()` with the `justOne` set to true; i.e. `db.collection.remove( <query>, true )` or `db.collection.remove( <query>, { justOne: true } )`.
`db.collection.findOneAndDelete()`	Equivalent to `db.collection.findAndModify()` method with `remove` set to true.
`db.collection.findOneAndReplace()`	Equivalent to `db.collection.findAndModify()` method with `update` set to a replacement document.
`db.collection.findOneAndUpdate()`	Equivalent to `db.collection.findAndModify()` method with `update` set to a document that specifies modifications using `update operators`.
`db.collection.insertMany()`	Equivalent to `db.collection.insert()` method with an array of documents as the parameter.
`db.collection.insertOne()`	Equivalent to `db.collection.insert()` method with a single document as the parameter.
`db.collection.replaceOne()`	Equivalent to `db.collection.update( <query>, <update> )` method with a replacement document as the `<update>` parameter.
`db.collection.updateMany()`	Equivalent to `db.collection.update( <query>, <update>, { multi: true, ... })` method with an `<update>` document that specifies modifications using `update operators` and the `multi` option set to true.
`db.collection.updateOne()`	Equivalent to `db.collection.update( <query>, <update> )` method with an `<update>` document that specifies modifications using `update operators`.

WiredTiger and `fsyncLock`

Starting in MongoDB 3.2, the WiredTiger storage engine supports the `fsync` command with the `lock` option or the `mongo` shell method `db.fsyncLock()`. That is, for the WiredTiger storage engine, these operations can guarantee that the data files do not change, ensuring consistency for the purposes of creating backups.

Platform Support

Starting in 3.2, 32-bit binaries are deprecated and will be unavailable in future releases.

MongoDB 3.2 deprecates support for Red Hat Enterprise Linux 5.

`$type` Operator Accepts String Aliases

`$type` operator accepts string aliases for the BSON types in addition to the numbers corresponding to the BSON types.

`explain()` Support for `distinct()` Operation

`db.collection.explain()` adds support for `db.collection.distinct()` method. For more information, see `db.collection.explain()`.

Deprecation of the HTTP Interface

Starting in 3.2, MongoDB deprecates its HTTP interface.

`keysExamined` Count Includes the Last Scanned Key

For `explain` operations run in `executionStats` or `allPlansExecution` mode, the `explain output` contains the `keysExamined` statistic, representing the number of index keys examined during index scans.

Prior to 3.2, `keysExamined` count in some queries did not include the last scanned key. As of 3.2 this error has been corrected. For more information, see :data:‘ ~explain.executionStats.executionStages.inputStage.keysExamined‘.

The diagnostic logs and the system profiler report on this statistic.

Geospatial Optimization

MongoDB 3.2 introduces version 3 of *2dsphere indexes* (page 543), which index *GeoJSON geometries* (page 554) at a finer gradation. The new version improves performance of *2dsphere index* (page 543) queries over smaller regions. In addition, for both *2d indexes* (page 557) and *2dsphere indexes* (page 543), the performance of geoNear queries has been improved for dense datasets.

See also:

2dsphere Index Version 3 Compatibility (page 899)

Diagnostic Data Capture

To facilitate analysis of the MongoDB server behavior by MongoDB engineers, MongoDB 3.2 introduces a diagnostic data collection mechanism for logging server statistics to diagnostic files at periodic intervals. By default, the mechanism captures data at 1 second intervals. To modify the interval, see `diagnosticDataCollectionPeriodMillis`.

MongoDB creates a `diagnostic.data` directory under the `mongod` instance's `--dbpath` or `storage.dbPath`. The diagnostic data is stored in files under this directory.

The maximum size of the diagnostic files is configurable with the `diagnosticDataCollectionFileSizeMB`, and the maximum size of the `diagnostic.data` directory is configurable with `diagnosticDataCollectionDirectorySizeMB`.

The default values for the capture interval and the maximum sizes are chosen to provide useful data to MongoDB engineers with minimal impact on performance and storage size. Typically, these values will only need modifications as requested by MongoDB engineers for specific diagnostic purposes.

Write Concern

- For replica sets using `protocolVersion: 1` (page 718), secondaries acknowledge write operations after the secondary members have written to their respective on-disk *journals* (page 606), regardless of the *j* (page 181) option.
- For replica sets using `protocolVersion: 1` (page 718), `w: "majority"` (page 180) implies *j: true* (page 181).
- With `j: true` (page 181), MongoDB returns only after the requested number of members, including the primary, have written to the journal. Previously `j: true` (page 181) write concern in a replica set only requires the *primary* to write to the journal, regardless of the *w: <value>* (page 180) write concern.

`journalCommitInterval` for WiredTiger

MongoDB 3.2 adds support for specifying the journal commit interval for the WiredTiger storage engine. See *journalCommitInterval* option. In previous versions, the option is applicable to MMAPv1 storage engine only.

For the corresponding configuration file setting, MongoDB 3.2 adds the `storage.journal.commitIntervalMs` setting and deprecates `storage.mmapv1.journal.commitIntervalMs`. The deprecated `storage.mmapv1.journal.commitIntervalMs` setting acts as an alias to the new `storage.journal.commitIntervalMs` setting.

Changes Affecting Compatibility

Some MongoDB 3.2 changes can affect compatibility and may require user actions. For a detailed list of compatibility changes, see *Compatibility Changes in MongoDB 3.2* (page 898).

Compatibility Changes in MongoDB 3.2

On this page

- Default Storage Engine Change (page 898)
- Index Changes (page 898)
- Aggregation Compatibility Changes (page 899)
- SpiderMonkey Compatibility Changes (page 899)
- Driver Compatibility Changes (page 900)
- General Compatibility Changes (page 900)
- Additional Information (page 900)

The following 3.2 changes can affect the compatibility with older versions of MongoDB. See also *Release Notes for MongoDB 3.2* (page 865) for the list of the 3.2 changes.

Default Storage Engine Change Starting in 3.2, MongoDB uses the WiredTiger as the default storage engine. Previous versions used the MMAPv1 as the default storage engine.

For existing deployments, if you do not specify the `--storageEngine` or the `storage.engine` setting, MongoDB automatically determines the storage engine used to create the data files in the `--dbpath` or `storage.dbPath`.

For new deployments, to use MMAPv1, you must explicitly specify the storage engine setting either:

- On the command line with the `--storageEngine` option:

```
mongod --storageEngine mmapv1
```

- Or in a `configuration file`, using the `storage.engine` setting:

```
storage:
   engine: mmapv1
```

Index Changes

Version 0 Indexes MongoDB 3.2 disallows the creation of version 0 indexes (i.e. `{v: 0}`). If version 0 indexes exist, MongoDB 3.2 outputs a warning log message, specifying the collection and the index.

Starting in MongoDB 2.0, MongoDB started automatically upgrading `v: 0` indexes during *initial sync* (page 658), `mongorestore` or `reIndex` operations.

If a version 0 index exists, you can use any of the aforementioned operations as well as drop and recreate the index to upgrade to the `v: 1` version.

For example, if upon startup, a warning message indicated that an index `index { v: 0, key: { x: 1.0 }, name: "x_1", ns: "test.legacyOrders" }` is a version 0 index, to upgrade to the appropriate version, you can drop and recreate the index:

1. Drop the index either by name:

   ```
   use test
   db.legacyOrders.dropIndex( "x_1" )
   ```

   or by key:

   ```
   use test
   db.legacyOrders.dropIndex( { x: 1 } )
   ```

2. Recreate the index without the version option `v`:

   ```
   db.legacyOrders.createIndex( { x: 1 } )
   ```

Text Index Version 3 Compatibility *Text index (version 3)* (page 893) is incompatible with earlier versions of MongoDB. Earlier versions of MongoDB will not start if *text index (version 3)* (page 533) exists in the database.

`2dsphere` Index Version 3 Compatibility *2dsphere index (version 3)* (page 897) is incompatible with earlier versions of MongoDB. Earlier versions of MongoDB will not start if `2dsphere` index (version 3) exists in the database.

Aggregation Compatibility Changes

- `$avg` accumulator returns null when run against a non-existent field. Previous versions returned `0`.
- `$substr` errors when the result is an invalid UTF-8. Previous versions output the invalid UTF-8 result.
- Array elements are no longer treated as literals in the aggregation pipeline. Instead, each element of an array is now parsed as an expression. To treat the element as a literal instead of an expression, use the `$literal` operator to create a literal value.

SpiderMonkey Compatibility Changes

JavaScript Changes in MongoDB 3.2

On this page

- Modernized JavaScript Implementation (ES6) (page 900)
- Changes to the `mongo` Shell (page 900)
- Removed Non-Standard V8 Features (page 900)

In MongoDB 3.2, the javascript engine used for both the `mongo` shell and for server-side javascript in `mongod` changed from V8 to SpiderMonkey[412].

To confirm which JavaScript engine you are using, you can use either `interpreterVersion()` method in the `mongo` shell and the `javascriptEngine` field in the output of `db.serverBuildInfo()`

In MongoDB 3.2, this will appear as `MozJS-38` and `mozjs`, respectively.

Modernized JavaScript Implementation (ES6) SpiderMonkey brings with it increased support for features defined in the 6th edition of ECMAScript[413], abbreviated as ES6. ES6 adds many new language features, including:

- arrow functions[414],
- destructuring assignment[415],
- for-of loops[416], and
- generators[417].

Changes to the `mongo` Shell MongoDB 3.2 will return JavaScript and BSON `undefined` (page 18) values intact if saved into a collection. Previously, the `mongo` shell would convert `undefined` values into `null`.

MongoDB 3.2 also adds the `disableJavaScriptJIT` parameter to `mongod`, which allows you to disable the JavaScript engine's JIT acceleration. The `mongo` shell has a corresponding *`--disableJavaScriptJIT`* flag.

Removed Non-Standard V8 Features SpiderMonkey does **not** allow the non-standard `Error.captureStackTrace()` function that prior versions of MongoDB supported. If you must record stack traces, you can capture the `Error().stack` property as a workaround.

MongoDB 3.2 changes the JavaScript engine from V8 to SpiderMonkey. The change allows the use of more modern JavaScript language features, and comes along with minor `mongo` shell improvements and compatibility changes.

See *JavaScript Changes in MongoDB 3.2* (page 899) for more information about this change.

Driver Compatibility Changes A driver upgrade is necessary to support the `find` and `getMore` commands.

General Compatibility Changes

- In MongoDB 3.2, `cursor.showDiskLoc()` is deprecated in favor of `cursor.showRecordId()`, and both return a new document format.
- MongoDB 3.2 renamed the `serverStatus.repl.slaves` field to to `repl.replicationProgress`. See: the **db.serverStatus()** *server-status-repl* reference for more information.
- The default changed from *`--moveParanoia`* to *`--noMoveParanoia`*.
- MongoDB 3.2 replica set members with `1 vote` (page 721) cannot sync from members with `0 votes` (page 721).
- `mongooplog` is deprecated starting in MongoDB 3.2.

Additional Information See also *Release Notes for MongoDB 3.2* (page 865).

[412] https://developer.mozilla.org/en-US/docs/SpiderMonkey
[413] http://www.ecma-international.org/ecma-262/6.0/index.html
[414] http://www.ecma-international.org/ecma-262/6.0/index.html#sec-arrow-function-definitions
[415] http://www.ecma-international.org/ecma-262/6.0/index.html#sec-destructuring-assignment
[416] http://www.ecma-international.org/ecma-262/6.0/index.html#sec-for-in-and-for-of-statements
[417] http://www.ecma-international.org/ecma-262/6.0/index.html#sec-generator-function-definitions

Upgrade Process

Upgrade MongoDB to 3.2

On this page

- Upgrade Recommendations and Checklists (page 901)
- Upgrade Standalone `mongod` Instance to MongoDB 3.2 (page 901)
- Upgrade a Replica Set to 3.2 (page 902)
- Upgrade a Sharded Cluster to 3.2 (page 902)
- Additional Resources (page 904)

Before you attempt any upgrade, please familiarize yourself with the content of this document.

If you need guidance on upgrading to 3.2, MongoDB offers 3.2 upgrade services[418] to help ensure a smooth transition without interruption to your MongoDB application.

Upgrade Recommendations and Checklists When upgrading, consider the following:

Upgrade Requirements To upgrade an existing MongoDB deployment to 3.2, you must be running a 3.0-series release.

To upgrade from a 2.6-series release, you *must* upgrade to the latest 3.0-series release before upgrading to 3.2. For the procedure to upgrade from the 2.6-series to a 3.0-series release, see *Upgrade MongoDB to 3.0* (page 953).

Preparedness Before beginning your upgrade, see the *Compatibility Changes in MongoDB 3.2* (page 898) document to ensure that your applications and deployments are compatible with MongoDB 3.2. Resolve the incompatibilities in your deployment before starting the upgrade.

Before upgrading MongoDB, always test your application in a staging environment before deploying the upgrade to your production environment.

Upgrade Standalone `mongod` Instance to MongoDB 3.2 The following steps outline the procedure to upgrade a standalone `mongod` from version 3.0 to 3.2. To upgrade from version 2.6 to 3.2, *upgrade to the latest 3.0-series release* (page 953) *first*, and then use the following procedure to upgrade from 3.0 to 3.2.

Upgrade with Package Manager If you installed MongoDB from the MongoDB `apt`, `yum`, `dnf`, or `zypper` repositories, you should upgrade to 3.2 using your package manager. Follow the appropriate *installation instructions* (page 22) for your Linux system. This will involve adding a repository for the new release, then performing the actual upgrade.

Manual Upgrade Otherwise, you can manually upgrade MongoDB:

Step 1: Download 3.2 binaries. Download binaries of the latest release in the 3.2 series from the MongoDB Download Page[419]. See *Install MongoDB* (page 21) for more information.

[418] https://www.mongodb.com/contact/mongodb-3-2-upgrade-services?jmp=docs
[419] http://www.mongodb.org/downloads?jmp=docs

Step 2: Replace with 3.2 binaries Shut down your `mongod` instance. Replace the existing binary with the 3.2 `mongod` binary and restart `mongod`.

Note: MongoDB 3.2 generates core dumps on some `mongod` failures. For production environments, you may prefer to turn off core dumps for the operating system, if not already.

Upgrade a Replica Set to 3.2

Prerequisites All replica set members must be running version 3.0 before you can upgrade them to version 3.2. To upgrade a replica set from an earlier MongoDB version, *upgrade all members of the replica set to the latest 3.0-series release* (page 953) *first*, and then follow the procedure to upgrade from MongoDB 3.0 to 3.2.

Upgrade Binaries You can upgrade from MongoDB 3.0 to 3.2 using a "rolling" upgrade to minimize downtime by upgrading the members individually while the other members are available:

Step 1: Upgrade secondary members of the replica set. Upgrade the *secondary* (page **??**) members of the replica set one at a time:

- Shut down the `mongod` instance and replace the 3.0 binary with the 3.2 binary.
- Restart the member and wait for the member to recover to `SECONDARY` state before upgrading the next secondary member. To check the member's state, issue `rs.status()` in the `mongo` shell.

Step 2: Step down the replica set primary. Connect a `mongo` shell to the primary and use `rs.stepDown()` to step down the primary and force an election of a new primary:

Step 3: Upgrade the primary. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, upgrade the stepped-down primary:

- Shut down the stepped-down primary and replace the `mongod` binary with the 3.2 binary.
- Restart.

Step 4: Upgrade the replication protocol. Connect a `mongo` shell to the current primary and upgrade the replication protocol

```
cfg = rs.conf();
cfg.protocolVersion=1;
rs.reconfig(cfg);
```

Replica set failover is not instant and will render the set unavailable to accept writes until the failover process completes. This may take 30 seconds or more: schedule the upgrade procedure during a scheduled maintenance window.

Note: MongoDB 3.2 generates core dumps on some `mongod` failures. For production environments, you may prefer to turn off core dumps for the operating system, if not already.

Upgrade a Sharded Cluster to 3.2

Prerequisites

- **Version 3.0 or Greater** To upgrade a sharded cluster to 3.2, **all** members of the cluster must be at least version 3.0. The upgrade process checks all components of the cluster and will produce warnings if any component is running version earlier than 3.0.
- **Stop Metadata Changes during the Upgrade** During the upgrade, ensure that clients do not make changes to the collection metadata. For example, during the upgrade, do **not** perform any of the following operations:
  - `sh.enableSharding()`
  - `sh.shardCollection()`
  - `sh.addShard()`
  - `db.createCollection()`
  - `db.collection.drop()`
  - `db.dropDatabase()`
  - any operation that creates a database
  - any other operation that modifies the cluster metadata in any way.

  See the *Sharding Reference* (page 822) for a complete list of sharding commands. Not all commands on the *Sharding Reference* (page 822) page modify the cluster metadata.
- *Disable the balancer* (page 802)
- **Back up the `config` Database** *Optional but Recommended.* As a precaution, take a backup of the `config` database *before* upgrading the sharded cluster.

Upgrade Binaries

Step 1: Disable the Balancer. Disable the balancer as described in *Disable the Balancer* (page 802).

Step 2: Upgrade the shards. Upgrade the shards one at a time. If the shards are replica sets, for each shard:

1. Upgrade the *secondary* (page **??**) members of the replica set one at a time:
   - Shut down the `mongod` instance and replace the 3.0 binary with the 3.2 binary.
   - Restart the member and wait for the member to recover to `SECONDARY` state before upgrading the next secondary member. To check the member's state, issue `rs.status()` in the `mongo` shell.
2. Step down the replica set primary.

   Connect a `mongo` shell to the primary and use `rs.stepDown()` to step down the primary and force an election of a new primary:

   ```
   rs.stepDown()
   ```

3. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, upgrade the stepped-down primary:
   - Shut down the stepped-down primary and replace the `mongod` binary with the 3.2 binary.
   - Restart.
4. Connect a `mongo` shell to the current primary and upgrade the `replication protocol` (page 718) for the shard:

```
cfg = rs.conf();
cfg.protocolVersion=1;
rs.reconfig(cfg);
```

Step 3: Upgrade the config servers. Upgrade the config servers one at a time in reverse order of the `configDB` or *`--configdb`* setting for the `mongos`. That is, if the `mongos` has the following *`--configdb`* listing:

```
mongos --configdb confserver1:port1,confserver2:port2,confserver3:port2
```

Upgrade first `confserver3`, then `confserver2`, and lastly `confserver1`.

Starting with the last config server listed in the `configDB` setting:

1. Stop the config server and replace with the 3.2 binary.
2. Start the 3.2 binary with both the `--configsvr` and `--port` options:

   ```
   mongod --configsvr --port <port> --dbpath <path>
   ```

   If using a `configuration file`, specify `sharding.clusterRole: configsvr` and `net.port` in the file:

   ```
   sharding:
      clusterRole: configsvr
   net:
      port: <port>
   storage:
      dbpath: <path>
   ```

Repeat for the config server listed *second* in the `configDB` setting, and finally the config server listed *first* in the `configDB` setting.

Step 4: Upgrade the `mongos` instances. Replace each `mongos` instance with the 3.2 binary and restart.

```
mongos --configdb <cfgsvr1:port1>,<cfgsvr2:port2>,<cfgsvr3:port3>
```

Step 5: Re-enable the balancer. Re-enable the balancer as described in *Enable the Balancer* (page 803).

Note: MongoDB 3.2 generates core dumps on some `mongod` failures. For production environments, you may prefer to turn off core dumps for the operating system, if not already.

Once the sharded cluster binaries have been upgraded to 3.2, existing config servers will continue to run as mirrored `mongod` instances. For instructions on upgrading existing config servers to a replica set, see *Upgrade Config Servers to Replica Set* (page 780) (requires MongoDB version 3.2.4 or later versions).

Additional Resources

- Getting ready for MongoDB 3.2? Get our help.[420]

[420] https://www.mongodb.com/contact/mongodb-3-2-upgrade-services?jmp=docs

Downgrade MongoDB from 3.2

On this page

- Downgrade Recommendations and Checklist (page 905)
- Prerequisites (page 905)
- Downgrade a Standalone `mongod` Instance (page 906)
- Downgrade a 3.2 Replica Set (page 906)
- Downgrade a 3.2 Sharded Cluster (page 907)

Before you attempt any downgrade, familiarize yourself with the content of this document, particularly the *Downgrade Recommendations and Checklist* (page 905) and the procedure for *downgrading sharded clusters* (page 907).

Downgrade Recommendations and Checklist When downgrading, consider the following:

Downgrade Path To downgrade, use the latest version in the 3.0-series.

Preparedness

- *Remove or downgrade version 3 text indexes* (page 905) before downgrading MongoDB 3.2 to 3.0.
- *Remove or downgrade version 3 2dsphere indexes* (page 906) before downgrading MongoDB 3.2 to 3.0.

Procedures Follow the downgrade procedures:

- To downgrade sharded clusters, see *Downgrade a 3.2 Sharded Cluster* (page 907).
- To downgrade replica sets, see *Downgrade a 3.2 Replica Set* (page 906).
- To downgrade a standalone MongoDB instance, see *Downgrade a Standalone mongod Instance* (page 906).

Prerequisites

Text Index Version Check If you have *version 3* text indexes (i.e. the default version for text indexes in MongoDB 3.2), drop the *version 3* text indexes before downgrading MongoDB. After the downgrade, recreate the dropped text indexes.

To determine the version of your `text` indexes, run `db.collection.getIndexes()` to view index specifications. For text indexes, the method returns the version information in the field `textIndexVersion`. For example, the following shows that the `text` index on the `quotes` collection is version 3.

```
{
   "v" : 1,
   "key" : {
      "_fts" : "text",
      "_ftsx" : 1
   },
   "name" : "quote_text_translation.quote_text",
   "ns" : "test.quotes",
   "weights" : {
      "quote" : 1,
      "translation.quote" : 1
   },
```

```
   "default_language" : "english",
   "language_override" : "language",
   "textIndexVersion" : 3
}
```

`2dsphere` Index Version Check If you have *version 3* `2dsphere` indexes (i.e. the default version for `2dsphere` indexes in MongoDB 3.2), drop the *version 3* `2dsphere` indexes before downgrading MongoDB. After the downgrade, recreate the `2dsphere` indexes.

To determine the version of your `2dsphere` indexes, run `db.collection.getIndexes()` to view index specifications. For `2dsphere` indexes, the method returns the version information in the field `2dsphereIndexVersion`. For example, the following shows that the `2dsphere` index on the `locations` collection is version 3.

```
{
   "v" : 1,
   "key" : {
      "geo" : "2dsphere"
   },
   "name" : "geo_2dsphere",
   "ns" : "test.locations",
   "sparse" : true,
   "2dsphereIndexVersion" : 3
}
```

Partial Indexes Check Before downgrading MongoDB, drop any partial indexes.

Downgrade a Standalone `mongod` Instance The following steps outline the procedure to downgrade a standalone `mongod` from version 3.2 to 3.0.

Step 1: Download the latest 3.0 binaries. For the downgrade, use the latest release in the 3.0 series.

Step 2: Restart with the latest 3.0 `mongod` instance.

Important: If your `mongod` instance is using the *WiredTiger* (page 595) storage engine, you must include the `--storageEngine` option (or `storage.engine` if using the configuration file) with the 3.0 binary.

Shut down your `mongod` instance. Replace the existing binary with the downloaded `mongod` binary and restart.

Downgrade a 3.2 Replica Set The following steps outline a "rolling" downgrade process for the replica set. The "rolling" downgrade process minimizes downtime by downgrading the members individually while the other members are available:

Step 1: Downgrade the protocolVersion. Connect a `mongo` shell to the current primary and downgrade the replication protocol:

```
cfg = rs.conf();
cfg.protocolVersion=0;
rs.reconfig(cfg);
```

Step 2: Downgrade secondary members of the replica set. Downgrade each *secondary* member of the replica set, one at a time:

1. Shut down the `mongod`. See *Stop mongod Processes* (page 324) for instructions on safely terminating `mongod` processes.
2. Replace the 3.2 binary with the 3.0 binary and restart.

   Important: If your `mongod` instance is using the *WiredTiger* (page 595) storage engine, you must include the `--storageEngine` option (or `storage.engine` if using the configuration file) with the 3.0 binary.

3. Wait for the member to recover to `SECONDARY` state before downgrading the next secondary. To check the member's state, use the `rs.status()` method in the `mongo` shell.

Step 3: Step down the primary. Use `rs.stepDown()` in the `mongo` shell to step down the *primary* and force the normal *failover* (page 644) procedure.

```
rs.stepDown()
```

`rs.stepDown()` expedites the failover procedure and is preferable to shutting down the primary directly.

Step 4: Replace and restart former primary `mongod`. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, shut down the previous primary and replace the `mongod` binary with the 3.0 binary and start the new instance.

Important: If your `mongod` instance is using the *WiredTiger* (page 595) storage engine, you must include the `--storageEngine` option (or `storage.engine` if using the configuration file) with the 3.0 binary.

Replica set failover is not instant but will render the set unavailable to writes and interrupt reads until the failover process completes. Typically this takes 10 seconds or more. You may wish to plan the downgrade during a predetermined maintenance window.

Downgrade a 3.2 Sharded Cluster

Requirements While the downgrade is in progress, you cannot make changes to the collection metadata. For example, during the downgrade, do **not** do any of the following:

- `sh.enableSharding()`
- `sh.shardCollection()`
- `sh.addShard()`
- `db.createCollection()`
- `db.collection.drop()`
- `db.dropDatabase()`
- any operation that creates a database
- any other operation that modifies the cluster meta-data in any way. See *Sharding Reference* (page 822) for a complete list of sharding commands. Note, however, that not all commands on the *Sharding Reference* (page 822) page modifies the cluster meta-data.

Downgrade a Sharded Cluster with SCCC Config Servers

Step 1: Disable the Balancer. Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).

Step 2: Downgrade each shard, one at a time. For each replica set shard:

1. Downgrade the protocolVersion.
2. Downgrade the `mongod` secondaries *before* downgrading the primary.
3. To downgrade the primary, run `replSetStepDown` and then downgrade.

For details on downgrading a replica set, see *Downgrade a 3.2 Replica Set* (page 906).

Step 3: Downgrade the SCCC config servers. If the sharded cluster uses 3 mirrored `mongod` instances for the config servers, downgrade all three instances in reverse order of their listing in the *`--configdb`* option for `mongos`. For example, if `mongos` has the following *`--configdb`* listing:

```
--configdb confserver1,confserver2,confserver3
```

Downgrade first `confserver3`, then `confserver2`, and lastly, `confserver1`. If your `mongod` instance is using the *WiredTiger* (page 595) storage engine, you must include the *`--storageEngine`* option (or `storage.engine` if using the configuration file) with the 3.0 binary.

```
mongod --configsvr --dbpath <path> --port <port> --storageEngine <storageEngine>
```

Step 4: Downgrade the `mongos` instances. Downgrade the binaries and restart.

Step 5: Re-enable the balancer. Once the downgrade of sharded cluster components is complete, *re-enable the balancer* (page 803).

Downgrade a Sharded Cluster with CSRS Config Servers

Step 1: Disable the Balancer. Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).

Step 2: Prepare CSRS Config Servers for downgrade. If the sharded cluster uses *CSRS* (page 742):

1. *Remove secondary members from the replica set* (page 682) to have only a primary and two secondaries and only the primary can vote and be eligible to be primary; i.e. the other two members have `0` for `votes` (page 721) and `priority` (page 720).

   Connect a `mongo` shell to the primary and run:

   ```
   rs.reconfig(
      {
         "_id" : <name>,
         "configsvr" : true,
         "protocolVersion" : NumberLong(1),
         "members" : [
            {
               "_id" : 0,
               "host" : "<host1>:<port1>",
               "priority" : 1,
               "votes" : 1
   ```

```
            },
            {
               "_id" : 1,
               "host" : "<host2>:<port2>",
               "priority" : 0,
               "votes" : 0
            },
            {
               "_id" : 2,
               "host" : "<host3>:<port3>",
               "priority" : 0,
               "votes" : 0
            }
         ]
      }
)
```

2. Step down the primary using `replSetStepDown` against the `admin` database. Ensure enough time for the secondaries to catch up.

   Connect a `mongo` shell to the primary and run:

   ```
   db.adminCommand( { replSetStepDown: 360, secondaryCatchUpPeriodSecs: 300, force: true })
   ```

3. Shut down all members of the config server replica set, the `mongos` instances, and the shards.

4. Restart each config server as standalone 3.2 `mongod`; i.e. without the *`--replSet`* or, if using a configuration file, `replication.replSetName`.

   ```
   mongod --configsvr --dbpath <path> --port <port> --storageEngine <storageEngine>
   ```

Step 3: Update the protocolVersion for each shard. Restart each replica set shard and update the protocolVersion.

Connect a `mongo` shell to the current primary and downgrade the replication protocol:

```
cfg = rs.conf();
cfg.protocolVersion=0;
rs.reconfig(cfg);
```

Step 4: Downgrade the `mongos` instances.

Important: As the config servers changed from a replica set to three mirrored `mongod` instances, update the the *`--configsvr`* setting. All `mongos` must use the same *`--configsvr`* string.

Downgrade the binaries and restart.

Step 5: Downgrade Config Servers. Downgrade the binaries and restart. Downgrade in reverse order of their listing in the *`--configdb`* option for `mongos`.

If your `mongod` instance is using the *WiredTiger* (page 595) storage engine, you must include the *`--storageEngine`* option (or `storage.engine` if using the configuration file) with the 3.0 binary.

```
mongod --configsvr --dbpath <path> --port <port> --storageEngine <storageEngine>
```

Step 6: Downgrade each shard, one at a time. For each replica set shard, downgrade the `mongod` binaries and restart. If your `mongod` instance is using the *WiredTiger* (page 595) storage engine, you must include the *`--storageEngine`* option (or `storage.engine` if using the configuration file) with the 3.0 binary.

1. Downgrade the `mongod` secondaries *before* downgrading the primary.
2. To downgrade the primary, run `replSetStepDown` and then downgrade.

For details on downgrading a replica set, see *Downgrade a 3.2 Replica Set* (page 906).

Step 7: Re-enable the balancer. Once the downgrade of sharded cluster components is complete, *re-enable the balancer* (page 803).

See *Upgrade MongoDB to 3.2* (page 901) for full upgrade instructions.

Known Issues in 3.2.1

List of known issues in the 3.2.1 release:

- Clients may fail to discover new primaries when clock skew between nodes is greater than `electionTimeout`: SERVER-21744[421]
- `fromMigrate` flag never set for deletes in oplog: SERVER-21678[422]
- Running `explain` with a *read preference* (page 727) in a v3.2 `mongo` shell connected to a v3.0 `mongos` or in a v3.0 `mongo` shell connected to a v3.0 `mongos` but with v3.2 shards is incompatible: SERVER-21661[423]
- Results of the connPoolStats command are no longer correct: SERVER-21597[424]
- ApplyOps background index creation may deadlock: SERVER-21583[425]
- Performance regression for `w:majority` writes with replica set protocolVersion 1: SERVER-21581[426]
- Performance regression on unicode-aware text processing logic (text index v3): SERVER-19936[427]
- Results from the `$indexStats` operator do not take into account queries which use the `$match` or `mapReduce` functions: SERVER-22048[428]

Known Issues in 3.2.0

List of known issues in the 3.2.0 release:

- findAndModify operations not captured by the profiler: SERVER-21772[429]
- `getMore` command does not set `"nreturned"` operation counter: SERVER-21750[430]
- Clients may fail to discover new primaries when clock skew between nodes is greater than `electionTimeout`: SERVER-21744[431]
- `fromMigrate` flag never set for deletes in oplog: SERVER-21678[432]
- Running `explain` with a *read preference* (page 727) in a v3.2 `mongo` shell connected to a v3.0 `mongos` or in a v3.0 `mongo` shell connected to a v3.0 `mongos` but with v3.2 shards is incompatible: SERVER-21661[433]

[421] https://jira.mongodb.org/browse/SERVER-21744
[422] https://jira.mongodb.org/browse/SERVER-21678
[423] https://jira.mongodb.org/browse/SERVER-21661
[424] https://jira.mongodb.org/browse/SERVER-21597
[425] https://jira.mongodb.org/browse/SERVER-21583
[426] https://jira.mongodb.org/browse/SERVER-21581
[427] https://jira.mongodb.org/browse/SERVER-19936
[428] https://jira.mongodb.org/browse/SERVER-22048
[429] https://jira.mongodb.org/browse/SERVER-21772
[430] https://jira.mongodb.org/browse/SERVER-21750
[431] https://jira.mongodb.org/browse/SERVER-21744
[432] https://jira.mongodb.org/browse/SERVER-21678
[433] https://jira.mongodb.org/browse/SERVER-21661

- Results of the connPoolStats command are no longer correct: SERVER-21597[434]
- ApplyOps background index creation may deadlock: SERVER-21583[435]
- Performance regression for `w:majority` writes with replica set protocolVersion 1: SERVER-21581[436]
- Performance regression on unicode-aware text processing logic (text index v3): SERVER-19936[437]
- Severe performance regression in insert workload under Windows with WiredTiger: SERVER-21792[438]
- Results from the `$indexStats` operator do not take into account queries which use the `$match` or `mapReduce` functions: SERVER-22048[439]

Download

To download MongoDB 3.2, go to the downloads page[440].

See also:

- All Third Party License Notices[441].
- All JIRA issues resolved in 3.2[442].

Additional Resources

- Getting ready for MongoDB 3.2? Get our help.[443]

15.2 Previous Stable Releases

15.2.1 Release Notes for MongoDB 3.0

On this page

- Minor Releases (page 912)
- Major Changes (page 940)
- Replica Sets (page 942)
- Sharded Clusters (page 943)
- Security Improvements (page 943)
- Improvements (page 943)
- MongoDB Enterprise Features (page 945)
- Additional Information (page 945)
- Additional Resources (page 967)

[434] https://jira.mongodb.org/browse/SERVER-21597
[435] https://jira.mongodb.org/browse/SERVER-21583
[436] https://jira.mongodb.org/browse/SERVER-21581
[437] https://jira.mongodb.org/browse/SERVER-19936
[438] https://jira.mongodb.org/browse/SERVER-21792
[439] https://jira.mongodb.org/browse/SERVER-22048
[440] http://www.mongodb.org/downloads
[441] https://github.com/mongodb/mongo/blob/v3.2/distsrc/THIRD-PARTY-NOTICES
[442] http://bit.ly/1XXomL9
[443] https://www.mongodb.com/contact/mongodb-3-2-upgrade-services?jmp=docs

March 3, 2015

MongoDB 3.0 is now available. Key features include support for the WiredTiger storage engine, pluggable storage engine API, `SCRAM-SHA-1` authentication mechanism, and improved `explain` functionality.

MongoDB Ops Manager, which includes Automation, Backup, and Monitoring, is now also available. See the Ops Manager documentation[444] and the Ops Manager release notes[445] for more information.

Minor Releases

3.0 Changelog

On this page

- 3.0.10 Changelog (page 912)
- 3.0.9 Changelog (page 913)
- 3.0.8 Changelog (page 915)
- 3.0.7 Changelog (page 917)
- 3.0.6 Changelog (page 920)
- 3.0.5 Changelog (page 922)
- 3.0.4 Changelog (page 925)
- 3.0.3 Changelog (page 927)
- 3.0.2 Changelog (page 931)
- 3.0.1 Changelog (page 933)

3.0.10 Changelog

Sharding

- SERVER-18671[446] SecondaryPreferred can end up using unversioned connections
- SERVER-22569[447] Initialization of eooElement static local variable isn't thread safe with MSVC 2013

Query SERVER-22535[448] Some index operations (drop index, abort index build, update TTL config) on collection during active migration can cause migration to skip documents

Storage

- SERVER-19800[449] DataSizeChange forces an int into a bool
- SERVER-22634[450] Data size change for oplog deletes can overflow 32-bit int

WiredTiger

- SERVER-22554[451] WiredTiger data handles not closed when collection is dropped

[444] http://docs.opsmanager.mongodb.com/current/
[445] http://docs.opsmanager.mongodb.com/current/release-notes/application/
[446] https://jira.mongodb.org/browse/SERVER-18671
[447] https://jira.mongodb.org/browse/SERVER-22569
[448] https://jira.mongodb.org/browse/SERVER-22535
[449] https://jira.mongodb.org/browse/SERVER-19800
[450] https://jira.mongodb.org/browse/SERVER-22634
[451] https://jira.mongodb.org/browse/SERVER-22554

MMAP

- SERVER-22261[452] MMAPv1 LSNFile may be updated ahead of what is synced to data files

Build and Packaging

- SERVER-22042[453] If ssl libraries not present, configure fails with a misleading error about boost
- SERVER-22350[454] Package generation failure doesn't fail compile tasks

Tools TOOLS-1039[455] mongoexport problem on data with quotes

Internals

- SERVER-22292[456] Use more reliable mechanism in the mongo shell to wait for process to terminate on windows
- SERVER-22328[457] bench_test_crud_commands.js fails due to resource contention from other resmoke jobs and low timeout values

3.0.9 Changelog

Security SERVER-21724[458] Backup role can't read system.profile

Sharding

- SERVER-19266[459] An error document is returned with result set
- SERVER-21382[460] Sharding migration transfers all document deletions
- SERVER-22114[461] Mongos can accumulate multiple copies of ChunkManager when a shard restarts

Replication

- SERVER-18219[462] "control reaches end of non-void function" errors in GCC with WCE retry loop
- SERVER-21583[463] ApplyOps background index creation may deadlock
- SERVER-22109[464] Invariant failure when running applyOps to create an index with a bad ns field

[452] https://jira.mongodb.org/browse/SERVER-22261
[453] https://jira.mongodb.org/browse/SERVER-22042
[454] https://jira.mongodb.org/browse/SERVER-22350
[455] https://jira.mongodb.org/browse/TOOLS-1039
[456] https://jira.mongodb.org/browse/SERVER-22292
[457] https://jira.mongodb.org/browse/SERVER-22328
[458] https://jira.mongodb.org/browse/SERVER-21724
[459] https://jira.mongodb.org/browse/SERVER-19266
[460] https://jira.mongodb.org/browse/SERVER-21382
[461] https://jira.mongodb.org/browse/SERVER-22114
[462] https://jira.mongodb.org/browse/SERVER-18219
[463] https://jira.mongodb.org/browse/SERVER-21583
[464] https://jira.mongodb.org/browse/SERVER-22109

Query

- SERVER-19128[465] Fatal assertion during secondary index build
- SERVER-19996[466] Queries which specify sort and batch size can generate results out of order, if documents concurrently updated
- SERVER-20083[467] Add log statement at default log level for when an index filter is set or cleared successfully
- SERVER-21602[468] Reduce execution time of cursor_timeout.js
- SERVER-21776[469] Move per-operation log lines for queries out of the QUERY log component

Write Operations SERVER-21647[470] $rename changes field ordering

Aggregation SERVER-7656[471] Optimize aggregation on sharded setup if first stage is exact match on shard key

Storage

- SERVER-20858[472] Invariant failure in OplogStones for non-capped oplog creation
- SERVER-20866[473] Race condition in oplog insert transaction rollback
- SERVER-21545[474] collMod and invalid parameter triggers fassert on dropCollection on mmapv1
- SERVER-22014[475] index_bigkeys_nofail.js triggers spurious failures when run in parallel with other tests

WiredTiger

- SERVER-20961[476] Large amounts of create and drop collections can cause listDatabases to be slow under WiredTiger
- SERVER-22129[477] WiredTiger changes for MongoDB 3.0.9

Operations SERVER-20358[478] Usernames can contain NULL characters

Build and Packaging

- SERVER-17747[479] FreeBSD 11.0-CURRENT build issue
- SERVER-18162[480] Fail to start with non-existing /var/run/mongodb/
- SERVER-18953[481] Generate debug symbols on OS X

[465] https://jira.mongodb.org/browse/SERVER-19128
[466] https://jira.mongodb.org/browse/SERVER-19996
[467] https://jira.mongodb.org/browse/SERVER-20083
[468] https://jira.mongodb.org/browse/SERVER-21602
[469] https://jira.mongodb.org/browse/SERVER-21776
[470] https://jira.mongodb.org/browse/SERVER-21647
[471] https://jira.mongodb.org/browse/SERVER-7656
[472] https://jira.mongodb.org/browse/SERVER-20858
[473] https://jira.mongodb.org/browse/SERVER-20866
[474] https://jira.mongodb.org/browse/SERVER-21545
[475] https://jira.mongodb.org/browse/SERVER-22014
[476] https://jira.mongodb.org/browse/SERVER-20961
[477] https://jira.mongodb.org/browse/SERVER-22129
[478] https://jira.mongodb.org/browse/SERVER-20358
[479] https://jira.mongodb.org/browse/SERVER-17747
[480] https://jira.mongodb.org/browse/SERVER-18162
[481] https://jira.mongodb.org/browse/SERVER-18953

Internals

- SERVER-18373[482] MONGO_COMPILER_UNREACHABLE should terminate if violated
- SERVER-19110[483] Ignore failed operations in mixed_storage_version_replication.js
- SERVER-21934[484] Add extra information to OSX stack traces to facilitate addr2line translation
- SERVER-21960[485] Include symbol name in stacktrace json when available
- SERVER-22013[486] coll_mod_bad_spec.js tries to pass filter to getCollectionInfos on v3.0 branch
- SERVER-22054[487] Authentication failure reports incorrect IP address
- SERVER-22191[488] Race condition in CurOp constructor (<=3.0 only)
- TOOLS-1002[489] oplog_rollover test is flaky

3.0.8 Changelog

Security SERVER-21278[490] Remove executable bit from mongod.lock

Sharding

- SERVER-20407[491] findAndModify on mongoS upserts to the wrong shard
- SERVER-20839[492] trace_missing_docs_test.js compares Timestamp instances using < operator in mongo shell

Query

- SERVER-2454[493] Queries that are killed during a yield should return error to user instead of partial result set
- SERVER-21227[494] MultiPlanStage::invalidate() should not flag and drop invalidated WorkingSetMembers
- SERVER-21275[495] Document not found due to WT commit visibility issue

Storage

- SERVER-20650[496] Backport MongoRocks changes to 3.0
- SERVER-21543[497] Lengthen delay before deleting old journal files

[482] https://jira.mongodb.org/browse/SERVER-18373
[483] https://jira.mongodb.org/browse/SERVER-19110
[484] https://jira.mongodb.org/browse/SERVER-21934
[485] https://jira.mongodb.org/browse/SERVER-21960
[486] https://jira.mongodb.org/browse/SERVER-22013
[487] https://jira.mongodb.org/browse/SERVER-22054
[488] https://jira.mongodb.org/browse/SERVER-22191
[489] https://jira.mongodb.org/browse/TOOLS-1002
[490] https://jira.mongodb.org/browse/SERVER-21278
[491] https://jira.mongodb.org/browse/SERVER-20407
[492] https://jira.mongodb.org/browse/SERVER-20839
[493] https://jira.mongodb.org/browse/SERVER-2454
[494] https://jira.mongodb.org/browse/SERVER-21227
[495] https://jira.mongodb.org/browse/SERVER-21275
[496] https://jira.mongodb.org/browse/SERVER-20650
[497] https://jira.mongodb.org/browse/SERVER-21543

WiredTiger

- SERVER-20303[498] Negative scaling at low thread count under WiredTiger when inserting large documents
- SERVER-21063[499] MongoDB with WiredTiger can build very deep trees
- SERVER-21442[500] WiredTiger changes for MongoDB 3.0.8
- SERVER-21553[501] Oplog grows to 3x configured size

Build and Packaging

- SERVER-10512[502] Add scons flag to set -fno-omit-frame-pointer
- SERVER-19755[503] scons should require c++11 on 3.0
- SERVER-20699[504] Add build manifest to every build
- SERVER-20830[505] set push and docs_tickets tasks as not available for patch testing
- SERVER-20834[506] Perf tasks should only require compiling once before execution
- SERVER-21209[507] PIDFILEPATH computation in init scripts fails to handle comments after values
- SERVER-21477[508] 3.0.7 RPMs missing for yum RHEL server versions

Tools

- TOOLS-702[509] bsondump does not keep attribut order
- TOOLS-920[510] mongodump issue with temporary map/reduce collections
- TOOLS-939[511] Error restoring database "insertion error: EOF"

Internals

- SERVER-8728[512] jstests/profile1.js is a race and fails randomly
- SERVER-20521[513] Update Mongo-perf display names in Evergreen to sort better
- SERVER-20527[514] Delete resmoke.py from the 3.0 branch
- SERVER-20876[515] Hang in scenario with sharded ttl collection under WiredTiger
- SERVER-21027[516] Reduced performance of index lookups after removing documents from collection

[498] https://jira.mongodb.org/browse/SERVER-20303
[499] https://jira.mongodb.org/browse/SERVER-21063
[500] https://jira.mongodb.org/browse/SERVER-21442
[501] https://jira.mongodb.org/browse/SERVER-21553
[502] https://jira.mongodb.org/browse/SERVER-10512
[503] https://jira.mongodb.org/browse/SERVER-19755
[504] https://jira.mongodb.org/browse/SERVER-20699
[505] https://jira.mongodb.org/browse/SERVER-20830
[506] https://jira.mongodb.org/browse/SERVER-20834
[507] https://jira.mongodb.org/browse/SERVER-21209
[508] https://jira.mongodb.org/browse/SERVER-21477
[509] https://jira.mongodb.org/browse/TOOLS-702
[510] https://jira.mongodb.org/browse/TOOLS-920
[511] https://jira.mongodb.org/browse/TOOLS-939
[512] https://jira.mongodb.org/browse/SERVER-8728
[513] https://jira.mongodb.org/browse/SERVER-20521
[514] https://jira.mongodb.org/browse/SERVER-20527
[515] https://jira.mongodb.org/browse/SERVER-20876
[516] https://jira.mongodb.org/browse/SERVER-21027

- SERVER-21099[517] Improve logging in SecureRandom and PseudoRandom classes
- SERVER-21150[518] Basic startup logging should be done as early as possible in initAndListen
- SERVER-21208[519] "server up" check in perf.yml is in the wrong place
- SERVER-21305[520] Lock 'timeAcquiringMicros' value is much higher than the actual time spent
- SERVER-21433[521] Perf.yml project should kill unwanted processes before starting tests
- SERVER-21533[522] Lock manager is not fair in the presence of compatible requests which can be granted immediately

3.0.7 Changelog

Security

- SERVER-13647[523] `root` (page 494) role does not contain sufficient privileges for a `mongorestore` of a system with security enabled
- SERVER-15893[524] `root` (page 494) role should be able to run validate on system collections
- SERVER-19131[525] `clusterManager` (page 489) role does not have permission for adding tag ranges
- SERVER-19284[526] Should not be able to create role with same name as builtin role
- SERVER-20394[527] Remove non-integer test case from `iteration_count_control.js`
- SERVER-20401[528] Publicly expose `net.ssl.disabledProtocols`

Sharding

- SERVER-17886[529] dbKillCursors op asserts on mongos when at log level 3
- SERVER-20191[530] multi-updates/remove can make successive queries skip shard version checking
- SERVER-20460[531] `listIndexes` on 3.0 `mongos` with 2.6 `mongod` instances returns erroneous "not authorized"
- SERVER-20557[532] Active window setting is not being processed correctly

[517] https://jira.mongodb.org/browse/SERVER-21099
[518] https://jira.mongodb.org/browse/SERVER-21150
[519] https://jira.mongodb.org/browse/SERVER-21208
[520] https://jira.mongodb.org/browse/SERVER-21305
[521] https://jira.mongodb.org/browse/SERVER-21433
[522] https://jira.mongodb.org/browse/SERVER-21533
[523] https://jira.mongodb.org/browse/SERVER-13647
[524] https://jira.mongodb.org/browse/SERVER-15893
[525] https://jira.mongodb.org/browse/SERVER-19131
[526] https://jira.mongodb.org/browse/SERVER-19284
[527] https://jira.mongodb.org/browse/SERVER-20394
[528] https://jira.mongodb.org/browse/SERVER-20401
[529] https://jira.mongodb.org/browse/SERVER-17886
[530] https://jira.mongodb.org/browse/SERVER-20191
[531] https://jira.mongodb.org/browse/SERVER-20460
[532] https://jira.mongodb.org/browse/SERVER-20557

Replication

- SERVER-20262[533] Replica set nodes can get stuck in a state where they will not step themselves down
- SERVER-20473[534] calling setMaintenanceMode(true) while running for election crashes server

Query

- SERVER-17895[535] Server should not clear collection plan cache periodically when write operations are issued
- SERVER-19412[536] NULL PlanStage in getStageByType causes segfault during stageDebug command
- SERVER-19725[537] NULL pointer crash in `QueryPlanner::plan` with `$near` operator
- SERVER-20139[538] Enable CachedPlanStage replanning by default in 3.0
- SERVER-20219[539] Add startup warning to 3.0 if have indexes with partialFilterExpression option
- SERVER-20347[540] Document is not found when searching on a field indexed by a hash index using a `$in` clause with regular expression
- SERVER-20364[541] Cursor is not closed when querying `system.profile` collection with `clusterMonitor` (page 490) role

Write Operations

- SERVER-11746[542] Improve shard version checking for versioned (single) updates after yield
- SERVER-19361[543] Insert of document with duplicate `_id` fields should be forbidden
- SERVER-20531[544] Mongodb server crash: Invariant failure res.existing

Storage

- SERVER-18624[545] `listCollections` command should not be O(n^2) on MMAPv1
- SERVER-20617[546] `wt_nojournal_toggle.js` failing intermittently in noPassthrough_WT
- SERVER-20638[547] Reading the profiling level shouldn't create databases that don't exist

WiredTiger

- SERVER-18250[548] Once enabled journal cannot be disabled under WiredTiger
- SERVER-20008[549] Stress test deadlock in WiredTiger

[533] https://jira.mongodb.org/browse/SERVER-20262
[534] https://jira.mongodb.org/browse/SERVER-20473
[535] https://jira.mongodb.org/browse/SERVER-17895
[536] https://jira.mongodb.org/browse/SERVER-19412
[537] https://jira.mongodb.org/browse/SERVER-19725
[538] https://jira.mongodb.org/browse/SERVER-20139
[539] https://jira.mongodb.org/browse/SERVER-20219
[540] https://jira.mongodb.org/browse/SERVER-20347
[541] https://jira.mongodb.org/browse/SERVER-20364
[542] https://jira.mongodb.org/browse/SERVER-11746
[543] https://jira.mongodb.org/browse/SERVER-19361
[544] https://jira.mongodb.org/browse/SERVER-20531
[545] https://jira.mongodb.org/browse/SERVER-18624
[546] https://jira.mongodb.org/browse/SERVER-20617
[547] https://jira.mongodb.org/browse/SERVER-20638
[548] https://jira.mongodb.org/browse/SERVER-18250
[549] https://jira.mongodb.org/browse/SERVER-20008

- SERVER-20091[550] Poor query throughput and erratic behavior at high connection counts under WiredTiger
- SERVER-20159[551] Out of memory on index build during initial sync even with low cacheSize parameter
- SERVER-20176[552] Deletes with `j:true` slower on WT than MMAPv1
- SERVER-20204[553] Segmentation fault during index build on 3.0 secondary

Operations

- SERVER-14750[554] Convert RPM and DEB mongod.conf files to new YAML format
- SERVER-18506[555] Balancer section of printShardingStatus should respect passed-in configDB

Build and Packaging

- SERVER-18516[556] ubuntu/debian packaging : Release files report wrong Codename
- SERVER-18581[557] The Ubuntu package should start the mongod with group=mongodb
- SERVER-18749[558] Ubuntu startup files have an inconsistent directory for dbpath and logs
- SERVER-18793[559] Enterprise RPM build issues
- SERVER-19088[560] The –cache flag should force –build-fast-and-loose=off
- SERVER-19509[561] The nproc ulimits are different across packages
- SERVER-19661[562] Build fails: error: expected expression

Tools

- TOOLS-767[563] `mongorestore`: error parsing metadata: call of reflect.Value.Set on zero Value
- TOOLS-847[564] `mongorestore` exits in response to SIGHUP, even when run under nohup
- TOOLS-874[565] `mongoimport` $date close to epoch not working
- TOOLS-916[566] `mongoexport` throws reflect.Value.Type errors

Internals

- SERVER-18178[567] Fix `mr_drop.js` test to not fail from nondeterministic collection drop timing

[550] https://jira.mongodb.org/browse/SERVER-20091
[551] https://jira.mongodb.org/browse/SERVER-20159
[552] https://jira.mongodb.org/browse/SERVER-20176
[553] https://jira.mongodb.org/browse/SERVER-20204
[554] https://jira.mongodb.org/browse/SERVER-14750
[555] https://jira.mongodb.org/browse/SERVER-18506
[556] https://jira.mongodb.org/browse/SERVER-18516
[557] https://jira.mongodb.org/browse/SERVER-18581
[558] https://jira.mongodb.org/browse/SERVER-18749
[559] https://jira.mongodb.org/browse/SERVER-18793
[560] https://jira.mongodb.org/browse/SERVER-19088
[561] https://jira.mongodb.org/browse/SERVER-19509
[562] https://jira.mongodb.org/browse/SERVER-19661
[563] https://jira.mongodb.org/browse/TOOLS-767
[564] https://jira.mongodb.org/browse/TOOLS-847
[565] https://jira.mongodb.org/browse/TOOLS-874
[566] https://jira.mongodb.org/browse/TOOLS-916
[567] https://jira.mongodb.org/browse/SERVER-18178

- SERVER-19819[568] Update perf.yml to use new mongo-perf release
- SERVER-19820[569] Update perf.yml to use mongo-perf check script
- SERVER-19899[570] Mongo-perf analysis script – Check for per thread level regressions
- SERVER-19901[571] Mongo-perf analysis script – Compare to tagged baseline
- SERVER-19902[572] Mongo-perf analysis script – Use noise data for regression comparison instead of fixed percentage
- SERVER-20035[573] Updated perf_regresison_check.py script to output report.json summarizing results
- SERVER-20121[574] XorShift PRNG should use unsigned arithmetic
- SERVER-20216[575] Extend optional Command properties to SASL
- SERVER-20316[576] Relax thread level comparisons on mongo-perf check script
- SERVER-20322[577] Wiredtiger develop can lose records following stop even with log enabled
- SERVER-20383[578] Cleanup old connections after every ChunkManagerTest
- SERVER-20429[579] Canceled lock attempts should unblock pending requests
- SERVER-20464[580] Add units of measurement to log output of perf regression analysis
- SERVER-20691[581] Improve SASL and SCRAM compatibility
- TOOLS-894[582] `mongoimport --upsert --type json` with _id being an object does not work most of the times
- TOOLS-898[583] Mongo tools attempt to connect as ipv6 rather than ipv4 by default, when built with go 1.5

3.0.6 Changelog

Security SERVER-19538[584] Segfault when calling dbexit in SSLManager with auditing enabled

Querying

- SERVER-19553[585] Mongod shouldn't use sayPiggyBack to send KillCursor messages

[568] https://jira.mongodb.org/browse/SERVER-19819
[569] https://jira.mongodb.org/browse/SERVER-19820
[570] https://jira.mongodb.org/browse/SERVER-19899
[571] https://jira.mongodb.org/browse/SERVER-19901
[572] https://jira.mongodb.org/browse/SERVER-19902
[573] https://jira.mongodb.org/browse/SERVER-20035
[574] https://jira.mongodb.org/browse/SERVER-20121
[575] https://jira.mongodb.org/browse/SERVER-20216
[576] https://jira.mongodb.org/browse/SERVER-20316
[577] https://jira.mongodb.org/browse/SERVER-20322
[578] https://jira.mongodb.org/browse/SERVER-20383
[579] https://jira.mongodb.org/browse/SERVER-20429
[580] https://jira.mongodb.org/browse/SERVER-20464
[581] https://jira.mongodb.org/browse/SERVER-20691
[582] https://jira.mongodb.org/browse/TOOLS-894
[583] https://jira.mongodb.org/browse/TOOLS-898
[584] https://jira.mongodb.org/browse/SERVER-19538
[585] https://jira.mongodb.org/browse/SERVER-19553

Replication

- SERVER-19719[586] Failure to rollback noPadding should not cause fatal error
- SERVER-19644[587] Seg Fault on cloneCollection (specifically gridfs)

WiredTiger

- SERVER-19673[588] Excessive memory allocated by WiredTiger journal
- SERVER-19987[589] Limit the size of the per-session cursor cache
- SERVER-19751[590] WiredTiger panic halt in eviction-server
- SERVER-19744[591] WiredTiger changes for MongoDB 3.0.6
- SERVER-19573[592] MongoDb crash due to segfault
- SERVER-19522[593] Capped collection insert rate declines over time under WiredTiger

MMAPv1 SERVER-19805[594] MMap memory mapped file address allocation code cannot handle addresses non-aligned to memory mapped granularity size

Networking

- SERVER-19389[595] Remove wire level endianness check

Aggregation Framework

- SERVER-19553[596] Mongod shouldn't use sayPiggyBack to send KillCursor messages
- SERVER-19464[597] $sort stage in aggregation doesn't call scoped connections done ()

Build and Testing

- SERVER-19650[598] update YML files to tag system/test command types
- SERVER-19236[599] clang-format the v3.0 branch
- SERVER-19540[600] Add perf.yml file to 3.0 branch for mongo-perf regressions

Internal Code

- SERVER-19856[601] Register for PRESHUTDOWN notifications on Windows Vista+

[586] https://jira.mongodb.org/browse/SERVER-19719
[587] https://jira.mongodb.org/browse/SERVER-19644
[588] https://jira.mongodb.org/browse/SERVER-19673
[589] https://jira.mongodb.org/browse/SERVER-19987
[590] https://jira.mongodb.org/browse/SERVER-19751
[591] https://jira.mongodb.org/browse/SERVER-19744
[592] https://jira.mongodb.org/browse/SERVER-19573
[593] https://jira.mongodb.org/browse/SERVER-19522
[594] https://jira.mongodb.org/browse/SERVER-19805
[595] https://jira.mongodb.org/browse/SERVER-19389
[596] https://jira.mongodb.org/browse/SERVER-19553
[597] https://jira.mongodb.org/browse/SERVER-19464
[598] https://jira.mongodb.org/browse/SERVER-19650
[599] https://jira.mongodb.org/browse/SERVER-19236
[600] https://jira.mongodb.org/browse/SERVER-19540
[601] https://jira.mongodb.org/browse/SERVER-19856

Tools

`mongoexport` and `bsondump`

- TOOLS-848[602] Can't handle some regexes

mongoimport

- TOOLS-874[603] `mongoimport $date` close to epoch not working

mongotop

- TOOLS-864[604] `mongotop` "i/o timeout error"

3.0.5 Changelog

Querying

- SERVER-19489[605] Assertion failure and segfault in WorkingSet::free in 3.0.5-rc0
- SERVER-18461[606] Range predicates comparing against a BinData value should be covered, but are not in 2.6
- SERVER-17815[607] Plan ranking tie breaker is computed incorrectly
- SERVER-17259[608] Coverity analysis defect 56350: Dereference null return value
- SERVER-18926[609] Full text search extremely slow and uses a lot of memory under WiredTiger

Replication

- SERVER-19375[610] choosing syncsource should compare against last fetched optime rather than last applied
- SERVER-19298[611] Use userCreateNS w/options consistently in cloner
- SERVER-18994[612] producer thread can continue producing after a node becomes primary
- SERVER-18455[613] master/slave keepalives are not silent on slaves
- SERVER-18280[614] ReplicaSetMonitor should use electionId to avoid talking to old primaries
- SERVER-17689[615] Server crash during initial replication sync

Sharding SERVER-18955[616] mongoS doesn't set batch size (and keeps the old one, 0) on getMore if performed on first _cursor->more()

[602] https://jira.mongodb.org/browse/TOOLS-848
[603] https://jira.mongodb.org/browse/TOOLS-874
[604] https://jira.mongodb.org/browse/TOOLS-864
[605] https://jira.mongodb.org/browse/SERVER-19489
[606] https://jira.mongodb.org/browse/SERVER-18461
[607] https://jira.mongodb.org/browse/SERVER-17815
[608] https://jira.mongodb.org/browse/SERVER-17259
[609] https://jira.mongodb.org/browse/SERVER-18926
[610] https://jira.mongodb.org/browse/SERVER-19375
[611] https://jira.mongodb.org/browse/SERVER-19298
[612] https://jira.mongodb.org/browse/SERVER-18994
[613] https://jira.mongodb.org/browse/SERVER-18455
[614] https://jira.mongodb.org/browse/SERVER-18280
[615] https://jira.mongodb.org/browse/SERVER-17689
[616] https://jira.mongodb.org/browse/SERVER-18955

Storage

- SERVER-19283[617] WiredTiger changes for MongoDB 3.0.5
- SERVER-18874[618] Backport changes to RocksDB from mongo-partners repo
- SERVER-18838[619] DB fails to recover creates and drops after system crash
- SERVER-17370[620] Clean up storage engine-specific index and collection options
- SERVER-15901[621] Cleanup unused locks on the lock manager

WiredTiger

- SERVER-19513[622] Truncating a capped collection may not unindex deleted documents in WiredTiger
- SERVER-19283[623] WiredTiger changes for MongoDB 3.0.5
- SERVER-19189[624] Improve performance under high number of threads with WT
- SERVER-19178[625] In WiredTiger capped collection truncates, avoid walking lists of deleted items
- SERVER-19052[626] Remove sizeStorer recalculations at startup with WiredTiger
- SERVER-18926[627] Full text search extremely slow and uses a lot of memory under WiredTiger
- SERVER-18902[628] Retrieval of large documents slower on WiredTiger than MMAPv1
- SERVER-18875[629] Oplog performance on WT degrades over time after accumulation of deleted items
- SERVER-18838[630] DB fails to recover creates and drops after system crash
- SERVER-18829[631] Cache usage exceeds configured maximum during index builds under WiredTiger
- SERVER-18321[632] Speed up background index build with WiredTiger LSM
- SERVER-17689[633] Server crash during initial replication sync
- SERVER-17386[634] Cursor cache causes excessive memory utilization in WiredTiger
- SERVER-17254[635] WT: drop collection while concurrent oplog tailing may greatly reduce throughput
- SERVER-17078[636] show databases taking extraordinarily long with wiredTiger

[617] https://jira.mongodb.org/browse/SERVER-19283
[618] https://jira.mongodb.org/browse/SERVER-18874
[619] https://jira.mongodb.org/browse/SERVER-18838
[620] https://jira.mongodb.org/browse/SERVER-17370
[621] https://jira.mongodb.org/browse/SERVER-15901
[622] https://jira.mongodb.org/browse/SERVER-19513
[623] https://jira.mongodb.org/browse/SERVER-19283
[624] https://jira.mongodb.org/browse/SERVER-19189
[625] https://jira.mongodb.org/browse/SERVER-19178
[626] https://jira.mongodb.org/browse/SERVER-19052
[627] https://jira.mongodb.org/browse/SERVER-18926
[628] https://jira.mongodb.org/browse/SERVER-18902
[629] https://jira.mongodb.org/browse/SERVER-18875
[630] https://jira.mongodb.org/browse/SERVER-18838
[631] https://jira.mongodb.org/browse/SERVER-18829
[632] https://jira.mongodb.org/browse/SERVER-18321
[633] https://jira.mongodb.org/browse/SERVER-17689
[634] https://jira.mongodb.org/browse/SERVER-17386
[635] https://jira.mongodb.org/browse/SERVER-17254
[636] https://jira.mongodb.org/browse/SERVER-17078

Networking

- SERVER-19255[637] Listener::waitUntilListening may return before listening has started

Build and Packaging

- SERVER-18911[638] Update source tarball push
- SERVER-18910[639] Path in distribution does not contain version
- SERVER-18371[640] Add SSL library config detection
- SERVER-17782[641] Generate source tarballs with pre-interpolated version metadata files from SCons
- SERVER-17568[642] Report most-vexing parse warnings as errors on MSVC
- SERVER-17329[643] Improve management of server version in build system
- SERVER-18977[644] Initscript does not stop a running mongod daemon
- SERVER-18911[645] Update source tarball push

Shell

- SERVER-18795[646] db.printSlaveReplicationInfo()/rs.printSlaveReplicationInfo() can not work with ARBITER role

Logging and Diagnostics

- SERVER-19054[647] Don't be too chatty about periodic tasks taking a few ms
- SERVER-18979[648] Duplicate uassert & fassert codes
- SERVER-19382[649] mongod enterprise crash running as snmp sub-agent

Internal Code and Testing

- SERVER-19353[650] Compilation failure with GCC 5.1
- SERVER-19298[651] Use userCreateNS w/options consistently in cloner
- SERVER-19255[652] Listener::waitUntilListening may return before listening has started
- SERVER-17728[653] typeid(glvalue) warns on clang 3.6

[637] https://jira.mongodb.org/browse/SERVER-19255
[638] https://jira.mongodb.org/browse/SERVER-18911
[639] https://jira.mongodb.org/browse/SERVER-18910
[640] https://jira.mongodb.org/browse/SERVER-18371
[641] https://jira.mongodb.org/browse/SERVER-17782
[642] https://jira.mongodb.org/browse/SERVER-17568
[643] https://jira.mongodb.org/browse/SERVER-17329
[644] https://jira.mongodb.org/browse/SERVER-18977
[645] https://jira.mongodb.org/browse/SERVER-18911
[646] https://jira.mongodb.org/browse/SERVER-18795
[647] https://jira.mongodb.org/browse/SERVER-19054
[648] https://jira.mongodb.org/browse/SERVER-18979
[649] https://jira.mongodb.org/browse/SERVER-19382
[650] https://jira.mongodb.org/browse/SERVER-19353
[651] https://jira.mongodb.org/browse/SERVER-19298
[652] https://jira.mongodb.org/browse/SERVER-19255
[653] https://jira.mongodb.org/browse/SERVER-17728

- SERVER-17567[654] Unconditional export of parseNumberFromStringWithBase
- SERVER-19540[655] Add perf.yml file to 3.0 branch for mongo-perf regressions
- SERVER-18068[656] Coverity analysis defect 72413: Resource leak
- SERVER-17259[657] Coverity analysis defect 56350: Dereference null return value
- SERVER-15017[658] benchRun might return incorrect stats values
- SERVER-19525[659] use of wrong type for size count of rolling back insert

3.0.4 Changelog

Security

- SERVER-18475[660] `authSchemaUpgrade` fails when the `system.users` (page 377) contains non `MONGODB-CR` users
- SERVER-18312[661] Upgrade PCRE to latest

Querying

- SERVER-18364[662] Ensure non-negation predicates get chosen over negation predicates for multikey index bounds construction
- SERVER-16265[663] Add query details to getmore entry in profiler and `db.currentOp()`
- SERVER-15225[664] `CachedPlanStage` should execute for trial period and re-plan if query performs poorly
- SERVER-13875[665] `ensureIndex()` of `2dsphere` index breaks after upgrading to 2.6 (with the new `createIndex` command)

Replication

- SERVER-18566[666] Primary member can trip fatal assertion if stepping down while running findAndModify op resulting in an upsert
- SERVER-18511[667] Report upstream progress when initial sync completes
- SERVER-18409[668] Retry failed heartbeats before marking a node as DOWN
- SERVER-18326[669] Rollback attempted during initial sync is fatal

[654] https://jira.mongodb.org/browse/SERVER-17567
[655] https://jira.mongodb.org/browse/SERVER-19540
[656] https://jira.mongodb.org/browse/SERVER-18068
[657] https://jira.mongodb.org/browse/SERVER-17259
[658] https://jira.mongodb.org/browse/SERVER-15017
[659] https://jira.mongodb.org/browse/SERVER-19525
[660] https://jira.mongodb.org/browse/SERVER-18475
[661] https://jira.mongodb.org/browse/SERVER-18312
[662] https://jira.mongodb.org/browse/SERVER-18364
[663] https://jira.mongodb.org/browse/SERVER-16265
[664] https://jira.mongodb.org/browse/SERVER-15225
[665] https://jira.mongodb.org/browse/SERVER-13875
[666] https://jira.mongodb.org/browse/SERVER-18566
[667] https://jira.mongodb.org/browse/SERVER-18511
[668] https://jira.mongodb.org/browse/SERVER-18409
[669] https://jira.mongodb.org/browse/SERVER-18326

- SERVER-17923[670] Creating/dropping multiple background indexes on the same collection can cause fatal error on secondaries
- SERVER-17913[671] New primary should log voters at default log level
- SERVER-17807[672] drain ops before restarting initial sync
- SERVER-15252[673] Write unit tests of ScatterGatherRunner
- SERVER-15192[674] Make all logOp listeners rollback-safe
- SERVER-18190[675] Secondary reads block replication

Sharding

- SERVER-18822[676] Sharded clusters with WiredTiger primaries may lose writes during chunk migration
- SERVER-18246[677] getmore on secondary in recovery mode can crash `mongos`

Storage SERVER-18442[678] better error message when attempting to change storage engine metadata options

WiredTiger

- SERVER-18647[679] WiredTiger changes for MongoDB 3.0.4
- SERVER-18646[680] Avoid WiredTiger checkpointing dead handles
- SERVER-18629[681] WiredTiger journal system syncs wrong directory
- SERVER-18460[682] Segfault during eviction under load
- SERVER-18316[683] Database with WT engine fails to recover after system crash
- SERVER-18315[684] Throughput drop during transaction pinned phase of checkpoints under WiredTiger
- SERVER-18213[685] Lots of `WriteConflict` during multi-upsert with WiredTiger storage engine
- SERVER-18079[686] Large performance drop with documents > 16k on Windows
- SERVER-17944[687] `WiredTigerRecordStore::truncate` spends a lot of time sleeping

HTTP Console SERVER-18117[688] Bring back the _replSet page in the html interface

[670] https://jira.mongodb.org/browse/SERVER-17923
[671] https://jira.mongodb.org/browse/SERVER-17913
[672] https://jira.mongodb.org/browse/SERVER-17807
[673] https://jira.mongodb.org/browse/SERVER-15252
[674] https://jira.mongodb.org/browse/SERVER-15192
[675] https://jira.mongodb.org/browse/SERVER-18190
[676] https://jira.mongodb.org/browse/SERVER-18822
[677] https://jira.mongodb.org/browse/SERVER-18246
[678] https://jira.mongodb.org/browse/SERVER-18442
[679] https://jira.mongodb.org/browse/SERVER-18647
[680] https://jira.mongodb.org/browse/SERVER-18646
[681] https://jira.mongodb.org/browse/SERVER-18629
[682] https://jira.mongodb.org/browse/SERVER-18460
[683] https://jira.mongodb.org/browse/SERVER-18316
[684] https://jira.mongodb.org/browse/SERVER-18315
[685] https://jira.mongodb.org/browse/SERVER-18213
[686] https://jira.mongodb.org/browse/SERVER-18079
[687] https://jira.mongodb.org/browse/SERVER-17944
[688] https://jira.mongodb.org/browse/SERVER-18117

Build and Packaging

- SERVER-18894[689] OSX SSL builds should use unique filename
- SERVER-18421[690] Create SSL Builder for OS X
- SERVER-18312[691] Upgrade PCRE to latest
- SERVER-13596[692] Support –prefix rpm installation

Internal Code SERVER-6826[693] Potential memory leak in `ConnectionString::connect`

Testing

- SERVER-18318[694] Disable `jsCore_small_oplog` suite in Windows
- SERVER-17336[695] fix `core/compact_keeps_indexes.js` in a master/slave test configuration
- SERVER-13237[696] `benchRun` should use a thread-safe random number generator
- SERVER-18097[697] Remove `mongosTest_auth` and `mongosTest_WT` tasks from evergreen.yml

3.0.3 Changelog

Security

- SERVER-18290[698] Adding a read role for a user doesn't seem to propagate to secondary until restart
- SERVER-18239[699] `dumpauth.js` uses ambiguous `--db`/`--collection` args
- SERVER-18169[700] Regression: Auth enabled arbiter cannot be shutdown using command
- SERVER-18140[701] Allow `getParameter` to be executed locally against an arbiter in an authenticated replica set
- SERVER-18051[702] OpenSSL internal error when using SCRAM-SHA1 authentication in FIPS mode
- SERVER-18021[703] Allow `serverStatus` to be executed locally against an arbiter in an authenticated replica set
- SERVER-17908[704] Allow `getCmdLineOpts` to be executed locally against an arbiter in an authenticated replica set
- SERVER-17832[705] Memory leak when `mongod` configured with SSL required and handle insecure connection

[689] https://jira.mongodb.org/browse/SERVER-18894
[690] https://jira.mongodb.org/browse/SERVER-18421
[691] https://jira.mongodb.org/browse/SERVER-18312
[692] https://jira.mongodb.org/browse/SERVER-13596
[693] https://jira.mongodb.org/browse/SERVER-6826
[694] https://jira.mongodb.org/browse/SERVER-18318
[695] https://jira.mongodb.org/browse/SERVER-17336
[696] https://jira.mongodb.org/browse/SERVER-13237
[697] https://jira.mongodb.org/browse/SERVER-18097
[698] https://jira.mongodb.org/browse/SERVER-18290
[699] https://jira.mongodb.org/browse/SERVER-18239
[700] https://jira.mongodb.org/browse/SERVER-18169
[701] https://jira.mongodb.org/browse/SERVER-18140
[702] https://jira.mongodb.org/browse/SERVER-18051
[703] https://jira.mongodb.org/browse/SERVER-18021
[704] https://jira.mongodb.org/browse/SERVER-17908
[705] https://jira.mongodb.org/browse/SERVER-17832

- SERVER-17812[706] LockPinger has audit-related GLE failure
- SERVER-17591[707] Add SSL flag to select supported protocols
- SERVER-16073[708] Allow disabling SSL Ciphers via hidden flag: `sslCipherConfig`
- SERVER-12235[709] Don't require a database read on every new localhost connection when auth is on

Querying

- SERVER-18304[710] duplicates on FindAndModify with remove option
- SERVER-17815[711] Plan ranking tie breaker is computed incorrectly

Replication

- SERVER-18211[712] MongoDB fails to correctly roll back collection creation
- SERVER-17273[713] Add support for `secondaryCatchupPeriodSecs` to `rs.stepdown()` shell helper

Sharding

- SERVER-17812[714] LockPinger has audit-related GLE failure
- SERVER-17749[715] `collMod usePowerOf2Sizes` fails on `mongos`
- SERVER-16987[716] `sh.getRecentMigrations()` shows aborted migration as success

Storage

- SERVER-18211[717] MongoDB fails to correctly roll back collection creation
- SERVER-18111[718] mongod allows user inserts into system.profile collection
- SERVER-17939[719] Backport mongo-rocks updates to v3.0 branch
- SERVER-17745[720] Improve dirty page estimation in mmapv1 on Windows

WiredTiger

- SERVER-18205[721] WiredTiger changes for MongoDB 3.0.3
- SERVER-18192[722] Crash running WiredTiger with "cache_resident=true"

[706] https://jira.mongodb.org/browse/SERVER-17812
[707] https://jira.mongodb.org/browse/SERVER-17591
[708] https://jira.mongodb.org/browse/SERVER-16073
[709] https://jira.mongodb.org/browse/SERVER-12235
[710] https://jira.mongodb.org/browse/SERVER-18304
[711] https://jira.mongodb.org/browse/SERVER-17815
[712] https://jira.mongodb.org/browse/SERVER-18211
[713] https://jira.mongodb.org/browse/SERVER-17273
[714] https://jira.mongodb.org/browse/SERVER-17812
[715] https://jira.mongodb.org/browse/SERVER-17749
[716] https://jira.mongodb.org/browse/SERVER-16987
[717] https://jira.mongodb.org/browse/SERVER-18211
[718] https://jira.mongodb.org/browse/SERVER-18111
[719] https://jira.mongodb.org/browse/SERVER-17939
[720] https://jira.mongodb.org/browse/SERVER-17745
[721] https://jira.mongodb.org/browse/SERVER-18205
[722] https://jira.mongodb.org/browse/SERVER-18192

- SERVER-18014[723] Dropping a collection can block creating a new collection for an extended time under WiredTiger
- SERVER-17907[724] B-tree eviction blocks access to collection for extended period under WiredTiger
- SERVER-17892[725] Explicitly turn checksum on for all collections/indexes in WiredTiger by default

Indexing

- SERVER-18087[726] index_retry.js and index_no_retry.js not checking for presence of "progress" field in currentOp() result
- SERVER-17882[727] Update with key too large to index crashes WiredTiger/RockDB secondary

Write Ops

- SERVER-18111[728] mongod allows user inserts into system.profile collection

Networking

- SERVER-17832[729] Memory leak when MongoD configured with SSL required and handle insecure connection
- SERVER-17591[730] Add SSL flag to select supported protocols
- SERVER-16073[731] Allow disabling SSL Ciphers via hidden flag: sslCipherConfig

Concurrency

- SERVER-18304[732] duplicates on FindAndModify with remove option
- SERVER-16636[733] Deadlock detection should check cycles for stability or should be disabled

Geo

- SERVER-17835[734] Aggregation geoNear deprecated uniqueDocs warning
- SERVER-9220[735] allow more than two values in the coordinate-array when using 2dsphere index

Aggregation Framework

- SERVER-17835[736] Aggregation geoNear deprecated uniqueDocs warning

[723] https://jira.mongodb.org/browse/SERVER-18014
[724] https://jira.mongodb.org/browse/SERVER-17907
[725] https://jira.mongodb.org/browse/SERVER-17892
[726] https://jira.mongodb.org/browse/SERVER-18087
[727] https://jira.mongodb.org/browse/SERVER-17882
[728] https://jira.mongodb.org/browse/SERVER-18111
[729] https://jira.mongodb.org/browse/SERVER-17832
[730] https://jira.mongodb.org/browse/SERVER-17591
[731] https://jira.mongodb.org/browse/SERVER-16073
[732] https://jira.mongodb.org/browse/SERVER-18304
[733] https://jira.mongodb.org/browse/SERVER-16636
[734] https://jira.mongodb.org/browse/SERVER-17835
[735] https://jira.mongodb.org/browse/SERVER-9220
[736] https://jira.mongodb.org/browse/SERVER-17835

MapReduce

- SERVER-17889[737] Using eval command to run mapReduce with non-inline "out" option triggers fatal assertion failure

Admin

- SERVER-18290[738] Adding a read role for a user doesn't seem to propagate to secondary until restart
- SERVER-18169[739] Regression: Auth enabled arbiter cannot be shutdown using command
- SERVER-17820[740] Windows service stop can lead to mongod abrupt termination due to long shutdown time

Build and Packaging

- SERVER-18344[741] logs should be sent to updated logkeeper server
- SERVER-18299[742] Use ld wrapper for compiling Enterprise GO tools in RHEL 5
- SERVER-18082[743] Change `smoke.py` buildlogger command line options to environment variables
- SERVER-17730[744] Parsing of Variables on Windows doesn't respect windows conventions
- SERVER-17694[745] support `RPATH=value` in top-level `SConstruct`
- SERVER-17465[746] `--use-system-tcmalloc` does not support `tcmalloc setParameters` and extension
- SERVER-17961[747] *THIRD-PARTY-NOTICES.windows'* needs to be updated
- SERVER-17780[748] Init script sets process ulimit to different value compared to documentation

JavaScript

- SERVER-17453[749] warn that db.eval() / eval command is deprecated

Shell

- SERVER-17951[750] db.currentOp() fails with read preference set
- SERVER-17273[751] Add support for secondaryCatchupPeriodSecs to rs.stepdown shell helper
- SERVER-16987[752] sh.getRecentMigrations shows aborted migration as success

[737] https://jira.mongodb.org/browse/SERVER-17889
[738] https://jira.mongodb.org/browse/SERVER-18290
[739] https://jira.mongodb.org/browse/SERVER-18169
[740] https://jira.mongodb.org/browse/SERVER-17820
[741] https://jira.mongodb.org/browse/SERVER-18344
[742] https://jira.mongodb.org/browse/SERVER-18299
[743] https://jira.mongodb.org/browse/SERVER-18082
[744] https://jira.mongodb.org/browse/SERVER-17730
[745] https://jira.mongodb.org/browse/SERVER-17694
[746] https://jira.mongodb.org/browse/SERVER-17465
[747] https://jira.mongodb.org/browse/SERVER-17961
[748] https://jira.mongodb.org/browse/SERVER-17780
[749] https://jira.mongodb.org/browse/SERVER-17453
[750] https://jira.mongodb.org/browse/SERVER-17951
[751] https://jira.mongodb.org/browse/SERVER-17273
[752] https://jira.mongodb.org/browse/SERVER-16987

Testing

- SERVER-18302[753] remove test buildlogger instance
- SERVER-18262[754] setup_multiversion_mongodb should retry links download on timeouts
- SERVER-18239[755] dumpauth.js uses ambiguous –db/–collection args
- SERVER-18229[756] Smoke.py with PyMongo 3.0.1 fails to run certain tests
- SERVER-18073[757] Fix smoke.py to work with pymongo 3.0
- SERVER-17998[758] Ignore socket exceptions in initial_sync_unsupported_auth_schema.js test
- SERVER-18293[759] ASAN tests should run on larger instance size
- SERVER-17761[760] RestAdminAccess/NoAdminAccess objects leak at shutdown

3.0.2 Changelog

Security

- SERVER-17719[761] `mongo` Shell crashes if -p is missing and user matches
- SERVER-17705[762] Fix credentials field inconsistency in HTTP interface
- SERVER-17671[763] Refuse to complete initial sync from nodes with 2.4-style auth data
- SERVER-17669[764] Remove auth prompt in webserver when auth is not enabled
- SERVER-17647[765] Compute BinData length in v8
- SERVER-17529[766] Can't list collections when `mongos` is running 3.0 and config servers are running 2.6 and auth is on

Query and Indexing

- SERVER-8188[767] Configurable idle cursor timeout
- SERVER-17469[768] `2d` nearSphere queries scan entire collection
- SERVER-17642[769] `WriteConfictException` during background index create

[753] https://jira.mongodb.org/browse/SERVER-18302
[754] https://jira.mongodb.org/browse/SERVER-18262
[755] https://jira.mongodb.org/browse/SERVER-18239
[756] https://jira.mongodb.org/browse/SERVER-18229
[757] https://jira.mongodb.org/browse/SERVER-18073
[758] https://jira.mongodb.org/browse/SERVER-17998
[759] https://jira.mongodb.org/browse/SERVER-18293
[760] https://jira.mongodb.org/browse/SERVER-17761
[761] https://jira.mongodb.org/browse/SERVER-17719
[762] https://jira.mongodb.org/browse/SERVER-17705
[763] https://jira.mongodb.org/browse/SERVER-17671
[764] https://jira.mongodb.org/browse/SERVER-17669
[765] https://jira.mongodb.org/browse/SERVER-17647
[766] https://jira.mongodb.org/browse/SERVER-17529
[767] https://jira.mongodb.org/browse/SERVER-8188
[768] https://jira.mongodb.org/browse/SERVER-17469
[769] https://jira.mongodb.org/browse/SERVER-17642

Replication

- SERVER-17677[770] Replica Set member backtraces sometimes when removed from replica set
- SERVER-17672[771] `serverStatus` command with `{oplog: 1}` option can trigger segmentation fault in `mongod`
- SERVER-17822[772] `OpDebug::writeConflicts` should be a 64-bit type

Sharding SERVER-17805[773] `logOp` / `OperationObserver` should always check shardversion

Storage SERVER-17613[774] Unable to start `mongod` after unclean shutdown

WiredTiger

- SERVER-17713[775] WiredTiger using zlib compression can create invalid compressed stream
- SERVER-17642[776] WriteConfictException during background index create
- SERVER-17587[777] Node crash scenario results in uncrecoverable error on subsequent startup under WiredTiger
- SERVER-17562[778] Invariant failure: `s->commit_transaction(s, NULL)` resulted in status `BadValue 22`
- SERVER-17551[779] mongod fatal assertion after "hazard pointer table full" message
- SERVER-17532[780] Duplicate key error message does not contain index name anymore
- SERVER-17471[781] WiredTiger Mutex on Windows can block the server
- SERVER-17382[782] rc10/wiredTiger multi collection/DB bulk insert slow than rc8 in initial insertion phase
- SERVER-16804[783] *`mongod --repair`* fails because `verify()` returns EBUSY under WiredTiger

MMAPv1

- SERVER-17616[784] Removing or inserting documents with large indexed arrays consumes excessive memory
- SERVER-17313[785] Segfault in `BtreeLogic::_insert` when inserting into previously-dropped namespace

RocksDB SERVER-17706[786] Sync new mongo+rocks changes to v3.0 branch

[770] https://jira.mongodb.org/browse/SERVER-17677
[771] https://jira.mongodb.org/browse/SERVER-17672
[772] https://jira.mongodb.org/browse/SERVER-17822
[773] https://jira.mongodb.org/browse/SERVER-17805
[774] https://jira.mongodb.org/browse/SERVER-17613
[775] https://jira.mongodb.org/browse/SERVER-17713
[776] https://jira.mongodb.org/browse/SERVER-17642
[777] https://jira.mongodb.org/browse/SERVER-17587
[778] https://jira.mongodb.org/browse/SERVER-17562
[779] https://jira.mongodb.org/browse/SERVER-17551
[780] https://jira.mongodb.org/browse/SERVER-17532
[781] https://jira.mongodb.org/browse/SERVER-17471
[782] https://jira.mongodb.org/browse/SERVER-17382
[783] https://jira.mongodb.org/browse/SERVER-16804
[784] https://jira.mongodb.org/browse/SERVER-17616
[785] https://jira.mongodb.org/browse/SERVER-17313
[786] https://jira.mongodb.org/browse/SERVER-17706

HTTP Console

- SERVER-17729[787] Cannot start `mongod httpinterface`: sockets higher than 1023 not supported
- SERVER-17705[788] Fix credentials field inconsistency in HTTP interface
- SERVER-17669[789] Remove auth prompt in webserver when auth is not enabled

Admin

- SERVER-17570[790] MongoDB 3.0 NT Service shutdown race condition with `db.serverShutdown()`
- SERVER-17699[791] "locks" section empty in diagnostic log and profiler output for some operations
- SERVER-17337[792] RPM Init script breaks with quotes in `yaml` config file
- SERVER-16731[793] Remove unused DBPATH init script variable

Networking SERVER-17652[794] Cannot start mongod due to "sockets higher than 1023 not being supported"

Testing

- SERVER-17826[795] Ignore ismaster exceptions in `'initial_sync_unsupported_auth_schema.js` test
- SERVER-17808[796] Ensure availability in `initial_sync_unsupported_auth_schema.js` test
- SERVER-17433[797] ASAN leak in small oplog suite `write_result.js`

3.0.1 Changelog

Security

- SERVER-17507[798] MongoDB3 enterprise AuditLog
- SERVER-17379[799] Change "or" to "and" in webserver localhost exception check
- SERVER-16944[800] dbAdminAnyDatabase should have full parity with dbAdmin for a given database
- SERVER-16849[801] On mongos we always invalidate the user cache once, even if no user definitions are changing
- SERVER-16452[802] Failed login attempts should log source IP address

[787] https://jira.mongodb.org/browse/SERVER-17729
[788] https://jira.mongodb.org/browse/SERVER-17705
[789] https://jira.mongodb.org/browse/SERVER-17669
[790] https://jira.mongodb.org/browse/SERVER-17570
[791] https://jira.mongodb.org/browse/SERVER-17699
[792] https://jira.mongodb.org/browse/SERVER-17337
[793] https://jira.mongodb.org/browse/SERVER-16731
[794] https://jira.mongodb.org/browse/SERVER-17652
[795] https://jira.mongodb.org/browse/SERVER-17826
[796] https://jira.mongodb.org/browse/SERVER-17808
[797] https://jira.mongodb.org/browse/SERVER-17433
[798] https://jira.mongodb.org/browse/SERVER-17507
[799] https://jira.mongodb.org/browse/SERVER-17379
[800] https://jira.mongodb.org/browse/SERVER-16944
[801] https://jira.mongodb.org/browse/SERVER-16849
[802] https://jira.mongodb.org/browse/SERVER-16452

Querying

- SERVER-17395[803] Add FSM tests to stress yielding
- SERVER-17387[804] invalid projection for findAndModify triggers fassert() failure
- SERVER-14723[805] Crash during query planning for geoNear with multiple 2dsphere indices
- SERVER-17486[806] Crash when parsing invalid polygon coordinates

Replication

- SERVER-17515[807] copyDatabase fails to replicate indexes to secondary
- SERVER-17499[808] Using eval command to run getMore on aggregation cursor trips fatal assertion
- SERVER-17487[809] cloner dropDups removes _id entries belonging to other records
- SERVER-17302[810] consider blacklist in shouldChangeSyncSource

Sharding

- SERVER-17398[811] Deadlock in MigrateStatus::startCommit
- SERVER-17300[812] Balancer tries to create config.tags index multiple times
- SERVER-16849[813] On mongos we always invalidate the user cache once, even if no user definitions are changing
- SERVER-5004[814] balancer should check for stopped between chunk moves in current round

Indexing

- SERVER-17521[815] improve createIndex validation of empty name
- SERVER-17436[816] MultiIndexBlock may access deleted collection after recovering from yield

Aggregation Framework SERVER-17224[817] Aggregation pipeline with 64MB document can terminate server

Write Ops

- SERVER-17489[818] in bulk ops, only mark last operation with commit=synchronous
- SERVER-17276[819] WriteConflictException retry loops needed for collection creation on upsert

[803] https://jira.mongodb.org/browse/SERVER-17395
[804] https://jira.mongodb.org/browse/SERVER-17387
[805] https://jira.mongodb.org/browse/SERVER-14723
[806] https://jira.mongodb.org/browse/SERVER-17486
[807] https://jira.mongodb.org/browse/SERVER-17515
[808] https://jira.mongodb.org/browse/SERVER-17499
[809] https://jira.mongodb.org/browse/SERVER-17487
[810] https://jira.mongodb.org/browse/SERVER-17302
[811] https://jira.mongodb.org/browse/SERVER-17398
[812] https://jira.mongodb.org/browse/SERVER-17300
[813] https://jira.mongodb.org/browse/SERVER-16849
[814] https://jira.mongodb.org/browse/SERVER-5004
[815] https://jira.mongodb.org/browse/SERVER-17521
[816] https://jira.mongodb.org/browse/SERVER-17436
[817] https://jira.mongodb.org/browse/SERVER-17224
[818] https://jira.mongodb.org/browse/SERVER-17489
[819] https://jira.mongodb.org/browse/SERVER-17276

Concurrency

- SERVER-17501[820] Increase journalling capacity limits
- SERVER-17416[821] Deadlock between MMAP V1 journal lock and oplog collection lock
- SERVER-17395[822] Add FSM tests to stress yielding

Storage

- SERVER-17515[823] copyDatabase fails to replicate indexes to secondary
- SERVER-17436[824] MultiIndexBlock may access deleted collection after recovering from yield
- SERVER-17416[825] Deadlock between MMAP V1 journal lock and oplog collection lock
- SERVER-17381[826] Rename rocksExperiment to RocksDB
- SERVER-17369[827] [Rocks] Fix the calculation of nextPrefix
- SERVER-17345[828] WiredTiger -> session.truncate: the start cursor position is after the stop cursor position
- SERVER-17331[829] RocksDB configuring and monitoring
- SERVER-17323[830] MMAPV1Journal lock counts are changing during WT run
- SERVER-17319[831] invariant at shutdown rc9, rc10, rc11 with wiredTiger
- SERVER-17293[832] Server crash setting wiredTigerEngineRuntimeConfig:"eviction=(threads_max=8)"

WiredTiger

- SERVER-17510[833] "Didn't find RecordId in WiredTigerRecordStore" on collections after an idle period
- SERVER-17506[834] Race between inserts and checkpoints can lose records under WiredTiger
- SERVER-17487[835] cloner dropDups removes _id entries belonging to other records
- SERVER-17481[836] WiredTigerRecordStore::validate should call WT_SESSION::verify
- SERVER-17451[837] WiredTiger unable to start if crash leaves 0-length journal file
- SERVER-17378[838] WiredTiger's compact code can return 'Operation timed out' error (invariant failure)
- SERVER-17345[839] WiredTiger -> session.truncate: the start cursor position is after the stop cursor position

[820] https://jira.mongodb.org/browse/SERVER-17501
[821] https://jira.mongodb.org/browse/SERVER-17416
[822] https://jira.mongodb.org/browse/SERVER-17395
[823] https://jira.mongodb.org/browse/SERVER-17515
[824] https://jira.mongodb.org/browse/SERVER-17436
[825] https://jira.mongodb.org/browse/SERVER-17416
[826] https://jira.mongodb.org/browse/SERVER-17381
[827] https://jira.mongodb.org/browse/SERVER-17369
[828] https://jira.mongodb.org/browse/SERVER-17345
[829] https://jira.mongodb.org/browse/SERVER-17331
[830] https://jira.mongodb.org/browse/SERVER-17323
[831] https://jira.mongodb.org/browse/SERVER-17319
[832] https://jira.mongodb.org/browse/SERVER-17293
[833] https://jira.mongodb.org/browse/SERVER-17510
[834] https://jira.mongodb.org/browse/SERVER-17506
[835] https://jira.mongodb.org/browse/SERVER-17487
[836] https://jira.mongodb.org/browse/SERVER-17481
[837] https://jira.mongodb.org/browse/SERVER-17451
[838] https://jira.mongodb.org/browse/SERVER-17378
[839] https://jira.mongodb.org/browse/SERVER-17345

- SERVER-17319[840] invariant at shutdown rc9, rc10, rc11 with wiredTiger

MMAPv1

- SERVER-17501[841] Increase journalling capacity limits
- SERVER-17416[842] Deadlock between MMAP V1 journal lock and oplog collection lock
- SERVER-17388[843] Invariant failure in MMAPv1 when disk full

RocksDB

- SERVER-17381[844] Rename rocksExperiment to RocksDB
- SERVER-17369[845] [Rocks] Fix the calculation of nextPrefix
- SERVER-17331[846] RocksDB configuring and monitoring

Shell and Administration

- SERVER-17226[847] 'top' command with 64MB result document can terminate server
- SERVER-17405[848] getLog command masserts when given number
- SERVER-17347[849] .explain() should be included in the shell's DBCollection help

Build and Packaging

- SERVER-17484[850] Migrate server MCI config into server repo
- SERVER-17463[851] Python error when specifying absolute path to scons cacheDir
- SERVER-17460[852] LIBDEPS_v8_SYSLIBDEP typo
- SERVER-14166[853] Semantics of the –osx-version-min flag should be improved
- SERVER-17517[854] mongodb-org rpm packages no longer "provide" mongo-10gen-server

Logging SERVER-16452[855] Failed login attempts should log source IP address

[840] https://jira.mongodb.org/browse/SERVER-17319
[841] https://jira.mongodb.org/browse/SERVER-17501
[842] https://jira.mongodb.org/browse/SERVER-17416
[843] https://jira.mongodb.org/browse/SERVER-17388
[844] https://jira.mongodb.org/browse/SERVER-17381
[845] https://jira.mongodb.org/browse/SERVER-17369
[846] https://jira.mongodb.org/browse/SERVER-17331
[847] https://jira.mongodb.org/browse/SERVER-17226
[848] https://jira.mongodb.org/browse/SERVER-17405
[849] https://jira.mongodb.org/browse/SERVER-17347
[850] https://jira.mongodb.org/browse/SERVER-17484
[851] https://jira.mongodb.org/browse/SERVER-17463
[852] https://jira.mongodb.org/browse/SERVER-17460
[853] https://jira.mongodb.org/browse/SERVER-14166
[854] https://jira.mongodb.org/browse/SERVER-17517
[855] https://jira.mongodb.org/browse/SERVER-16452

Platform

- SERVER-17252[856] Upgrade PCRE Version from 8.30 to Latest
- SERVER-14166[857] Semantics of the –osx-version-min flag should be improved

Internal Code SERVER-17338[858] NULL pointer crash when running copydb against stepped-down 2.6 primary

Testing

- SERVER-17443[859] get_replication_info_helper.js should assert.soon rather than assert for log messages
- SERVER-17442[860] increase tolerance for shutdown timeout in stepdown.js to fix windows build break
- SERVER-17395[861] Add FSM tests to stress yielding

3.0.11 – Mar 31, 2016

- Fixed issue in MongoDB 3.0.9 and MongoDB 3.0.10 sharded clusters where during chunk migration, insert and update operations to documents in the migrating chunk are not reflected in the destination shard: SERVER-23425[862]

3.0.10 – Mar 8, 2016

- Fixed issues where read preference of `secondaryPreferred` (page 729) can end up using unversioned connections: SERVER-18671[863]
- Fixed issue with MMAPv1 journaling where the "last sequence number" file (`lsn` file) may be ahead of what is synced to the data files: SERVER-22261[864].
- Fixed issue where a data size change for oplog deletes can overflow 32-bit int: SERVER-22634[865]
- Fixed issue with high fragmentation on WiredTiger databases under write workloads: SERVER-22898[866].
- All issues closed in 3.0.10[867]

3.0.9 – Jan 26, 2016

- Fixed issue where queries which specify sort and batch size can return results out of order if documents are concurrently updated. SERVER-19996[868]
- Fixed performance issue where large amounts of create and drop collections can cause `listDatabases` to be slow under WiredTiger. SERVER-20961[869]

[856] https://jira.mongodb.org/browse/SERVER-17252
[857] https://jira.mongodb.org/browse/SERVER-14166
[858] https://jira.mongodb.org/browse/SERVER-17338
[859] https://jira.mongodb.org/browse/SERVER-17443
[860] https://jira.mongodb.org/browse/SERVER-17442
[861] https://jira.mongodb.org/browse/SERVER-17395
[862] https://jira.mongodb.org/browse/SERVER-23425
[863] https://jira.mongodb.org/browse/SERVER-18671
[864] https://jira.mongodb.org/browse/SERVER-22261
[865] https://jira.mongodb.org/browse/SERVER-22634
[866] https://jira.mongodb.org/browse/SERVER-22898
[867] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.0.10%20AND%20resolution%20%3D%
[868] https://jira.mongodb.org/browse/SERVER-19996
[869] https://jira.mongodb.org/browse/SERVER-20961

- Modified the authentication failure message to include the client IP address. SERVER-22054[870]
- All issues closed in 3.0.9[871]

3.0.8 – Dec 15, 2015

- Fixed issue where `findAndModify` on `mongos` can upsert to the wrong shard. SERVER-20407[872].
- Fixed WiredTiger commit visibility issue which caused document not found. SERVER-21275[873].
- Fixed issue where the oplog can grow to 3x configured size. SERVER-21553[874]
- All issues closed in 3.0.8[875]

3.0.7 – Oct 13, 2015

- Improvements to WiredTiger memory handling and performance: SERVER-20159[876], SERVER-20204[877], SERVER-20091[878], and SERVER-20176[879].
- Fixed issue whereby reconfig during a pending step down may prevent a primary from stepping down: SERVER-20262[880].
- Additional privileges for built-in roles: SERVER-19131[881], SERVER-15893[882], and SERVER-13647[883].
- All issues closed in 3.0.7[884]

3.0.6 – August 24, 2015

- Improvements to WiredTiger Stability SERVER-19751[885], SERVER-19673[886], and SERVER-19573[887].
- Fixed issue with the interaction between SSL and Auditing. SERVER-19538[888].
- Fixed issue with aggregation `$sort` on sharded systems SERVER-19464[889].
- All issues closed in 3.0.6[890]

[870] https://jira.mongodb.org/browse/SERVER-22054
[871] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.0.9%20AND%20resolution%20%3D%2
[872] https://jira.mongodb.org/browse/SERVER-20407
[873] https://jira.mongodb.org/browse/SERVER-21275
[874] https://jira.mongodb.org/browse/SERVER-21553
[875] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.0.8%20AND%20resolution%20%3D%2
[876] https://jira.mongodb.org/browse/SERVER-20159
[877] https://jira.mongodb.org/browse/SERVER-20204
[878] https://jira.mongodb.org/browse/SERVER-20091
[879] https://jira.mongodb.org/browse/SERVER-20176
[880] https://jira.mongodb.org/browse/SERVER-20262
[881] https://jira.mongodb.org/browse/SERVER-19131
[882] https://jira.mongodb.org/browse/SERVER-15893
[883] https://jira.mongodb.org/browse/SERVER-13647
[884] https://jira.mongodb.org/issues/?jql=project%20in%20(SERVER%2C%20TOOLS)%20AND%20fixVersion%20%3D%203.0.7%20AND%20resolution%20%3D%2
[885] https://jira.mongodb.org/browse/SERVER-19751
[886] https://jira.mongodb.org/browse/SERVER-19673
[887] https://jira.mongodb.org/browse/SERVER-19573
[888] https://jira.mongodb.org/browse/SERVER-19538
[889] https://jira.mongodb.org/browse/SERVER-19464
[890] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%223.0.6%22%20AND%20project%20%3D%20SERVER%20AND%20resolution%20%3D%20Fixed

3.0.5 – July 28, 2015

- Improvements to WiredTiger for capped collections and replication (SERVER-19178[891], SERVER-18875[892] and SERVER-19513[893]).
- Additional WiredTiger improvements for performance (SERVER-19189[894]) and improvements related to cache and session use (SERVER-18829[895] SERVER-17836[896]).
- Performance improvements for longer running queries, particularly `$text` and `$near` queries SERVER-18926[897].
- All issues closed in 3.0.5[898]

3.0.4 – June 16, 2015

- Fix missed writes with concurrent inserts during chunk migration from shards with WiredTiger primaries: SERVER-18822[899]
- Resolve write conflicts with multi-update updates with `upsert=true` with the Wired Tiger Storage engine: SERVER-18213[900]
- Fix case where secondary reads could block replication: SERVER-18190[901]
- Improve performance on Windows with WiredTiger and documents larger than 16kb: SERVER-18079[902]
- Fix issue where WiredTiger data files are not correctly recovered following unexpected system restarts: SERVER-18316[903]
- All issues closed in 3.0.4[904]

3.0.3 – May 12, 2015

- Deprecate `db.eval()` and add warnings: SERVER-17453[905]
- Fix potential for abrupt termination with the Windows service stop operation: SERVER-17802[906]
- Fix crash caused by update with a *key too large to index* on WiredTiger and RocksDB storage engines: SERVER-17882[907]
- Remove inconsistent support for `mapReduce` in `eval` environment: SERVER-17889[908]
- All issues closed in 3.0.3[909]

[891] https://jira.mongodb.org/browse/SERVER-19178
[892] https://jira.mongodb.org/browse/SERVER-18875
[893] https://jira.mongodb.org/browse/SERVER-19513
[894] https://jira.mongodb.org/browse/SERVER-19189
[895] https://jira.mongodb.org/browse/SERVER-18829
[896] https://jira.mongodb.org/browse/SERVER-17836
[897] https://jira.mongodb.org/browse/SERVER-18926
[898] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%223.0.5%22%20AND%20project%20%3D%20SERVER%20AND%20resolution%20%3D%20Fixed
[899] https://jira.mongodb.org/browse/SERVER-18822
[900] https://jira.mongodb.org/browse/SERVER-18213
[901] https://jira.mongodb.org/browse/SERVER-18190
[902] https://jira.mongodb.org/browse/SERVER-18079
[903] https://jira.mongodb.org/browse/SERVER-18316
[904] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%223.0.4%22%20AND%20project%20%3D%20SERVER%20AND%20resolution%20%3D%20Fixed
[905] https://jira.mongodb.org/browse/SERVER-17453
[906] https://jira.mongodb.org/browse/SERVER-17802
[907] https://jira.mongodb.org/browse/SERVER-17882
[908] https://jira.mongodb.org/browse/SERVER-17889
[909] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%223.0.3%22%20AND%20project%20%3D%20SERVER%20AND%20resolution%20%3D%20Fixed

3.0.2 – April 9, 2015

- Fix inefficient query plans for `2d $nearSphere`: SERVER-17469[910]
- Fix problem starting `mongod` during repair operations with WiredTiger: SERVER-17652[911] and SERVER-17729[912]
- Resolved invalid compression stream error with WiredTiger and `zlib` block compression: SERVER-17713[913]
- Fix memory use issue for inserts into large indexed arrays: SERVER-17616[914]
- All issues closed in 3.0.2[915]

3.0.1 – March 17, 2015

- Fixed race condition in WiredTiger between inserts and checkpoints that could result in lost records: SERVER-17506[916].
- Resolved issue in WiredTiger's capped collections implementation that caused a server crash: SERVER-17345[917].
- Fixed issue is initial sync with duplicate `_id` entries: SERVER-17487[918].
- Fixed deadlock condition in MMAPv1 between the journal lock and the oplog collection lock: SERVER-17416[919].
- All issues closed in 3.0.1[920]

Major Changes

Pluggable Storage Engine API

MongoDB 3.0 introduces a pluggable storage engine API that allows third parties to develop storage engines for MongoDB.

WiredTiger

MongoDB 3.0 introduces support for the WiredTiger[921] storage engine. With the support for WiredTiger, MongoDB now supports two storage engines:

- MMAPv1, the storage engine available in previous versions of MongoDB and the default storage engine for MongoDB 3.0, and
- WiredTiger[922], available only in the 64-bit versions of MongoDB 3.0.

[910] https://jira.mongodb.org/browse/SERVER-17469
[911] https://jira.mongodb.org/browse/SERVER-17652
[912] https://jira.mongodb.org/browse/SERVER-17729
[913] https://jira.mongodb.org/browse/SERVER-17713
[914] https://jira.mongodb.org/browse/SERVER-17616
[915] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%223.0.2%22%20AND%20project%20%3D%20SERVER%20AND%20resolution%20%3D%20Fixed
[916] https://jira.mongodb.org/browse/SERVER-17506
[917] https://jira.mongodb.org/browse/SERVER-17345
[918] https://jira.mongodb.org/browse/SERVER-17487
[919] https://jira.mongodb.org/browse/SERVER-17416
[920] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%223.0.1%22%20AND%20project%20%3D%20SERVER%20AND%20resolution%20%3D%20Fixed
[921] http://wiredtiger.com
[922] http://wiredtiger.com

WiredTiger Usage WiredTiger is an alternate to the default MMAPv1 storage engine. WiredTiger supports all MongoDB features, including operations that report on server, database, and collection statistics. Switching to WiredTiger, however, requires a change to the *on-disk storage format* (page 946). For instructions on changing the storage engine to WiredTiger, see the appropriate sections in the *Upgrade MongoDB to 3.0* (page 953) documentation.

MongoDB 3.0 replica sets and sharded clusters can have members with different storage engines; however, performance can vary according to workload. For details, see the appropriate sections in the *Upgrade MongoDB to 3.0* (page 953) documentation.

The WiredTiger storage engine requires the latest official MongoDB drivers. For more information, see *WiredTiger and Driver Version Compatibility* (page 946).

See also:

Support for touch Command (page 947), *WiredTiger Storage Engine* (page 595) documentation

WiredTiger Configuration To configure the behavior and properties of the WiredTiger storage engine, see `storage.wiredTiger` configuration options. You can set *WiredTiger options on the command line*.

See also:

WiredTiger Storage Engine (page 595)

WiredTiger Concurrency and Compression The 3.0 WiredTiger storage engine provides document-level locking and compression.

By default, WiredTiger compresses collection data using the *snappy* compression library. WiredTiger uses *prefix compression* on all indexes by default.

See also:

WiredTiger (page 298) section in the *Production Notes* (page 296), the blog post New Compression Options in MongoDB 3.0[923]

MMAPv1 Improvements

MMAPv1 Concurrency Improvement In version 3.0, the MMAPv1 storage engine adds support for collection-level locking.

MMAPv1 Configuration Changes To support multiple storage engines, some configuration settings for MMAPv1 have changed. See *Configuration File Options Changes* (page 946).

MMAPv1 Record Allocation Behavior Changes MongoDB 3.0 no longer implements dynamic record allocation and deprecates *paddingFactor*. The default allocation strategy for collections in instances that use MMAPv1 is *power of 2 allocation* (page 604), which has been improved to better handle large document sizes. In 3.0, the `usePowerOf2Sizes` flag is ignored, so the power of 2 strategy is used for all collections that do not have `noPadding` flag set.

For collections with workloads that consist only of inserts or in-place updates (such as incrementing counters), you can disable the power of 2 strategy. To disable the power of 2 strategy for a collection, use the `collMod` command with the `noPadding` flag or the `db.createCollection()` method with the `noPadding` option.

[923] https://www.mongodb.com/blog/post/new-compression-options-mongodb-30?jmp=docs

Warning: Do not set `noPadding` if the workload includes removes or any updates that may cause documents to grow. For more information, see *No Padding Allocation Strategy* (page 605).

When low on disk space, MongoDB 3.0 no longer errors on all writes but only when the required disk allocation fails. As such, MongoDB now allows in-place updates and removes when low on disk space.

See also:

Dynamic Record Allocation (page 947)

Replica Sets

Increased Number of Replica Set Members

In MongoDB 3.0, replica sets can have up to 50 members. [924] The following drivers support the larger replica sets:

- C# (.NET) Driver 1.10
- Java Driver 2.13
- Python Driver (PyMongo) 3.0
- Ruby Driver 2.0
- Node.JS Driver 2.0

The C, C++, Perl, and PHP drivers, as well as the earlier versions of the Ruby, Python, and Node.JS drivers, discover and monitor replica set members serially, and thus are not suitable for use with large replica sets.

Replica Set Step Down Behavior Changes

The process that a *primary* member of a *replica set* uses to step down has the following changes:

- Before stepping down, `replSetStepDown` will attempt to terminate long running user operations that would block the primary from stepping down, such as an index build, a write operation or a map-reduce job.
- To help prevent rollbacks, the `replSetStepDown` will wait for an electable secondary to catch up to the state of the primary before stepping down. Previously, a primary would wait for a secondary to catch up to within 10 seconds of the primary (i.e. a secondary with a replication lag of 10 seconds or less) before stepping down.
- `replSetStepDown` now allows users to specify a `secondaryCatchUpPeriodSecs` parameter to specify how long the primary should wait for a secondary to catch up before stepping down.

Other Replica Set Operational Changes

- Initial sync builds indexes more efficiently for each collection and applies oplog entries in batches using threads.
- Definition of *w: "majority"* (page 180) write concern changed to mean majority of *voting* nodes.
- Stronger restrictions on *Replica Set Configuration* (page 717). For details, see *Replica Set Configuration Validation* (page 947).
- For pre-existing collections on secondary members, MongoDB 3.0 no longer automatically builds missing `_id` indexes.

[924] The maximum number of *voting* members remains at 7.

See also:

Replication Changes (page 947) in *Compatibility Changes in MongoDB 3.0* (page 945)

Sharded Clusters

MongoDB 3.0 provides the following enhancements to sharded clusters:

- Adds a new `sh.removeTagRange()` helper to improve management of sharded collections with tags. The new `sh.removeTagRange()` method acts as a complement to `sh.addTagRange()`.
- Provides a more predictable read preference behavior. `mongos` instances no longer pin connections to members of replica sets when performing read operations. Instead, `mongos` reevaluates *read preferences* (page 651) for every operation to provide a more predictable read preference behavior when read preferences change.
- Provides a new `writeConcern` setting to configure the *write concern* (page 179) of chunk migration operations. You can configure the `writeConcern` setting for the *balancer* (page 799) as well as for `moveChunk` and `cleanupOrphaned` commands.
- Improves visibility of balancer operations. `sh.status()` includes information about the state of the balancer. See `sh.status()` for details.

See also:

Sharded Cluster Setting (page 948) in *Compatibility Changes in MongoDB 3.0* (page 945)

Security Improvements

MongoDB 3.0 includes the following security enhancements:

- MongoDB 3.0 adds a new *SCRAM-SHA-1* (page 399) challenge-response user authentication mechanism. `SCRAM-SHA-1` requires a driver upgrade if your current driver version does not support `SCRAM-SHA-1`. For the driver versions that support `SCRAM-SHA-1`, see *Upgrade Drivers* (page 958).
- Increases restrictions when using the *Localhost Exception* (page 396) to access MongoDB. For details, see *Localhost Exception Changed* (page 949).

See also:

Security Changes (page 948)

Improvements

New Query Introspection System

MongoDB 3.0 includes a new query introspection system that provides an improved output format and a finer-grained introspection into both query plan and query execution.

For details, see the new `db.collection.explain()` method and the new `explain` command as well as the updated `cursor.explain()` method.

For information on the format of the new output, see `https://docs.mongodb.org/manual/reference/explain-result`

Enhanced Logging

To improve usability of the log messages for diagnosis, MongoDB categorizes some log messages under specific components, or operations, and provides the ability to set the verbosity level for these components. For information, see `https://docs.mongodb.org/manual/reference/log-messages`.

MongoDB Tools Enhancements

All MongoDB tools except for `mongosniff` and `mongoperf` are now written in Go and maintained as a separate project.

- New options for parallelized `mongodump` and `mongorestore`. You can control the number of collections that `mongorestore` will restore at a time with the `--numParallelCollections` option.
- New options `-excludeCollection` and `--excludeCollectionsWithPrefix` for `mongodump` to exclude collections.
- `mongorestore` can now accept BSON data input from standard input in addition to reading BSON data from file.
- `mongostat` and `mongotop` can now return output in JSON format with the `--json` option.
- Added configurable *write concern* to `mongoimport`, `mongorestore`, and `mongofiles`. Use the `--writeConcern` option. The default writeConcern has been changed to 'w:majority'.
- `mongofiles` now allows you to configure the GridFS prefix with the `--prefix` option so that you can use custom namespaces and store multiple GridFS namespaces in a single database.

See also:

MongoDB Tools Changes (page 948)

Indexes

- Background index builds will no longer automatically interrupt if `dropDatabase`, `drop`, `dropIndexes` operations occur for the database or collection affected by the index builds. The `dropDatabase`, `drop`, and `dropIndexes` commands will still fail with the error message `a background operation is currently running`, as in 2.6.
- If you specify multiple indexes to the `createIndexes` command,
  - the command only scans the collection once, and
  - if at least one index is to be built in the foreground, the operation will build all the specified indexes in the foreground.
- For sharded collections, indexes can now *cover queries* (page 106) that execute against the `mongos` if the index includes the shard key.

See also:

Indexes (page 950) in *Compatibility Changes in MongoDB 3.0* (page 945)

Query Enhancements

MongoDB 3.0 includes the following query enhancements:

- For geospatial queries, adds support for "big" polygons for `$geoIntersects` and `$geoWithin` queries. "Big" polygons are single-ringed GeoJSON polygons with areas greater than that of a single hemisphere. See `$geometry`, `$geoIntersects`, and `$geoWithin` for details.
- For `aggregate()`, adds a new `$dateToString` operator to facilitate converting a date to a formatted string.
- Adds the `$eq` query operator to query for equality conditions.

See also:

2d Indexes and Geospatial Near Queries (page 950)

Distributions and Supported Versions

Most non-Enterprise MongoDB distributions now include support for TLS/SSL. Previously, only MongoDB Enterprise distributions came with TLS/SSL support included; for non-Enterprise distributions, you had to build MongoDB locally with the `--ssl` flag (i.e. `scons --ssl`).

32-bit MongoDB builds are available for testing, but are not for production use. 32-bit MongoDB builds do not include the WiredTiger storage engine.

MongoDB builds for Solaris do not support the WiredTiger storage engine.

MongoDB builds are available for Windows Server 2003 and Windows Vista (as "64-bit Legacy"), but the minimum officially supported Windows version is Windows Server 2008.

See also:

Platform Support (page 952), *Deprecation of 32-bit Versions* (page 21)

Package Repositories

Non-Enterprise MongoDB Linux packages for 3.0 and later are in a new repository. Follow the appropriate *Linux installation instructions* (page 22) to install the 3.0 packages from the new location.

MongoDB Enterprise Features

Auditing

Auditing (page 466) in MongoDB Enterprise can filter on *any field in the audit message* (page 506), including the fields returned in the *param* (page 506) document. This enhancement, along with the `auditAuthorizationSuccess` parameter, enables auditing to filter on CRUD operations. However, enabling `auditAuthorizationSuccess` to audit of all authorization successes degrades performance more than auditing only the authorization failures.

Additional Information

Changes Affecting Compatibility

Compatibility Changes in MongoDB 3.0

On this page

- Storage Engine (page 946)
- Replication Changes (page 947)
- MongoDB Tools Changes (page 948)
- Sharded Cluster Setting (page 948)
- Security Changes (page 948)
- Indexes (page 950)
- Driver Compatibility Changes (page 950)
- General Compatibility Changes (page 951)

The following 3.0 changes can affect the compatibility with older versions of MongoDB. See *Release Notes for MongoDB 3.0* (page 911) for the full list of the 3.0 changes.

Storage Engine

Configuration File Options Changes With the introduction of additional storage engines in 3.0, some `configuration file options` have changed:

Previous Setting	New Setting
`storage.journal.commitIntervalMs`	`storage.mmapv1.journal.commitIntervalMs`
`storage.journal.debugFlags`	`storage.mmapv1.journal.debugFlags`
`storage.nsSize`	`storage.mmapv1.nsSize`
`storage.preallocDataFiles`	`storage.mmapv1.preallocDataFiles`
`storage.quota.enforced`	`storage.mmapv1.quota.enforced`
`storage.quota.maxFilesPerDB`	`storage.mmapv1.quota.maxFilesPerDB`
`storage.smallFiles`	`storage.mmapv1.smallFiles`

3.0 `mongod` instances are backward compatible with existing configuration files, but will issue warnings when if you attempt to use the old settings.

Data Files Must Correspond to Configured Storage Engine The files in the `dbPath` directory must correspond to the configured storage engine (i.e. `--storageEngine`). `mongod` will not start if `dbPath` contains data files created by a storage engine other than the one specified by `--storageEngine`.

See also:

Change Storage Engine to WiredTiger sections in *Upgrade MongoDB to 3.0* (page 953)

WiredTiger and Driver Version Compatibility For MongoDB 3.0 deployments that use the WiredTiger storage engine, the following operations return no output when issued in previous versions of the `mongo` shell or drivers:

- `db.getCollectionNames()`
- `db.collection.getIndexes()`
- `show collections`
- `show tables`

Use the 3.0 `mongo` shell or the *3.0 compatible version* (page 950) of the official drivers when connecting to 3.0 `mongod` instances that use WiredTiger. The 2.6.8 `mongo` shell is also compatible with 3.0 `mongod` instances that use WiredTiger.

`db.fsyncLock()` is not Compatible with WiredTiger With WiredTiger the `db.fsyncLock()` and `db.fsyncUnlock()` operations *cannot* guarantee that the data files do not change. As a result, do not use these methods to ensure consistency for the purposes of creating backups.

Support for `touch` Command If a storage engine does not support the `touch`, then the `touch` command will return an error.

- The MMAPv1 storage engine supports `touch`.
- The WiredTiger storage engine *does not* support `touch`.

Dynamic Record Allocation MongoDB 3.0 no longer supports dynamic record allocation and deprecates *paddingFactor*.

MongoDB 3.0 deprecates the `newCollectionsUsePowerOf2Sizes` parameter such that you can no longer use the parameter to disable the power of 2 sizes allocation for a collection. Instead, use the `collMod` command with the `noPadding` flag or the `db.createCollection()` method with the `noPadding` option. Only set `noPadding` for collections with workloads that consist only of inserts or in-place updates (such as incrementing counters).

Warning: Only set `noPadding` to `true` for collections whose workloads have *no* update operations that cause documents to grow, such as for collections with workloads that are insert-only. For more information, see *No Padding Allocation Strategy* (page 605).

For more information, see *MMAPv1 Record Allocation Behavior Changes* (page 941).

Replication Changes

Replica Set Oplog Format Change MongoDB 3.0 is not compatible with oplog entries generated by versions of MongoDB before 2.2.1. If you upgrade from one of these versions, you must wait for new oplog entries to overwrite *all* old oplog entries generated by one of these versions before upgrading to 3.0.0 or earlier.

Secondaries may abort if they replay a pre-2.6 oplog with an index build operation that would fail on a 2.6 or later primary.

Replica Set Configuration Validation MongoDB 3.0 provides a stricter validation of *replica set configuration settings* (page 717) and replica sets invalid replica set configurations.

Stricter validations include:

- Arbiters can only have `1` vote. Previously, arbiters could also have a value of `0` for `members[n].votes` (page 721). If an arbiter has any value other than `1` for `members[n].votes` (page 721), you must fix the setting.
- Non-arbiter members can **only** have value of `0` or `1` for `members[n].votes` (page 721). If a non-arbiter member has any other value for `members[n].votes` (page 721), you must fix the setting.
- `_id` (page 718) in the *Replica Set Configuration* (page 717) must specify the same name as that specified by `--replSet` or `replication.replSetName`. Otherwise, you must fix the setting.
- Disallows `0` for `settings.getLastErrorDefaults` (page 722) value. If `settings.getLastErrorDefaults` (page 722) value is `0`, you must fix the setting.
- `settings` (page 721) can only contain the recognized settings. Previously, MongoDB ignored unrecognized settings. If `settings` (page 721) contains unrecognized settings, you must remove the unrecognized settings.

To fix the settings before upgrading to MongoDB 3.0, connect to the primary and `reconfigure` your replica set to valid configuration settings.

If you have already upgraded to MongoDB 3.0, you must *downgrade to MongoDB 2.6* (page 961) first and then fix the settings. Once you have `reconfigured` the replica set, you can re-upgrade to MongoDB 3.0.

Change of `w: majority` Semantics A write concern with a *w: majority* (page 180) value is satisfied when a majority of the *voting* members replicates a write operation. In previous versions, *majority* referred a majority of all voting and non-voting members of the set.

Remove `local.slaves` Collection MongoDB 3.0 removes the `local.slaves` collection that tracked the secondaries' replication progress. To track the replication progress, use the `rs.status()` method.

Replica Set State Change The `FATAL` replica set state does not exist as of 3.0.0.

HTTP Interface The HTTP Interface (i.e. `net.http.enabled`) no longer reports replication data.

MongoDB Tools Changes

Require a Running MongoDB Instance The 3.0 versions of MongoDB tools, `mongodump`, `mongorestore`, `mongoexport`, `mongoimport`, `mongofiles`, and `mongooplog`, must connect to running MongoDB instances and these tools *cannot* directly modify the data files with `--dbpath` as in previous versions. Ensure that you start your `mongod` instance(s) before using these tools.

Removed Options

- Removed `--dbpath`, `--journal`, and `--filter` options for `mongodump`, `mongorestore`, `mongoimport`, `mongoexport`, and `bsondump`.
- Removed `--locks` option for `mongotop`.
- Removed `--noobjcheck` option for `bsondump` and `mongorestore`.
- Removed `--csv` option for `mongoexport`. Use the new `--type` option to specify the export format type (`csv` or `json`).

See also:

MongoDB Tools Enhancements (page 944)

Sharded Cluster Setting

Remove `releaseConnectionsAfterResponse` Parameter MongoDB now always releases connections after response. `releaseConnectionsAfterResponse` parameter is no longer available.

Security Changes

MongoDB 2.4 User Model Removed MongoDB 3.0 completely removes support for the deprecated 2.4 user model. MongoDB 3.0 will exit with an error message if there is user data with the 2.4 schema, i.e. if `authSchema` version is less than `3`.

To verify the version of your existing 2.6 schema, query the `system.version` collection in the `admin` database:

Note: You must have privileges to query the collection.

```
use admin
db.system.version.find( { _id: "authSchema" })
```

If you are currently using `auth` and you have schema version 2 or 3, the query returns the `currentVersion` of the existing `authSchema`.

If you do not currently have any users *or* you are using `authSchema` version 1, the query will not return any result.

If your `authSchema` version is less than `3` or the query does not return any results, see *Upgrade User Authorization Data to 2.6 Format* (page 1014) to upgrade the `authSchema` version before upgrading to MongoDB 3.0.

After upgrading MongoDB to 3.0 from 2.6, to use the new `SCRAM-SHA-1` challenge-response mechanism if you have existing user data, you will need to upgrade the authentication schema a second time. This upgrades the `MONGODB-CR` user model to `SCRAM-SHA-1` user model. See *Upgrade to SCRAM-SHA-1* (page 957) for details.

Localhost Exception Changed In 3.0, the localhost exception changed so that these connections *only* have access to create the first user on the `admin` database. In previous versions, connections that gained access using the localhost exception had unrestricted access to the MongoDB instance.

See *Localhost Exception* (page 396) for more information.

`db.addUser()` Removed 3.0 removes the legacy `db.addUser()` method. Use `db.createUser()` and `db.updateUser()` instead.

TLS/SSL Configuration Option Changes MongoDB 3.0 introduced new `net.ssl.allowConnectionsWithoutCertificates` configuration file setting and `--sslAllowConnectionsWithoutCertificates` command line option for `mongod` and `mongos`. These options replace previous `net.ssl.weakCertificateValidation` and `--sslWeakCertificateValidation` options, which became aliases. Update your configuration to ensure future compatibility.

TLS/SSL Certificates Validation By default, when running in SSL mode, MongoDB instances will *only* start if its certificate (i.e. `net.ssl.PemKeyFile`) is valid. You can disable this behavior with the `net.ssl.allowInvalidCertificates` setting or the `--sslAllowInvalidCertificates` command line option.

To start the `mongo` shell with `--ssl`, you must explicitly specify either the `--sslCAFile` or `--sslAllowInvalidCertificates` option at startup. See *TLS/SSL Configuration for Clients* (page 455) for more information.

TLS/SSL Certificate Hostname Validation By default, MongoDB validates the hostnames of hosts attempting to connect using certificates against the hostnames listed in those certificates. In certain deployment situations this behavior may be undesirable. It is now possible to disable such hostname validation without disabling validation of the rest of the certificate information with the `net.ssl.allowInvalidHostnames` setting or the `--sslAllowInvalidHostnames` command line option.

SSLv3 Ciphers Disabled In light of vulnerabilities in legacy SSL ciphers[925], these ciphers have been explicitly disabled in MongoDB. No configuration changes are necessary.

`mongo` Shell Version Compatibility Versions of the `mongo` shell before 3.0 are *not* compatible with 3.0 deployments of MongoDB that enforce access control. If you have a 3.0 MongoDB deployment that requires access control, you must use 3.0 versions of the `mongo` shell.

HTTP Status Interface and REST API Compatibility Neither the HTTP status interface nor the REST API support the *SCRAM-SHA-1* (page 399) challenge-response user authentication mechanism introduced in version 3.0.

Indexes

Remove `dropDups` Option `dropDups` option is no longer available for `createIndex()`, `ensureIndex()`, and `createIndexes`.

Changes to Restart Behavior during Background Indexing For 3.0 `mongod` instances, if a background index build is in progress when the `mongod` process terminates, when the instance restarts the index build will restart as foreground index build. If the index build encounters any errors, such as a duplicate key error, the `mongod` will exit with an error.

To start the `mongod` after a failed index build, use the `storage.indexBuildRetry` or *`--noIndexBuildRetry`* to skip the index build on start up.

`2d` Indexes and Geospatial Near Queries For `$near` queries that use a *2d* (page 557) index:

- MongoDB no longer uses a default limit of 100 documents.
- Specifying a `batchSize()` is no longer analogous to specifying a `limit()`.

For `$nearSphere` queries that use a *2d* (page 557) index, MongoDB no longer uses a default limit of 100 documents.

Driver Compatibility Changes Each officially supported driver has release a version that includes support for all new features introduced in MongoDB 3.0. Upgrading to one of these version is strongly recommended as part of the upgrade process.

A driver upgrade is **necessary** in certain scenarios due to changes in functionality:

- Use of the `SCRAM-SHA-1` authentication method
- Use of functionality that calls `listIndexes` or `listCollections`

The minimum 3.0-compatible driver versions are:

[925] https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-3566

Driver Language	Minimum 3.0-Compatible Version
C[926]	1.1.0[927]
C++[928]	1.0.0[929]
C#[930]	1.10[931]
Java[932]	2.13[933]
Node.js[934]	1.4.29[935]
Perl[936]	0.708.0.0[937]
PHP[938]	1.6[939]
Python[940]	2.8[941]
Motor[942]	0.4[943]
Ruby[944]	1.12[945]
Scala[946]	2.8.0[947]

General Compatibility Changes

`findAndModify` Return Document In MongoDB 3.0, when performing an update with `findAndModify` that also specifies `upsert: true` *and* either the `new` option is not set or `new: false`, `findAndModify` returns `null` in the `value` field if the `query` does not match any document, regardless of the `sort` specification.

In previous versions, `findAndModify` returns an empty document `{}` in the `value` field if a `sort` is specified for the update, and `upsert: true`, and the `new` option is not set or `new: false`.

`upsert:true` with a Dotted `_id` Query When you execute an `update()` with `upsert: true` and the query matches no existing document, MongoDB will refuse to insert a new document if the query specifies conditions on the `_id` field using *dot notation* (page 9).

This restriction ensures that the order of fields embedded in the `_id` document is well-defined and not bound to the order specified in the query

If you attempt to insert a document in this way, MongoDB will raise an error.

For example, consider the following update operation. Since the update operation specifies `upsert:true` and the query specifies conditions on the `_id` field using dot notation, then the update will result in an error when constructing the document to insert.

[926] https://docs.mongodb.org/ecosystem/drivers/c
[927] https://github.com/mongodb/mongo-c-driver/releases
[928] https://github.com/mongodb/mongo-cxx-driver
[929] https://github.com/mongodb/mongo-cxx-driver/releases
[930] https://docs.mongodb.org/ecosystem/drivers/csharp
[931] https://github.com/mongodb/mongo-csharp-driver/releases
[932] https://docs.mongodb.org/ecosystem/drivers/java
[933] https://github.com/mongodb/mongo-java-driver/releases
[934] https://docs.mongodb.org/ecosystem/drivers/node-js
[935] https://github.com/mongodb/node-mongodb-native/releases
[936] https://docs.mongodb.org/ecosystem/drivers/perl
[937] http://search.cpan.org/dist/MongoDB/
[938] https://docs.mongodb.org/ecosystem/drivers/php
[939] http://pecl.php.net/package/mongo
[940] https://docs.mongodb.org/ecosystem/drivers/python
[941] https://pypi.python.org/pypi/pymongo/
[942] https://docs.mongodb.org/ecosystem/drivers/python
[943] https://pypi.python.org/pypi/motor/
[944] https://docs.mongodb.org/ecosystem/drivers/ruby
[945] https://rubygems.org/gems/mongo
[946] https://docs.mongodb.org/ecosystem/drivers/scala
[947] https://github.com/mongodb/casbah/releases

```
db.collection.update( { "_id.name": "Robert Frost", "_id.uid": 0 },
   { "categories": ["poet", "playwright"] },
   { upsert: true } )
```

Deprecate Access to `system.indexes` and `system.namespaces` MongoDB 3.0 deprecates *direct* access to `system.indexes` and `system.namespaces` collections. Use the `createIndexes` and `listIndexes` commands instead. See also *WiredTiger and Driver Version Compatibility* (page 946).

Collection Name Validation MongoDB 3.0 more consistently enforces the `collection naming restrictions`. Ensure your application does not create or depend on invalid collection names.

Platform Support Commercial support is no longer provided for MongoDB on 32-bit platforms (Linux and Windows). Linux RPM and DEB packages are also no longer available. However, binary archives are still available.

Linux Package Repositories Non-Enterprise MongoDB Linux packages for 3.0 and later are in a new repository. Follow the appropriate *Linux installation instructions* (page 22) to install the 3.0 packages from the new location.

Removed/Deprecated Commands The following commands and methods are no longer available in MongoDB 3.0:

- `closeAllDatabases`
- `getoptime`
- `text`
- `indexStats`, `db.collection.getIndexStats()`, and `db.collection.indexStats()`

The following commands and methods are deprecated in MongoDB 3.0:

- `diagLogging`
- `eval`, `db.eval()`
- `db.collection.copyTo()`

In addition, you cannot use the now deprecated `eval` command or the `db.eval()` method to invoke `mapReduce` or `db.collection.mapReduce()`.

Date and Timestamp Comparison Order MongoDB 3.0 no longer treats the *Timestamp* (page 15) and the *Date* (page 15) data types as equivalent for comparison purposes. Instead, the *Timestamp* (page 15) data type has a higher comparison/sort order (i.e. is "greater") than the *Date* (page 15) data type. If your application relies on the equivalent comparison/sort order of Date and Timestamp objects, modify your application accordingly before upgrading.

Server Status Output Change The `serverStatus` command and the `db.serverStatus()` method no longer return `workingSet`, `indexCounters`, and `recordStats` sections in the output.

Unix Socket Permissions Change Unix domain socket file permission now defaults to `0700`. To change the permission, MongoDB provides the `net.unixDomainSocket.filePermissions` setting as well as the `--filePermission` option.

cloneCollection The `cloneCollection` command and the `db.cloneCollection()` method will now return an error if the collection already exists, instead of inserting into it.

Some changes in 3.0 can affect *compatibility* (page 945) and may require user actions. For a detailed list of compatibility changes, see *Compatibility Changes in MongoDB 3.0* (page 945).

Upgrade Process

	On this page
Upgrade MongoDB to 3.0	• Upgrade Recommendations and Checklists (page 953) • Upgrade MongoDB Processes (page 953) • Upgrade Existing `MONGODB-CR` Users to Use `SCRAM-SHA-1` (page 957) • General Upgrade Procedure (page 957)

In the general case, the upgrade from MongoDB 2.6 to 3.0 is a binary-compatible "drop-in" upgrade: shut down the `mongod` instances and replace them with `mongod` instances running 3.0. **However**, before you attempt any upgrade please familiarize yourself with the content of this document, particularly the procedure for *upgrading sharded clusters* (page 955).

If you need guidance on upgrading to 3.0, MongoDB offers consulting[948] to help ensure a smooth transition without interruption to your MongoDB application.

Upgrade Recommendations and Checklists

When upgrading, consider the following:

Upgrade Requirements To upgrade an existing MongoDB deployment to 3.0, you must be running 2.6. If you're running a version of MongoDB before 2.6, you *must* upgrade to 2.6 before upgrading to 3.0. See *Upgrade MongoDB to 2.6* (page 1010) for the procedure to upgrade from 2.4 to 2.6. Once upgraded to MongoDB 2.6, you **cannot** downgrade to any version earlier than MongoDB 2.4.

If your existing MongoDB deployment is already running with authentication and authorization, your user data model `authSchema` must be at least version 3. To verify the version of your existing `authSchema`, see *MongoDB 2.4 User Model Removed* (page 949). To upgrade your `authSchema` version, see *Upgrade User Authorization Data to 2.6 Format* (page 1014) for details.

Preparedness Before upgrading MongoDB, always test your application in a staging environment before deploying the upgrade to your production environment.

Some changes in MongoDB 3.0 require manual checks and intervention. Before beginning your upgrade, see the *Compatibility Changes in MongoDB 3.0* (page 945) document to ensure that your applications and deployments are compatible with MongoDB 3.0. Resolve the incompatibilities in your deployment before starting the upgrade.

Downgrade Limitations Once upgraded to MongoDB 3.0, you **cannot** downgrade to a version lower than **2.6.8**.

If you upgrade to 3.0 and have run `authSchemaUpgrade`, you **cannot** downgrade to 2.6 without disabling `--auth` or restoring a pre-upgrade backup, as `authSchemaUpgrade` discards the `MONGODB-CR` credentials used in 2.6. See *Upgrade Existing MONGODB-CR Users to Use SCRAM-SHA-1* (page 957).

Upgrade MongoDB Processes

[948] https://www.mongodb.com/products/consulting?jmp=docs#major_version_upgrade

Upgrade Standalone `mongod` Instance to MongoDB 3.0 The following steps outline the procedure to upgrade a standalone `mongod` from version 2.6 to 3.0. To upgrade from version 2.4 to 3.0, *upgrade to version 2.6* (page 1010) *first*, and then use the following procedure to upgrade from 2.6 to 3.0.

Upgrade Binaries If you installed MongoDB from the MongoDB `apt`, `yum`, or `zypper` repositories, you should upgrade to 3.0 using your package manager. Follow the appropriate *installation instructions* (page 22) for your Linux system. This will involve adding a repository for the new release, then performing the actual upgrade.

Otherwise, you can manually upgrade MongoDB:

Step 1: Download 3.0 binaries. Download binaries of the latest release in the 3.0 series from the MongoDB Download Page[949]. See *Install MongoDB* (page 21) for more information.

Step 2: Replace 2.6 binaries. Shut down your `mongod` instance. Replace the existing binary with the 3.0 `mongod` binary and restart `mongod`.

Change Storage Engine for Standalone to WiredTiger To change the storage engine for a standalone `mongod` instance to WiredTiger, see *Change Standalone to WiredTiger* (page 597).

Upgrade a Replica Set to 3.0

Prerequisites

- If the oplog contains entries generated by versions of MongoDB that precede version 2.2.1, you must wait for the entries to be overwritten by later versions *before* you can upgrade to MongoDB 3.0. For more information, see *Replica Set Oplog Format Change* (page 947)
- *Stricter validation in MongoDB 3.0* (page 947) of replica set configuration may invalidate previously-valid replica set configuration, preventing replica sets from starting in MongoDB 3.0. For more information, see *Replica Set Configuration Validation* (page 947).
- All replica set members must be running version 2.6 before you can upgrade them to version 3.0. To upgrade a replica set from an earlier MongoDB version, *upgrade all members of the replica set to version 2.6* (page 1010) *first*, and then follow the procedure to upgrade from MongoDB 2.6 to 3.0.

Upgrade Binaries You can upgrade from MongoDB 2.6 to 3.0 using a "rolling" upgrade to minimize downtime by upgrading the members individually while the other members are available:

Step 1: Upgrade secondary members of the replica set. Upgrade the *secondary* members of the set one at a time by shutting down the `mongod` and replacing the 2.6 binary with the 3.0 binary. After upgrading a `mongod` instance, wait for the member to recover to `SECONDARY` state before upgrading the next instance. To check the member's state, issue `rs.status()` in the `mongo` shell.

Step 2: Step down the replica set primary. Use `rs.stepDown()` in the `mongo` shell to step down the *primary* and force the set to *failover* (page 644). `rs.stepDown()` expedites the failover procedure and is preferable to shutting down the primary directly.

[949] http://www.mongodb.org/downloads?jmp=docs

Step 3: Upgrade the primary. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, shut down the previous primary and replace the `mongod` binary with the 3.0 binary and start the new instance.

Replica set failover is not instant and will render the set unavailable to accept writes until the failover process completes. This may take 30 seconds or more: schedule the upgrade procedure during a scheduled maintenance window.

Change Replica Set Storage Engine to WiredTiger To change the storage engine for a replica set to WiredTiger, see *Change Replica Set to WiredTiger* (page 598).

Upgrade a Sharded Cluster to 3.0 Only upgrade sharded clusters to 3.0 if **all** members of the cluster are currently running instances of 2.6. The only supported upgrade path for sharded clusters running 2.4 is via 2.6. The upgrade process checks all components of the cluster and will produce warnings if any component is running version 2.4.

Considerations The upgrade process does not require any downtime. However, while you upgrade the sharded cluster, ensure that clients do not make changes to the collection meta-data. For example, during the upgrade, do **not** do any of the following:

- `sh.enableSharding()`
- `sh.shardCollection()`
- `sh.addShard()`
- `db.createCollection()`
- `db.collection.drop()`
- `db.dropDatabase()`
- any operation that creates a database
- any other operation that modifies the cluster metadata in any way. See *Sharding Reference* (page 822) for a complete list of sharding commands. Note, however, that not all commands on the *Sharding Reference* (page 822) page modifies the cluster meta-data.

Upgrade Sharded Clusters *Optional but Recommended.* As a precaution, take a backup of the `config` database *before* upgrading the sharded cluster.

Step 1: Disable the Balancer. Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).

Step 2: Upgrade the cluster's meta data. Start a single 3.0 `mongos` instance with the `configDB` pointing to the cluster's config servers and with the `--upgrade` option.

To run a `mongos` with the `--upgrade` option, you can upgrade an existing `mongos` instance to 3.0, or if you need to avoid reconfiguring a production `mongos` instance, you can use a new 3.0 `mongos` that can reach all the config servers.

To upgrade the meta data, run:

```
mongos --configdb <configDB string> --upgrade
```

You can include the *--logpath* option to output the log messages to a file instead of the standard output. Also include any other options required to start mongos instances in your cluster, such as *--sslOnNormalPorts* or *--sslPEMKeyFile*.

The 3.0 mongos will output informational log messages.

```
<timestamp> I SHARDING [mongosMain] MongoS version 3.0.0 starting: ...
...
<timestamp> I SHARDING [mongosMain] starting upgrade of config server from v5 to v6
<timestamp> I SHARDING [mongosMain] starting next upgrade step from v5 to v6
<timestamp> I SHARDING [mongosMain] about to log new metadata event: ...
<timestamp> I SHARDING [mongosMain] checking that version of host ... is compatible with 2.6
...
<timestamp> I SHARDING [mongosMain] upgrade of config server to v6 successful
...
<timestamp> I SHARDING [mongosMain] distributed lock 'configUpgrade/...' unlocked.
<timestamp> I -        [mongosMain] Config database is at version v6
```

The mongos will exit upon completion of the *--upgrade* process.

The upgrade will prevent any chunk moves or splits from occurring during the upgrade process. If the data files have many sharded collections or if failed processes hold stale locks, acquiring the locks for all collections can take seconds or minutes. Watch the log for progress updates.

Step 3: Ensure `mongos --upgrade` process completes successfully. The mongos will exit upon completion of the meta data upgrade process. If successful, the process will log the following messages:

```
<timestamp> I SHARDING [mongosMain] upgrade of config server to v6 successful
...
<timestamp> I -        [mongosMain] Config database is at version v6
```

After a successful upgrade, restart the mongos instance. If mongos fails to start, check the log for more information.

If the mongos instance loses its connection to the config servers during the upgrade or if the upgrade is otherwise unsuccessful, you may always safely retry the upgrade.

Step 4: Upgrade the remaining `mongos` instances to 3.0. Upgrade and restart **without** the *--upgrade* option the other mongos instances in the sharded cluster.

After you have successfully upgraded *all* mongos instances, you can proceed to upgrade the other components in your sharded cluster.

Warning: Do not upgrade the mongod instances until after you have upgraded *all* the mongos instances.

Step 5: Upgrade the config servers. After you have successfully upgraded *all* mongos instances, upgrade all 3 mongod config server instances, leaving the *first* config server listed in the *mongos --configdb* argument to upgrade *last*.

Step 6: Upgrade the shards. Upgrade each shard, one at a time, upgrading the mongod secondaries before running replSetStepDown and upgrading the primary of each shard.

Step 7: Re-enable the balancer. Once the upgrade of sharded cluster components is complete, *Re-enable the balancer* (page 803).

Change Sharded Cluster Storage Engine to WiredTiger For a sharded cluster in MongoDB 3.0, you can choose to update the shards to use WiredTiger storage engine and have the config servers use MMAPv1. If you update the config servers to use WiredTiger, you must update all three config servers to use WiredTiger.

To change a sharded cluster to use WiredTiger, see *Change Sharded Cluster to WiredTiger* (page 599).

Upgrade Existing `MONGODB-CR` Users to Use `SCRAM-SHA-1` After upgrading the binaries, see *Upgrade to SCRAM-SHA-1* (page 957) for details on `SCRAM-SHA-1` upgrade scenarios.

General Upgrade Procedure **Except** as described on this page, moving between 2.6 and 3.0 is a drop-in replacement:

Step 1: Stop the existing `mongod` instance. For example, on Linux, run 2.6 `mongod` with the *`--shutdown`* option as follows:

```
mongod --dbpath /var/mongod/data --shutdown
```

Replace `/var/mongod/data` with your MongoDB `dbPath`. See also the *Stop mongod Processes* (page 324) for alternate methods of stopping a `mongod` instance.

Step 2: Start the new `mongod` instance. Ensure you start the 3.0 `mongod` with the same `dbPath`:

```
mongod --dbpath /var/mongod/data
```

Replace `/var/mongod/data` with your MongoDB `dbPath`.

Upgrade to `SCRAM-SHA-1`

On this page

- Overview (page 957)
- Considerations (page 958)
- Upgrade 2.6 `MONGODB-CR` Users to `SCRAM-SHA-1` (page 960)
- Result (page 960)
- Additional Resources (page 961)

Overview MongoDB 3.0 includes support for the *SCRAM-SHA-1* (page 399) challenge-response user authentication mechanism, which changes how MongoDB uses and stores user credentials.

For deployments that already contain user authentication data, to use the `SCRAM-SHA-1` mechanism, you must upgrade the authentication schema in addition to upgrading the MongoDB processes.

You may, alternatively, opt to continue to use the `MONGODB-CR` challenge-response mechanism and skip this upgrade.

See *Upgrade Scenarios* (page 957) for details.

Upgrade Scenarios The following scenarios are possible when upgrading from 2.6 to 3.0:

Continue to Use `MONGODB-CR` If you are upgrading from a 2.6 database with **existing** user authentication data, to continue to use `MONGODB-CR` for existing challenge-response users, **no upgrade to the existing user data is required**. However, new challenge-response users created in 3.0 will use the following authentication mechanism:

- If you populated MongoDB 3.0 user data by importing the 2.6 user authentication data, including user data, new challenge-response users created in MongoDB 3.0 will use `SCRAM-SHA1`.
- If you run MongoDB 3.0 binary against the 2.6 data files, including the user authentication data files, new challenge-response users created in MongoDB 3.0 will continue to use the `MONGODB-CR`.

You can execute the upgrade to `SCRAM-SHA-1` at any point in the future.

Use `SCRAM-SHA-1`

- If you are starting with a new 3.0 installation without any users or upgrading from a 2.6 database that has no users, to use `SCRAM-SHA-1`, **no user data upgrade is required**. All newly created users will have the correct format for `SCRAM-SHA-1`.
- If you are upgrading from a 2.6 database with **existing** user data, to use `SCRAM-SHA-1`, follow the steps in *Upgrade 2.6 MONGODB-CR Users to SCRAM-SHA-1* (page 960).

Important: Before you attempt any upgrade, familiarize yourself with the *Considerations* (page 958) as the upgrade to `SCRAM-SHA-1` is **irreversible** short of restoring from backups.

Recommendation `SCRAM-SHA-1` represents a significant improvement in security over `MONGODB-CR`, the previous default authentication mechanism: you are strongly urged to upgrade. For advantages of using `SCRAM-SHA-1`, see *SCRAM-SHA-1* (page 399).

Considerations

Backwards Incompatibility The procedure to upgrade to `SCRAM-SHA-1` **discards** the `MONGODB-CR` credentials used by 2.6. As such, the procedure is **irreversible**, short of restoring from backups.

The procedure also disables `MONGODB-CR` as an authentication mechanism.

Upgrade Binaries Before upgrading the authentication model, you should first upgrade MongoDB binaries to 3.0. For sharded clusters, ensure that **all** cluster components are 3.0.

Upgrade Drivers You must upgrade all drivers used by applications that will connect to upgraded database instances to version that support `SCRAM-SHA-1`. The minimum driver versions that support `SCRAM-SHA-1` are:

Driver Language	Version
C[950]	1.1.0[951]
C++[952]	1.0.0[953]
C#[954]	1.10[955]
Java[956]	2.13[957]
Node.js[958]	1.4.29[959]
Perl[960]	0.708.0.0[961]
PHP[962]	1.6[963]
Python[964]	2.8[965]
Motor[966]	0.4[967]
Ruby[968]	1.12[969]
Scala[970]	2.8.0[971]

See the MongoDB Drivers Page[972] for links to download upgraded drivers.

Requirements To upgrade the authentication model, you must have a user in the `admin` database with the role `userAdminAnyDatabase` (page 493).

Timing Because downgrades are more difficult after you upgrade the user authentication model, once you upgrade the MongoDB binaries to version 3.0, allow your MongoDB deployment to run for a day or two before following this procedure.

This allows 3.0 some time to "burn in" and decreases the likelihood of downgrades occurring after the user privilege model upgrade. The user authentication and access control will continue to work as it did in 2.6.

If you decide to upgrade the user authentication model immediately instead of waiting the recommended "burn in" period, then for sharded clusters, you must wait at least 10 seconds after upgrading the sharded clusters to run the authentication upgrade command.

Replica Sets For a replica set, it is only necessary to run the upgrade process on the *primary* as the changes will automatically replicate to the secondaries.

[950] https://docs.mongodb.org/ecosystem/drivers/c
[951] https://github.com/mongodb/mongo-c-driver/releases
[952] https://github.com/mongodb/mongo-cxx-driver
[953] https://github.com/mongodb/mongo-cxx-driver/releases
[954] https://docs.mongodb.org/ecosystem/drivers/csharp
[955] https://github.com/mongodb/mongo-csharp-driver/releases
[956] https://docs.mongodb.org/ecosystem/drivers/java
[957] https://github.com/mongodb/mongo-java-driver/releases
[958] https://docs.mongodb.org/ecosystem/drivers/node-js
[959] https://github.com/mongodb/node-mongodb-native/releases
[960] https://docs.mongodb.org/ecosystem/drivers/perl
[961] http://search.cpan.org/dist/MongoDB/
[962] https://docs.mongodb.org/ecosystem/drivers/php
[963] http://pecl.php.net/package/mongo
[964] https://docs.mongodb.org/ecosystem/drivers/python
[965] https://pypi.python.org/pypi/pymongo/
[966] https://docs.mongodb.org/ecosystem/drivers/python
[967] https://pypi.python.org/pypi/motor/
[968] https://docs.mongodb.org/ecosystem/drivers/ruby
[969] https://rubygems.org/gems/mongo
[970] https://docs.mongodb.org/ecosystem/drivers/scala
[971] https://github.com/mongodb/casbah/releases
[972] https://docs.mongodb.org/ecosystem/drivers

Sharded Clusters For a sharded cluster, connect to one `mongos` instance and run the upgrade procedure to upgrade the cluster's authentication data. By default, the procedure will upgrade the authentication data of the shards as well.

To override this behavior, run `authSchemaUpgrade` with the `upgradeShards: false` option. If you choose to override, you must run the upgrade procedure on the `mongos` first, and then run the procedure on the *primary* members of each shard.

For a sharded cluster, do **not** run the upgrade process directly against the *config servers* (page 742). Instead, perform the upgrade process using one `mongos` instance to interact with the config database.

Upgrade 2.6 `MONGODB-CR` Users to `SCRAM-SHA-1`

Warning: The procedure to upgrade to `SCRAM-SHA-1` **discards** the `MONG`
such, the procedure is **irreversible**, short of restoring from backups.
The procedure also disables `MONGODB-CR` as an authentication mechanism.

Important: To use the `SCRAM-SHA-1` authentication mechanism, a driver upgrade is **necessary** if your current driver version does not support `SCRAM-SHA-1`. See *required driver versions* (page 958) for details.

Step 1: Connect to the MongoDB instance. Connect and authenticate to the `mongod` instance for a single deployment, the primary `mongod` for a replica set, or a `mongos` for a sharded cluster as an `admin` database user with the role `userAdminAnyDatabase` (page 493).

Step 2: Upgrade authentication schema. Use the `authSchemaUpgrade` command in the `admin` database to update the user data using the `mongo` shell.

Run `authSchemaUpgrade` command.

```
db.adminCommand({authSchemaUpgrade: 1});
```

In case of error, you may safely rerun the `authSchemaUpgrade` command.

Sharded cluster `authSchemaUpgrade` consideration. For a sharded cluster *without shard local users* (page 395), `authSchemaUpgrade` will, by default, upgrade the authorization data of the shards as well, completing the upgrade.

You can, however, override this behavior by including `upgradeShards: false` in the command, as in the following example:

```
db.adminCommand(
   {authSchemaUpgrade: 1, upgradeShards: false }
);
```

If you override the default behavior or your cluster has shard local users, after running `authSchemaUpgrade` on a `mongos` instance, you will need to connect to the primary for each shard and repeat the upgrade process after upgrading on the `mongos`.

Result After this procedure is complete, all users in the database will have `SCRAM-SHA-1`-style credentials, and any subsequently-created users will also have this type of credentials.

Additional Resources

- Blog Post: Improved Password-Based Authentication in MongoDB 3.0: SCRAM Explained (Part 1)[973]
- Blog Post: Improved Password-Based Authentication in MongoDB 3.0: SCRAM Explained (Part 2)[974]

Downgrade MongoDB from 3.0

On this page

- Downgrade Recommendations and Checklist (page 961)
- Downgrade MongoDB Processes (page 961)
- General Downgrade Procedure (page 966)

Before you attempt any downgrade, familiarize yourself with the content of this document, particularly the *Downgrade Recommendations and Checklist* (page 961) and the procedure for *downgrading sharded clusters* (page 963).

Downgrade Recommendations and Checklist When downgrading, consider the following:

Downgrade Path Once upgraded to MongoDB 3.0, you **cannot** downgrade to a version lower than **2.6.8**.

Important: If you upgrade to MongoDB 3.0 and have run `authSchemaUpgrade`, you **cannot** downgrade to the 2.6 series without disabling `--auth`.

Procedures Follow the downgrade procedures:

- To downgrade sharded clusters, see *Downgrade a 3.0 Sharded Cluster* (page 963).
- To downgrade replica sets, see *Downgrade a 3.0 Replica Set* (page 962).
- To downgrade a standalone MongoDB instance, see *Downgrade a Standalone mongod Instance* (page 961).

Note: *Optional*. Consider `compacting` collections after downgrading. Otherwise, older versions will not be able to reuse free space regions larger than 2MB created while running 3.0. This can result in wasted space but no data loss following the downgrade.

Downgrade MongoDB Processes

Downgrade a Standalone `mongod` Instance If you have changed the storage engine to `WiredTiger`, change the storage engine to MMAPv1 before downgrading to 2.6.

Change Storage Engine to MMAPv1 To change storage engine to MMAPv1 for a standalone `mongod` instance, you will need to manually export and upload the data using `mongodump` and `mongorestore`.

Step 1: Ensure 3.0 `mongod` is running with WiredTiger.

[973] https://www.mongodb.com/blog/post/improved-password-based-authentication-mongodb-30-scram-explained-part-1?jmp=docs
[974] https://www.mongodb.com/blog/post/improved-password-based-authentication-mongodb-30-scram-explained-part-2?jmp=docs

Step 2: Export the data using `mongodump`.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 3: Create data directory for MMAPv1. Create a new data directory for MMAPv1. Ensure that the user account running `mongod` has read and write permissions for the new directory.

Step 4: Restart the `mongod` with MMAPv1. Restart the 3.0 `mongod`, specifying the newly created data directory for MMAPv1 as the *`--dbpath`*. You do not have to specify *`--storageEngine`* as MMAPv1 is the default.

```
mongod --dbpath <newMMAPv1DBPath>
```

Specify additional options as appropriate.

Step 5: Upload the exported data using `mongorestore`.

```
mongorestore <exportDataDestination>
```

Specify additional options as appropriate. See `mongorestore` for available options.

Downgrade Binaries The following steps outline the procedure to downgrade a standalone `mongod` from version 3.0 to 2.6.

Once upgraded to MongoDB 3.0, you **cannot** downgrade to a version lower than **2.6.8**.

Step 1: Download 2.6 binaries. Download binaries of the latest release in the 2.6 series from the MongoDB Download Page[975]. See *Install MongoDB* (page 21) for more information.

Step 2: Replace with 2.6 binaries. Shut down your `mongod` instance. Replace the existing binary with the 2.6 `mongod` binary and restart `mongod`.

Downgrade a 3.0 Replica Set If you have changed the storage engine to WiredTiger, change the storage engine to MMAPv1 before downgrading to 2.6.

Change Storage Engine to MMAPv1 You can update members to use the MMAPv1 storage engine in a rolling manner.

Note: When running a replica set with mixed storage engines, performance can vary according to workload.

To change the storage engine to MMAPv1 for an existing secondary replica set member, remove the member's data and perform an *initial sync* (page 699):

Step 1: Shutdown the secondary member. Stop the `mongod` instance for the secondary member.

[975] http://www.mongodb.org/downloads

Step 2: Prepare data directory for MMAPv1. Prepare *--dbpath* directory for initial sync.

For the stopped secondary member, either delete the content of the data directory or create a new data directory. If creating a new directory, ensure that the user account running `mongod` has read and write permissions for the new directory.

Step 3: Restart the secondary member with MMAPv1. Restart the 3.0 `mongod`, specifying the MMAPv1 data directory as the *--dbpath*. Specify additional options as appropriate for the member. You do not have to specify *--storageEngine* since MMAPv1 is the default.

```
mongod --dbpath <preparedMMAPv1DBPath>
```

Since no data exists in the `--dbpath`, the `mongod` will perform an initial sync. The length of the initial sync process depends on the size of the database and network connection between members of the replica set.

Repeat for the remaining the secondary members. Once all the secondary members have switched to MMAPv1, step down the primary, and update the stepped-down member.

Downgrade Binaries Once upgraded to MongoDB 3.0, you **cannot** downgrade to a version lower than **2.6.8**.

The following steps outline a "rolling" downgrade process for the replica set. The "rolling" downgrade process minimizes downtime by downgrading the members individually while the other members are available:

Step 1: Downgrade secondary members of the replica set. Downgrade each *secondary* member of the replica set, one at a time:

1. Shut down the `mongod`. See *Stop mongod Processes* (page 324) for instructions on safely terminating `mongod` processes.
2. Replace the 3.0 binary with the 2.6 binary and restart.
3. Wait for the member to recover to `SECONDARY` state before downgrading the next secondary. To check the member's state, use the `rs.status()` method in the `mongo` shell.

Step 2: Step down the primary. Use `rs.stepDown()` in the `mongo` shell to step down the *primary* and force the normal *failover* (page 644) procedure.

```
rs.stepDown()
```

`rs.stepDown()` expedites the failover procedure and is preferable to shutting down the primary directly.

Step 3: Replace and restart former primary `mongod`. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, shut down the previous primary and replace the `mongod` binary with the 2.6 binary and start the new instance.

Replica set failover is not instant but will render the set unavailable to writes and interrupt reads until the failover process completes. Typically this takes 10 seconds or more. You may wish to plan the downgrade during a predetermined maintenance window.

Downgrade a 3.0 Sharded Cluster

Requirements While the downgrade is in progress, you cannot make changes to the collection meta-data. For example, during the downgrade, do **not** do any of the following:

- `sh.enableSharding()`
- `sh.shardCollection()`
- `sh.addShard()`
- `db.createCollection()`
- `db.collection.drop()`
- `db.dropDatabase()`
- any operation that creates a database
- any other operation that modifies the cluster meta-data in any way. See *Sharding Reference* (page 822) for a complete list of sharding commands. Note, however, that not all commands on the *Sharding Reference* (page 822) page modifies the cluster meta-data.

Change Storage Engine to MMAPv1 If you have changed the storage engine to WiredTiger, change the storage engine to MMAPv1 before downgrading to 2.6.

Change Shards to Use MMAPv1 To change the storage engine to MMAPv1, refer to the procedure in *Change Storage Engine to MMAPv1 for replica set members* (page 962) and *Change Storage Engine to MMAPv1 for standalone mongod* (page 961) as appropriate for your shards.

Change Config Servers to Use MMAPv1

Note: During this process, only two config servers will be running at any given time to ensure that the sharded cluster's metadata is *read only*.

Step 1: Disable the Balancer. Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).

Step 2: Stop the last config server listed in the `mongos`' `configDB` setting.

Step 3: Export data of the second config server listed in the `mongos`' `configDB` setting.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 4: For the second config server, create a new data directory for MMAPv1. Ensure that the user account running `mongod` has read and write permissions for the new directory.

Step 5: Restart the second config server with MMAPv1. Specify the newly created MMAPv1 data directory as the *`--dbpath`* as well as any additional options as appropriate.

```
mongod --dbpath <newMMAPv1DBPath> --configsvr
```

Step 6: Upload the exported data using `mongorestore` to the second config server.

```
mongorestore <exportDataDestination>
```

Specify additional options as appropriate. See `mongorestore` for available options.

Step 7: Shut down the second config server.

Step 8: Restart the third config server.

Step 9: Export data of the third config server.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 10: For the third config server, create a new data directory for MMAPv1. Ensure that the user account running `mongod` has read and write permissions for the new directory.

Step 11: Restart the third config server with MMAPv1. Specify the newly created MMAPv1 data directory as the *`--dbpath`* as well as any additional options as appropriate.

```
mongod --dbpath <newMMAPv1DBPath> --configsvr
```

Step 12: Upload the exported data using `mongorestore` to the third config server.

```
mongorestore <exportDataDestination>
```

Specify additional options as appropriate. See `mongorestore` for available options

Step 13: Export data of the first config server listed in the `mongos`' `configDB` setting.

```
mongodump --out <exportDataDestination>
```

Specify additional options as appropriate, such as username and password if running with authorization enabled. See `mongodump` for available options.

Step 14: For the first config server, create data directory for MMAPv1. Ensure that the user account running `mongod` has read and write permissions for the new directory.

Step 15: Restart the first config server with MMAPv1. Specify the newly created MMAPv1 data directory as the *`--dbpath`* as well as any additional options as appropriate.

```
mongod --dbpath <newMMAPv1DBPath> --configsvr
```

Step 16: Upload the exported data using `mongorestore` to the first config server.

```
mongorestore <exportDataDestination>
```

Specify additional options as appropriate. See `mongorestore` for available options

Step 17: Enable writes to the sharded cluster's metadata. Restart the **second** config server, specifying the newly created MMAPv1 data directory as the *--dbpath*. Specify additional options as appropriate.

```
mongod --dbpath <newMMAPv1DBPath> --configsvr
```

Once all three config servers are up, the sharded cluster's metadata is available for writes.

Step 18: Re-enable the balancer. Once all three config servers are up and running with WiredTiger, *Re-enable the balancer* (page 803).

Downgrade Binaries Once upgraded to MongoDB 3.0, you **cannot** downgrade to a version lower than **2.6.8**.

The downgrade procedure for a sharded cluster reverses the order of the upgrade procedure. The version `v6` config database is backwards compatible with MongoDB 2.6.

Step 1: Disable the Balancer. Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).

Step 2: Downgrade each shard, one at a time. For each shard:

1. Downgrade the `mongod` secondaries *before* downgrading the primary.
2. To downgrade the primary, run `replSetStepDown` and downgrade.

Step 3: Downgrade the config servers. Downgrade all 3 `mongod` config server instances, leaving the *first* system in the *mongos --configdb* argument to downgrade *last*.

Step 4: Downgrade the `mongos` instances. Downgrade and restart each `mongos`, one at a time. The downgrade process is a binary drop-in replacement.

Step 5: Re-enable the balancer. Once the upgrade of sharded cluster components is complete, *re-enable the balancer* (page 803).

General Downgrade Procedure **Except** as described on this page, moving between 2.6 and 3.0 is a drop-in replacement:

Step 1: Stop the existing `mongod` instance. For example, on Linux, run 3.0 `mongod` with the *--shutdown* option as follows:

```
mongod --dbpath /var/mongod/data --shutdown
```

Replace `/var/mongod/data` with your MongoDB `dbPath`. See also the *Stop mongod Processes* (page 324) for alternate methods of stopping a `mongod` instance.

Step 2: Start the new `mongod` instance. Ensure you start the 2.6 `mongod` with the same `dbPath`:

```
mongod --dbpath /var/mongod/data
```

Replace `/var/mongod/data` with your MongoDB `dbPath`.

See *Upgrade MongoDB to 3.0* (page 953) for full upgrade instructions.

Download

To download MongoDB 3.0, go to the downloads page[976].

See also:

- All Third Party License Notices[977].
- All JIRA issues resolved in 3.0[978].

Additional Resources

- Blog Post: Announcing MongoDB 3.0[979]
- Whitepaper: What's New in MongoDB 3.0[980]
- Webinar: What's New in MongoDB 3.0[981]

15.2.2 Release Notes for MongoDB 2.6

On this page

- Minor Releases (page 967)
- Major Changes (page 994)
- Security Improvements (page 996)
- Query Engine Improvements (page 996)
- Improvements (page 997)
- Operational Changes (page 998)
- MongoDB Enterprise Features (page 998)
- Additional Information (page 999)

April 8, 2014

MongoDB 2.6 is now available. Key features include aggregation enhancements, text-search integration, query-engine improvements, a new write-operation protocol, and security enhancements.

Minor Releases

2.6 Changelog

[976] http://www.mongodb.org/downloads
[977] https://github.com/mongodb/mongo/blob/v3.0/distsrc/THIRD-PARTY-NOTICES
[978] http://bit.ly/1CpOu6t
[979] http://www.mongodb.com/blog/post/announcing-mongodb-30?jmp=docs
[980] https://www.mongodb.com/lp/white-paper/mongodb-3.0?jmp=docs
[981] https://www.mongodb.com/webinar/Whats-New-in-MongoDB-3-0?jmp=docs

On this page

- 2.6.12 Changelog (page 968)
- 2.6.11 – Changes (page 969)
- 2.6.10 – Changes (page 971)
- 2.6.9 – Changes (page 973)
- 2.6.8 – Changes (page 973)
- 2.6.7 – Changes (page 975)
- 2.6.6 – Changes (page 976)
- 2.6.5 – Changes (page 978)
- 2.6.4 – Changes (page 981)
- 2.6.3 – Changes (page 985)
- 2.6.2 – Changes (page 985)
- 2.6.1 – Changes (page 989)

2.6.12 Changelog

Security SERVER-19284[982] Should not be able to create role with same name as builtin role

Sharding

- SERVER-17886[983] dbKillCursors op asserts on mongos when at log level 3
- SERVER-19266[984] An error document is returned with result set
- SERVER-20191[985] multi-updates/remove can make successive queries skip shard version checking
- SERVER-20839[986] trace_missing_docs_test.js compares Timestamp instances using < operator in mongo shell

Query

- SERVER-2454[987] Queries that are killed during a yield should return error to user instead of partial result set
- SERVER-16042[988] Optimise $all/$and to select smallest subset as initial index bounds
- SERVER-19725[989] NULL pointer crash in QueryPlanner::plan with $near operator
- SERVER-20083[990] Add log statement at default log level for when an index filter is set or cleared successfully
- SERVER-20829[991] RUNNER_DEAD on document delete during update by _id or find by _id
- SERVER-21227[992] MultiPlanStage::invalidate() should not flag and drop invalidated WorkingSetMembers
- SERVER-21602[993] Reduce execution time of cursor_timeout.js

[982] https://jira.mongodb.org/browse/SERVER-19284
[983] https://jira.mongodb.org/browse/SERVER-17886
[984] https://jira.mongodb.org/browse/SERVER-19266
[985] https://jira.mongodb.org/browse/SERVER-20191
[986] https://jira.mongodb.org/browse/SERVER-20839
[987] https://jira.mongodb.org/browse/SERVER-2454
[988] https://jira.mongodb.org/browse/SERVER-16042
[989] https://jira.mongodb.org/browse/SERVER-19725
[990] https://jira.mongodb.org/browse/SERVER-20083
[991] https://jira.mongodb.org/browse/SERVER-20829
[992] https://jira.mongodb.org/browse/SERVER-21227
[993] https://jira.mongodb.org/browse/SERVER-21602

- SERVER-22195[994] queryoptimizer3.js failing on 2.6
- SERVER-22535[995] Some index operations (drop index, abort index build, update TTL config) on collection during active migration can cause migration to skip documents

Write Operations SERVER-21647[996] $rename changes field ordering

Storage SERVER-21543[997] Lengthen delay before deleting old journal files

MMAP

- SERVER-22261[998] MMAPv1 LSNFile may be updated ahead of what is synced to data files

Operations SERVER-13985[999] printShardingStatus uses group/JS

Build and Packaging

- SERVER-18432[1000] alert when passing unused variables to scons
- SERVER-18793[1001] Enterprise RPM build issues
- SERVER-19509[1002] The nproc ulimits are different across packages
- SERVER-20583[1003] migrate all windows-64 vs2010 builders in evergreen to use new version of the distribution
- SERVER-20830[1004] set push and docs_tickets tasks as not available for patch testing
- SERVER-21864[1005] streamline artifact signing procedure to support coherent release process

Internals

- SERVER-20121[1006] XorShift PRNG should use unsigned arithmetic
- SERVER-20401[1007] Publicly expose net.ssl.disabledProtocols

2.6.11 – Changes

Querying

- SERVER-19553[1008] `mongod` shouldn't use `sayPiggyBack` to send `killCursor` messages
- SERVER-18620[1009] Reduce frequency of "`staticYield can't unlock`" log message

[994] https://jira.mongodb.org/browse/SERVER-22195
[995] https://jira.mongodb.org/browse/SERVER-22535
[996] https://jira.mongodb.org/browse/SERVER-21647
[997] https://jira.mongodb.org/browse/SERVER-21543
[998] https://jira.mongodb.org/browse/SERVER-22261
[999] https://jira.mongodb.org/browse/SERVER-13985
[1000] https://jira.mongodb.org/browse/SERVER-18432
[1001] https://jira.mongodb.org/browse/SERVER-18793
[1002] https://jira.mongodb.org/browse/SERVER-19509
[1003] https://jira.mongodb.org/browse/SERVER-20583
[1004] https://jira.mongodb.org/browse/SERVER-20830
[1005] https://jira.mongodb.org/browse/SERVER-21864
[1006] https://jira.mongodb.org/browse/SERVER-20121
[1007] https://jira.mongodb.org/browse/SERVER-20401
[1008] https://jira.mongodb.org/browse/SERVER-19553
[1009] https://jira.mongodb.org/browse/SERVER-18620

- SERVER-18461[1010] Range predicates comparing against a BinData value should be covered, but are not in 2.6
- SERVER-17815[1011] Plan ranking tie breaker is computed incorrectly
- SERVER-16265[1012] Add query details to getmore entry in profiler and `db.currentOp()`
- SERVER-15217[1013] v2.6 query plan ranking test "`NonCoveredIxisectFetchesLess`" relies on order of deleted record list
- SERVER-14070[1014] Compound index not providing sort if equality predicate given on sort field

Replication

- SERVER-18280[1015] `ReplicaSetMonitor` should use `electionId` to avoid talking to old primaries
- SERVER-18795[1016] `db.printSlaveReplicationInfo()`/`rs.printSlaveReplicationInfo()` can not work with `ARBITER` role

Sharding

- SERVER-19464[1017] `$sort` stage in aggregation doesn't call scoped connections done ()
- SERVER-18955[1018] `mongos` doesn't set batch size (and keeps the old one, 0) on getMore if performed on first `_cursor->more()`

Indexing

- SERVER-19559[1019] Document growth of "key too large" document makes it disappear from the index
- SERVER-16348[1020] `Assertion failure n >= 0 && n < static_cast<int>(_files.size()) src/mongo/db/storage/extent_manager.cpp 109`
- SERVER-13875[1021] `ensureIndex()` of `2dsphere` index breaks after upgrading to 2.6 (with the new `createIndex` command)

Networking

SERVER-19389[1022] Remove wire level endianness check

Build and Testing

- SERVER-18097[1023] Remove `mongosTest_auth` and `mongosTest_WT` tasks from `evergreen.yml`
- SERVER-18068[1024] Coverity analysis defect 72413: Resource leak
- SERVER-18371[1025] Add SSL library config detection

[1010] https://jira.mongodb.org/browse/SERVER-18461
[1011] https://jira.mongodb.org/browse/SERVER-17815
[1012] https://jira.mongodb.org/browse/SERVER-16265
[1013] https://jira.mongodb.org/browse/SERVER-15217
[1014] https://jira.mongodb.org/browse/SERVER-14070
[1015] https://jira.mongodb.org/browse/SERVER-18280
[1016] https://jira.mongodb.org/browse/SERVER-18795
[1017] https://jira.mongodb.org/browse/SERVER-19464
[1018] https://jira.mongodb.org/browse/SERVER-18955
[1019] https://jira.mongodb.org/browse/SERVER-19559
[1020] https://jira.mongodb.org/browse/SERVER-16348
[1021] https://jira.mongodb.org/browse/SERVER-13875
[1022] https://jira.mongodb.org/browse/SERVER-19389
[1023] https://jira.mongodb.org/browse/SERVER-18097
[1024] https://jira.mongodb.org/browse/SERVER-18068
[1025] https://jira.mongodb.org/browse/SERVER-18371

2.6.10 – Changes

Security

- SERVER-18312[1026] Upgrade PCRE to latest
- SERVER-17812[1027] LockPinger has audit-related GLE failure
- SERVER-17647[1028] Compute BinData length in v8
- SERVER-17591[1029] Add SSL flag to select supported protocols
- SERVER-16849[1030] On mongos we always invalidate the user cache once, even if no user definitions are changing
- SERVER-11980[1031] Improve user cache invalidation enforcement on mongos

Querying

- SERVER-18364[1032] Ensure non-negation predicates get chosen over negation predicates for multikey index bounds construction
- SERVER-17815[1033] Plan ranking tie breaker is computed incorrectly
- SERVER-16256[1034] $all clause with elemMatch uses wider bounds than needed

Replication

- SERVER-18211[1035] MongoDB fails to correctly roll back collection creation
- SERVER-17771[1036] Reconfiguring a replica set to remove a node causes a segmentation fault on 2.6.8
- SERVER-13542[1037] Expose electionId on primary in isMaster

Sharding

- SERVER-17812[1038] LockPinger has audit-related GLE failure
- SERVER-17805[1039] logOp / OperationObserver should always check shardversion
- SERVER-17749[1040] `collMod usePowerOf2Sizes` fails on `mongos`
- SERVER-11980[1041] Improve user cache invalidation enforcement on `mongos`

[1026] https://jira.mongodb.org/browse/SERVER-18312
[1027] https://jira.mongodb.org/browse/SERVER-17812
[1028] https://jira.mongodb.org/browse/SERVER-17647
[1029] https://jira.mongodb.org/browse/SERVER-17591
[1030] https://jira.mongodb.org/browse/SERVER-16849
[1031] https://jira.mongodb.org/browse/SERVER-11980
[1032] https://jira.mongodb.org/browse/SERVER-18364
[1033] https://jira.mongodb.org/browse/SERVER-17815
[1034] https://jira.mongodb.org/browse/SERVER-16256
[1035] https://jira.mongodb.org/browse/SERVER-18211
[1036] https://jira.mongodb.org/browse/SERVER-17771
[1037] https://jira.mongodb.org/browse/SERVER-13542
[1038] https://jira.mongodb.org/browse/SERVER-17812
[1039] https://jira.mongodb.org/browse/SERVER-17805
[1040] https://jira.mongodb.org/browse/SERVER-17749
[1041] https://jira.mongodb.org/browse/SERVER-11980

Storage

- SERVER-18211[1042] MongoDB fails to correctly roll back collection creation
- SERVER-17653[1043] ERROR: socket XXX is higher than 1023; not supported on 2.6.*

Indexing SERVER-17018[1044] Assertion failure false `src/mongo/db/structure/btree/key.cpp` Line 433 on remove operation

Write Ops

- SERVER-18111[1045] `mongod` allows user inserts into `system.profile` collection
- SERVER-13542[1046] Expose `electionId` on primary in `isMaster`

Networking

- SERVER-18096[1047] Shard primary incorrectly reuses closed sockets after relinquish and re-election
- SERVER-17591[1048] Add SSL flag to select supported protocols

Build and Packaging

- SERVER-18344[1049] logs should be sent to updated logkeeper server
- SERVER-18082[1050] Change smoke.py buildlogger command line options to environment variables
- SERVER-18312[1051] Upgrade PCRE to latest
- SERVER-17780[1052] Init script sets process ulimit to different value compared to documentation
- SERVER-16563[1053] Debian repo component mismatch - mongodb/10gen

Shell SERVER-17951[1054] db.currentOp() fails with read preference set

Testing

- SERVER-18262[1055] setup_multiversion_mongodb should retry links download on timeouts
- SERVER-18229[1056] `smoke.py` with PyMongo 3.0.1 fails to run certain tests
- SERVER-18073[1057] Fix `smoke.py` to work with PyMongo 3.0

[1042] https://jira.mongodb.org/browse/SERVER-18211
[1043] https://jira.mongodb.org/browse/SERVER-17653
[1044] https://jira.mongodb.org/browse/SERVER-17018
[1045] https://jira.mongodb.org/browse/SERVER-18111
[1046] https://jira.mongodb.org/browse/SERVER-13542
[1047] https://jira.mongodb.org/browse/SERVER-18096
[1048] https://jira.mongodb.org/browse/SERVER-17591
[1049] https://jira.mongodb.org/browse/SERVER-18344
[1050] https://jira.mongodb.org/browse/SERVER-18082
[1051] https://jira.mongodb.org/browse/SERVER-18312
[1052] https://jira.mongodb.org/browse/SERVER-17780
[1053] https://jira.mongodb.org/browse/SERVER-16563
[1054] https://jira.mongodb.org/browse/SERVER-17951
[1055] https://jira.mongodb.org/browse/SERVER-18262
[1056] https://jira.mongodb.org/browse/SERVER-18229
[1057] https://jira.mongodb.org/browse/SERVER-18073

2.6.9 – Changes

Security SERVER-16073[1058] Create hidden `net.ssl.sslCipherConfig` flag

Querying

- SERVER-14723[1059] Crash during query planning for `geoNear` with multiple `2dsphere` indexes
- SERVER-14071[1060] For queries with `sort()`, bad non-blocking plan can be cached if there are zero results
- SERVER-8188[1061] Configurable idle cursor timeout

Replication and Sharding

- SERVER-17429[1062] the message logged when changing sync target due to stale data should format OpTimes in a consistent way
- SERVER-17441[1063] `mongos` crash right after "not master" error

Storage SERVER-15907[1064] Use `ftruncate` rather than `fallocate` when running on `tmpfs`

Aggregation Framework

- SERVER-17426[1065] Aggregation framework query by `_id` returns duplicates in sharded cluster (orphan documents)
- SERVER-17224[1066] Aggregation pipeline with 64MB document can terminate server

Build and Platform

- SERVER-17484[1067] Migrate server MCI config into server repo
- SERVER-17252[1068] Upgrade PCRE Version from 8.30 to Latest

Diagnostics and Internal Code

- SERVER-17226[1069] `top` command with 64MB result document can terminate server
- SERVER-17338[1070] NULL pointer crash when running `copydb` against stepped-down 2.6 primary
- SERVER-14992[1071] Query for Windows 7 File Allocation Fix, and other hotfixes

2.6.8 – Changes

[1058] https://jira.mongodb.org/browse/SERVER-16073
[1059] https://jira.mongodb.org/browse/SERVER-14723
[1060] https://jira.mongodb.org/browse/SERVER-14071
[1061] https://jira.mongodb.org/browse/SERVER-8188
[1062] https://jira.mongodb.org/browse/SERVER-17429
[1063] https://jira.mongodb.org/browse/SERVER-17441
[1064] https://jira.mongodb.org/browse/SERVER-15907
[1065] https://jira.mongodb.org/browse/SERVER-17426
[1066] https://jira.mongodb.org/browse/SERVER-17224
[1067] https://jira.mongodb.org/browse/SERVER-17484
[1068] https://jira.mongodb.org/browse/SERVER-17252
[1069] https://jira.mongodb.org/browse/SERVER-17226
[1070] https://jira.mongodb.org/browse/SERVER-17338
[1071] https://jira.mongodb.org/browse/SERVER-14992

Security and Networking

- SERVER-17278[1072] BSON BinData validation enforcement
- SERVER-17022[1073] No SSL Session Caching may not be respected
- SERVER-17264[1074] improve bson validation

Query and Aggregation

- SERVER-16655[1075] Geo predicate is unable to use compound 2dsphere index if it is root of `$or` clause
- SERVER-16527[1076] `2dsphere` explain reports "works" for `nscanned` & `nscannedObjects`
- SERVER-15802[1077] Query optimizer should always use equality predicate over unique index when possible
- SERVER-14044[1078] Incorrect `{$meta: 'text'}` reference in aggregation `$sort` error message

Replication

- SERVER-16599[1079] `copydb` and `clone` commands can crash the server if a primary steps down
- SERVER-16315[1080] Replica set nodes should not threaten to veto nodes whose config version is higher than their own
- SERVER-16274[1081] secondary `fasserts` trying to replicate an index
- SERVER-15471[1082] Better error message when replica is not found in `GhostSync::associateSlave`

Sharding

- SERVER-17191[1083] Spurious warning during upgrade of sharded cluster
- SERVER-17163[1084] Fatal error "logOp but not primary" in `MigrateStatus::go`
- SERVER-16984[1085] `UpdateLifecycleImpl` can return empty `collectionMetadata` even if `ns` is sharded
- SERVER-10904[1086] Possible for `_master` and `_slaveConn` to be pointing to different connections even with primary read pref

Storage

- SERVER-17087[1087] Add listCollections command functionality to 2.6 shell & client
- SERVER-14572[1088] Increase C runtime stdio file limit

[1072] https://jira.mongodb.org/browse/SERVER-17278
[1073] https://jira.mongodb.org/browse/SERVER-17022
[1074] https://jira.mongodb.org/browse/SERVER-17264
[1075] https://jira.mongodb.org/browse/SERVER-16655
[1076] https://jira.mongodb.org/browse/SERVER-16527
[1077] https://jira.mongodb.org/browse/SERVER-15802
[1078] https://jira.mongodb.org/browse/SERVER-14044
[1079] https://jira.mongodb.org/browse/SERVER-16599
[1080] https://jira.mongodb.org/browse/SERVER-16315
[1081] https://jira.mongodb.org/browse/SERVER-16274
[1082] https://jira.mongodb.org/browse/SERVER-15471
[1083] https://jira.mongodb.org/browse/SERVER-17191
[1084] https://jira.mongodb.org/browse/SERVER-17163
[1085] https://jira.mongodb.org/browse/SERVER-16984
[1086] https://jira.mongodb.org/browse/SERVER-10904
[1087] https://jira.mongodb.org/browse/SERVER-17087
[1088] https://jira.mongodb.org/browse/SERVER-14572

Tools

- SERVER-17216[1089] 2.6 `mongostat` cannot be used with 3.0 `mongod`
- SERVER-14190[1090] `mongorestore parseMetadataFile` passes non-null terminated string to 'fromjson'

Build and Packaging

- SERVER-14803[1091] Support static libstdc++ builds for non-Linux builds
- SERVER-15400[1092] Create Windows Enterprise Zip File with vcredist and dependent dlls

Usability SERVER-14756[1093] The YAML `storage.quota.enforced` option is not found

Testing SERVER-16421[1094] `sharding_rs2.js` should clean up data on all replicas

2.6.7 – Changes

Stability

- SERVER-16237[1095] Don't check the shard version if the primary server is down

Querying

- SERVER-16408[1096] `max_time_ms.js` should not run in parallel suite.

Replication

- SERVER-16732[1097] `SyncSourceFeedback::replHandshake()` may perform an illegal erase from a `std::map` in some circumstances

Sharding

- SERVER-16683[1098] Decrease mongos memory footprint when shards have several tags
- SERVER-15766[1099] prefix_shard_key.js depends on primary allocation to particular shards
- SERVER-14306[1100] `mongos` can cause shards to hit the in-memory sort limit by requesting more results than needed.

[1089] https://jira.mongodb.org/browse/SERVER-17216
[1090] https://jira.mongodb.org/browse/SERVER-14190
[1091] https://jira.mongodb.org/browse/SERVER-14803
[1092] https://jira.mongodb.org/browse/SERVER-15400
[1093] https://jira.mongodb.org/browse/SERVER-14756
[1094] https://jira.mongodb.org/browse/SERVER-16421
[1095] https://jira.mongodb.org/browse/SERVER-16237
[1096] https://jira.mongodb.org/browse/SERVER-16408
[1097] https://jira.mongodb.org/browse/SERVER-16732
[1098] https://jira.mongodb.org/browse/SERVER-16683
[1099] https://jira.mongodb.org/browse/SERVER-15766
[1100] https://jira.mongodb.org/browse/SERVER-14306

Packaging

- SERVER-16081[1101] `/etc/init.d/mongod` startup script fails, with dirname message

2.6.6 – Changes

Security

- SERVER-15673[1102] Disable SSLv3 ciphers
- SERVER-15515[1103] New test for mixed version replSet, 2.4 primary, user updates
- SERVER-15500[1104] New test for system.user operations

Stability

- SERVER-12061[1105] Do not silently ignore read errors when syncing a replica set node
- SERVER-12058[1106] Primary should abort if encountered problems writing to the oplog

Querying

- SERVER-16291[1107] Cannot set/list/clear index filters on the secondary
- SERVER-15958[1108] The "isMultiKey" value is not correct in the output of aggregation explain plan
- SERVER-15899[1109] Querying against path in document containing long array of subdocuments with nested arrays causes stack overflow
- SERVER-15696[1110] `$regex`, `$in` and `$sort` with index returns too many results
- SERVER-15639[1111] Text queries can return incorrect results and leak memory when multiple predicates given on same text index prefix field
- SERVER-15580[1112] Evaluating candidate query plans with concurrent writes on same collection may crash `mongod`
- SERVER-15528[1113] Distinct queries can scan many index keys without yielding read lock
- SERVER-15485[1114] CanonicalQuery::canonicalize can leak a LiteParsedQuery
- SERVER-15403[1115] `$min` and `$max` equal errors in 2.6 but not in 2.4
- SERVER-15233[1116] Cannot run `planCacheListQueryShapes` on a Secondary
- SERVER-14799[1117] `count` with hint doesn't work when hint is a document

[1101] https://jira.mongodb.org/browse/SERVER-16081
[1102] https://jira.mongodb.org/browse/SERVER-15673
[1103] https://jira.mongodb.org/browse/SERVER-15515
[1104] https://jira.mongodb.org/browse/SERVER-15500
[1105] https://jira.mongodb.org/browse/SERVER-12061
[1106] https://jira.mongodb.org/browse/SERVER-12058
[1107] https://jira.mongodb.org/browse/SERVER-16291
[1108] https://jira.mongodb.org/browse/SERVER-15958
[1109] https://jira.mongodb.org/browse/SERVER-15899
[1110] https://jira.mongodb.org/browse/SERVER-15696
[1111] https://jira.mongodb.org/browse/SERVER-15639
[1112] https://jira.mongodb.org/browse/SERVER-15580
[1113] https://jira.mongodb.org/browse/SERVER-15528
[1114] https://jira.mongodb.org/browse/SERVER-15485
[1115] https://jira.mongodb.org/browse/SERVER-15403
[1116] https://jira.mongodb.org/browse/SERVER-15233
[1117] https://jira.mongodb.org/browse/SERVER-14799

Replication

- SERVER-16107[1118] 2.6 `mongod` crashes with segfault when added to a 2.8 replica set with >= 12 nodes.
- SERVER-15994[1119] `listIndexes` and `listCollections` can be run on secondaries without slaveOk bit
- SERVER-15849[1120] do not forward replication progress for nodes that are no longer part of a replica set
- SERVER-15491[1121] `SyncSourceFeedback` can crash due to a `SocketException` in `authenticateInternalUser`

Sharding

- SERVER-15318[1122] `copydb` should not use exhaust flag when used against `mongos`
- SERVER-14728[1123] Shard depends on string comparison of replica set connection string
- SERVER-14506[1124] special top chunk logic can move max chunk to a shard with incompatible tag
- SERVER-14299[1125] For sharded limit=N queries with sort, mongos can request >N results from shard
- SERVER-14080[1126] Have migration result reported in the changelog correctly
- SERVER-12472[1127] Fail MoveChunk if an index is needed on TO shard and data exists

Storage

- SERVER-16283[1128] Can't start new wiredtiger node with log file or config file in data directory - false detection of old `mmapv1` files
- SERVER-15986[1129] Starting with different storage engines in the same dbpath should error/warn
- SERVER-14057[1130] Changing TTL expiration time with collMod does not correctly update index definition

Indexing and write Operations

- SERVER-14287[1131] ensureIndex can abort reIndex and lose indexes
- SERVER-14886[1132] Updates against paths composed with array index notation and positional operator fail with error

Data Aggregation SERVER-15552[1133] Errors writing to temporary collections during `mapReduce` command execution should be operation-fatal

[1118] https://jira.mongodb.org/browse/SERVER-16107
[1119] https://jira.mongodb.org/browse/SERVER-15994
[1120] https://jira.mongodb.org/browse/SERVER-15849
[1121] https://jira.mongodb.org/browse/SERVER-15491
[1122] https://jira.mongodb.org/browse/SERVER-15318
[1123] https://jira.mongodb.org/browse/SERVER-14728
[1124] https://jira.mongodb.org/browse/SERVER-14506
[1125] https://jira.mongodb.org/browse/SERVER-14299
[1126] https://jira.mongodb.org/browse/SERVER-14080
[1127] https://jira.mongodb.org/browse/SERVER-12472
[1128] https://jira.mongodb.org/browse/SERVER-16283
[1129] https://jira.mongodb.org/browse/SERVER-15986
[1130] https://jira.mongodb.org/browse/SERVER-14057
[1131] https://jira.mongodb.org/browse/SERVER-14287
[1132] https://jira.mongodb.org/browse/SERVER-14886
[1133] https://jira.mongodb.org/browse/SERVER-15552

Build and Packaging

- SERVER-14184[1134] Unused preprocessor macros from s2 conflict on OS X Yosemite
- SERVER-14015[1135] S2 Compilation on GCC 4.9/Solaris fails
- SERVER-16017[1136] Suse11 enterprise packages fail due to unmet dependencies
- SERVER-15598[1137] Ubuntu 14.04 Enterprise packages depend on unavailable libsnmp15 package
- SERVER-13595[1138] Red Hat init.d script error: YAML config file parsing

Logging and Diagnostics

- SERVER-13471[1139] Increase log level of "did reduceInMemory" message in map/reduce
- SERVER-16324[1140] Command execution log line displays "`query not recording (too large)`" instead of abbreviated command object
- SERVER-10069[1141] Improve errorcodes.py so it captures multiline messages

Testing and Internals

- SERVER-15632[1142] `MultiHostQueryOp::PendingQueryContext::doBlockingQuery` can leak a cursor object
- SERVER-15629[1143] `GeoParser::parseMulti{Line|Polygon}` does not clear objects owned by out parameter
- SERVER-16316[1144] Remove unsupported behavior in shard3.js
- SERVER-14763[1145] Update jstests/sharding/split_large_key.js
- SERVER-14249[1146] Add tests for querying oplog via mongodump using –dbpath
- SERVER-13726[1147] indexbg_drop.js

2.6.5 – Changes

Security

- SERVER-15465[1148] OpenSSL crashes on stepdown
- SERVER-15360[1149] User document changes made on a 2.4 primary and replicated to a 2.6 secondary don't make the 2.6 secondary invalidate its user cache

[1134] https://jira.mongodb.org/browse/SERVER-14184
[1135] https://jira.mongodb.org/browse/SERVER-14015
[1136] https://jira.mongodb.org/browse/SERVER-16017
[1137] https://jira.mongodb.org/browse/SERVER-15598
[1138] https://jira.mongodb.org/browse/SERVER-13595
[1139] https://jira.mongodb.org/browse/SERVER-13471
[1140] https://jira.mongodb.org/browse/SERVER-16324
[1141] https://jira.mongodb.org/browse/SERVER-10069
[1142] https://jira.mongodb.org/browse/SERVER-15632
[1143] https://jira.mongodb.org/browse/SERVER-15629
[1144] https://jira.mongodb.org/browse/SERVER-16316
[1145] https://jira.mongodb.org/browse/SERVER-14763
[1146] https://jira.mongodb.org/browse/SERVER-14249
[1147] https://jira.mongodb.org/browse/SERVER-13726
[1148] https://jira.mongodb.org/browse/SERVER-15465
[1149] https://jira.mongodb.org/browse/SERVER-15360

- SERVER-14887[1150] Allow user document changes made on a 2.4 primary to replicate to a 2.6 secondary
- SERVER-14727[1151] Details of SASL failures aren't logged
- SERVER-12551[1152] Audit DML/CRUD operations

Stability SERVER-9032[1153] mongod fails when launched with misconfigured locale

Querying

- SERVER-15287[1154] Query planner sort analysis incorrectly allows index key pattern plugin fields to provide sort
- SERVER-15286[1155] Assertion in date indexes when opposite-direction-sorted and double "or" filtered
- SERVER-15279[1156] Disable hash-based index intersection (AND_HASH) by default
- SERVER-15152[1157] When evaluating plans, some index candidates cause complete index scan
- SERVER-15015[1158] Assertion failure when combining `$max` and `$min` and reverse index scan
- SERVER-15012[1159] Server crashes on indexed rooted $or queries using a 2d index
- SERVER-14969[1160] Dropping index during active aggregation operation can crash server
- SERVER-14961[1161] Plan ranker favors intersection plans if predicate generates empty range index scan
- SERVER-14892[1162] Invalid `{$elemMatch: {$where}}` query causes memory leak
- SERVER-14706[1163] Queries that use negated $type predicate over a field may return incomplete results when an index is present on that field
- SERVER-13104[1164] Plan enumerator doesn't enumerate all possibilities for a nested `$or`
- SERVER-14984[1165] Server aborts when running `$centerSphere` query with `NaN` radius
- SERVER-14981[1166] Server aborts when querying against `2dsphere` index with `coarsestIndexedLevel:0`
- SERVER-14831[1167] Text search trips assertion when default language only supported in `textIndexVersion=1` used

Replication

- SERVER-15038[1168] Multiple background index builds may not interrupt cleanly for commands, on secondaries

[1150] https://jira.mongodb.org/browse/SERVER-14887
[1151] https://jira.mongodb.org/browse/SERVER-14727
[1152] https://jira.mongodb.org/browse/SERVER-12551
[1153] https://jira.mongodb.org/browse/SERVER-9032
[1154] https://jira.mongodb.org/browse/SERVER-15287
[1155] https://jira.mongodb.org/browse/SERVER-15286
[1156] https://jira.mongodb.org/browse/SERVER-15279
[1157] https://jira.mongodb.org/browse/SERVER-15152
[1158] https://jira.mongodb.org/browse/SERVER-15015
[1159] https://jira.mongodb.org/browse/SERVER-15012
[1160] https://jira.mongodb.org/browse/SERVER-14969
[1161] https://jira.mongodb.org/browse/SERVER-14961
[1162] https://jira.mongodb.org/browse/SERVER-14892
[1163] https://jira.mongodb.org/browse/SERVER-14706
[1164] https://jira.mongodb.org/browse/SERVER-13104
[1165] https://jira.mongodb.org/browse/SERVER-14984
[1166] https://jira.mongodb.org/browse/SERVER-14981
[1167] https://jira.mongodb.org/browse/SERVER-14831
[1168] https://jira.mongodb.org/browse/SERVER-15038

- SERVER-14887[1169] Allow user document changes made on a 2.4 primary to replicate to a 2.6 secondary
- SERVER-14805[1170] Use multithreaded oplog replay during initial sync

Sharding

- SERVER-15056[1171] Sharded connection cleanup on setup error can crash mongos
- SERVER-13702[1172] Commands without optional query may target to wrong shards on mongos
- SERVER-15156[1173] MongoDB upgrade 2.4 to 2.6 check returns error in `config.changelog collection`

Storage

- SERVER-15369[1174] explicitly zero .ns files on creation
- SERVER-15319[1175] Verify 2.8 freelist is upgrade-downgrade safe with 2.6
- SERVER-15111[1176] partially written journal last section causes recovery to fail

Indexing

- SERVER-14848[1177] Port `index_id_desc.js` to v2.6 and master branches
- SERVER-14205[1178] ensureIndex failure reports `ok: 1` on some failures

Write Operations

- SERVER-15106[1179] Incorrect nscanned and nscannedObjects for idhack updates in 2.6.4 profiler or slow query log
- SERVER-15029[1180] The `$rename` modifier uses incorrect dotted source path
- SERVER-14829[1181] `UpdateIndexData::clear()` should reset all member variables

Data Aggregation

- SERVER-15087[1182] Server crashes when running concurrent mapReduce and dropDatabase commands
- SERVER-14969[1183] Dropping index during active aggregation operation can crash server
- SERVER-14168[1184] Warning logged when incremental MR collections are unsuccessfully dropped on secondaries

[1169] https://jira.mongodb.org/browse/SERVER-14887
[1170] https://jira.mongodb.org/browse/SERVER-14805
[1171] https://jira.mongodb.org/browse/SERVER-15056
[1172] https://jira.mongodb.org/browse/SERVER-13702
[1173] https://jira.mongodb.org/browse/SERVER-15156
[1174] https://jira.mongodb.org/browse/SERVER-15369
[1175] https://jira.mongodb.org/browse/SERVER-15319
[1176] https://jira.mongodb.org/browse/SERVER-15111
[1177] https://jira.mongodb.org/browse/SERVER-14848
[1178] https://jira.mongodb.org/browse/SERVER-14205
[1179] https://jira.mongodb.org/browse/SERVER-15106
[1180] https://jira.mongodb.org/browse/SERVER-15029
[1181] https://jira.mongodb.org/browse/SERVER-14829
[1182] https://jira.mongodb.org/browse/SERVER-15087
[1183] https://jira.mongodb.org/browse/SERVER-14969
[1184] https://jira.mongodb.org/browse/SERVER-14168

Packaging

- SERVER-14679[1185] (CentOS 7/RHEL 7) `init.d` script should create directory for `pid` file if it is missing
- SERVER-14023[1186] Support for RHEL 7 Enterprise `.rpm` packages
- SERVER-13243[1187] Support for Ubuntu 14 "Trusty" Enterprise `.deb` packages
- SERVER-11077[1188] Support for Debian 7 Enterprise `.deb` packages
- SERVER-10642[1189] Generate Community and Enterprise packages for SUSE 11

Logging and Diagnostics

- SERVER-14964[1190] nscanned not written to the logs at `logLevel` 1 unless `slowms` exceeded or profiling enabled
- SERVER-12551[1191] Audit DML/CRUD operations
- SERVER-14904[1192] Adjust dates in `tool/exportimport_date.js` to account for different timezones

Internal Code and Testing

- SERVER-13770[1193] `Helpers::removeRange` should check all runner states
- SERVER-14284[1194] jstests should not leave profiler enabled at test run end
- SERVER-14076[1195] remove test `replset_remove_node.js`
- SERVER-14778[1196] Hide function and data pointers for natively-injected v8 functions

2.6.4 – Changes

Security

- SERVER-14701[1197] The "backup" auth role should allow running the "collstats" command for all resources
- SERVER-14518[1198] Allow disabling hostname validation for SSL
- SERVER-14268[1199] Potential information leak
- SERVER-14170[1200] Cannot read from secondary if both audit and auth are enabled in a sharded cluster
- SERVER-13833[1201] userAdminAnyDatabase role should be able to create indexes on admin.system.users and admin.system.roles

[1185] https://jira.mongodb.org/browse/SERVER-14679
[1186] https://jira.mongodb.org/browse/SERVER-14023
[1187] https://jira.mongodb.org/browse/SERVER-13243
[1188] https://jira.mongodb.org/browse/SERVER-11077
[1189] https://jira.mongodb.org/browse/SERVER-10642
[1190] https://jira.mongodb.org/browse/SERVER-14964
[1191] https://jira.mongodb.org/browse/SERVER-12551
[1192] https://jira.mongodb.org/browse/SERVER-14904
[1193] https://jira.mongodb.org/browse/SERVER-13770
[1194] https://jira.mongodb.org/browse/SERVER-14284
[1195] https://jira.mongodb.org/browse/SERVER-14076
[1196] https://jira.mongodb.org/browse/SERVER-14778
[1197] https://jira.mongodb.org/browse/SERVER-14701
[1198] https://jira.mongodb.org/browse/SERVER-14518
[1199] https://jira.mongodb.org/browse/SERVER-14268
[1200] https://jira.mongodb.org/browse/SERVER-14170
[1201] https://jira.mongodb.org/browse/SERVER-13833

- SERVER-12512[1202] Add role-based, selective audit logging.
- SERVER-9482[1203] Add build flag for sslFIPSMode

Querying

- SERVER-14625[1204] Query planner can construct incorrect bounds for negations inside $elemMatch
- SERVER-14607[1205] hash intersection of fetched and non-fetched data can discard data from a result
- SERVER-14532[1206] Improve logging in the case of plan ranker ties
- SERVER-14350[1207] Server crash when $centerSphere has non-positive radius
- SERVER-14317[1208] Dead code in IDHackRunner::applyProjection
- SERVER-14311[1209] skipping of index keys is not accounted for in plan ranking by the index scan stage
- SERVER-14123[1210] some operations can create BSON object larger than the 16MB limit
- SERVER-14034[1211] Sorted $in query with large number of elements can't use merge sort
- SERVER-13994[1212] do not aggressively pre-fetch data for parallelCollectionScan

Replication

- SERVER-14665[1213] Build failure for v2.6 in closeall.js caused by access violation reading _me
- SERVER-14505[1214] cannot dropAllIndexes when index builds in progress assertion failure
- SERVER-14494[1215] Dropping collection during active background index build on secondary triggers segfault
- SERVER-13822[1216] Running resync before replset config is loaded can crash `mongod`
- SERVER-11776[1217] Replication 'isself' check should allow mapped ports

Sharding

- SERVER-14551[1218] Runner yield during migration cleanup (removeRange) results in fassert
- SERVER-14431[1219] Invalid chunk data after splitting on a key that's too large
- SERVER-14261[1220] stepdown during migration range delete can abort `mongod`
- SERVER-14032[1221] v2.6 `mongos` doesn't verify _id is present for config server upserts

[1202] https://jira.mongodb.org/browse/SERVER-12512
[1203] https://jira.mongodb.org/browse/SERVER-9482
[1204] https://jira.mongodb.org/browse/SERVER-14625
[1205] https://jira.mongodb.org/browse/SERVER-14607
[1206] https://jira.mongodb.org/browse/SERVER-14532
[1207] https://jira.mongodb.org/browse/SERVER-14350
[1208] https://jira.mongodb.org/browse/SERVER-14317
[1209] https://jira.mongodb.org/browse/SERVER-14311
[1210] https://jira.mongodb.org/browse/SERVER-14123
[1211] https://jira.mongodb.org/browse/SERVER-14034
[1212] https://jira.mongodb.org/browse/SERVER-13994
[1213] https://jira.mongodb.org/browse/SERVER-14665
[1214] https://jira.mongodb.org/browse/SERVER-14505
[1215] https://jira.mongodb.org/browse/SERVER-14494
[1216] https://jira.mongodb.org/browse/SERVER-13822
[1217] https://jira.mongodb.org/browse/SERVER-11776
[1218] https://jira.mongodb.org/browse/SERVER-14551
[1219] https://jira.mongodb.org/browse/SERVER-14431
[1220] https://jira.mongodb.org/browse/SERVER-14261
[1221] https://jira.mongodb.org/browse/SERVER-14032

- SERVER-13648[1222] better stats from migration cleanup
- SERVER-12750[1223] `mongos` shouldn't accept initial query with "exhaust" flag set
- SERVER-9788[1224] `mongos` does not re-evaluate read preference once a valid replica set member is chosen
- SERVER-9526[1225] Log messages regarding chunks not very informative when the shard key is of type BinData

Storage

- SERVER-14198[1226] Std::set<pointer> and Windows Heap Allocation Reuse produces non-deterministic results
- SERVER-13975[1227] Creating index on collection named "system" can cause server to abort
- SERVER-13729[1228] Reads & Writes are blocked during data file allocation on Windows
- SERVER-13681[1229] `mongod` B stalls during background flush on Windows

Indexing SERVER-14494[1230] Dropping collection during active background index build on secondary triggers segfault

Write Ops

- SERVER-14257[1231] "remove" command can cause process termination by throwing unhandled exception if profiling is enabled
- SERVER-14024[1232] Update fails when query contains part of a DBRef and results in an insert (upsert:true)
- SERVER-13764[1233] debug mechanisms report incorrect nscanned / nscannedObjects for updates

Networking SERVER-13734[1234] Remove catch (...) from handleIncomingMsg

Geo

- SERVER-14039[1235] $nearSphere query with 2d index, skip, and limit returns incomplete results
- SERVER-13701[1236] Query using 2d index throws exception when using explain()

Text Search

- SERVER-14738[1237] Updates to documents with text-indexed fields may lead to incorrect entries
- SERVER-14027[1238] Renaming collection within same database fails if wildcard text index present

[1222] https://jira.mongodb.org/browse/SERVER-13648
[1223] https://jira.mongodb.org/browse/SERVER-12750
[1224] https://jira.mongodb.org/browse/SERVER-9788
[1225] https://jira.mongodb.org/browse/SERVER-9526
[1226] https://jira.mongodb.org/browse/SERVER-14198
[1227] https://jira.mongodb.org/browse/SERVER-13975
[1228] https://jira.mongodb.org/browse/SERVER-13729
[1229] https://jira.mongodb.org/browse/SERVER-13681
[1230] https://jira.mongodb.org/browse/SERVER-14494
[1231] https://jira.mongodb.org/browse/SERVER-14257
[1232] https://jira.mongodb.org/browse/SERVER-14024
[1233] https://jira.mongodb.org/browse/SERVER-13764
[1234] https://jira.mongodb.org/browse/SERVER-13734
[1235] https://jira.mongodb.org/browse/SERVER-14039
[1236] https://jira.mongodb.org/browse/SERVER-13701
[1237] https://jira.mongodb.org/browse/SERVER-14738
[1238] https://jira.mongodb.org/browse/SERVER-14027

Tools

- SERVER-14212[1239] `mongorestore` may drop system users and roles
- SERVER-14048[1240] `mongodump` against `mongos` can't send dump to standard output

Admin

- SERVER-14556[1241] Default dbpath for `mongod --configsvr` changes in 2.6
- SERVER-14355[1242] Allow dbAdmin role to manually create system.profile collections

Packaging SERVER-14283[1243] Parameters in installed config file are out of date

JavaScript

- SERVER-14254[1244] Do not store native function pointer as a property in function prototype
- SERVER-13798[1245] v8 garbage collection can cause crash due to independent lifetime of DBClient and Cursor objects
- SERVER-13707[1246] mongo shell may crash when converting invalid regular expression

Shell

- SERVER-14341[1247] negative opcounter values in serverStatus
- SERVER-14107[1248] Querying for a document containing a value of either type Javascript or JavascriptWithScope crashes the shell

Usability SERVER-13833[1249] userAdminAnyDatabase role should be able to create indexes on admin.system.users and admin.system.roles

Logging and Diagnostics

- SERVER-12512[1250] Add role-based, selective audit logging.
- SERVER-14341[1251] negative opcounter values in serverStatus

[1239] https://jira.mongodb.org/browse/SERVER-14212
[1240] https://jira.mongodb.org/browse/SERVER-14048
[1241] https://jira.mongodb.org/browse/SERVER-14556
[1242] https://jira.mongodb.org/browse/SERVER-14355
[1243] https://jira.mongodb.org/browse/SERVER-14283
[1244] https://jira.mongodb.org/browse/SERVER-14254
[1245] https://jira.mongodb.org/browse/SERVER-13798
[1246] https://jira.mongodb.org/browse/SERVER-13707
[1247] https://jira.mongodb.org/browse/SERVER-14341
[1248] https://jira.mongodb.org/browse/SERVER-14107
[1249] https://jira.mongodb.org/browse/SERVER-13833
[1250] https://jira.mongodb.org/browse/SERVER-12512
[1251] https://jira.mongodb.org/browse/SERVER-14341

Testing

- SERVER-14731[1252] plan_cache_ties.js sometimes fails
- SERVER-14147[1253] make index_multi.js retry on connection failure
- SERVER-13615[1254] sharding_rs2.js intermittent failure due to reliance on opcounters

2.6.3 – Changes

- SERVER-14302[1255] Fixed: "Equality queries on `_id` with projection may return no results on sharded collections"
- SERVER-14304[1256] Fixed: "Equality queries on `_id` with projection on `_id` may return orphan documents on sharded collections"

2.6.2 – Changes

Security

- SERVER-13727[1257] The `backup` (page 491) authorization role now includes privileges to run the `collStats` command.
- SERVER-13804[1258] The built-in role `restore` (page 492) now has privileges on `system.roles` collection.
- SERVER-13612[1259] Fixed: "SSL-enabled server appears not to be sending the list of supported certificate issuers to the client"
- SERVER-13753[1260] Fixed: "`mongod` may terminate if x.509 authentication certificate is invalid"
- SERVER-13945[1261] For *replica set/sharded cluster member authentication* (page 430), now matches x.509 cluster certificates by attributes instead of by substring comparison.
- SERVER-13868[1262] Now marks V1 users as probed on databases that do not have surrogate user documents.
- SERVER-13850[1263] Now ensures that the user cache entry is up to date before using it to determine a user's roles in user management commands on `mongos`.
- SERVER-13588[1264] Fixed: "Shell prints startup warning when auth enabled"

Querying

- SERVER-13731[1265] Fixed: "Stack overflow when parsing deeply nested `$not` query"
- SERVER-13890[1266] Fixed: "Index bounds builder constructs invalid bounds for multiple negations joined by an `$or`"

[1252] https://jira.mongodb.org/browse/SERVER-14731
[1253] https://jira.mongodb.org/browse/SERVER-14147
[1254] https://jira.mongodb.org/browse/SERVER-13615
[1255] https://jira.mongodb.org/browse/SERVER-14302
[1256] https://jira.mongodb.org/browse/SERVER-14304
[1257] https://jira.mongodb.org/browse/SERVER-13727
[1258] https://jira.mongodb.org/browse/SERVER-13804
[1259] https://jira.mongodb.org/browse/SERVER-13612
[1260] https://jira.mongodb.org/browse/SERVER-13753
[1261] https://jira.mongodb.org/browse/SERVER-13945
[1262] https://jira.mongodb.org/browse/SERVER-13868
[1263] https://jira.mongodb.org/browse/SERVER-13850
[1264] https://jira.mongodb.org/browse/SERVER-13588
[1265] https://jira.mongodb.org/browse/SERVER-13731
[1266] https://jira.mongodb.org/browse/SERVER-13890

- SERVER-13752[1267] Verified assertion on empty `$in` clause and sort on second field in a compound index.
- SERVER-13337[1268] Re-enabled `idhack` for queries with projection.
- SERVER-13715[1269] Fixed: "Aggregation pipeline execution can fail with $or and blocking sorts"
- SERVER-13714[1270] Fixed: "non-top-level indexable `$not` triggers query planning bug"
- SERVER-13769[1271] Fixed: "`distinct` command on indexed field with geo predicate fails to execute"
- SERVER-13675[1272] Fixed "Plans with differing performance can tie during plan ranking"
- SERVER-13899[1273] Fixed: "'Whole index scan' query solutions can use incompatible indexes, return incorrect results"
- SERVER-13852[1274] Fixed "IndexBounds::endKeyInclusive not initialized by constructor"
- SERVER-14073[1275] planSummary no longer truncated at 255 characters
- SERVER-14174[1276] Fixed: "If ntoreturn is a limit (rather than batch size) extra data gets buffered during plan ranking"
- SERVER-13789[1277] Some nested queries no longer trigger an assertion error
- SERVER-14064[1278] Added planSummary information for `count` command log message.
- SERVER-13960[1279] Queries containing `$or` no longer miss results if multiple clauses use the same index.
- SERVER-14180[1280] Fixed: "Crash with 'and' clause, `$elemMatch`, and nested `$mod` or regex"
- SERVER-14176[1281] Natural order sort specification no longer ignored if query is specified.
- SERVER-13754[1282] Bounds no longer combined for `$or` queries that can use merge sort.

Geospatial SERVER-13687[1283] Results of `$near` query on compound multi-key 2dsphere index are now sorted by distance.

Write Operations SERVER-13802[1284] Insert field validation no longer stops at first `Timestamp()` field.

Replication

- SERVER-13993[1285] Fixed: "log a message when `shouldChangeSyncTarget()` believes a node should change sync targets"

[1267] https://jira.mongodb.org/browse/SERVER-13752
[1268] https://jira.mongodb.org/browse/SERVER-13337
[1269] https://jira.mongodb.org/browse/SERVER-13715
[1270] https://jira.mongodb.org/browse/SERVER-13714
[1271] https://jira.mongodb.org/browse/SERVER-13769
[1272] https://jira.mongodb.org/browse/SERVER-13675
[1273] https://jira.mongodb.org/browse/SERVER-13899
[1274] https://jira.mongodb.org/browse/SERVER-13852
[1275] https://jira.mongodb.org/browse/SERVER-14073
[1276] https://jira.mongodb.org/browse/SERVER-14174
[1277] https://jira.mongodb.org/browse/SERVER-13789
[1278] https://jira.mongodb.org/browse/SERVER-14064
[1279] https://jira.mongodb.org/browse/SERVER-13960
[1280] https://jira.mongodb.org/browse/SERVER-14180
[1281] https://jira.mongodb.org/browse/SERVER-14176
[1282] https://jira.mongodb.org/browse/SERVER-13754
[1283] https://jira.mongodb.org/browse/SERVER-13687
[1284] https://jira.mongodb.org/browse/SERVER-13802
[1285] https://jira.mongodb.org/browse/SERVER-13993

- SERVER-13976[1286] Fixed: "Cloner needs to detect failure to create collection"

Sharding

- SERVER-13616[1287] Resolved: "'type 7' (OID) error when acquiring distributed lock for first time"
- SERVER-13812[1288] Now catches exception thrown by `getShardsForQuery` for geo query.
- SERVER-14138[1289] `mongos` will now correctly target multiple shards for nested field shard key predicates.
- SERVER-11332[1290] Fixed: "Authentication requests delayed if first config server is unresponsive"

Map/Reduce

- SERVER-14186[1291] Resolved: "`rs.stepDown` during mapReduce causes fassert in logOp"
- SERVER-13981[1292] Temporary map/reduce collections are now correctly replicated to secondaries.

Storage

- SERVER-13750[1293] `convertToCapped` on empty collection no longer aborts after `invariant()` failure.
- SERVER-14056[1294] Moving large collection across databases with renameCollection no longer triggers fatal assertion.
- SERVER-14082[1295] Fixed: "Excessive freelist scanning for MaxBucket"
- SERVER-13737[1296] CollectionOptions parser now skips non-numeric for "size"/"max" elements if values non-numeric.

Build and Packaging

- SERVER-13950[1297] MongoDB Enterprise now includes required dependency list.
- SERVER-13862[1298] Support for mongodb-org-server installation 2.6.1-1 on RHEL5 via RPM.
- SERVER-13724[1299] Added SCons flag to override treating all warnings as errors.

Diagnostics

- SERVER-13587[1300] Resolved: "`ndeleted` (page 381) in `system.profile` documents reports 1 too few documents removed"
- SERVER-13368[1301] Improved exposure of timing information in `currentOp`.

[1286] https://jira.mongodb.org/browse/SERVER-13976
[1287] https://jira.mongodb.org/browse/SERVER-13616
[1288] https://jira.mongodb.org/browse/SERVER-13812
[1289] https://jira.mongodb.org/browse/SERVER-14138
[1290] https://jira.mongodb.org/browse/SERVER-11332
[1291] https://jira.mongodb.org/browse/SERVER-14186
[1292] https://jira.mongodb.org/browse/SERVER-13981
[1293] https://jira.mongodb.org/browse/SERVER-13750
[1294] https://jira.mongodb.org/browse/SERVER-14056
[1295] https://jira.mongodb.org/browse/SERVER-14082
[1296] https://jira.mongodb.org/browse/SERVER-13737
[1297] https://jira.mongodb.org/browse/SERVER-13950
[1298] https://jira.mongodb.org/browse/SERVER-13862
[1299] https://jira.mongodb.org/browse/SERVER-13724
[1300] https://jira.mongodb.org/browse/SERVER-13587
[1301] https://jira.mongodb.org/browse/SERVER-13368

Administration SERVER-13954[1302] `security.javascriptEnabled` option is now available in the YAML configuration file.

Tools

- SERVER-10464[1303] `mongodump` can now query `oplog.$main` and `oplog.rs` when using *`--dbpath`*.
- SERVER-13760[1304] `mongoexport` can now handle large timestamps on Windows.

Shell

- SERVER-13865[1305] Shell now returns correct `WriteResult` for compatibility-mode upsert with non-OID equality predicate on `_id` field.
- SERVER-13037[1306] Fixed typo in error message for "compatibility mode".

Internal Code

- SERVER-13794[1307] Fixed: "Unused snapshot history consuming significant heap space"
- SERVER-13446[1308] Removed Solaris builds dependency on ILLUMOS libc.
- SERVER-14092[1309] MongoDB upgrade 2.4 to 2.6 check no longer returns an error in internal collections.
- SERVER-14000[1310] Added new lsb file location for Debian 7.1

Testing

- SERVER-13723[1311] Stabilized `tags.js` after a change in its timeout when it was ported to use write commands.
- SERVER-13494[1312] Fixed: "`setup_multiversion_mongodb.py` doesn't download 2.4.10 because of non-numeric version sorting"
- SERVER-13603[1313] Fixed: "Test suites with options tests fail when run with *`--nopreallocj`*"
- SERVER-13948[1314] Fixed: "`awaitReplication()` failures related to getting a config version from master causing test failures"
- SERVER-13839[1315] Fixed `sync2.js` failure.
- SERVER-13972[1316] Fixed `connections_opened.js` failure.
- SERVER-13712[1317] Reduced peak disk usage of test suites.

[1302] https://jira.mongodb.org/browse/SERVER-13954
[1303] https://jira.mongodb.org/browse/SERVER-10464
[1304] https://jira.mongodb.org/browse/SERVER-13760
[1305] https://jira.mongodb.org/browse/SERVER-13865
[1306] https://jira.mongodb.org/browse/SERVER-13037
[1307] https://jira.mongodb.org/browse/SERVER-13794
[1308] https://jira.mongodb.org/browse/SERVER-13446
[1309] https://jira.mongodb.org/browse/SERVER-14092
[1310] https://jira.mongodb.org/browse/SERVER-14000
[1311] https://jira.mongodb.org/browse/SERVER-13723
[1312] https://jira.mongodb.org/browse/SERVER-13494
[1313] https://jira.mongodb.org/browse/SERVER-13603
[1314] https://jira.mongodb.org/browse/SERVER-13948
[1315] https://jira.mongodb.org/browse/SERVER-13839
[1316] https://jira.mongodb.org/browse/SERVER-13972
[1317] https://jira.mongodb.org/browse/SERVER-13712

- SERVER-14249[1318] Added tests for querying oplog via `mongodump` using `--dbpath`
- SERVER-10462[1319] Fixed: "Windows file locking related buildbot failures"

2.6.1 – Changes

Stability SERVER-13739[1320] Repair database failure can delete database files

Build and Packaging

- SERVER-13287[1321] Addition of debug symbols has doubled compile time
- SERVER-13563[1322] Upgrading from 2.4.x to 2.6.0 via `yum` clobbers configuration file
- SERVER-13691[1323] yum and apt "stable" repositories contain release candidate 2.6.1-rc0 packages
- SERVER-13515[1324] Cannot install MongoDB as a service on Windows

Querying

- SERVER-13066[1325] Negations over multikey fields do not use index
- SERVER-13495[1326] Concurrent `GETMORE` and `KILLCURSORS` operations can cause race condition and server crash
- SERVER-13503[1327] The `$where` operator should not be allowed under `$elemMatch`
- SERVER-13537[1328] Large skip and and limit values can cause crash in blocking sort stage
- SERVER-13557[1329] Incorrect negation of $elemMatch value in 2.6
- SERVER-13562[1330] Queries that use tailable cursors do not stream results if skip() is applied
- SERVER-13566[1331] Using the OplogReplay flag with extra predicates can yield incorrect results
- SERVER-13611[1332] Missing sort order for compound index leads to unnecessary in-memory sort
- SERVER-13618[1333] Optimization for sorted $in queries not applied to reverse sort order
- SERVER-13661[1334] Increase the maximum allowed depth of query objects
- SERVER-13664[1335] Query with `$elemMatch` using a compound multikey index can generate incorrect results
- SERVER-13677[1336] Query planner should traverse through $all while handling $elemMatch object predicates

[1318] https://jira.mongodb.org/browse/SERVER-14249
[1319] https://jira.mongodb.org/browse/SERVER-10462
[1320] https://jira.mongodb.org/browse/SERVER-13739
[1321] https://jira.mongodb.org/browse/SERVER-13287
[1322] https://jira.mongodb.org/browse/SERVER-13563
[1323] https://jira.mongodb.org/browse/SERVER-13691
[1324] https://jira.mongodb.org/browse/SERVER-13515
[1325] https://jira.mongodb.org/browse/SERVER-13066
[1326] https://jira.mongodb.org/browse/SERVER-13495
[1327] https://jira.mongodb.org/browse/SERVER-13503
[1328] https://jira.mongodb.org/browse/SERVER-13537
[1329] https://jira.mongodb.org/browse/SERVER-13557
[1330] https://jira.mongodb.org/browse/SERVER-13562
[1331] https://jira.mongodb.org/browse/SERVER-13566
[1332] https://jira.mongodb.org/browse/SERVER-13611
[1333] https://jira.mongodb.org/browse/SERVER-13618
[1334] https://jira.mongodb.org/browse/SERVER-13661
[1335] https://jira.mongodb.org/browse/SERVER-13664
[1336] https://jira.mongodb.org/browse/SERVER-13677

- SERVER-13766[1337] Dropping index or collection while $or query is yielding triggers fatal assertion

Geospatial

- SERVER-13666[1338] `$near` queries with out-of-bounds points in legacy format can lead to crashes
- SERVER-13540[1339] The `geoNear` command no longer returns distance in radians for legacy point
- SERVER-13486[1340]: The `geoNear` command can create too large BSON objects for aggregation.

Replication

- SERVER-13500[1341] Changing replica set configuration can crash running members
- SERVER-13589[1342] Background index builds from a 2.6.0 primary fail to complete on 2.4.x secondaries
- SERVER-13620[1343] Replicated data definition commands will fail on secondaries during background index build
- SERVER-13496[1344] Creating index with same name but different spec in mixed version replicaset can abort replication

Sharding

- SERVER-12638[1345] Initial sharding with hashed shard key can result in duplicate split points
- SERVER-13518[1346] The `_id` field is no longer automatically generated by `mongos` when missing
- SERVER-13777[1347] Migrated ranges waiting for deletion do not report cursors still open

Security

- SERVER-9358[1348] Log rotation can overwrite previous log files
- SERVER-13644[1349] Sensitive credentials in startup options are not redacted and may be exposed
- SERVER-13441[1350] Inconsistent error handling in user management shell helpers

Write Operations

- SERVER-13466[1351] Error message in collection creation failure contains incorrect namespace
- SERVER-13499[1352] Yield policy for batch-inserts should be the same as for batch-updates/deletes
- SERVER-13516[1353] Array updates on documents with more than 128 BSON elements may crash `mongod`

[1337] https://jira.mongodb.org/browse/SERVER-13766
[1338] https://jira.mongodb.org/browse/SERVER-13666
[1339] https://jira.mongodb.org/browse/SERVER-13540
[1340] https://jira.mongodb.org/browse/SERVER-13486
[1341] https://jira.mongodb.org/browse/SERVER-13500
[1342] https://jira.mongodb.org/browse/SERVER-13589
[1343] https://jira.mongodb.org/browse/SERVER-13620
[1344] https://jira.mongodb.org/browse/SERVER-13496
[1345] https://jira.mongodb.org/browse/SERVER-12638
[1346] https://jira.mongodb.org/browse/SERVER-13518
[1347] https://jira.mongodb.org/browse/SERVER-13777
[1348] https://jira.mongodb.org/browse/SERVER-9358
[1349] https://jira.mongodb.org/browse/SERVER-13644
[1350] https://jira.mongodb.org/browse/SERVER-13441
[1351] https://jira.mongodb.org/browse/SERVER-13466
[1352] https://jira.mongodb.org/browse/SERVER-13499
[1353] https://jira.mongodb.org/browse/SERVER-13516

2.6.12 – Mar 24, 2016

- Fixed issue with MMAPv1 journaling where the "last sequence number" file (`lsn` file) may be ahead of what is synced to the data files: SERVER-22261[1354].
- Fixed issue that permitted the creation of new role with the same name as a *built-in role* (page 485): SERVER-19284[1355].
- Fixed issue where some index operations running during an active migration may cause the migration operation to skip some documents: SERVER-22535[1356].
- All issues closed in 2.6.12[1357]

2.6.11 – Aug 12, 2015

- Improvements to query plan ranking SERVER-17815[1358]
- Improved ability for `mongos` to detect replica set failover and correctly route read operations to the new primary SERVER-18280[1359]
- Improved reporting of queries in `getMore` operation in `db.currentOp()` and the database profiler SERVER-16265[1360]
- All issues closed in 2.6.11[1361]

2.6.10 – May 19, 2015

- Improve user cache invalidation enforcement on `mongos` SERVER-11980[1362]
- Provide correct rollbacks for collection creation SERVER-18211[1363]
- Allow user inserts into the `system.profile` collection SERVER-18211[1364]
- Fix to query system to ensure non-negation predicates get chosen over negation predicates for multikey index bounds construction SERVER-18364[1365]
- All issues closed in 2.6.10[1366]

2.6.9 – March 24, 2015

- Resolve connection handling related crash with `mongos` instances SERVER-17441[1367]
- Add server parameter to configure idle cursor timeout SERVER-8188[1368]

[1354] https://jira.mongodb.org/browse/SERVER-22261
[1355] https://jira.mongodb.org/browse/SERVER-19284
[1356] https://jira.mongodb.org/browse/SERVER-22535
[1357] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.12%22%20AND%20project%20%3D%20SERVER
[1358] https://jira.mongodb.org/browse/SERVER-17815
[1359] https://jira.mongodb.org/browse/SERVER-18280
[1360] https://jira.mongodb.org/browse/SERVER-16265
[1361] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.11%22%20AND%20project%20%3D%20SERVER
[1362] https://jira.mongodb.org/browse/SERVER-11980
[1363] https://jira.mongodb.org/browse/SERVER-18211
[1364] https://jira.mongodb.org/browse/SERVER-18211
[1365] https://jira.mongodb.org/browse/SERVER-18364
[1366] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.10%22%20AND%20project%20%3D%20SERVER
[1367] https://jira.mongodb.org/browse/SERVER-17441
[1368] https://jira.mongodb.org/browse/SERVER-8188

- Remove duplicated (orphan) documents from aggregation pipelines with `_id` queries in sharded clusters SERVER-17426[1369]
- Fixed crash in `geoNear` queries with multiple `2dsphere` indexes SERVER-14723[1370]
- All issues closed in 2.6.9[1371]

2.6.8 – February 25, 2015

- Add `listCollections` command functionality to 2.6 shell and client SERVER-17087[1372]
- `copydb`/`clone` commands can crash the server if a primary steps down SERVER-16599[1373]
- Secondary fasserts trying to replicate an index SERVER-16274[1374]
- Query optimizer should always use equality predicate over unique index when possible SERVER-15802[1375]
- All issues closed in 2.6.8[1376]

2.6.7 – January 13, 2015

- Decreased `mongos` memory footprint when shards have several tags SERVER-16683[1377]
- Removed check for shard version if the primary server is down SERVER-16237[1378]
- Fixed: `/etc/init.d/mongod` startup script failure with dirname message SERVER-16081[1379]
- Fixed: `mongos` can cause shards to hit the in-memory sort limit by requesting more results than needed SERVER-14306[1380]
- All issues closed in 2.6.7[1381]

2.6.6 – December 09, 2014

- Fixed: Evaluating candidate query plans with concurrent writes on same collection may crash `mongod` SERVER-15580[1382]
- Fixed: 2.6 `mongod` crashes with segfault when added to a 2.8 replica set with 12 or more members SERVER-16107[1383]
- Fixed: `$regex`, `$in` and `$sort` with index returns too many results SERVER-15696[1384]

[1369] https://jira.mongodb.org/browse/SERVER-17426
[1370] https://jira.mongodb.org/browse/SERVER-14723
[1371] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.9%22%20AND%20project%20%3D%20SERVER
[1372] https://jira.mongodb.org/browse/SERVER-17087
[1373] https://jira.mongodb.org/browse/SERVER-16599
[1374] https://jira.mongodb.org/browse/SERVER-16274
[1375] https://jira.mongodb.org/browse/SERVER-15802
[1376] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.8%22%20AND%20project%20%3D%20SERVER
[1377] https://jira.mongodb.org/browse/SERVER-16683
[1378] https://jira.mongodb.org/browse/SERVER-16237
[1379] https://jira.mongodb.org/browse/SERVER-16081
[1380] https://jira.mongodb.org/browse/SERVER-14306
[1381] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.7%22%20AND%20project%20%3D%20SERVER
[1382] https://jira.mongodb.org/browse/SERVER-15580
[1383] https://jira.mongodb.org/browse/SERVER-16107
[1384] https://jira.mongodb.org/browse/SERVER-15696

- Change: `moveChunk` will fail if there is data on the target shard and a required index does not exist. SERVER-12472[1385]
- Primary should abort if encountered problems writing to the oplog SERVER-12058[1386]
- All issues closed in 2.6.6[1387]

2.6.5 – October 07, 2014

- `$rename` now uses correct dotted source paths SERVER-15029[1388]
- Partially written journal last section does not affect recovery SERVER-15111[1389]
- Explicitly zero `.ns` files on creation SERVER-15369[1390]
- Plan ranker will no longer favor intersection plans if predicate generates empty range index scan SERVER-14961[1391]
- Generate Community and Enterprise packages for SUSE 11 SERVER-10642[1392]
- All issues closed in 2.6.5[1393]

2.6.4 – August 11, 2014

- Fix for `text` index where under specific circumstances, in-place updates to a `text`-indexed field may result in incorrect/incomplete results SERVER-14738[1394]
- Check the size of the split point before performing a manual split chunk operation SERVER-14431[1395]
- Ensure read preferences are re-evaluated by drawing secondary connections from a global pool and releasing back to the pool at the end of a query/command SERVER-9788[1396]
- Allow read from secondaries when both audit and authorization are enabled in a sharded cluster SERVER-14170[1397]
- All issues closed in 2.6.4[1398]

2.6.3 – June 19, 2014

- Equality queries on `_id` with projection may return no results on sharded collections SERVER-14302[1399].
- Equality queries on `_id` with projection on `_id` may return orphan documents on sharded collections SERVER-14304[1400].

[1385] https://jira.mongodb.org/browse/SERVER-12472
[1386] https://jira.mongodb.org/browse/SERVER-12058
[1387] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.6%22%20AND%20project%20%3D%20SERVER
[1388] https://jira.mongodb.org/browse/SERVER-15029
[1389] https://jira.mongodb.org/browse/SERVER-15111
[1390] https://jira.mongodb.org/browse/SERVER-15369
[1391] https://jira.mongodb.org/browse/SERVER-14961
[1392] https://jira.mongodb.org/browse/SERVER-10642
[1393] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.5%22%20AND%20project%20%3D%20SERVER
[1394] https://jira.mongodb.org/browse/SERVER-14738
[1395] https://jira.mongodb.org/browse/SERVER-14431
[1396] https://jira.mongodb.org/browse/SERVER-9788
[1397] https://jira.mongodb.org/browse/SERVER-14170
[1398] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.4%22%20AND%20project%20%3D%20SERVER
[1399] https://jira.mongodb.org/browse/SERVER-14302
[1400] https://jira.mongodb.org/browse/SERVER-14304

- All issues closed in 2.6.3[1401].

2.6.2 – June 16, 2014

- Query plans with differing performance can tie during plan ranking SERVER-13675[1402].
- `mongod` may terminate if x.509 authentication certificate is invalid SERVER-13753[1403].
- Temporary map/reduce collections are incorrectly replicated to secondaries SERVER-13981[1404].
- `mongos` incorrectly targets multiple shards for nested field shard key predicates SERVER-14138[1405].
- `rs.stepDown()` during mapReduce causes `fassert` when writing to op log SERVER-14186[1406].
- All issues closed in 2.6.2[1407].

2.6.1 – May 5, 2014

- Fix to install MongoDB service on Windows with the `--install` option SERVER-13515[1408].
- Allow direct upgrade from 2.4.x to 2.6.0 via `yum` SERVER-13563[1409].
- Fix issues with background index builds on secondaries: SERVER-13589[1410] and SERVER-13620[1411].
- Redact credential information passed as startup options SERVER-13644[1412].
- *2.6.1 Changelog* (page 989).
- All issues closed in 2.6.1[1413].

Major Changes

The following changes in MongoDB affect both the standard and Enterprise editions:

Aggregation Enhancements

The aggregation pipeline adds the ability to return result sets of any size, either by returning a cursor or writing the output to a collection. Additionally, the aggregation pipeline supports variables and adds new operations to handle sets and redact data.

- The `db.collection.aggregate()` now returns a cursor, which enables the aggregation pipeline to return result sets of any size.
- Aggregation pipelines now support an `explain` operation to aid analysis of aggregation operations.
- Aggregation can now use a more efficient external-disk-based sorting process.

[1401] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.3%22%20AND%20project%20%3D%20SERVER
[1402] https://jira.mongodb.org/browse/SERVER-13675
[1403] https://jira.mongodb.org/browse/SERVER-13753
[1404] https://jira.mongodb.org/browse/SERVER-13981
[1405] https://jira.mongodb.org/browse/SERVER-14138
[1406] https://jira.mongodb.org/browse/SERVER-14186
[1407] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.2%22%20AND%20project%20%3D%20SERVER
[1408] https://jira.mongodb.org/browse/SERVER-13515
[1409] https://jira.mongodb.org/browse/SERVER-13563
[1410] https://jira.mongodb.org/browse/SERVER-13589
[1411] https://jira.mongodb.org/browse/SERVER-13620
[1412] https://jira.mongodb.org/browse/SERVER-13644
[1413] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.6.1%22%20AND%20project%20%3D%20SERVER

- New pipeline stages:
  - `$out` stage to output to a collection.
  - `$redact` stage to allow additional control to accessing the data.
- New or modified operators:
  - `set expression operators`.
  - `$let` and `$map` operators to allow for the use of variables.
  - `$literal` operator and `$size` operator.
  - `$cond` expression now accepts either an object or an array.

Text Search Integration

Text search is now enabled by default, and the query system, including the aggregation pipeline `$match` stage, includes the `$text` operator, which resolves text-search queries.

MongoDB 2.6 includes an updated *text index* (page 533) format and deprecates the `text` command.

Insert and Update Improvements

Improvements to the update and insert systems include additional operations and improvements that increase consistency of modified data.

- MongoDB preserves the order of the document fields following write operations *except* for the following cases:
  - The `_id` field is always the first field in the document.
  - Updates that include `renaming` of field names may result in the reordering of fields in the document.
- New or enhanced update operators:
  - `$bit` operator supports bitwise `xor` operation.
  - `$min` and `$max` operators that perform conditional update depending on the relative size of the specified value and the current value of a field.
  - `$push` operator has enhanced support for the `$sort`, `$slice`, and `$each` modifiers and supports a new `$position` modifier.
  - `$currentDate` operator to set the value of a field to the current date.
- The `$mul` operator for multiplicative increments for insert and update operations.

See also:

Update Operator Syntax Validation (page 1005)

New Write Operation Protocol

A new write protocol integrates write operations with write concerns. The protocol also provides improved support for bulk operations.

MongoDB 2.6 adds the write commands `insert`, `update`, and `delete`, which provide the basis for the improved bulk insert. All officially supported MongoDB drivers support the new write commands.

The `mongo` shell now includes methods to perform bulk-write operations. See `Bulk()` for more information.

See also:

Write Method Acknowledgements (page 1002)

MSI Package for MongoDB Available for Windows

MongoDB now distributes MSI packages for Microsoft Windows. This is the recommended method for MongoDB installation under Windows.

Security Improvements

MongoDB 2.6 enhances support for secure deployments through improved SSL support, x.509-based authentication, an improved authorization system with more granular controls, as well as centralized credential storage, and improved user management tools.

Specifically these changes include:

- A new *authorization model* (page 433) that provides the ability to create custom *User-Defined Roles* (page 440) and the ability to specify user privileges at a collection-level granularity.
- Global user management, which stores all user and user-defined role data in the `admin` database and provides a new set of commands for managing users and roles.
- x.509 certificate authentication for *client authentication* (page 403) as well as for *internal authentication* (page 430) of sharded cluster and/or replica set members. x.509 authentication is only available for deployments using SSL.
- Enhanced SSL Support:
  - *Rolling upgrades of clusters* (page 458) to use SSL.
  - *MongoDB Tools* (page 458) support connections to `mongod` and `mongos` instances using SSL connections.
  - *Prompt for passphrase* (page 455) by `mongod` or `mongos` at startup.
  - Require the use of strong SSL ciphers, with a minimum 128-bit key length for all connections. The strong-cipher requirement prevents an old or malicious client from forcing use of a weak cipher.
- MongoDB disables the http interface by default, limiting *network exposure* (page 474). To enable the interface, see `enabled`.

See also:

New Authorization Model (page 1003), *SSL Certificate Hostname Validation* (page 1004), and *Security Checklist* (page 391).

Query Engine Improvements

- MongoDB can now use *index intersection* (page 581) to fulfill queries supported by more than one index.
- *Index Filters* (page 108) to limit which indexes can become the winning plan for a query.
- `https://docs.mongodb.org/manual/reference/method/js-plan-cache` methods to view and clear the *query plans* (page 108) cached by the query optimizer.
- MongoDB can now use `count()` with `hint()`. See `count()` for details.

Improvements

Geospatial Enhancements

- *2dsphere indexes version 2* (page 543).
- Support for *MultiPoint* (page 555), *MultiLineString* (page 556), *MultiPolygon* (page 556), and *GeometryCollection* (page 556).
- Support for geospatial query clauses in `$or` expressions.

See also:

2dsphere Index Version 2 (page 1004), *$maxDistance Changes* (page 1007), *Deprecated $uniqueDocs* (page 1007), *Stronger Validation of Geospatial Queries* (page 1007)

Index Build Enhancements

- *Background index build* (page 577) allowed on secondaries. If you initiate a background index build on a *primary*, the secondaries will replicate the index build in the background.
- Automatic rebuild of interrupted index builds after a restart.
  - If a standalone or a primary instance terminates during an index build *without a clean shutdown*, `mongod` now restarts the index build when the instance restarts. If the instance shuts down cleanly or if a user kills the index build, the interrupted index builds do not automatically restart upon the restart of the server.
  - If a secondary instance terminates during an index build, the `mongod` instance will now restart the interrupted index build when the instance restarts.

  To disable this behavior, use the `--noIndexBuildRetry` command-line option.
- `ensureIndex()` now wraps a new `createIndex` command.
- The `dropDups` option to `ensureIndex()` and `createIndex` is deprecated.

See also:

Enforce Index Key Length Limit (page 1000)

Enhanced Sharding and Replication Administration

- New `cleanupOrphaned` command to remove *orphaned documents* from a shard.
- New `mergeChunks` command to combine contiguous chunks located on a single shard. See `mergeChunks` and *Merge Chunks in a Sharded Cluster* (page 810).
- New `rs.printReplicationInfo()` and `rs.printSlaveReplicationInfo()` methods to provide a formatted report of the status of a replica set from the perspective of the primary and the secondary, respectively.

Configuration Options YAML File Format

MongoDB 2.6 supports a YAML-based configuration file format in addition to the previous configuration file format. See the documentation of the `Configuration File` for more information.

Operational Changes

Storage

`usePowerOf2Sizes` is now the default allocation strategy for all new collections. The new allocation strategy uses more storage relative to total document size but results in lower levels of storage fragmentation and more predictable storage capacity planning over time.

To use the previous *exact-fit allocation strategy*:

- For a specific collection, use `collMod` with `usePowerOf2Sizes` set to `false`.
- For all new collections on an entire `mongod` instance, set `newCollectionsUsePowerOf2Sizes` to `false`.

  New collections include those: created during *initial sync* (page 658), as well as those created by the `mongorestore` and `mongoimport` tools, by running `mongod` with the `--repair` option, as well as the `restoreDatabase` command.

See /core/storage[1414] for more information about MongoDB's storage system.

Networking

- Removed upward limit for the `maxIncomingConnections` for `mongod` and `mongos`. Previous versions capped the maximum possible `maxIncomingConnections` setting at `20,000` connections.
- Connection pools for a `mongos` instance may be used by multiple MongoDB servers. This can reduce the number of connections needed for high-volume workloads and reduce resource consumption in sharded clusters.
- The C++ driver now monitors *replica set* health with the `isMaster` command instead of `replSetGetStatus`. This allows the C++ driver to support systems that require authentication.
- New `cursor.maxTimeMS()` and corresponding `maxTimeMS` option for commands to specify a time limit.

Tool Improvements

- `mongo` shell supports a global */etc/mongorc.js*.
- All MongoDB `executable files` now support the `--quiet` option to suppress all logging output except for error messages.
- `mongoimport` uses the input filename, without the file extension if any, as the collection name if run without the `-c` or `--collection` specification.
- `mongoexport` can now constrain export data using `--skip` and `--limit`, as well as order the documents in an export using the `--sort` option.
- `mongostat` can support the use of `--rowcount` (`-n`) with the `--discover` option to produce the specified number of output lines.
- Add strict mode representation for `data_numberlong` (page 19) for use by `mongoexport` and `mongoimport`.

MongoDB Enterprise Features

The following changes are specific to MongoDB Enterprise Editions:

[1414] https://docs.mongodb.org/v2.6/core/storage

MongoDB Enterprise for Windows

MongoDB Enterprise for Windows (page 69) is now available. It includes support for Kerberos, SSL, and SNMP.

MongoDB Enterprise for Windows does **not** include LDAP support for authentication. However, MongoDB Enterprise for Linux supports using LDAP authentication with an ActiveDirectory server.

MongoDB Enterprise for Windows includes OpenSSL version 1.0.1g.

Auditing

MongoDB Enterprise adds *auditing* (page 466) capability for `mongod` and `mongos` instances. See *Auditing* (page 466) for details.

LDAP Support for Authentication

MongoDB Enterprise provides support for proxy authentication of users. This allows administrators to configure a MongoDB cluster to authenticate users by proxying authentication requests to a specified Lightweight Directory Access Protocol (LDAP) service. See *Authenticate Using SASL and LDAP with OpenLDAP* (page 420) and *Authenticate Using SASL and LDAP with ActiveDirectory* (page 417) for details.

MongoDB Enterprise for Windows does **not** include LDAP support for authentication. However, MongoDB Enterprise for Linux supports using LDAP authentication with an ActiveDirectory server.

MongoDB does **not** support LDAP authentication in mixed sharded cluster deployments that contain both version 2.4 and version 2.6 shards. See *Upgrade MongoDB to 2.6* (page 1010) for upgrade instructions.

Expanded SNMP Support

MongoDB Enterprise has greatly expanded its SNMP support to provide SNMP access to nearly the full range of metrics provided by `db.serverStatus()`.

See also:

SNMP Changes (page 1005)

Additional Information

Changes Affecting Compatibility

Compatibility Changes in MongoDB 2.6

On this page

- Index Changes (page 1000)
- Write Method Acknowledgements (page 1002)
- `db.collection.aggregate()` Change (page 1002)
- Write Concern Validation (page 1003)
- Security Changes (page 1003)
- `2dsphere` Index Version 2 (page 1004)
- Log Messages (page 1004)
- Package Configuration Changes (page 1004)
- Remove Method Signature Change (page 1005)
- Update Operator Syntax Validation (page 1005)
- Updates Enforce Field Name Restrictions (page 1005)
- Query and Sort Changes (page 1006)
- Replica Set/Sharded Cluster Validation (page 1009)
- Time Format Changes (page 1010)
- Other Resources (page 1010)

The following 2.6 changes can affect the compatibility with older versions of MongoDB. See *Release Notes for MongoDB 2.6* (page 967) for the full list of the 2.6 changes.

Index Changes

Enforce Index Key Length Limit

Description MongoDB 2.6 implements a stronger enforcement of the limit on `index key`.

Creating indexes will error if an index key in an existing document exceeds the limit:

- `db.collection.ensureIndex()`, `db.collection.reIndex()`, `compact`, and `repairDatabase` will error and not create the index. Previous versions of MongoDB would create the index but not index such documents.
- Because `db.collection.reIndex()`, `compact`, and `repairDatabase` drop *all* the indexes from a collection and then recreate them sequentially, the error from the index key limit prevents these operations from rebuilding any remaining indexes for the collection and, in the case of the `repairDatabase` command, from continuing with the remainder of the process.

Inserts will error:

- `db.collection.insert()` and other operations that perform inserts (e.g. `db.collection.save()` and `db.collection.update()` with `upsert` that result in inserts) will fail to insert if the new document has an indexed field whose corresponding index entry exceeds the limit. Previous versions of MongoDB would insert but not index such documents.
- `mongorestore` and `mongoimport` will fail to insert if the new document has an indexed field whose corresponding index entry exceeds the limit.

Updates will error:

- `db.collection.update()` and `db.collection.save()` operations on an indexed field will error if the updated value causes the index entry to exceed the limit.
- If an existing document contains an indexed field whose index entry exceeds the limit, updates on other fields that result in the relocation of a document on disk will error.

Chunk Migration will fail:

- Migrations will fail for a chunk that has a document with an indexed field whose index entry exceeds the limit.
- If left unfixed, the chunk will repeatedly fail migration, effectively ceasing chunk balancing for that collection. Or, if chunk splits occur in response to the migration failures, this response would lead to unnecessarily large number of chunks and an overly large config databases.

Secondary members of replica sets will warn:

- Secondaries will continue to replicate documents with an indexed field whose corresponding index entry exceeds the limit on initial sync but will print warnings in the logs.
- Secondaries allow index build and rebuild operations on a collection that contains an indexed field whose corresponding index entry exceeds the limit but with warnings in the logs.
- With *mixed version* replica sets where the secondaries are version 2.6 and the primary is version 2.4, secondaries will replicate documents inserted or updated on the 2.4 primary, but will print error messages in the log if the documents contain an indexed field whose corresponding index entry exceeds the limit.

Solution Run `db.upgradeCheckAllDBs()` to find current keys that violate this limit and correct as appropriate. Preferably, run the test before upgrading; i.e. connect the 2.6 `mongo` shell to your MongoDB 2.4 database and run the method.

If you have an existing data set and want to disable the default index key length validation so that you can upgrade before resolving these indexing issues, use the `failIndexKeyTooLong` parameter.

Index Specifications Validate Field Names

Description In MongoDB 2.6, create and re-index operations fail when the index key refers to an empty field, e.g. `"a..b" : 1` or the field name starts with a dollar sign (`$`).

- `db.collection.ensureIndex()` will not create a new index with an invalid or empty key name.
- `db.collection.reIndex()`, `compact`, and `repairDatabase` will error if an index exists with an invalid or empty key name.
- Chunk migration will fail if an index exists with an invalid or empty key name.

Previous versions of MongoDB allow the index.

Solution Run `db.upgradeCheckAllDBs()` to find current keys that violate this limit and correct as appropriate. Preferably, run the test before upgrading; i.e. connect the 2.6 `mongo` shell to your MongoDB 2.4 database and run the method.

`ensureIndex` and Existing Indexes

Description `db.collection.ensureIndex()` now errors:

- if you try to create an existing index but with different options; e.g. in the following example, the second `db.collection.ensureIndex()` will error.

```
db.mycollection.ensureIndex( { x: 1 } )
db.mycollection.ensureIndex( { x: 1 }, { unique: 1 } )
```

- if you specify an index name that already exists but the key specifications differ; e.g. in the following example, the second `db.collection.ensureIndex()` will error.

```
db.mycollection.ensureIndex( { a: 1 }, { name: "myIdx" } )
db.mycollection.ensureIndex( { z: 1 }, { name: "myIdx" } )
```

Previous versions did not create the index but did not error.

Write Method Acknowledgements

Description The `mongo` shell write methods `db.collection.insert()`, `db.collection.update()`, `db.collection.save()` and `db.collection.remove()` now integrate the *write concern* (page 179) directly into the method rather than with a separate `getLastError` command to provide *acknowledgement of writes* (page 180) whether run interactively in the `mongo` shell or non-interactively in a script. In previous versions, these methods exhibited a "fire-and-forget" behavior. [1415]

- Existing scripts for the `mongo` shell that used these methods will now wait for acknowledgement, which take **longer** than the previous "fire-and-forget" behavior.
- The write methods now return a `WriteResult` object that contains the results of the operation, including any write errors and write concern errors, and obviates the need to call `getLastError` command to get the status of the results. See `db.collection.insert()`, `db.collection.update()`, `db.collection.save()` and `db.collection.remove()` for details.
- In sharded environments, `mongos` no longer supports "fire-and-forget" behavior. This limits throughput when writing data to sharded clusters.

Solution Scripts that used these `mongo` shell methods for bulk write operations with "fire-and-forget" behavior should use the `Bulk()` methods.

In sharded environments, applications using any driver or `mongo` shell should use `Bulk()` methods for optimal performance when inserting or modifying groups of documents.

For example, instead of:

```
for (var i = 1; i <= 1000000; i++) {
    db.test.insert( { x : i } );
}
```

In MongoDB 2.6, replace with `Bulk()` operation:

```
var bulk = db.test.initializeUnorderedBulkOp();

for (var i = 1; i <= 1000000; i++) {
    bulk.insert( { x : i} );
}

bulk.execute( { w: 1 } );
```

Bulk method returns a `BulkWriteResult` object that contains the result of the operation.

See also:

New Write Operation Protocol (page 995), `Bulk()`, `Bulk.execute()`, `db.collection.initializeUnorderedBulkOp()`, `db.collection.initializeOrderedBulkOp()`

`db.collection.aggregate()` Change

Description The `db.collection.aggregate()` method in the `mongo` shell defaults to returning a cursor to the results set. This change enables the aggregation pipeline to return result sets of any size and requires cursor iteration to access the result set. For example:

```
var myCursor = db.orders.aggregate( [
    {
      $group: {
         _id: "$cust_id",
```

[1415] In previous versions, when using the `mongo` shell interactively, the `mongo` shell automatically called the `getLastError` command after a write method to provide acknowledgment of the write. Scripts, however, would observe "fire-and-forget" behavior in previous versions unless the scripts included an **explicit** call to the `getLastError` command after a write method.

```
            total: { $sum: "$price" }
        }
    }
] );

myCursor.forEach( function(x) { printjson (x); } );
```

Previous versions returned a single document with a field `results` that contained an array of the result set, subject to the *BSON Document size* limit. Accessing the result set in the previous versions of MongoDB required accessing the `results` field and iterating the array. For example:

```
var returnedDoc = db.orders.aggregate( [
    {
      $group: {
         _id: "$cust_id",
         total: { $sum: "$price" }
      }
    }
] );

var myArray = returnedDoc.result; // access the result field

myArray.forEach( function(x) { printjson (x); } );
```

Solution Update scripts that currently expect `db.collection.aggregate()` to return a document with a `results` array to handle cursors instead.

See also:

Aggregation Enhancements (page 994), `db.collection.aggregate()`,

Write Concern Validation

Description Specifying a write concern that includes `j: true` to a `mongod` or `mongos` instance running with *`--nojournal`* option now errors. Previous versions would ignore the `j: true`.

Solution Either remove the `j: true` specification from the write concern when issued against a `mongod` or `mongos` instance with *`--nojournal`* or run `mongod` or `mongos` with journaling.

Security Changes

New Authorization Model

Description MongoDB 2.6 *authorization model* (page 433) changes how MongoDB stores and manages user privilege information:

- Before the upgrade, MongoDB 2.6 requires at least one user in the admin database.
- MongoDB versions using older models cannot create/modify users or create user-defined roles.

Solution Ensure that at least one user exists in the admin database. If no user exists in the admin database, add a user. Then upgrade to MongoDB 2.6. Finally, upgrade the user privilege model. See *Upgrade MongoDB to 2.6* (page 1010).

Important: Before upgrading the authorization model, you should first upgrade MongoDB binaries to 2.6. For sharded clusters, ensure that **all** cluster components are 2.6. If there are users in any database, be sure you have at least one user in the `admin` database with the role `userAdminAnyDatabase` (page 493) **before** upgrading the MongoDB binaries.

See also:

Security Improvements (page 996)

SSL Certificate Hostname Validation

Description The SSL certificate validation now checks the Common Name (`CN`) and the Subject Alternative Name (`SAN`) fields to ensure that either the `CN` or one of the `SAN` entries matches the hostname of the server. As a result, if you currently use SSL and *neither* the `CN` nor any of the `SAN` entries of your current SSL certificates match the hostnames, upgrading to version 2.6 will cause the SSL connections to fail.

Solution To allow for the continued use of these certificates, MongoDB provides the `allowInvalidCertificates` setting. The setting is available for:

- `mongod` and `mongos` to bypass the validation of SSL certificates on other servers in the cluster.
- `mongo` shell, *MongoDB tools that support SSL* (page 458), and the C++ driver to bypass the validation of server certificates.

When using the `allowInvalidCertificates` setting, MongoDB logs as a warning the use of the invalid certificates.

> **Warning:** The `allowInvalidCertificates` setting bypasses the other certificate validation, such as checks for expiration and valid signatures.

`2dsphere` Index Version 2

Description MongoDB 2.6 introduces a version 2 of the *2dsphere index* (page 543). If a document lacks a `2dsphere` index field (or the field is `null` or an empty array), MongoDB does not add an entry for the document to the `2dsphere` index. For inserts, MongoDB inserts the document but does not add to the `2dsphere` index.

Previous version would not insert documents where the `2dsphere` index field is a `null` or an empty array. For documents that lack the `2dsphere` index field, previous versions would insert and index the document.

Solution To revert to old behavior, create the `2dsphere` index with `{ "2dsphereIndexVersion" : 1 }` to create a version 1 index. However, version 1 index cannot use the new GeoJSON geometries.

See also:

2dsphere (Version 2) (page 543)

Log Messages

Timestamp Format Change

Description Each message now starts with the timestamp format given in *Time Format Changes* (page 1010). Previous versions used the `ctime` format.

Solution MongoDB adds a new option `--timeStampFormat` which supports timestamp format in `ctime`, `iso8601-utc`, and `iso8601-local` (new default).

Package Configuration Changes

Default `bindIp` for RPM/DEB Packages

Description In the official MongoDB packages in RPM (Red Hat, CentOS, Fedora Linux, and derivatives) and DEB (Debian, Ubuntu, and derivatives), the default `bindIp` value attaches MongoDB components to the localhost interface *only*. These packages set this default in the default configuration file (i.e. `/etc/mongod.conf`.)

Solution If you use one of these packages and have *not* modified the default `/etc/mongod.conf` file, you will need to set `bindIp` before or during the upgrade.

There is no default `bindIp` setting in any other official MongoDB packages.

SNMP Changes

Description

- The IANA enterprise identifier for MongoDB changed from 37601 to 34601.
- MongoDB changed the MIB field name `globalopcounts` to `globalOpcounts`.

Solution

- Users of SNMP monitoring must modify their SNMP configuration (i.e. MIB) from 37601 to 34601.
- Update references to `globalopcounts` to `globalOpcounts`.

Remove Method Signature Change

Description `db.collection.remove()` requires a query document as a parameter. In previous versions, the method invocation without a query document deleted all documents in a collection.

Solution For existing `db.collection.remove()` invocations without a query document, modify the invocations to include an empty document `db.collection.remove({})`.

Update Operator Syntax Validation

Description

- `Update operators (e.g $set)` must specify a non-empty operand expression. For example, the following expression is now invalid:

```
{ $set: { } }
```

- `Update operators (e.g $set)` cannot repeat in the update statement. For example, the following expression is invalid:

```
{ $set: { a: 5 }, $set: { b: 5 } }
```

Updates Enforce Field Name Restrictions

Description

- Updates cannot use `update operators (e.g $set)` to target fields with empty field names (i.e. `""`).
- Updates no longer support saving field names that contain a dot (`.`) or a field name that starts with a dollar sign (`$`).

Solution

- For existing documents that have fields with empty names `""`, replace the whole document. See `db.collection.update()` and `db.collection.save()` for details on replacing an existing document.
- For existing documents that have fields with names that contain a dot (`.`), either replace the whole document or `unset` the field. To find fields whose names contain a dot, run `db.upgradeCheckAllDBs()`.
- For existing documents that have fields with names that start with a dollar sign (`$`), `unset` or `rename` those fields. To find fields whose names start with a dollar sign, run `db.upgradeCheckAllDBs()`.

See *New Write Operation Protocol* (page 995) for the changes to the write operation protocol, and *Insert and Update Improvements* (page 995) for the changes to the insert and update operations. Also consider the documentation of the `Restrictions on Field Names`.

Query and Sort Changes

Enforce Field Name Restrictions

Description Queries cannot specify conditions on fields with names that start with a dollar sign (`$`).

Solution `Unset` or `rename` existing fields whose names start with a dollar sign (`$`). Run `db.upgradeCheckAllDBs()` to find fields whose names start with a dollar sign.

Sparse Index and Incomplete Results

Description If a *sparse index* (page 574) results in an incomplete result set for queries and sort operations, MongoDB will not use that index unless a `hint()` explicitly specifies the index.

For example, the query `{ x: { $exists: false } }` will no longer use a sparse index on the `x` field, unless explicitly hinted.

Solution To override the behavior to use the sparse index and return incomplete results, explicitly specify the index with a `hint()`.

See *Sparse Index On A Collection Cannot Return Complete Results* (page 575) for an example that details the new behavior.

`sort()` Specification Values

Description The `sort()` method **only** accepts the following values for the sort keys:

- `1` to specify ascending order for a field,
- `-1` to specify descending order for a field, or
- `$meta` expression to specify sort by the text search score.

Any other value will result in an error.

Previous versions also accepted either `true` or `false` for ascending.

Solution Update sort key values that use `true` or `false` to `1`.

`skip()` and `_id` Queries

Description Equality match on the `_id` field obeys `skip()`.

Previous versions ignored `skip()` when performing an equality match on the `_id` field.

`explain()` Retains Query Plan Cache

Description `explain()` no longer clears the *query plans* (page 108) cached for that *query shape*.

In previous versions, `explain()` would have the side effect of clearing the query plan cache for that query shape.

See also:

The `PlanCache()` reference.

Geospatial Changes

`$maxDistance` Changes

Description

- For `$near` queries on GeoJSON data, if the queries specify a `$maxDistance`, `$maxDistance` must be inside of the `$near` document.

  In previous version, `$maxDistance` could be either inside or outside the `$near` document.

- `$maxDistance` must be a positive value.

Solution

- Update any existing `$near` queries on GeoJSON data that currently have the `$maxDistance` outside the `$near` document
- Update any existing queries where `$maxDistance` is a negative value.

Deprecated `$uniqueDocs`

Description MongoDB 2.6 deprecates `$uniqueDocs`, and geospatial queries no longer return duplicated results when a document matches the query multiple times.

Stronger Validation of Geospatial Queries

Description MongoDB 2.6 enforces a stronger validation of geospatial queries, such as validating the options or GeoJSON specifications, and errors if the geospatial query is invalid. Previous versions allowed/ignored invalid options.

Query Operator Changes

`$not` Query Behavior Changes

Description

- Queries with `$not` expressions on an indexed field now match:
  - Documents that are missing the indexed field. Previous versions would not return these documents using the index.
  - Documents whose indexed field value is a different type than that of the specified value. Previous versions would not return these documents using the index.

  For example, if a collection `orders` contains the following documents:

```
{ _id: 1, status: "A", cust_id: "123", price: 40 }
{ _id: 2, status: "A", cust_id: "xyz", price: "N/A" }
{ _id: 3, status: "D", cust_id: "xyz" }
```

If the collection has an index on the `price` field:

```
db.orders.ensureIndex( { price: 1 } )
```

The following query uses the index to search for documents where `price` is not greater than or equal to `50`:

```
db.orders.find( { price: { $not: { $gte: 50 } } } )
```

In 2.6, the query returns the following documents:

```
{ "_id" : 3, "status" : "D", "cust_id" : "xyz" }
{ "_id" : 1, "status" : "A", "cust_id" : "123", "price" : 40 }
{ "_id" : 2, "status" : "A", "cust_id" : "xyz", "price" : "N/A" }
```

In previous versions, indexed plans would only return matching documents where the type of the field matches the type of the query predicate:

```
{ "_id" : 1, "status" : "A", "cust_id" : "123", "price" : 40 }
```

If using a collection scan, previous versions would return the same results as those in 2.6.

- MongoDB 2.6 allows chaining of `$not` expressions.

`null` Comparison Queries

Description

- `$lt` and `$gt` comparisons to `null` no longer match documents that are missing the field.
- `null` equality conditions on array elements (e.g. `"a.b": null`) no longer match document missing the nested field `a.b` (e.g. `a: [ 2, 3 ]`).
- `null` equality queries (i.e. `field: null`) now match fields with values `undefined`.

`$all` Operator Behavior Change

Description

- The `$all` operator is now equivalent to an `$and` operation of the specified values. This change in behavior can allow for more matches than previous versions when passed an array of a single nested array (e.g. `[ [ "A" ] ]`). When passed an array of a nested array, `$all` can now match documents where the field contains the nested array as an element (e.g. `field: [ [ "A" ], ... ]`), *or* the field equals the nested array (e.g. `field: [ "A", "B" ]`). Earlier version could only match documents where the field contains the nested array.
- The `$all` operator returns no match if the array field contains nested arrays (e.g. `field: [ "a", ["b"] ]`) and `$all` on the nested field is the element of the nested array (e.g. `"field.1": { $all: [ "b" ] }`). Previous versions would return a match.

`$mod` Operator Enforces Strict Syntax

Description The `$mod` operator now only accepts an array with exactly two elements, and errors when passed an array with fewer or more elements. See *mod-not-enough-elements* and *mod-too-many-elements* for details.

In previous versions, if passed an array with one element, the `$mod` operator uses `0` as the second element, and if passed an array with more than two elements, the `$mod` ignores all but the first two elements. Previous versions do return an error when passed an empty array.

Solution Ensure that the array passed to `$mod` contains exactly two elements:

- If the array contains the a single element, add `0` as the second element.
- If the array contains more than two elements, remove the extra elements.

`$where` Must Be Top-Level

Description `$where` expressions can now only be at top level and cannot be nested within another expression, such as `$elemMatch`.

Solution Update existing queries that nest `$where`.

`$exists` and `notablescan` If the MongoDB server has disabled collection scans, i.e. `notablescan`, then `$exists` queries that have no *indexed solution* will error.

`MinKey` and `MaxKey` Queries

Description Equality match for either `MinKey` or `MaxKey` no longer match documents missing the field.

Nested Array Queries with $elemMatch

Description The `$elemMatch` query operator no longer traverses recursively into nested arrays.

For example, if a collection `test` contains the following document:

```
{ "_id": 1, "a" : [ [ 1, 2, 5 ] ] }
```

In 2.6, the following `$elemMatch` query does *not* match the document:

```
db.test.find( { a: { $elemMatch: { $gt: 1, $lt: 5 } } } )
```

Solution Update existing queries that rely upon the old behavior.

Text Search Compatibility MongoDB does not support the use of the `$text` query operator in mixed sharded cluster deployments that contain both version 2.4 and version 2.6 shards. See *Upgrade MongoDB to 2.6* (page 1010) for upgrade instructions.

Replica Set/Sharded Cluster Validation

Shard Name Checks on Metadata Refresh

Description For sharded clusters, MongoDB 2.6 disallows a shard from refreshing the metadata if the shard name has not been explicitly set.

For mixed sharded cluster deployments that contain both version 2.4 and version 2.6 shards, this change can cause errors when migrating chunks **from** version 2.4 shards **to** version 2.6 shards if the shard name is unknown to the version 2.6 shards. MongoDB does not support migrations in mixed sharded cluster deployments.

Solution Upgrade all components of the cluster to 2.6. See *Upgrade MongoDB to 2.6* (page 1010).

Replica Set Vote Configuration Validation

Description MongoDB now deprecates giving any *replica set* member more than a single vote. During configuration, `local.system.replset.members[n].votes` should only have a value of 1 for voting members and 0 for non-voting members. MongoDB treats values other than 1 or 0 as a value of 1 and produces a warning message.

Solution Update `local.system.replset.members[n].votes` with values other than 1 or 0 to 1 or 0 as appropriate.

Time Format Changes MongoDB now uses `iso8601-local` when formatting time data in many outputs. This format follows the template `YYYY-MM-DDTHH:mm:ss.mmm<+/-Offset>`. For example, `2014-03-04T20:13:38.944-0500`.

This change impacts all clients using *Extended JSON* (page 16) in *Strict mode*, such as `mongoexport` and the REST and HTTP Interfaces[1416].

Other Resources

- All backwards incompatible changes (JIRA)[1417].
- *Release Notes for MongoDB 2.6* (page 967).
- *Upgrade MongoDB to 2.6* (page 1010) for the upgrade process.

Some changes in 2.6 can affect *compatibility* (page 999) and may require user actions. The 2.6 `mongo` shell provides a `db.upgradeCheckAllDBs()` method to perform a check for upgrade preparedness for some of these changes.

See *Compatibility Changes in MongoDB 2.6* (page 999) for a detailed list of compatibility changes.

See also:

All Backwards incompatible changes (JIRA)[1418].

Upgrade Process

	On this page
Upgrade MongoDB to 2.6	• Upgrade Recommendations and Checklists (page 1010) • Upgrade MongoDB Processes (page 1012) • Upgrade Procedure (page 1014)

In the general case, the upgrade from MongoDB 2.4 to 2.6 is a binary-compatible "drop-in" upgrade: shut down the `mongod` instances and replace them with `mongod` instances running 2.6. **However**, before you attempt any upgrade, familiarize yourself with the content of this document, particularly the *Upgrade Recommendations and Checklists* (page 1010), the procedure for *upgrading sharded clusters* (page 1012), and the considerations for *reverting to 2.4 after running 2.6* (page 1016).

Upgrade Recommendations and Checklists When upgrading, consider the following:

[1416] https://docs.mongodb.org/ecosystem/tools/http-interfaces

[1417] https://jira.mongodb.org/issues/?jql=project%20%3D%20SERVER%20AND%20fixVersion%20in%20(%222.5.0%22%2C%20%222.5.1%22%2C%20%222.5.2%22%2C%20%222.6.0-rc1%22%2C%20%222.6.0-rc2%22)%20AND%20%22Backwards%20Compatibility%22%20in%20%20(%22Major%20Change%22%2C%20%22Minor%20Change%22%2C%20%22Major%20Change%20or%20Minor%20Change%22)

[1418] https://jira.mongodb.org/issues/?jql=project%20%3D%20SERVER%20AND%20fixVersion%20in%20(%222.5.0%22%2C%20%222.5.1%22%2C%20%222.5.2%22%2C%20%222.6.0-rc1%22%2C%20%222.6.0-rc2%22%2C%20%222.6.0-rc3%22)%20AND%20%22Backwards%20Compatibility%22%20in%20(%20%22Minor%20Change%22%2C%20%22Major%20Change%20or%20Minor%20Change%22)

Upgrade Requirements To upgrade an existing MongoDB deployment to 2.6, you must be running 2.4. If you're running a version of MongoDB before 2.4, you *must* upgrade to 2.4 before upgrading to 2.6. See *Upgrade MongoDB to 2.4* (page 1037) for the procedure to upgrade from 2.2 to 2.4.

If you use MongoDB Cloud Manager[1419] Backup, ensure that you're running *at least* version `v20131216.1` of the Backup agent before upgrading. Version `1.4.0` of the backup agent followed `v20131216.1`

Preparedness Before upgrading MongoDB always test your application in a staging environment before deploying the upgrade to your production environment.

To begin the upgrade procedure, connect a 2.6 `mongo` shell to your MongoDB 2.4 `mongos` or `mongod` and run the `db.upgradeCheckAllDBs()` to check your data set for compatibility. This is a preliminary automated check. Assess and resolve all issues identified by `db.upgradeCheckAllDBs()`.

Some changes in MongoDB 2.6 require manual checks and intervention. See *Compatibility Changes in MongoDB 2.6* (page 999) for an explanation of these changes. Resolve all incompatibilities in your deployment before continuing.

For a deployment that uses authentication and authorization, be sure you have at least one user in the `admin` database with the role `userAdminAnyDatabase` (page 493) **before** upgrading the MongoDB binaries. For deployments currently using authentication and authorization, see the *consideration for deployments that use authentication and authorization* (page 1011).

Authentication MongoDB 2.6 includes significant changes to the authorization model, which requires changes to the way that MongoDB stores users' credentials. As a result, in addition to upgrading MongoDB processes, if your deployment uses authentication and authorization, after upgrading all MongoDB process to 2.6 you **must** also upgrade the authorization model.

Before beginning the upgrade process for a deployment that uses authentication and authorization:

- Ensure that at least one user exists in the `admin` database with the role `userAdminAnyDatabase` (page 493).
- If your application performs CRUD operations on the `<database>.system.users` collection or uses a `db.addUser()`-like method, then you **must** upgrade those drivers (i.e. client libraries) **before** `mongod` or `mongos` instances.
- You must fully complete the upgrade procedure for *all* MongoDB processes before upgrading the authorization model.

After you begin to upgrade a MongoDB deployment that uses authentication to 2.6, you *cannot* modify existing user data until you complete the *authorization user schema upgrade* (page 1014).

See *Upgrade User Authorization Data to 2.6 Format* (page 1014) for a complete discussion of the upgrade procedure for the authorization model including additional requirements and procedures.

Downgrade Limitations Once upgraded to MongoDB 2.6, you **cannot** downgrade to **any** version earlier than MongoDB 2.4. If you created `text` or `2dsphere` indexes while running 2.6, you can only downgrade to MongoDB 2.4.10 or later.

Package Upgrades If you installed MongoDB from the MongoDB `apt` or `yum` repositories, upgrade to 2.6 using the package manager.

For Debian, Ubuntu, and related operating systems, type these commands:

[1419] https://cloud.mongodb.com/?jmp=docs

```
sudo apt-get update
sudo apt-get install mongodb-org
```

For Red Hat Enterprise, CentOS, Fedora, or Amazon Linux:

```
sudo yum install mongodb-org
```

If you did not install the `mongodb-org` package, and installed a subset of MongoDB components replace `mongodb-org` in the commands above with the appropriate package names.

See installation instructions for *Ubuntu* (page 33), *RHEL* (page 23), *Debian* (page 36), or *other Linux Systems* (page 39) for a list of the available packages and complete MongoDB installation instructions.

Upgrade MongoDB Processes

Upgrade Standalone `mongod` Instance to MongoDB 2.6 The following steps outline the procedure to upgrade a standalone `mongod` from version 2.4 to 2.6. To upgrade from version 2.2 to 2.6, *upgrade to version 2.4* (page 1037) *first*, and then follow the procedure to upgrade from 2.4 to 2.6.

1. Download binaries of the latest release in the 2.6 series from the MongoDB Download Page[1420]. See *Install MongoDB* (page 21) for more information.
2. Shut down your `mongod` instance. Replace the existing binary with the 2.6 `mongod` binary and restart `mongod`.

Upgrade a Replica Set to 2.6 The following steps outline the procedure to upgrade a replica set from MongoDB 2.4 to MongoDB 2.6. To upgrade from MongoDB 2.2 to 2.6, *upgrade all members of the replica set to version 2.4* (page 1037) *first*, and then follow the procedure to upgrade from MongoDB 2.4 to 2.6.

You can upgrade from MongoDB 2.4 to 2.6 using a "rolling" upgrade to minimize downtime by upgrading the members individually while the other members are available:

Step 1: Upgrade secondary members of the replica set. Upgrade the *secondary* members of the set one at a time by shutting down the `mongod` and replacing the 2.4 binary with the 2.6 binary. After upgrading a `mongod` instance, wait for the member to recover to `SECONDARY` state before upgrading the next instance. To check the member's state, issue `rs.status()` in the `mongo` shell.

Step 2: Step down the replica set primary. Use `rs.stepDown()` in the `mongo` shell to step down the *primary* and force the set to *failover* (page 644). `rs.stepDown()` expedites the failover procedure and is preferable to shutting down the primary directly.

Step 3: Upgrade the primary. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, shut down the previous primary and replace the `mongod` binary with the 2.6 binary and start the new instance.

Replica set failover is not instant but will render the set unavailable accept writes until the failover process completes. Typically this takes 30 seconds or more: schedule the upgrade procedure during a scheduled maintenance window.

Upgrade a Sharded Cluster to 2.6 Only upgrade sharded clusters to 2.6 if **all** members of the cluster are currently running instances of 2.4. The only supported upgrade path for sharded clusters running 2.2 is via 2.4. The upgrade process checks all components of the cluster and will produce warnings if any component is running version 2.2.

[1420] http://www.mongodb.org/downloads

Considerations The upgrade process does not require any downtime. However, while you upgrade the sharded cluster, ensure that clients do not make changes to the collection meta-data. For example, during the upgrade, do **not** do any of the following:

- `sh.enableSharding()`
- `sh.shardCollection()`
- `sh.addShard()`
- `db.createCollection()`
- `db.collection.drop()`
- `db.dropDatabase()`
- any operation that creates a database
- any other operation that modifies the cluster metadata in any way. See *Sharding Reference* (page 822) for a complete list of sharding commands. Note, however, that not all commands on the *Sharding Reference* (page 822) page modifies the cluster meta-data.

Upgrade Sharded Clusters *Optional but Recommended.* As a precaution, take a backup of the `config` database *before* upgrading the sharded cluster.

Step 1: Disable the Balancer. Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).

Step 2: Upgrade the cluster's meta data. Start a single 2.6 `mongos` instance with the `configDB` pointing to the cluster's config servers and with the *`--upgrade`* option.

To run a `mongos` with the *`--upgrade`* option, you can upgrade an existing `mongos` instance to 2.6, or if you need to avoid reconfiguring a production `mongos` instance, you can use a new 2.6 `mongos` that can reach all the config servers.

To upgrade the meta data, run:

```
mongos --configdb <configDB string> --upgrade
```

You can include the *`--logpath`* option to output the log messages to a file instead of the standard output. Also include any other options required to start `mongos` instances in your cluster, such as *`--sslOnNormalPorts`* or *`--sslPEMKeyFile`*.

The `mongos` will exit upon completion of the *`--upgrade`* process.

The upgrade will prevent any chunk moves or splits from occurring during the upgrade process. If the data files have many sharded collections or if failed processes hold stale locks, acquiring the locks for all collections can take seconds or minutes. Watch the log for progress updates.

Step 3: Ensure `mongos --upgrade` process completes successfully. The `mongos` will exit upon completion of the meta data upgrade process. If successful, the process will log the following messages:

```
upgrade of config server to v5 successful
Config database is at version v5
```

After a successful upgrade, restart the `mongos` instance. If `mongos` fails to start, check the log for more information.

If the `mongos` instance loses its connection to the config servers during the upgrade or if the upgrade is otherwise unsuccessful, you may always safely retry the upgrade.

Step 4: Upgrade the remaining `mongos` instances to v2.6. Upgrade and restart **without** the `--upgrade` option the other `mongos` instances in the sharded cluster. After upgrading all the `mongos`, see *Complete Sharded Cluster Upgrade* (page 1014) for information on upgrading the other cluster components.

Complete Sharded Cluster Upgrade After you have successfully upgraded *all* `mongos` instances, you can upgrade the other instances in your MongoDB deployment.

Warning: Do not upgrade `mongod` instances until after you have upgraded *all* `mongos` instances.

While the balancer is still disabled, upgrade the components of your sharded cluster in the following order:

- Upgrade all 3 `mongod` config server instances, leaving the *first* system in the `mongos --configdb` argument to upgrade *last*.
- Upgrade each shard, one at a time, upgrading the `mongod` secondaries before running `replSetStepDown` and upgrading the primary of each shard.

When this process is complete, *re-enable the balancer* (page 803).

Upgrade Procedure Once upgraded to MongoDB 2.6, you **cannot** downgrade to **any** version earlier than MongoDB 2.4. If you have `text` or `2dsphere` indexes, you can only downgrade to MongoDB 2.4.10 or later.

Except as described on this page, moving between 2.4 and 2.6 is a drop-in replacement:

Step 1: Stop the existing `mongod` instance. For example, on Linux, run 2.4 `mongod` with the `--shutdown` option as follows:

```
mongod --dbpath /var/mongod/data --shutdown
```

Replace `/var/mongod/data` with your MongoDB `dbPath`. See also the *Stop mongod Processes* (page 324) for alternate methods of stopping a `mongod` instance.

Step 2: Start the new `mongod` instance. Ensure you start the 2.6 `mongod` with the same `dbPath`:

```
mongod --dbpath /var/mongod/data
```

Replace `/var/mongod/data` with your MongoDB `dbPath`.

On this page

Upgrade User Authorization Data to 2.6 Format

- Considerations (page 1014)
- Requirements (page 1015)
- Procedure (page 1015)
- Result (page 1016)

MongoDB 2.6 includes significant changes to the authorization model, which requires changes to the way that MongoDB stores users' credentials. As a result, in addition to upgrading MongoDB processes, if your deployment uses authentication and authorization, after upgrading all MongoDB process to 2.6 you **must** also upgrade the authorization model.

Considerations

Complete all other Upgrade Requirements Before upgrading the authorization model, you should first upgrade MongoDB binaries to 2.6. For sharded clusters, ensure that **all** cluster components are 2.6. If there are users in any database, be sure you have at least one user in the `admin` database with the role `userAdminAnyDatabase` (page 493) **before** upgrading the MongoDB binaries.

Timing Because downgrades are more difficult after you upgrade the user authorization model, once you upgrade the MongoDB binaries to version 2.6, allow your MongoDB deployment to run a day or two **without** upgrading the user authorization model.

This allows 2.6 some time to "burn in" and decreases the likelihood of downgrades occurring after the user privilege model upgrade. The user authentication and access control will continue to work as it did in 2.4, **but** it will be impossible to create or modify users or to use user-defined roles until you run the authorization upgrade.

If you decide to upgrade the user authorization model immediately instead of waiting the recommended "burn in" period, then for sharded clusters, you must wait at least 10 seconds after upgrading the sharded clusters to run the authorization upgrade script.

Replica Sets For a replica set, it is only necessary to run the upgrade process on the *primary* as the changes will automatically replicate to the secondaries.

Sharded Clusters For a sharded cluster, connect to a `mongos` and run the upgrade procedure to upgrade the cluster's authorization data. By default, the procedure will upgrade the authorization data of the shards as well.

To override this behavior, run the upgrade command with the additional parameter `upgradeShards: false`. If you choose to override, you must run the upgrade procedure on the `mongos` first, and then run the procedure on the *primary* members of each shard.

For a sharded cluster, do **not** run the upgrade process directly against the *config servers* (page 742). Instead, perform the upgrade process using one `mongos` instance to interact with the config database.

Requirements To upgrade the authorization model, you must have a user in the `admin` database with the role `userAdminAnyDatabase` (page 493).

Procedure

Step 1: Connect to MongoDB instance. Connect and authenticate to the `mongod` instance for a single deployment or a `mongos` for a sharded cluster as an `admin` database user with the role `userAdminAnyDatabase` (page 493).

Step 2: Upgrade authorization schema. Use the `authSchemaUpgrade` command in the `admin` database to update the user data using the `mongo` shell.

Run `authSchemaUpgrade` command.

```
db.getSiblingDB("admin").runCommand({authSchemaUpgrade: 1 });
```

In case of error, you may safely rerun the `authSchemaUpgrade` command.

Sharded cluster `authSchemaUpgrade` consideration. For a sharded cluster, `authSchemaUpgrade` will upgrade the authorization data of the shards as well and the upgrade is complete. You can, however, override this behavior by including `upgradeShards: false` in the command, as in the following example:

```
db.getSiblingDB("admin").runCommand({authSchemaUpgrade: 1,
upgradeShards: false });
```

If you override the behavior, after running `authSchemaUpgrade` on a `mongos` instance, you will need to connect to the primary for each shard and repeat the upgrade process after upgrading on the `mongos`.

Result All users in a 2.6 system are stored in the `admin.system.users` (page 377) collection. To manipulate these users, use the `user management methods`.

The upgrade procedure copies the version 2.4 `admin.system.users` collection to `admin.system.backup_users`.

The upgrade procedure leaves the version 2.4 `<database>.system.users` collection(s) intact.

Downgrade MongoDB from 2.6

On this page

- Downgrade Recommendations and Checklist (page 1016)
- Downgrade 2.6 User Authorization Model (page 1017)
- Downgrade Updated Indexes (page 1019)
- Downgrade MongoDB Processes (page 1020)
- Downgrade Procedure (page 1021)

Before you attempt any downgrade, familiarize yourself with the content of this document, particularly the *Downgrade Recommendations and Checklist* (page 1016) and the procedure for *downgrading sharded clusters* (page 1021).

Downgrade Recommendations and Checklist When downgrading, consider the following:

Downgrade Path Once upgraded to MongoDB 2.6, you **cannot** downgrade to **any** version earlier than MongoDB 2.4. If you created `text` or `2dsphere` indexes while running 2.6, you can only downgrade to MongoDB 2.4.10 or later.

Preparedness

- *Remove or downgrade version 2 text indexes* (page 1019) before downgrading MongoDB 2.6 to 2.4.
- *Remove or downgrade version 2 2dsphere indexes* (page 1020) before downgrading MongoDB 2.6 to 2.4.
- *Downgrade 2.6 User Authorization Model* (page 1017). If you have upgraded to the 2.6 user authorization model, you must downgrade the user model to 2.4 before downgrading MongoDB 2.6 to 2.4.

Procedures Follow the downgrade procedures:

- To downgrade sharded clusters, see *Downgrade a 2.6 Sharded Cluster* (page 1021).
- To downgrade replica sets, see *Downgrade a 2.6 Replica Set* (page 1020).
- To downgrade a standalone MongoDB instance, see *Downgrade 2.6 Standalone mongod Instance* (page 1020).

Downgrade 2.6 User Authorization Model If you have upgraded to the 2.6 user authorization model, you **must first** downgrade the user authorization model to 2.4 **before** before downgrading MongoDB 2.6 to 2.4.

Considerations

- For a replica set, it is only necessary to run the downgrade process on the *primary* as the changes will automatically replicate to the secondaries.
- For sharded clusters, although the procedure lists the downgrade of the cluster's authorization data first, you may downgrade the authorization data of the cluster or shards first.
- You *must* have the `admin.system.backup_users` and `admin.system.new_users` collections created during the upgrade process.
- **Important**. The downgrade process returns the user data to its state prior to upgrading to 2.6 authorization model. Any changes made to the user/role data using the 2.6 users model will be lost.

Access Control Prerequisites To downgrade the authorization model, you must connect as a user with the following *privileges*:

```
{ resource: { db: "admin", collection: "system.new_users" }, actions: [ "find", "insert", "update" ]
{ resource: { db: "admin", collection: "system.backup_users" }, actions: [  "find" ] }
{ resource: { db: "admin", collection: "system.users" }, actions: [ "find", "remove", "insert"] }
{ resource: { db: "admin", collection: "system.version" }, actions: [ "find", "update" ] }
```

If no user exists with the appropriate *privileges*, create an authorization model downgrade user:

Step 1: Connect as user with privileges to manage users and roles. Connect and authenticate as a user with `userAdminAnyDatabase` (page 493).

Step 2: Create a role with required privileges. Using the `db.createRole` method, create a *role* (page 440) with the required privileges.

```
use admin
db.createRole(
  {
    role: "downgradeAuthRole",
    privileges: [
      { resource: { db: "admin", collection: "system.new_users" }, actions: [ "find", "insert", "upda
      { resource: { db: "admin", collection: "system.backup_users" }, actions: [  "find" ] },
      { resource: { db: "admin", collection: "system.users" }, actions: [ "find", "remove", "insert"
      { resource: { db: "admin", collection: "system.version" }, actions: [ "find", "update" ] }
    ],
    roles: [ ]
  }
)
```

Step 3: Create a user with the new role. Create a user and assign the user the `downgradeRole`.

```
use admin
db.createUser(
   {
     user: "downgradeAuthUser",
     pwd: "somePass123",
     roles: [ { role: "downgradeAuthRole", db: "admin" } ]
```

```
    }
)
```

Note: Instead of creating a new user, you can also grant the role to an existing user. See `db.grantRolesToUser()` method.

Step 4: Authenticate as the new user. Authenticate as the newly created user.

```
use admin
db.auth( "downgradeAuthUser", "somePass123" )
```

The method returns `1` upon successful authentication.

Procedure The following downgrade procedure requires `<database>.system.users` collections used in version 2.4. to be intact for non-`admin` databases.

Step 1: Connect and authenticate to MongoDB instance. Connect and authenticate to the `mongod` instance for a single deployment or a `mongos` for a sharded cluster with the appropriate privileges. See *Access Control Prerequisites* (page 1017) for details.

Step 2: Create backup of 2.6 `admin.system.users` collection. Copy all documents in the `admin.system.users` (page 377) collection to the `admin.system.new_users` collection:

```
db.getSiblingDB("admin").system.users.find().forEach( function(userDoc) {
    status = db.getSiblingDB("admin").system.new_users.save( userDoc );
    if (status.hasWriteError()) {
        print(status.writeError);
    }
  }
);
```

Step 3: Update the version document for the `authSchema`.

```
db.getSiblingDB("admin").system.version.update(
   { _id: "authSchema" },
   { $set: { currentVersion: 2 } }
);
```

The method returns a `WriteResult` object with the status of the operation. Upon successful update, the `WriteResult` object should have `"nModified"` equal to `1`.

Step 4: Remove existing documents from the `admin.system.users` collection.

```
db.getSiblingDB("admin").system.users.remove( {} )
```

The method returns a `WriteResult` object with the number of documents removed in the `"nRemoved"` field.

Step 5: Copy documents from the `admin.system.backup_users` collection. Copy all documents from the `admin.system.backup_users`, created during the 2.6 upgrade, to `admin.system.users`.

```
db.getSiblingDB("admin").system.backup_users.find().forEach(
   function (userDoc) {
      status = db.getSiblingDB("admin").system.users.insert( userDoc );
      if (status.hasWriteError()) {
          print(status.writeError);
      }
   }
);
```

Step 6: Update the version document for the `authSchema`.

```
db.getSiblingDB("admin").system.version.update(
   { _id: "authSchema" },
   { $set: { currentVersion: 1 } }
)
```

For a sharded cluster, repeat the downgrade process by connecting to the *primary* replica set member for each shard.

Note: The cluster's `mongos` instances will fail to detect the authorization model downgrade until the user cache is refreshed. You can run `invalidateUserCache` on each `mongos` instance to refresh immediately, or you can wait until the cache is refreshed automatically at the end of the `user cache invalidation interval`. To run `invalidateUserCache`, you must have privilege with `invalidateUserCache` (page 502) action, which is granted by `userAdminAnyDatabase` (page 493) and `hostManager` (page 490) roles.

Result The downgrade process returns the user data to its state prior to upgrading to 2.6 authorization model. Any changes made to the user/role data using the 2.6 users model will be lost.

Downgrade Updated Indexes

Text Index Version Check If you have *version 2* text indexes (i.e. the default version for text indexes in MongoDB 2.6), drop the *version 2* text indexes before downgrading MongoDB. After the downgrade, enable text search and recreate the dropped text indexes.

To determine the version of your `text` indexes, run `db.collection.getIndexes()` to view index specifications. For text indexes, the method returns the version information in the field `textIndexVersion`. For example, the following shows that the `text` index on the `quotes` collection is version 2.

```
{
   "v" : 1,
   "key" : {
      "_fts" : "text",
      "_ftsx" : 1
   },
   "name" : "quote_text_translation.quote_text",
   "ns" : "test.quotes",
   "weights" : {
      "quote" : 1,
      "translation.quote" : 1
   },
   "default_language" : "english",
   "language_override" : "language",
   "textIndexVersion" : 2
}
```

`2dsphere` Index Version Check If you have *version 2* `2dsphere` indexes (i.e. the default version for `2dsphere` indexes in MongoDB 2.6), drop the *version 2* `2dsphere` indexes before downgrading MongoDB. After the downgrade, recreate the `2dsphere` indexes.

To determine the version of your `2dsphere` indexes, run `db.collection.getIndexes()` to view index specifications. For `2dsphere` indexes, the method returns the version information in the field `2dsphereIndexVersion`. For example, the following shows that the `2dsphere` index on the `locations` collection is version 2.

```
{
   "v" : 1,
   "key" : {
      "geo" : "2dsphere"
   },
   "name" : "geo_2dsphere",
   "ns" : "test.locations",
   "sparse" : true,
   "2dsphereIndexVersion" : 2
}
```

Downgrade MongoDB Processes

Downgrade 2.6 Standalone `mongod` Instance The following steps outline the procedure to downgrade a standalone `mongod` from version 2.6 to 2.4.

1. Download binaries of the latest release in the 2.4 series from the MongoDB Download Page[1421]. See *Install MongoDB* (page 21) for more information.
2. Shut down your `mongod` instance. Replace the existing binary with the 2.4 `mongod` binary and restart `mongod`.

Downgrade a 2.6 Replica Set The following steps outline a "rolling" downgrade process for the replica set. The "rolling" downgrade process minimizes downtime by downgrading the members individually while the other members are available:

Step 1: Downgrade each secondary member, one at a time. For each *secondary* in a replica set:

Replace and restart secondary `mongod` instances. First, shut down the `mongod`, then replace these binaries with the 2.4 binary and restart `mongod`. See *Stop mongod Processes* (page 324) for instructions on safely terminating `mongod` processes.

Allow secondary to recover. Wait for the member to recover to `SECONDARY` state before upgrading the next secondary.

To check the member's state, use the `rs.status()` method in the `mongo` shell.

Step 2: Step down the primary. Use `rs.stepDown()` in the `mongo` shell to step down the *primary* and force the normal *failover* (page 644) procedure.

```
rs.stepDown()
```

`rs.stepDown()` expedites the failover procedure and is preferable to shutting down the primary directly.

[1421] http://www.mongodb.org/downloads

Step 3: Replace and restart former primary `mongod`. When `rs.status()` shows that the primary has stepped down and another member has assumed `PRIMARY` state, shut down the previous primary and replace the `mongod` binary with the 2.4 binary and start the new instance.

Replica set failover is not instant but will render the set unavailable to writes and interrupt reads until the failover process completes. Typically this takes 10 seconds or more. You may wish to plan the downgrade during a predetermined maintenance window.

Downgrade a 2.6 Sharded Cluster

Requirements While the downgrade is in progress, you cannot make changes to the collection meta-data. For example, during the downgrade, do **not** do any of the following:

- `sh.enableSharding()`
- `sh.shardCollection()`
- `sh.addShard()`
- `db.createCollection()`
- `db.collection.drop()`
- `db.dropDatabase()`
- any operation that creates a database
- any other operation that modifies the cluster meta-data in any way. See *Sharding Reference* (page 822) for a complete list of sharding commands. Note, however, that not all commands on the *Sharding Reference* (page 822) page modifies the cluster meta-data.

Procedure The downgrade procedure for a sharded cluster reverses the order of the upgrade procedure.

1. Turn off the *balancer* (page 758) in the sharded cluster, as described in *Disable the Balancer* (page 802).
2. Downgrade each shard, one at a time. For each shard,
   (a) Downgrade the `mongod` secondaries *before* downgrading the primary.
   (b) To downgrade the primary, run `replSetStepDown` and downgrade.
3. Downgrade all 3 `mongod` config server instances, leaving the *first* system in the *`mongos --configdb`* argument to downgrade *last*.
4. Downgrade and restart each `mongos`, one at a time. The downgrade process is a binary drop-in replacement.
5. Turn on the balancer, as described in *Enable the Balancer* (page 803).

Downgrade Procedure Once upgraded to MongoDB 2.6, you **cannot** downgrade to **any** version earlier than MongoDB 2.4. If you have `text` or `2dsphere` indexes, you can only downgrade to MongoDB 2.4.10 or later.

Except as described on this page, moving between 2.4 and 2.6 is a drop-in replacement:

Step 1: Stop the existing `mongod` instance. For example, on Linux, run 2.6 `mongod` with the *`--shutdown`* option as follows:

```
mongod --dbpath /var/mongod/data --shutdown
```

Replace `/var/mongod/data` with your MongoDB `dbPath`. See also the *Stop mongod Processes* (page 324) for alternate methods of stopping a `mongod` instance.

Step 2: Start the new `mongod` instance. Ensure you start the 2.4 `mongod` with the same `dbPath`:

```
mongod --dbpath /var/mongod/data
```

Replace `/var/mongod/data` with your MongoDB `dbPath`.

See *Upgrade MongoDB to 2.6* (page 1010) for full upgrade instructions.

Download

To download MongoDB 2.6, go to the downloads page[1422].

Other Resources

- All JIRA issues resolved in 2.6[1423].
- All Third Party License Notices[1424].

15.2.3 Release Notes for MongoDB 2.4

March 19, 2013

On this page

- Minor Releases (page 1022)
- Major New Features (page 1029)
- Security Enhancements (page 1030)
- Performance Improvements (page 1030)
- Enterprise (page 1036)
- Additional Information (page 1037)

MongoDB 2.4 includes enhanced geospatial support, switch to V8 JavaScript engine, security enhancements, and text search (beta) and hashed index.

Minor Releases

2.4 Changelog

On this page

- 2.4.14 (page 1023)
- 2.4.13 - Changes (page 1023)
- 2.4.12 - Changes (page 1023)
- 2.4.11 - Changes (page 1023)
- 2.4.10 - Changes (page 1024)
- Previous Releases (page 1025)

[1422] http://www.mongodb.org/downloads
[1423] https://jira.mongodb.org/secure/IssueNavigator.jspa?reset=true&jqlQuery=project+%3D+SERVER+AND+fixVersion+in+%28%222.5.0%22%2C+%222.5.1%22%2
rc1%22%2C+%222.6.0-rc2%22%2C+%222.6.0-rc3%22%29
[1424] https://github.com/mongodb/mongo/blob/v2.6/distsrc/THIRD-PARTY-NOTICES

2.4.14

- Packaging: Init script sets process ulimit to different value compared to documentation (SERVER-17780[1425])
- Security: Compute BinData length in v8 (SERVER-17647[1426])
- Build: Upgrade PCRE Version from 8.30 to Latest (SERVER-17252[1427])

2.4.13 - Changes

- Security: Enforce BSON BinData length validation (SERVER-17278[1428])
- Security: Disable SSLv3 ciphers (SERVER-15673[1429])
- Networking: Improve BSON validation (SERVER-17264[1430])

2.4.12 - Changes

- Sharding: Sharded connection cleanup on setup error can crash mongos (SERVER-15056[1431])
- Sharding: "type 7" (OID) error when acquiring distributed lock for first time (SERVER-13616[1432])
- Storage: explicitly zero .ns files on creation (SERVER-15369[1433])
- Storage: partially written journal last section causes recovery to fail (SERVER-15111[1434])

2.4.11 - Changes

- Security: Potential information leak (SERVER-14268[1435])
- Replication: `_id` with `$prefix` field causes replication failure due to unvalidated insert (SERVER-12209[1436])
- Sharding: Invalid access: seg fault in `SplitChunkCommand::run` (SERVER-14342[1437])
- Indexing: Creating descending index on `_id` can corrupt namespace (SERVER-14833[1438])
- Text Search: Updates to documents with text-indexed fields may lead to incorrect entries (SERVER-14738[1439])
- Build: Add SCons flag to override treating all warnings as errors (SERVER-13724[1440])
- Packaging: Fix mongodb enterprise 2.4 init script to allow multiple processes per host (SERVER-14336[1441])
- JavaScript: Do not store native function pointer as a property in function prototype (SERVER-14254[1442])

[1425] https://jira.mongodb.org/browse/SERVER-17780
[1426] https://jira.mongodb.org/browse/SERVER-17647
[1427] https://jira.mongodb.org/browse/SERVER-17252
[1428] https://jira.mongodb.org/browse/SERVER-17278
[1429] https://jira.mongodb.org/browse/SERVER-15673
[1430] https://jira.mongodb.org/browse/SERVER-17264
[1431] https://jira.mongodb.org/browse/SERVER-15056
[1432] https://jira.mongodb.org/browse/SERVER-13616
[1433] https://jira.mongodb.org/browse/SERVER-15369
[1434] https://jira.mongodb.org/browse/SERVER-15111
[1435] https://jira.mongodb.org/browse/SERVER-14268
[1436] https://jira.mongodb.org/browse/SERVER-12209
[1437] https://jira.mongodb.org/browse/SERVER-14342
[1438] https://jira.mongodb.org/browse/SERVER-14833
[1439] https://jira.mongodb.org/browse/SERVER-14738
[1440] https://jira.mongodb.org/browse/SERVER-13724
[1441] https://jira.mongodb.org/browse/SERVER-14336
[1442] https://jira.mongodb.org/browse/SERVER-14254

2.4.10 - Changes

- Indexes: Fixed issue that can cause index corruption when building indexes concurrently (SERVER-12990[1443])
- Indexes: Fixed issue that can cause index corruption when shutting down secondary node during index build (SERVER-12956[1444])
- Indexes: Mongod now recognizes incompatible "future" text and geo index versions and exits gracefully (SERVER-12914[1445])
- Indexes: Fixed issue that can cause secondaries to fail replication when building the same index multiple times concurrently (SERVER-12662[1446])
- Indexes: Fixed issue that can cause index corruption on the tenth index in a collection if the index build fails (SERVER-12481[1447])
- Indexes: Introduced versioning for text and geo indexes to ensure backwards compatibility (SERVER-12175[1448])
- Indexes: Disallowed building indexes on the system.indexes collection, which can lead to initial sync failure on secondaries (SERVER-10231[1449])
- Sharding: Avoid frequent immediate balancer retries when config servers are out of sync (SERVER-12908[1450])
- Sharding: Add indexes to locks collection on config servers to avoid long queries in case of large numbers of collections (SERVER-12548[1451])
- Sharding: Fixed issue that can corrupt the config metadata cache when sharding collections concurrently (SERVER-12515[1452])
- Sharding: Don't move chunks created on collections with a hashed shard key if the collection already contains data (SERVER-9259[1453])
- Replication: Fixed issue where node appears to be down in a replica set during a compact operation (SERVER-12264[1454])
- Replication: Fixed issue that could cause delays in elections when a node is not vetoing an election (SERVER-12170[1455])
- Replication: Step down all primaries if multiple primaries are detected in replica set to ensure correct election result (SERVER-10793[1456])
- Replication: Upon clock skew detection, secondaries will switch to sync directly from the primary to avoid sync cycles (SERVER-8375[1457])
- Runtime: The SIGXCPU signal is now caught and mongod writes a log message and exits gracefully (SERVER-12034[1458])

[1443] https://jira.mongodb.org/browse/SERVER-12990
[1444] https://jira.mongodb.org/browse/SERVER-12956
[1445] https://jira.mongodb.org/browse/SERVER-12914
[1446] https://jira.mongodb.org/browse/SERVER-12662
[1447] https://jira.mongodb.org/browse/SERVER-12481
[1448] https://jira.mongodb.org/browse/SERVER-12175
[1449] https://jira.mongodb.org/browse/SERVER-10231
[1450] https://jira.mongodb.org/browse/SERVER-12908
[1451] https://jira.mongodb.org/browse/SERVER-12548
[1452] https://jira.mongodb.org/browse/SERVER-12515
[1453] https://jira.mongodb.org/browse/SERVER-9259
[1454] https://jira.mongodb.org/browse/SERVER-12264
[1455] https://jira.mongodb.org/browse/SERVER-12170
[1456] https://jira.mongodb.org/browse/SERVER-10793
[1457] https://jira.mongodb.org/browse/SERVER-8375
[1458] https://jira.mongodb.org/browse/SERVER-12034

- Runtime: Fixed issue where mongod fails to start on Linux when /sys/dev/block directory is not readable (SERVER-9248[1459])
- Windows: No longer zero-fill newly allocated files on systems other than Windows 7 or Windows Server 2008 R2 (SERVER-8480[1460])
- GridFS: Chunk size is decreased to 255 kB (from 256 kB) to avoid overhead with usePowerOf2Sizes option (SERVER-13331[1461])
- SNMP: Fixed MIB file validation under smilint (SERVER-12487[1462])
- Shell: Fixed issue in V8 memory allocation that could cause long-running shell commands to crash (SERVER-11871[1463])
- Shell: Fixed memory leak in the md5sumFile shell utility method (SERVER-11560[1464])

Previous Releases

- All 2.4.9 improvements[1465].
- All 2.4.8 improvements[1466].
- All 2.4.7 improvements[1467].
- All 2.4.6 improvements[1468].
- All 2.4.5 improvements[1469].
- All 2.4.4 improvements[1470].
- All 2.4.3 improvements[1471].
- All 2.4.2 improvements[1472]
- All 2.4.1 improvements[1473].

2.4.14 – April 28, 2015

- Init script sets process ulimit to different value compared to documentation SERVER-17780[1474]
- Compute BinData length in v8 SERVER-17647[1475]
- Upgrade PCRE Version from 8.30 to Latest SERVER-17252[1476]
- *2.4.14 Changelog* (page 1023).

[1459] https://jira.mongodb.org/browse/SERVER-9248
[1460] https://jira.mongodb.org/browse/SERVER-8480
[1461] https://jira.mongodb.org/browse/SERVER-13331
[1462] https://jira.mongodb.org/browse/SERVER-12487
[1463] https://jira.mongodb.org/browse/SERVER-11871
[1464] https://jira.mongodb.org/browse/SERVER-11560
[1465] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.9%22%20AND%20project%20%3D%20SERVER
[1466] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.8%22%20AND%20project%20%3D%20SERVER
[1467] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.7%22%20AND%20project%20%3D%20SERVER
[1468] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.6%22%20AND%20project%20%3D%20SERVER
[1469] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.5%22%20AND%20project%20%3D%20SERVER
[1470] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.4%22%20AND%20project%20%3D%20SERVER
[1471] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.3%22%20AND%20project%20%3D%20SERVER
[1472] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.2%22%20AND%20project%20%3D%20SERVER
[1473] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.1%22%20AND%20project%20%3D%20SERVER
[1474] https://jira.mongodb.org/browse/SERVER-17780
[1475] https://jira.mongodb.org/browse/SERVER-17647
[1476] https://jira.mongodb.org/browse/SERVER-17252

- All 2.4.14 improvements[1477].

2.4.13 – February 25, 2015

- Enforce BSON BinData length validation SERVER-17278[1478]
- Disable SSLv3 ciphers SERVER-15673[1479]
- Improve BSON validation SERVER-17264[1480]
- *2.4.13 Changelog* (page 1023).
- All 2.4.13 improvements[1481].

2.4.12 – October 16, 2014

- Partially written journal last section causes recovery to fail SERVER-15111[1482].
- Explicitly zero `.ns` files on creation SERVER-15369[1483].
- *2.4.12 Changelog* (page 1023).
- All 2.4.12 improvements[1484].

2.4.11 – August 18, 2014

- Fixed potential information leak: SERVER-14268[1485].
- Resolved issue were an `_id` with a `$prefix` field caused replication failure due to unvalidated insert SERVER-12209[1486].
- Addressed issue where updates to documents with text-indexed fields could lead to incorrect entries SERVER-14738[1487].
- Resolved issue where creating descending index on `_id` could corrupt namespace SERVER-14833[1488].
- *2.4.11 Changelog* (page 1023).
- All 2.4.11 improvements[1489].

2.4.10 – April 4, 2014

- Performs fast file allocation on Windows when available SERVER-8480[1490].

[1477] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.14%22%20AND%20project%20%3D%20SERVER
[1478] https://jira.mongodb.org/browse/SERVER-17278
[1479] https://jira.mongodb.org/browse/SERVER-15673
[1480] https://jira.mongodb.org/browse/SERVER-17264
[1481] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.13%22%20AND%20project%20%3D%20SERVER
[1482] https://jira.mongodb.org/browse/SERVER-15111
[1483] https://jira.mongodb.org/browse/SERVER-15369
[1484] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.12%22%20AND%20project%20%3D%20SERVER
[1485] https://jira.mongodb.org/browse/SERVER-14268
[1486] https://jira.mongodb.org/browse/SERVER-12209
[1487] https://jira.mongodb.org/browse/SERVER-14738
[1488] https://jira.mongodb.org/browse/SERVER-14833
[1489] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.11%22%20AND%20project%20%3D%20SERVER
[1490] https://jira.mongodb.org/browse/SERVER-8480

- Start elections if more than one primary is detected SERVER-10793[1491].
- Changes to allow safe downgrading from v2.6 to v2.4 SERVER-12914[1492], SERVER-12175[1493].
- Fixes for edge cases in index creation SERVER-12481[1494], SERVER-12956[1495].
- *2.4.10 Changelog* (page 1024).
- All 2.4.10 improvements[1496].

2.4.9 – January 10, 2014

- Fix for instances where `mongos` incorrectly reports a successful write SERVER-12146[1497].
- Make non-primary read preferences consistent with `slaveOK` versioning logic SERVER-11971[1498].
- Allow new sharded cluster connections to read from secondaries when primary is down SERVER-7246[1499].
- All 2.4.9 improvements[1500].

2.4.8 – November 1, 2013

- Increase future compatibility for 2.6 authorization features SERVER-11478[1501].
- Fix `dbhash` cache issue for config servers SERVER-11421[1502].
- All 2.4.8 improvements[1503].

2.4.7 – October 21, 2013

- Fixed over-aggressive caching of V8 Isolates SERVER-10596[1504].
- Removed extraneous initial count during mapReduce SERVER-9907[1505].
- Cache results of dbhash command SERVER-11021[1506].
- Fixed memory leak in aggregation SERVER-10554[1507].
- All 2.4.7 improvements[1508].

[1491] https://jira.mongodb.org/browse/SERVER-10793
[1492] https://jira.mongodb.org/browse/SERVER-12914
[1493] https://jira.mongodb.org/browse/SERVER-12175
[1494] https://jira.mongodb.org/browse/SERVER-12481
[1495] https://jira.mongodb.org/browse/SERVER-12956
[1496] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.10%22%20AND%20project%20%3D%20SERVER
[1497] https://jira.mongodb.org/browse/SERVER-12146
[1498] https://jira.mongodb.org/browse/SERVER-11971
[1499] https://jira.mongodb.org/browse/SERVER-7246
[1500] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.9%22%20AND%20project%20%3D%20SERVER
[1501] https://jira.mongodb.org/browse/SERVER-11478
[1502] https://jira.mongodb.org/browse/SERVER-11421
[1503] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.8%22%20AND%20project%20%3D%20SERVER
[1504] https://jira.mongodb.org/browse/SERVER-10596
[1505] https://jira.mongodb.org/browse/SERVER-9907
[1506] https://jira.mongodb.org/browse/SERVER-11021
[1507] https://jira.mongodb.org/browse/SERVER-10554
[1508] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.7%22%20AND%20project%20%3D%20SERVER

2.4.6 – August 20, 2013

- Fix for possible loss of documents during the chunk migration process if a document in the chunk is very large SERVER-10478[1509].
- Fix for C++ client shutdown issues SERVER-8891[1510].
- Improved replication robustness in presence of high network latency SERVER-10085[1511].
- Improved Solaris support SERVER-9832[1512], SERVER-9786[1513], and SERVER-7080[1514].
- All 2.4.6 improvements[1515].

2.4.5 – July 3, 2013

- Fix for CVE-2013-4650 Improperly grant user system privileges on databases other than local SERVER-9983[1516].
- Fix for CVE-2013-3969 Remotely triggered segmentation fault in Javascript engine SERVER-9878[1517].
- Fix to prevent identical background indexes from being built SERVER-9856[1518].
- Config server performance improvements SERVER-9864[1519] and SERVER-5442[1520].
- Improved initial sync resilience to network failure SERVER-9853[1521].
- All 2.4.5 improvements[1522].

2.4.4 – June 4, 2013

- Performance fix for Windows version SERVER-9721[1523]
- Fix for config upgrade failure SERVER-9661[1524].
- Migration to Cyrus SASL library for MongoDB Enterprise SERVER-8813[1525].
- All 2.4.4 improvements[1526].

2.4.3 – April 23, 2013

- Fix for `mongo` shell ignoring modified object's `_id` field SERVER-9385[1527].

[1509] https://jira.mongodb.org/browse/SERVER-10478
[1510] https://jira.mongodb.org/browse/SERVER-8891
[1511] https://jira.mongodb.org/browse/SERVER-10085
[1512] https://jira.mongodb.org/browse/SERVER-9832
[1513] https://jira.mongodb.org/browse/SERVER-9786
[1514] https://jira.mongodb.org/browse/SERVER-7080
[1515] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.6%22%20AND%20project%20%3D%20SERVER
[1516] https://jira.mongodb.org/browse/SERVER-9983
[1517] https://jira.mongodb.org/browse/SERVER-9878
[1518] https://jira.mongodb.org/browse/SERVER-9856
[1519] https://jira.mongodb.org/browse/SERVER-9864
[1520] https://jira.mongodb.org/browse/SERVER-5442
[1521] https://jira.mongodb.org/browse/SERVER-9853
[1522] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.5%22%20AND%20project%20%3D%20SERVER
[1523] https://jira.mongodb.org/browse/SERVER-9721
[1524] https://jira.mongodb.org/browse/SERVER-9661
[1525] https://jira.mongodb.org/browse/SERVER-8813
[1526] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.4%22%20AND%20project%20%3D%20SERVER
[1527] https://jira.mongodb.org/browse/SERVER-9385

- Fix for race condition in log rotation SERVER-4739[1528].
- Fix for `copydb` command with authorization in a sharded cluster SERVER-9093[1529].
- All 2.4.3 improvements[1530].

2.4.2 – April 17, 2013

- Several V8 memory leak and performance fixes SERVER-9267[1531] and SERVER-9230[1532].
- Fix for upgrading sharded clusters SERVER-9125[1533].
- Fix for high volume connection crash SERVER-9014[1534].
- All 2.4.2 improvements[1535]

2.4.1 – April 17, 2013

- Fix for losing index changes during initial sync SERVER-9087[1536]
- All 2.4.1 improvements[1537].

Major New Features

The following changes in MongoDB affect both standard and Enterprise editions:

Text Search

Add support for text search of content in MongoDB databases as a *beta* feature. See *Text Indexes* (page 533) for more information.

Geospatial Support Enhancements

- Add new *2dsphere index* (page 543). The new index supports GeoJSON[1538] objects `Point`, `LineString`, and `Polygon`. See *2dsphere Indexes* (page 543) and `https://docs.mongodb.org/manual/applications/geospatial-indexes`.
- Introduce operators `$geometry`, `$geoWithin` and `$geoIntersects` to work with the GeoJSON data.

Hashed Index

Add new *hashed index* (page 564) to index documents using hashes of field values. When used to index a shard key, the hashed index ensures an evenly distributed shard key. See also *Hashed Shard Keys* (page 748).

[1528] https://jira.mongodb.org/browse/SERVER-4739
[1529] https://jira.mongodb.org/browse/SERVER-9093
[1530] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.3%22%20AND%20project%20%3D%20SERVER
[1531] https://jira.mongodb.org/browse/SERVER-9267
[1532] https://jira.mongodb.org/browse/SERVER-9230
[1533] https://jira.mongodb.org/browse/SERVER-9125
[1534] https://jira.mongodb.org/browse/SERVER-9014
[1535] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.2%22%20AND%20project%20%3D%20SERVER
[1536] https://jira.mongodb.org/browse/SERVER-9087
[1537] https://jira.mongodb.org/issues/?jql=fixVersion%20%3D%20%222.4.1%22%20AND%20project%20%3D%20SERVER
[1538] http://geojson.org/geojson-spec.html

Improvements to the Aggregation Framework

- Improve support for geospatial queries. See the `$geoWithin` operator and the `$geoNear` pipeline stage.
- Improve sort efficiency when the `$sort` stage immediately precedes a `$limit` in the pipeline.
- Add new operators `$millisecond` and `$concat` and modify how `$min` operator processes `null` values.

Changes to Update Operators

- Add new `$setOnInsert` operator for use with `upsert` .
- Enhance functionality of the `$push` operator, supporting its use with the `$each`, the `$sort`, and the `$slice` modifiers.

Additional Limitations for Map-Reduce and `$where` Operations

The `mapReduce` command, `group` command, and the `$where` operator expressions cannot access certain global functions or properties, such as `db`, that are available in the `mongo` shell. See the individual command or operator for details.

Improvements to `serverStatus` Command

Provide additional metrics and customization for the `serverStatus` command. See `db.serverStatus()` and `serverStatus` for more information.

Security Enhancements

- Introduce a role-based access control system User Privileges[1539] now use a new format for `Privilege Documents`.
- Enforce uniqueness of the user in user privilege documents per database. Previous versions of MongoDB did not enforce this requirement, and existing databases may have duplicates.
- Support encrypted connections using SSL certificates signed by a Certificate Authority. See *Configure mongod and mongos for TLS/SSL* (page 451).

For more information on security and risk management strategies, see *MongoDB Security Practices and Procedures* (page 391).

Performance Improvements

V8 JavaScript Engine

JavaScript Changes in MongoDB 2.4

On this page

- Improved Concurrency (page 1031)
- Modernized JavaScript Implementation (ES5) (page 1031)
- Removed Non-Standard SpiderMonkey Features (page 1031)

[1539] https://docs.mongodb.org/v2.4/reference/user-privileges

Consider the following impacts of *V8 JavaScript Engine* (page 1030) in MongoDB 2.4:

Tip

Use the new `interpreterVersion()` method in the `mongo` shell and the `javascriptEngine` field in the output of `db.serverBuildInfo()` to determine which JavaScript engine a MongoDB binary uses.

Improved Concurrency Previously, MongoDB operations that required the JavaScript interpreter had to acquire a lock, and a single `mongod` could only run a single JavaScript operation at a time. The switch to V8 improves concurrency by permitting multiple JavaScript operations to run at the same time.

Modernized JavaScript Implementation (ES5) The 5th edition of ECMAscript[1540], abbreviated as ES5, adds many new language features, including:

- standardized JSON[1541],
- strict mode[1542],
- function.bind()[1543],
- array extensions[1544], and
- getters and setters.

With V8, MongoDB supports the ES5 implementation of Javascript with the following exceptions.

Note: The following features do not work as expected on documents **returned from MongoDB queries**:

- `Object.seal()` throws an exception on documents returned from MongoDB queries.
- `Object.freeze()` throws an exception on documents returned from MongoDB queries.
- `Object.preventExtensions()` incorrectly allows the addition of new properties on documents returned from MongoDB queries.
- `enumerable` properties, when added to documents returned from MongoDB queries, are not saved during write operations.

See SERVER-8216[1545], SERVER-8223[1546], SERVER-8215[1547], and SERVER-8214[1548] for more information.

For objects that have not been returned from MongoDB queries, the features work as expected.

Removed Non-Standard SpiderMonkey Features V8 does **not** support the following *non-standard* SpiderMonkey[1549] JavaScript extensions, previously supported by MongoDB's use of SpiderMonkey as its JavaScript engine.

[1540] http://www.ecma-international.org/publications/standards/Ecma-262.htm
[1541] http://www.ecma-international.org/ecma-262/5.1/#sec-15.12.1
[1542] http://www.ecma-international.org/ecma-262/5.1/#sec-4.2.2
[1543] http://www.ecma-international.org/ecma-262/5.1/#sec-15.3.4.5
[1544] http://www.ecma-international.org/ecma-262/5.1/#sec-15.4.4.16
[1545] https://jira.mongodb.org/browse/SERVER-8216
[1546] https://jira.mongodb.org/browse/SERVER-8223
[1547] https://jira.mongodb.org/browse/SERVER-8215
[1548] https://jira.mongodb.org/browse/SERVER-8214
[1549] https://developer.mozilla.org/en-US/docs/SpiderMonkey

E4X Extensions V8 does not support the *non-standard* E4X[1550] extensions. E4X provides a native XML[1551] object to the JavaScript language and adds the syntax for embedding literal XML documents in JavaScript code.

You need to use alternative XML processing if you used any of the following constructors/methods:

- `XML()`
- `Namespace()`
- `QName()`
- `XMLList()`
- `isXMLName()`

Destructuring Assignment V8 does not support the non-standard destructuring assignments. Destructuring assignment "extract[s] data from arrays or objects using a syntax that mirrors the construction of array and object literals." - Mozilla docs[1552]

Example

The following destructuring assignment is **invalid** with V8 and throws a `SyntaxError`:

```
original = [4, 8, 15];
var [b, ,c] = a;  // <== destructuring assignment
print(b) // 4
print(c) // 15
```

`Iterator()`, `StopIteration()`, and Generators V8 does not support Iterator(), StopIteration(), and generators[1553].

`InternalError()` V8 does not support `InternalError()`. Use `Error()` instead.

`for each...in` Construct V8 does not support the use of for each...in[1554] construct. Use `for (var x in y)` construct instead.

Example

The following `for each (var x in y)` construct is **invalid** with V8:

```
var o = { name: 'MongoDB', version: 2.4 };

for each (var value in o) {
  print(value);
}
```

Instead, in version 2.4, you can use the `for (var x in y)` construct:

```
var o = { name: 'MongoDB', version: 2.4 };

for (var prop in o) {
  var value = o[prop];
```

[1550] https://developer.mozilla.org/en-US/docs/E4X
[1551] https://developer.mozilla.org/en-US/docs/E4X/Processing_XML_with_E4X
[1552] https://developer.mozilla.org/en-US/docs/JavaScript/New_in_JavaScript/1.7#Destructuring_assignment_(Merge_into_own_page.2Fsection)
[1553] https://developer.mozilla.org/en-US/docs/JavaScript/Guide/Iterators_and_Generators
[1554] https://developer.mozilla.org/en-US/docs/JavaScript/Reference/Statements/for_each...in

```
  print(value);
}
```

You can also use the array *instance* method `forEach()` with the ES5 method `Object.keys()`:

```
Object.keys(o).forEach(function (key) {
  var value = o[key];
  print(value);
});
```

Array Comprehension V8 does not support Array comprehensions[1555].

Use other methods such as the `Array` instance methods `map()`, `filter()`, or `forEach()`.

Example

With V8, the following array comprehension is **invalid**:

```
var a = { w: 1, x: 2, y: 3, z: 4 }

var arr = [i * i for each (i in a) if (i > 2)]
printjson(arr)
```

Instead, you can implement using the `Array` *instance* method `forEach()` and the ES5 method `Object.keys()`:

```
var a = { w: 1, x: 2, y: 3, z: 4 }

var arr = [];
Object.keys(a).forEach(function (key) {
  var val = a[key];
  if (val > 2) arr.push(val * val);
})
printjson(arr)
```

Note: The new logic uses the `Array` *instance* method `forEach()` and not the *generic* method `Array.forEach()`; V8 does **not** support `Array` *generic* methods. See *Array Generic Methods* (page 1035) for more information.

Multiple Catch Blocks V8 does not support multiple `catch` blocks and will throw a `SyntaxError`.

Example

The following multiple catch blocks is **invalid** with V8 and will throw `"SyntaxError: Unexpected token if"`:

```
try {
  something()
} catch (err if err instanceof SomeError) {
  print('some error')
} catch (err) {
  print('standard error')
}
```

[1555] https://developer.mozilla.org/en-US/docs/JavaScript/Guide/Predefined_Core_Objects#Array_comprehensions

Conditional Function Definition V8 will produce different outcomes than SpiderMonkey with conditional function definitions[1556].

Example

The following conditional function definition produces different outcomes in SpiderMonkey versus V8:

```
function test () {
   if (false) {
      function go () {};
   }
   print(typeof go)
}
```

With SpiderMonkey, the conditional function outputs `undefined`, whereas with V8, the conditional function outputs `function`.

If your code defines functions this way, it is highly recommended that you refactor the code. The following example refactors the conditional function definition to work in both SpiderMonkey and V8.

```
function test () {
  var go;
  if (false) {
    go = function () {}
  }
  print(typeof go)
}
```

The refactored code outputs `undefined` in both SpiderMonkey and V8.

Note: ECMAscript prohibits conditional function definitions. To force V8 to throw an `Error`, enable strict mode[1557].

```
function test () {
  'use strict';

  if (false) {
    function go () {}
  }
}
```

The JavaScript code throws the following syntax error:

```
SyntaxError: In strict mode code, functions can only be declared at top level or immediately within a
```

String Generic Methods V8 does not support String generics[1558]. String generics are a set of methods on the `String` class that mirror instance methods.

Example

The following use of the generic method `String.toLowerCase()` is **invalid** with V8:

```
var name = 'MongoDB';

var lower = String.toLowerCase(name);
```

[1556] https://developer.mozilla.org/en-US/docs/JavaScript/Guide/Functions
[1557] http://www.nczonline.net/blog/2012/03/13/its-time-to-start-using-javascript-strict-mode/
[1558] https://developer.mozilla.org/en-US/docs/JavaScript/Reference/Global_Objects/String#String_generic_methods

With V8, use the `String` instance method `toLowerCase()` available through an *instance* of the `String` class instead:

```
var name = 'MongoDB';

var lower = name.toLowerCase();
print(name + ' becomes ' + lower);
```

With V8, use the `String` *instance* methods instead of following *generic* methods:

`String.charAt()`	`String.quote()`	`String.toLocaleLowerCase()`
`String.charCodeAt()`	`String.replace()`	`String.toLocaleUpperCase()`
`String.concat()`	`String.search()`	`String.toLowerCase()`
`String.endsWith()`	`String.slice()`	`String.toUpperCase()`
`String.indexOf()`	`String.split()`	`String.trim()`
`String.lastIndexOf()`	`String.startsWith()`	`String.trimLeft()`
`String.localeCompare()`	`String.substr()`	`String.trimRight()`
`String.match()`	`String.substring()`	

Array Generic Methods V8 does not support Array generic methods[1559]. Array generics are a set of methods on the `Array` class that mirror instance methods.

Example

The following use of the generic method `Array.every()` is **invalid** with V8:

```
var arr = [4, 8, 15, 16, 23, 42];

function isEven (val) {
   return 0 === val % 2;
}

var allEven = Array.every(arr, isEven);
print(allEven);
```

With V8, use the `Array` instance method `every()` available through an *instance* of the `Array` class instead:

```
var allEven = arr.every(isEven);
print(allEven);
```

With V8, use the `Array` *instance* methods instead of the following *generic* methods:

`Array.concat()`	`Array.lastIndexOf()`	`Array.slice()`
`Array.every()`	`Array.map()`	`Array.some()`
`Array.filter()`	`Array.pop()`	`Array.sort()`
`Array.forEach()`	`Array.push()`	`Array.splice()`
`Array.indexOf()`	`Array.reverse()`	`Array.unshift()`
`Array.join()`	`Array.shift()`	

Array Instance Method `toSource()` V8 does not support the `Array` instance method toSource()[1560]. Use the `Array` instance method `toString()` instead.

[1559] https://developer.mozilla.org/en-US/docs/JavaScript/Reference/Global_Objects/Array#Array_generic_methods
[1560] https://developer.mozilla.org/en-US/docs/JavaScript/Reference/Global_Objects/Array/toSource

`uneval()` V8 does not support the non-standard method `uneval()`. Use the standardized JSON.stringify()[1561] method instead.

Change default JavaScript engine from SpiderMonkey to V8. The change provides improved concurrency for JavaScript operations, modernized JavaScript implementation, and the removal of non-standard SpiderMonkey features, and affects all JavaScript behavior including the commands `mapReduce`, `group`, and `eval` and the query operator `$where`.

See *JavaScript Changes in MongoDB 2.4* (page 1030) for more information about all changes .

BSON Document Validation Enabled by Default for `mongod` and `mongorestore`

Enable basic *BSON* object validation for `mongod` and `mongorestore` when writing to MongoDB data files. See `wireObjectCheck` for details.

Index Build Enhancements

- Add support for multiple concurrent index builds in the background by a single `mongod` instance. See *building indexes in the background* (page 577) for more information on background index builds.
- Allow the `db.killOp()` method to terminate a foreground index build.
- Improve index validation during index creation. See *Compatibility and Index Type Changes in MongoDB 2.4* (page 1044) for more information.

Set Parameters as Command Line Options

Provide `--setParameter` as a command line option for `mongos` and `mongod`. See `mongod` and `mongos` for list of available options for `setParameter`.

Changed Replication Behavior for Chunk Migration

By default, each document move during *chunk migration* (page 760) in a *sharded cluster* propagates to at least one secondary before the balancer proceeds with its next operation. See *Chunk Migration and Replication* (page 761).

Improved Chunk Migration Queue Behavior

Increase performance for moving multiple chunks off an overloaded shard. The balancer no longer waits for the current migration's delete phase to complete before starting the next chunk migration. See *Chunk Migration Queuing* (page 761) for details.

Enterprise

The following changes are specific to MongoDB Enterprise Editions:

[1561] https://developer.mozilla.org/en-US/docs/JavaScript/Reference/Global_Objects/JSON/stringify

SASL Library Change

In 2.4.4, MongoDB Enterprise uses Cyrus SASL. Earlier 2.4 Enterprise versions use GNU SASL (`libgsasl`). To upgrade to 2.4.4 MongoDB Enterprise or greater, you **must** install all package dependencies related to this change, including the appropriate Cyrus SASL `GSSAPI` library. See *Install MongoDB Enterprise* (page 49) for details of the dependencies.

New Modular Authentication System with Support for Kerberos

In 2.4, the MongoDB Enterprise now supports authentication via a Kerberos mechanism. See *Configure MongoDB with Kerberos Authentication on Linux* (page 409) for more information. For drivers that provide support for Kerberos authentication to MongoDB, refer to *Driver Support* (page 408).

For more information on security and risk management strategies, see *MongoDB Security Practices and Procedures* (page 391).

Additional Information

Platform Notes

For OS X, MongoDB 2.4 only supports OS X versions 10.6 (Snow Leopard) and later. There are no other platform support changes in MongoDB 2.4. See the downloads page[1562] for more information on platform support.

Upgrade Process

Upgrade MongoDB to 2.4

On this page

- Upgrade Recommendations and Checklist (page 1037)
- Upgrade Standalone `mongod` Instance to MongoDB 2.4 (page 1038)
- Upgrade a Replica Set from MongoDB 2.2 to MongoDB 2.4 (page 1038)
- Upgrade a Sharded Cluster from MongoDB 2.2 to MongoDB 2.4 (page 1038)
- Rolling Upgrade Limitation for 2.2.0 Deployments Running with `auth` Enabled (page 1042)
- Upgrade from 2.3 to 2.4 (page 1042)
- Downgrade MongoDB from 2.4 to Previous Versions (page 1042)

In the general case, the upgrade from MongoDB 2.2 to 2.4 is a binary-compatible "drop-in" upgrade: shut down the `mongod` instances and replace them with `mongod` instances running 2.4. **However**, before you attempt any upgrade please familiarize yourself with the content of this document, particularly the procedure for *upgrading sharded clusters* (page 1038) and the considerations for *reverting to 2.2 after running 2.4* (page 1042).

Upgrade Recommendations and Checklist When upgrading, consider the following:

- For all deployments using authentication, upgrade the drivers (i.e. client libraries), before upgrading the `mongod` instance or instances.
- To upgrade to 2.4 sharded clusters *must* upgrade following the *meta-data upgrade procedure* (page 1038).
- If you're using 2.2.0 and running with `authorization` enabled, you will need to upgrade first to 2.2.1 and then upgrade to 2.4. See *Rolling Upgrade Limitation for 2.2.0 Deployments Running with auth Enabled* (page 1042).

[1562] http://www.mongodb.org/downloads/

- If you have `system.users` documents (i.e. for `authorization`) that you created before 2.4 you *must* ensure that there are no duplicate values for the `user` field in the `system.users` collection in *any* database. If you *do* have documents with duplicate user fields, you must remove them before upgrading.

  See *Security Enhancements* (page 1030) for more information.

Upgrade Standalone `mongod` Instance to MongoDB 2.4

1. Download binaries of the latest release in the 2.4 series from the MongoDB Download Page[1563]. See *Install MongoDB* (page 21) for more information.
2. Shutdown your `mongod` instance. Replace the existing binary with the 2.4 `mongod` binary and restart `mongod`.

Upgrade a Replica Set from MongoDB 2.2 to MongoDB 2.4 You can upgrade to 2.4 by performing a "rolling" upgrade of the set by upgrading the members individually while the other members are available to minimize downtime. Use the following procedure:

1. Upgrade the *secondary* members of the set one at a time by shutting down the `mongod` and replacing the 2.2 binary with the 2.4 binary. After upgrading a `mongod` instance, wait for the member to recover to `SECONDARY` state before upgrading the next instance. To check the member's state, issue `rs.status()` in the `mongo` shell.
2. Use the `mongo` shell method `rs.stepDown()` to step down the *primary* to allow the normal *failover* (page 644) procedure. `rs.stepDown()` expedites the failover procedure and is preferable to shutting down the primary directly.

   Once the primary has stepped down and another member has assumed `PRIMARY` state, as observed in the output of `rs.status()`, shut down the previous primary and replace `mongod` binary with the 2.4 binary and start the new process.

   Note: Replica set failover is not instant but will render the set unavailable to read or accept writes until the failover process completes. Typically this takes 10 seconds or more. You may wish to plan the upgrade during a predefined maintenance window.

Upgrade a Sharded Cluster from MongoDB 2.2 to MongoDB 2.4

Important: Only upgrade sharded clusters to 2.4 if **all** members of the cluster are currently running instances of 2.2. The only supported upgrade path for sharded clusters running 2.0 is via 2.2.

Overview Upgrading a *sharded cluster* from MongoDB version 2.2 to 2.4 (or 2.3) requires that you run a 2.4 `mongos` with the `--upgrade` option, described in this procedure. The upgrade process does not require downtime.

The upgrade to MongoDB 2.4 adds epochs to the meta-data for all collections and chunks in the existing cluster. MongoDB 2.2 processes are capable of handling epochs, even though 2.2 did not require them. This procedure applies only to upgrades from version 2.2. Earlier versions of MongoDB do not correctly handle epochs. See *Cluster Meta-data Upgrade* (page 1039) for more information.

After completing the meta-data upgrade you can fully upgrade the components of the cluster. With the balancer disabled:

- Upgrade all `mongos` instances in the cluster.
- Upgrade all 3 `mongod` config server instances.
- Upgrade the `mongod` instances for each shard, one at a time.

[1563] http://www.mongodb.org/downloads

See *Upgrade Sharded Cluster Components* (page 1042) for more information.

Cluster Meta-data Upgrade

Considerations Beware of the following properties of the cluster upgrade process:

- Before you start the upgrade, ensure that the amount of free space on the filesystem for the *config database* (page 823) is at least 4 to 5 times the amount of space currently used by the *config database* (page 823) data files.

  Additionally, ensure that all indexes in the *config database* (page 823) are `{v:1}` indexes. If a critical index is a `{v:0}` index, chunk splits can fail due to known issues with the `{v:0}` format. `{v:0}` indexes are present on clusters created with MongoDB 2.0 or earlier.

  The duration of the metadata upgrade depends on the network latency between the node performing the upgrade and the three config servers. Ensure low latency between the upgrade process and the config servers.

- While the upgrade is in progress, you cannot make changes to the collection meta-data. For example, during the upgrade, do **not** perform:
  - `sh.enableSharding()`,
  - `sh.shardCollection()`,
  - `sh.addShard()`,
  - `db.createCollection()`,
  - `db.collection.drop()`,
  - `db.dropDatabase()`,
  - any operation that creates a database, or
  - any other operation that modifies the cluster meta-data in any way. See *Sharding Reference* (page 822) for a complete list of sharding commands. Note, however, that not all commands on the *Sharding Reference* (page 822) page modifies the cluster meta-data.
- Once you upgrade to 2.4 and complete the upgrade procedure **do not** use 2.0 `mongod` and `mongos` processes in your cluster. 2.0 process may re-introduce old meta-data formats into cluster meta-data.

The upgraded config database will require more storage space than before, to make backup and working copies of the `config.chunks` (page 826) and `config.collections` (page 826) collections. As always, if storage requirements increase, the `mongod` might need to pre-allocate additional data files. See *How can I get information on the storage use of a database?* (page 856) for more information.

Meta-data Upgrade Procedure Changes to the meta-data format for sharded clusters, stored in the *config database* (page 823), require a special meta-data upgrade procedure when moving to 2.4.

Do not perform operations that modify meta-data while performing this procedure. See *Upgrade a Sharded Cluster from MongoDB 2.2 to MongoDB 2.4* (page 1038) for examples of prohibited operations.

1. Before you start the upgrade, ensure that the amount of free space on the filesystem for the *config database* (page 823) is at least 4 to 5 times the amount of space currently used by the *config database* (page 823) data files.

   Additionally, ensure that all indexes in the *config database* (page 823) are `{v:1}` indexes. If a critical index is a `{v:0}` index, chunk splits can fail due to known issues with the `{v:0}` format. `{v:0}` indexes are present on clusters created with MongoDB 2.0 or earlier.

The duration of the metadata upgrade depends on the network latency between the node performing the upgrade and the three config servers. Ensure low latency between the upgrade process and the config servers.

To check the version of your indexes, use `db.collection.getIndexes()`.

If any index **on the config database** is `{v:0}`, you should rebuild those indexes by connecting to the `mongos` and either: rebuild all indexes using the `db.collection.reIndex()` method, or drop and rebuild specific indexes using `db.collection.dropIndex()` and then `db.collection.ensureIndex()`. If you need to upgrade the `_id` index to `{v:1}` use `db.collection.reIndex()`.

You may have `{v:0}` indexes on other databases in the cluster.

2. Turn off the *balancer* (page 758) in the *sharded cluster*, as described in *Disable the Balancer* (page 802).

Optional

For additional security during the upgrade, you can make a backup of the config database using `mongodump` or other backup tools.

---

3. Ensure there are no version 2.0 `mongod` or `mongos` processes still active in the sharded cluster. The automated upgrade process checks for 2.0 processes, but network availability can prevent a definitive check. Wait 5 minutes after stopping or upgrading version 2.0 `mongos` processes to confirm that none are still active.

4. Start a single 2.4 `mongos` process with `configDB` pointing to the sharded cluster's *config servers* (page 742) and with the *`--upgrade`* option. The upgrade process happens before the process becomes a daemon (i.e. before *`--fork`*.)

   You can upgrade an existing `mongos` instance to 2.4 or you can start a new *mongos* instance that can reach all config servers if you need to avoid reconfiguring a production `mongos`.

   Start the `mongos` with a command that resembles the following:

   ```
   mongos --configdb <config servers> --upgrade
   ```

   Without the *`--upgrade`* option 2.4 `mongos` processes will fail to start until the upgrade process is complete.

   The upgrade will prevent any chunk moves or splits from occurring during the upgrade process. If there are very many sharded collections or there are stale locks held by other failed processes, acquiring the locks for all collections can take seconds or minutes. See the log for progress updates.

5. When the `mongos` process starts successfully, the upgrade is complete. If the `mongos` process fails to start, check the log for more information.

   If the `mongos` terminates or loses its connection to the config servers during the upgrade, you may always safely retry the upgrade.

   However, if the upgrade failed during the short critical section, the `mongos` will exit and report that the upgrade will require manual intervention. To continue the upgrade process, you must follow the *Resync after an Interruption of the Critical Section* (page 1041) procedure.

Optional

If the `mongos` logs show the upgrade waiting for the upgrade lock, a previous upgrade process may still be active or may have ended abnormally. After 15 minutes of no remote activity `mongos` will force the upgrade lock. If you can verify that there are no running upgrade processes, you may connect to a 2.2 `mongos` process and force the lock manually:

```
mongo <mongos.example.net>

db.getMongo().getCollection("config.locks").findOne({ _id : "configUpgrade" })
```

If the process specified in the `process` field of this document is *verifiably* offline, run the following operation to force the lock.

```
db.getMongo().getCollection("config.locks").update({ _id : "configUpgrade" }, { $set : { state :
```

It is always more safe to wait for the `mongos` to verify that the lock is inactive, if you have any doubts about the activity of another upgrade operation. In addition to the `configUpgrade`, the `mongos` may need to wait for specific collection locks. Do not force the specific collection locks.

---

6. Upgrade and restart other `mongos` processes in the sharded cluster, *without* the `--upgrade` option.

   See *Upgrade Sharded Cluster Components* (page 1042) for more information.

7. *Re-enable the balancer* (page 802). You can now perform operations that modify cluster meta-data.

Once you have upgraded, *do not* introduce version 2.0 MongoDB processes into the sharded cluster. This can reintroduce old meta-data formats into the config servers. The meta-data change made by this upgrade process will help prevent errors caused by cross-version incompatibilities in future versions of MongoDB.

Resync after an Interruption of the Critical Section During the short critical section of the upgrade that applies changes to the meta-data, it is unlikely but possible that a network interruption can prevent all three config servers from verifying or modifying data. If this occurs, the *config servers* (page 742) must be re-synced, and there may be problems starting new `mongos` processes. The *sharded cluster* will remain accessible, but avoid all cluster meta-data changes until you resync the config servers. Operations that change meta-data include: adding shards, dropping databases, and dropping collections.

Note: Only perform the following procedure *if* something (e.g. network, power, etc.) interrupts the upgrade process during the short critical section of the upgrade. Remember, you may always safely attempt the *meta data upgrade procedure* (page 1039).

To resync the config servers:

1. Turn off the *balancer* (page 758) in the sharded cluster and stop all meta-data operations. If you are in the middle of an upgrade process (*Upgrade a Sharded Cluster from MongoDB 2.2 to MongoDB 2.4* (page 1038)), you have already disabled the balancer.
2. Shut down two of the three config servers, preferably the last two listed in the `configDB` string. For example, if your `configDB` string is `configA:27019,configB:27019,configC:27019`, shut down `configB` and `configC`. Shutting down the last two config servers ensures that most `mongos` instances will have uninterrupted access to cluster meta-data.
3. `mongodump` the data files of the active config server (`configA`).
4. Move the data files of the deactivated config servers (`configB` and `configC`) to a backup location.
5. Create new, empty *data directories*.
6. Restart the disabled config servers with `--dbpath` pointing to the now-empty data directory and `--port` pointing to an alternate port (e.g. `27020`).
7. Use `mongorestore` to repopulate the data files on the disabled documents from the active config server (`configA`) to the restarted config servers on the new port (`configB:27020,configC:27020`). These config servers are now re-synced.
8. Restart the restored config servers on the old port, resetting the port back to the old settings (`configB:27019` and `configC:27019`).
9. In some cases connection pooling may cause spurious failures, as the `mongos` disables old connections only after attempted use. 2.4 fixes this problem, but to avoid this issue in version 2.2, you can restart all `mongos`

instances (one-by-one, to avoid downtime) and use the `rs.stepDown()` method before restarting each of the shard *replica set primaries*.

10. The sharded cluster is now fully resynced; however before you attempt the upgrade process again, you must manually reset the upgrade state using a version 2.2 `mongos`. Begin by connecting to the 2.2 `mongos` with the `mongo` shell:

    ```
    mongo <mongos.example.net>
    ```

    Then, use the following operation to reset the upgrade process:

    ```
    db.getMongo().getCollection("config.version").update({ _id : 1 }, { $unset : { upgradeState : 1
    ```

11. Finally retry the upgrade process, as in *Upgrade a Sharded Cluster from MongoDB 2.2 to MongoDB 2.4* (page 1038).

Upgrade Sharded Cluster Components After you have successfully completed the meta-data upgrade process described in *Meta-data Upgrade Procedure* (page 1039), and the 2.4 `mongos` instance starts, you can upgrade the other processes in your MongoDB deployment.

While the balancer is still disabled, upgrade the components of your sharded cluster in the following order:

- Upgrade all `mongos` instances in the cluster, in any order.
- Upgrade all 3 `mongod` config server instances, upgrading the *first* system in the *`mongos --configdb`* argument *last*.
- Upgrade each shard, one at a time, upgrading the `mongod` secondaries before running `replSetStepDown` and upgrading the primary of each shard.

When this process is complete, you can now *re-enable the balancer* (page 802).

Rolling Upgrade Limitation for 2.2.0 Deployments Running with `auth` Enabled MongoDB *cannot* support deployments that mix 2.2.0 and 2.4.0, or greater, components. MongoDB version 2.2.1 and later processes *can* exist in mixed deployments with 2.4-series processes. Therefore you cannot perform a rolling upgrade from MongoDB 2.2.0 to MongoDB 2.4.0. To upgrade a cluster with 2.2.0 components, use one of the following procedures.

1. Perform a rolling upgrade of all 2.2.0 processes to the latest 2.2-series release (e.g. 2.2.3) so that there are no processes in the deployment that predate 2.2.1. When there are no 2.2.0 processes in the deployment, perform a rolling upgrade to 2.4.0.
2. Stop all processes in the cluster. Upgrade all processes to a 2.4-series release of MongoDB, and start all processes at the same time.

Upgrade from 2.3 to 2.4 If you used a `mongod` from the 2.3 or 2.4-rc (release candidate) series, you can safely transition these databases to 2.4.0 or later; *however*, if you created `2dsphere` or `text` indexes using a `mongod` before v2.4-rc2, you will need to rebuild these indexes. For example:

```
db.records.dropIndex( { loc: "2dsphere" } )
db.records.dropIndex( "records_text" )

db.records.ensureIndex( { loc: "2dsphere" } )
db.records.ensureIndex( { records: "text" } )
```

Downgrade MongoDB from 2.4 to Previous Versions For some cases the on-disk format of data files used by 2.4 and 2.2 `mongod` is compatible, and you can upgrade and downgrade if needed. However, several new features in 2.4 are incompatible with previous versions:

- 2dsphere indexes are incompatible with 2.2 and earlier mongod instances.
- text indexes are incompatible with 2.2 and earlier mongod instances.
- using a hashed index as a shard key are incompatible with 2.2 and earlier mongos instances.
- hashed indexes are incompatible with 2.0 and earlier mongod instances.

Important: Collections sharded using hashed shard keys, should **not** use 2.2 mongod instances, which cannot correctly support cluster operations for these collections.

If you completed the *meta-data upgrade for a sharded cluster* (page 1038), you can safely downgrade to 2.2 MongoDB processes. **Do not** use 2.0 processes after completing the upgrade procedure.

Note: In sharded clusters, once you have completed the *meta-data upgrade procedure* (page 1038), you cannot use 2.0 mongod or mongos instances in the same cluster.

If you complete the meta-data upgrade, you can safely downgrade components in any order. When upgrade again, always upgrade mongos instances before mongod instances.

Do not create 2dsphere or text indexes in a cluster that has 2.2 components.

Considerations and Compatibility If you upgrade to MongoDB 2.4, and then need to run MongoDB 2.2 with the same data files, consider the following limitations.

- If you use a hashed index as the shard key index, which is only possible under 2.4 you will not be able to query data in this sharded collection. Furthermore, a 2.2 mongos cannot properly route an insert operation for a collections sharded using a hashed index for the shard key index: any data that you insert using a 2.2 mongos, will not arrive on the correct shard and will not be reachable by future queries.
- If you *never* create an 2dsphere or text index, you can move between a 2.4 and 2.2 mongod for a given data set; however, after you create the first 2dsphere or text index with a 2.4 mongod you will need to run a 2.2 mongod with the *--upgrade* option and drop any 2dsphere or text index.

Upgrade and Downgrade Procedures

Basic Downgrade and Upgrade **Except** as described below, moving between 2.2 and 2.4 is a drop-in replacement:

- stop the existing mongod, using the *--shutdown* option as follows:

  ```
  mongod --dbpath /var/mongod/data --shutdown
  ```

  Replace /var/mongod/data with your MongoDB dbPath.

- start the new mongod processes with the same dbPath setting, for example:

  ```
  mongod --dbpath /var/mongod/data
  ```

  Replace /var/mongod/data with your MongoDB dbPath.

Downgrade to 2.2 After Creating a 2dsphere or text Index If you have created 2dsphere or text indexes while running a 2.4 mongod instance, you can downgrade at any time, by starting the 2.2 mongod with the *--upgrade* option as follows:

```
mongod --dbpath /var/mongod/data/ --upgrade
```

Then, you will need to drop any existing `2dsphere` or `text` indexes using `db.collection.dropIndex()`, for example:

```
db.records.dropIndex( { loc: "2dsphere" } )
db.records.dropIndex( "records_text" )
```

Warning: `--upgrade` will run `repairDatabase` on any database where you have created a `2dsphere` or `text` index, which will rebuild *all* indexes.

Troubleshooting Upgrade/Downgrade Operations If you do not use `--upgrade`, when you attempt to start a 2.2 `mongod` and you have created a `2dsphere` or `text` index, `mongod` will return the following message:

```
'need to upgrade database index_plugin_upgrade with pdfile version 4.6, new version: 4.5 Not upgradi
```

While running 2.4, to check the data file version of a MongoDB database, use the following operation in the shell:

```
db.getSiblingDB('<databasename>').stats().dataFileVersion
```

The major data file [1564] version for both 2.2 and 2.4 is `4`, the minor data file version for 2.2 is `5` and the minor data file version for 2.4 is `6` **after** you create a `2dsphere` or `text` index.

Compatibility and Index Type Changes in MongoDB 2.4

On this page

- New Index Types (page 1044)
- Index Type Validation (page 1044)

In 2.4 MongoDB includes two new features related to indexes that users upgrading to version 2.4 must consider, particularly with regard to possible downgrade paths. For more information on downgrades, see *Downgrade MongoDB from 2.4 to Previous Versions* (page 1042).

New Index Types In 2.4 MongoDB adds two new index types: `2dsphere` and `text`. These index types do not exist in 2.2, and for each database, creating a `2dsphere` or `text` index, will upgrade the data-file version and make that database incompatible with 2.2.

If you intend to downgrade, you should always drop all `2dsphere` and `text` indexes before moving to 2.2.

You can use the *downgrade procedure* (page 1042) to downgrade these databases and run 2.2 if needed, however this will run a full database repair (as with `repairDatabase`) for all affected databases.

Index Type Validation In MongoDB 2.2 and earlier you could specify invalid index types that did not exist. In these situations, MongoDB would create an ascending (e.g. `1`) index. Invalid indexes include index types specified by strings that do not refer to an existing index type, and all numbers other than `1` and `-1`. [1565]

In 2.4, creating any invalid index will result in an error. Furthermore, you cannot create a `2dsphere` or `text` index on a collection if its containing database has any invalid index types. [1]

Example

If you attempt to add an invalid index in MongoDB 2.4, as in the following:

[1564] The data file version (i.e. `pdfile version`) is independent and unrelated to the release version of MongoDB.

[1565] In 2.4, indexes that specify a type of `"1"` or `"-1"` (the strings `"1"` and `"-1"`) will continue to exist, despite a warning on start-up. **However**, a *secondary* in a replica set cannot complete an initial sync from a primary that has a `"1"` or `"-1"` index. Avoid all indexes with invalid types.

```
db.coll.ensureIndex( { field: "1" } )
```

MongoDB will return the following error document:

```
{
  "err" : "Unknown index plugin '1' in index { field: \"1\" }"
  "code": 16734,
  "n": <number>,
  "connectionId": <number>,
  "ok": 1
}
```

See *Upgrade MongoDB to 2.4* (page 1037) for full upgrade instructions.

Other Resources

- MongoDB Downloads[1566].
- All JIRA issues resolved in 2.4[1567].
- All Backwards incompatible changes[1568].
- All Third Party License Notices[1569].

15.2.4 Release Notes for MongoDB 2.2

On this page

- Upgrading (page 1045)
- Changes (page 1047)
- Licensing Changes (page 1054)
- Resources (page 1054)

Upgrading

MongoDB 2.2 is a production release series and succeeds the 2.0 production release series.

MongoDB 2.0 data files are compatible with 2.2-series binaries without any special migration process. However, always perform the upgrade process for replica sets and sharded clusters using the procedures that follow.

Synopsis

- `mongod`, 2.2 is a drop-in replacement for 2.0 and 1.8.

[1566] http://mongodb.org/downloads
[1567] https://jira.mongodb.org/secure/IssueNavigator.jspa?reset=true&jqlQuery=project+%3D+SERVER+AND+fixVersion+in+%28%222.3.2%22,+%222.3.1%22,+%222
rc0%22,+%222.4.0-rc1%22,+%222.4.0-rc2%22,+%222.4.0-rc3%22%29
[1568] https://jira.mongodb.org/issues/?jql=project%20%3D%20SERVER%20AND%20fixVersion%20in%20(%222.3.2%22%2C%20%222.3.1%22%2C%20%222.3.0%2
rc0%22%2C%20%222.4.0-rc1%22%2C%20%222.4.0-rc2%22%2C%20%222.4.0-rc3%22)%20AND%20%22Backwards%20Compatibility%22%20in%20(%22Major%
[1569] https://github.com/mongodb/mongo/blob/v2.4/distsrc/THIRD-PARTY-NOTICES

- Check your `driver` documentation for information regarding required compatibility upgrades, and always run the recent release of your driver.

  Typically, only users running with authentication, will need to upgrade drivers before continuing with the upgrade to 2.2.

- For all deployments using authentication, upgrade the drivers (i.e. client libraries), before upgrading the `mongod` instance or instances.

- For all upgrades of sharded clusters:

  – turn off the balancer during the upgrade process. See the *Disable the Balancer* (page 802) section for more information.

  – upgrade all `mongos` instances before upgrading any `mongod` instances.

Other than the above restrictions, 2.2 processes can interoperate with 2.0 and 1.8 tools and processes. You can safely upgrade the `mongod` and `mongos` components of a deployment one by one while the deployment is otherwise operational. Be sure to read the detailed upgrade procedures below before upgrading production systems.

Upgrading a Standalone `mongod`

1. Download binaries of the latest release in the 2.2 series from the MongoDB Download Page[1570].
2. Shutdown your `mongod` instance. Replace the existing binary with the 2.2 `mongod` binary and restart MongoDB.

Upgrading a Replica Set

You can upgrade to 2.2 by performing a "rolling" upgrade of the set by upgrading the members individually while the other members are available to minimize downtime. Use the following procedure:

1. Upgrade the *secondary* members of the set one at a time by shutting down the `mongod` and replacing the 2.0 binary with the 2.2 binary. After upgrading a `mongod` instance, wait for the member to recover to `SECONDARY` state before upgrading the next instance. To check the member's state, issue `rs.status()` in the `mongo` shell.
2. Use the `mongo` shell method `rs.stepDown()` to step down the *primary* to allow the normal *failover* (page 644) procedure. `rs.stepDown()` expedites the failover procedure and is preferable to shutting down the primary directly.

   Once the primary has stepped down and another member has assumed `PRIMARY` state, as observed in the output of `rs.status()`, shut down the previous primary and replace `mongod` binary with the 2.2 binary and start the new process.

   Note: Replica set failover is not instant but will render the set unavailable to read or accept writes until the failover process completes. Typically this takes 10 seconds or more. You may wish to plan the upgrade during a predefined maintenance window.

Upgrading a Sharded Cluster

Use the following procedure to upgrade a sharded cluster:

- *Disable the balancer* (page 802).

[1570] http://downloads.mongodb.org/

- Upgrade all `mongos` instances *first*, in any order.
- Upgrade all of the `mongod` config server instances using the *stand alone* (page 1046) procedure. To keep the cluster online, be sure that at all times at least one config server is up.
- Upgrade each shard's replica set, using the *upgrade procedure for replica sets* (page 1046) detailed above.
- re-enable the balancer.

Note: Balancing is not currently supported in *mixed* 2.0.x and 2.2.0 deployments. Thus you will want to reach a consistent version for all shards within a reasonable period of time, e.g. same-day. See SERVER-6902[1571] for more information.

Changes

Major Features

Aggregation Framework The aggregation framework makes it possible to do aggregation operations without needing to use *map-reduce*. The `aggregate` command exposes the aggregation framework, and the `aggregate()` helper in the `mongo` shell provides an interface to these operations. Consider the following resources for background on the aggregation framework and its use:

- Documentation: *Aggregation* (page 195)
- Reference: *Aggregation Reference* (page 225)

TTL Collections TTL collections remove expired data from a collection, using a special index and a background thread that deletes expired documents every minute. These collections are useful as an alternative to *capped collections* in some cases, such as for data warehousing and caching cases, including: machine generated event data, logs, and session information that needs to persist in a database for only a limited period of time.

For more information, see the *Expire Data from Collections by Setting TTL* (page 567) tutorial.

Concurrency Improvements MongoDB 2.2 increases the server's capacity for concurrent operations with the following improvements:

1. DB Level Locking[1572]
2. Improved Yielding on Page Faults[1573]
3. Improved Page Fault Detection on Windows[1574]

To reflect these changes, MongoDB now provides changed and improved reporting for concurrency and use. See *locks*, recordStats[1575], `db.currentOp()`, `mongotop`, and `mongostat`.

Improved Data Center Awareness with Tag Aware Sharding MongoDB 2.2 adds additional support for geographic distribution or other custom partitioning for sharded collections in *clusters*. By using this "tag aware" sharding, you can automatically ensure that data in a sharded database system is always on specific shards. For example, with tag aware sharding, you can ensure that data is closest to the application servers that use that data most frequently.

[1571] https://jira.mongodb.org/browse/SERVER-6902
[1572] https://jira.mongodb.org/browse/SERVER-4328
[1573] https://jira.mongodb.org/browse/SERVER-3357
[1574] https://jira.mongodb.org/browse/SERVER-4538
[1575] https://docs.mongodb.org/v2.2/reference/server-status

Shard tagging controls data location, and is complementary but separate from replica set tagging, which controls *read preference* (page 651) and *write concern* (page 179). For example, shard tagging can pin all "USA" data to one or more logical shards, while replica set tagging can control which `mongod` instances (e.g. "`production`" or "`reporting`") the application uses to service requests.

See the documentation for the following helpers in the `mongo` shell that support tagged sharding configuration:

- `sh.addShardTag()`
- `sh.addTagRange()`
- `sh.removeShardTag()`

Also, see *Tag Aware Sharding* (page 756) and *Manage Shard Tags* (page 816).

Fully Supported Read Preference Semantics All MongoDB clients and drivers now support full *read preferences* (page 651), including consistent support for a full range of *read preference modes* (page 728) and *tag sets* (page 653). This support extends to the `mongos` and applies identically to single replica sets and to the replica sets for each shard in a *sharded cluster*.

Additional read preference support now exists in the `mongo` shell using the `readPref()` cursor method.

Compatibility Changes

Authentication Changes MongoDB 2.2 provides more reliable and robust support for authentication clients, including drivers and `mongos` instances.

If your cluster runs with authentication:

- For all drivers, use the latest release of your driver and check its release notes.
- In sharded environments, to ensure that your cluster remains available during the upgrade process you **must** use the *upgrade procedure for sharded clusters* (page 1046).

`findAndModify` Returns Null Value for Upserts that Perform Inserts In version 2.2, for *upsert* that perform inserts with the `new` option set to `false`, `findAndModify` commands will now return the following output:

```
{ 'ok': 1.0, 'value': null }
```

In the `mongo` shell, upsert `findAndModify` operations that perform inserts (with `new` set to `false`.)only output a `null` value.

In version 2.0 these operations would return an empty document, e.g. `{ }`.

See: SERVER-6226[1576] for more information.

`mongodump` 2.2 Output Incompatible with Pre-2.2 `mongorestore` If you use the `mongodump` tool from the 2.2 distribution to create a dump of a database, you must use a 2.2 (or later) version of `mongorestore` to restore that dump.

See: SERVER-6961[1577] for more information.

[1576] https://jira.mongodb.org/browse/SERVER-6226
[1577] https://jira.mongodb.org/browse/SERVER-6961

`ObjectId().toString()` Returns String Literal `ObjectId("...")` In version 2.2, the `toString()` method returns the string representation of the *ObjectId()* object and has the format `ObjectId("...")`.

Consider the following example that calls the `toString()` method on the `ObjectId("507c7f79bcf86cd7994f6c0e")` object:

```
ObjectId("507c7f79bcf86cd7994f6c0e").toString()
```

The method now returns the *string* `ObjectId("507c7f79bcf86cd7994f6c0e")`.

Previously, in version 2.0, the method would return the *hexadecimal string* `507c7f79bcf86cd7994f6c0e`.

If compatibility between versions 2.0 and 2.2 is required, use *ObjectId().str*, which holds the hexadecimal string value in both versions.

`ObjectId().valueOf()` Returns hexadecimal string In version 2.2, the `valueOf()` method returns the value of the *ObjectId()* object as a lowercase hexadecimal string.

Consider the following example that calls the `valueOf()` method on the `ObjectId("507c7f79bcf86cd7994f6c0e")` object:

```
ObjectId("507c7f79bcf86cd7994f6c0e").valueOf()
```

The method now returns the *hexadecimal string* `507c7f79bcf86cd7994f6c0e`.

Previously, in version 2.0, the method would return the *object* `ObjectId("507c7f79bcf86cd7994f6c0e")`.

If compatibility between versions 2.0 and 2.2 is required, use *ObjectId().str* attribute, which holds the hexadecimal string value in both versions.

Behavioral Changes

Restrictions on Collection Names In version 2.2, collection names cannot:

- contain the `$`.
- be an empty string (i.e. `""`).

This change does not affect collections created with now illegal names in earlier versions of MongoDB.

These new restrictions are in addition to the existing restrictions on collection names which are:

- A collection name should begin with a letter or an underscore.
- A collection name cannot contain the null character.
- Begin with the `system.` prefix. MongoDB reserves `system.` for system collections, such as the `system.indexes` collection.
- The maximum size of a collection name is 128 characters, including the name of the database. However, for maximum flexibility, collections should have names less than 80 characters.

Collections names may have any other valid UTF-8 string.

See the SERVER-4442[1578] and the *faq-restrictions-on-collection-names* FAQ item.

[1578] https://jira.mongodb.org/browse/SERVER-4442

Restrictions on Database Names for Windows Database names running on Windows can no longer contain the following characters:

```
/\. "*<>:|?
```

The names of the data files include the database name. If you attempt to upgrade a database instance with one or more of these characters, `mongod` will refuse to start.

Change the name of these databases before upgrading. See SERVER-4584[1579] and SERVER-6729[1580] for more information.

`_id` Fields and Indexes on Capped Collections All *capped collections* now have an `_id` field by default, *if* they exist outside of the `local` database, and now have indexes on the `_id` field. This change only affects capped collections created with 2.2 instances and does not affect existing capped collections.

See: SERVER-5516[1581] for more information.

New `$elemMatch` Projection Operator The `$elemMatch` operator allows applications to narrow the data returned from queries so that the query operation will only return the first matching element in an array. See the `$elemMatch` reference and the SERVER-2238[1582] and SERVER-828[1583] issues for more information.

Windows Specific Changes

Windows XP is Not Supported As of 2.2, MongoDB does not support Windows XP. Please upgrade to a more recent version of Windows to use the latest releases of MongoDB. See SERVER-5648[1584] for more information.

Service Support for `mongos.exe` You may now run `mongos.exe` instances as a Windows Service. See the `mongos.exe` reference and *Configure a Windows Service for MongoDB Community Edition* (page 47) and SERVER-1589[1585] for more information.

Log Rotate Command Support MongoDB for Windows now supports log rotation by way of the `logRotate` database command. See SERVER-2612[1586] for more information.

New Build Using SlimReadWrite Locks for Windows Concurrency Labeled "2008+" on the Downloads Page[1587], this build for 64-bit versions of Windows Server 2008 R2 and for Windows 7 or newer, offers increased performance over the standard 64-bit Windows build of MongoDB. See SERVER-3844[1588] for more information.

Tool Improvements

Index Definitions Handled by `mongodump` and `mongorestore` When you specify the `--collection` option to `mongodump`, `mongodump` will now backup the definitions for all indexes that exist on the source database. When

[1579] https://jira.mongodb.org/browse/SERVER-4584
[1580] https://jira.mongodb.org/browse/SERVER-6729
[1581] https://jira.mongodb.org/browse/SERVER-5516
[1582] https://jira.mongodb.org/browse/SERVER-2238
[1583] https://jira.mongodb.org/browse/SERVER-828
[1584] https://jira.mongodb.org/browse/SERVER-5648
[1585] https://jira.mongodb.org/browse/SERVER-1589
[1586] https://jira.mongodb.org/browse/SERVER-2612
[1587] http://www.mongodb.org/downloads
[1588] https://jira.mongodb.org/browse/SERVER-3844

you attempt to restore this backup with `mongorestore`, the target `mongod` will rebuild all indexes. See SERVER-808[1589] for more information.

`mongorestore` now includes the *`--noIndexRestore`* option to provide the preceding behavior. Use *`--noIndexRestore`* to prevent `mongorestore` from building previous indexes.

`mongooplog` for Replaying Oplogs The `mongooplog` tool makes it possible to pull *oplog* entries from `mongod` instance and apply them to another `mongod` instance. You can use `mongooplog` to achieve point-in-time backup of a MongoDB data set. See the SERVER-3873[1590] case and the `mongooplog` reference.

Authentication Support for `mongotop` and `mongostat` `mongotop` and `mongostat` now contain support for username/password authentication. See SERVER-3875[1591] and SERVER-3871[1592] for more information regarding this change. Also consider the documentation of the following options for additional information:

- *`mongotop --username`*
- *`mongotop --password`*
- *`mongostat --username`*
- *`mongostat --password`*

Write Concern Support for `mongoimport` and `mongorestore` `mongoimport` now provides an option to halt the import if the operation encounters an error, such as a network interruption, a duplicate key exception, or a write error. The *`--stopOnError`* option will produce an error rather than silently continue importing data. See SERVER-3937[1593] for more information.

In `mongorestore`, the *`--w`* option provides support for configurable write concern.

`mongodump` Support for Reading from Secondaries You can now run `mongodump` when connected to a *secondary* member of a *replica set*. See SERVER-3854[1594] for more information.

`mongoimport` Support for full 16MB Documents Previously, `mongoimport` would only import documents that were less than 4 megabytes in size. This issue is now corrected, and you may use `mongoimport` to import documents that are at least 16 megabytes ins size. See SERVER-4593[1595] for more information.

`Timestamp()` Extended JSON format MongoDB extended JSON now includes a new `Timestamp()` type to represent the Timestamp type that MongoDB uses for timestamps in the *oplog* among other contexts.

This permits tools like `mongooplog` and `mongodump` to query for specific timestamps. Consider the following `mongodump` operation:

```
mongodump --db local --collection oplog.rs --query '{"ts":{"$gt":{"$timestamp" : {"t": 1344969612000,
```

See SERVER-3483[1596] for more information.

[1589] https://jira.mongodb.org/browse/SERVER-808
[1590] https://jira.mongodb.org/browse/SERVER-3873
[1591] https://jira.mongodb.org/browse/SERVER-3875
[1592] https://jira.mongodb.org/browse/SERVER-3871
[1593] https://jira.mongodb.org/browse/SERVER-3937
[1594] https://jira.mongodb.org/browse/SERVER-3854
[1595] https://jira.mongodb.org/browse/SERVER-4593
[1596] https://jira.mongodb.org/browse/SERVER-3483

Shell Improvements

Improved Shell User Interface 2.2 includes a number of changes that improve the overall quality and consistency of the user interface for the `mongo` shell:

- Full Unicode support.
- Bash-like line editing features. See SERVER-4312[1597] for more information.
- Multi-line command support in shell history. See SERVER-3470[1598] for more information.
- Windows support for the `edit` command. See SERVER-3998[1599] for more information.

Helper to load Server-Side Functions The `db.loadServerScripts()` loads the contents of the current database's `system.js` collection into the current `mongo` shell session. See SERVER-1651[1600] for more information.

Support for Bulk Inserts If you pass an array of *documents* to the `insert()` method, the `mongo` shell will now perform a bulk insert operation. See SERVER-3819[1601] and SERVER-2395[1602] for more information.

Note: For bulk inserts on sharded clusters, the `getLastError` command alone is insufficient to verify success. Applications should must verify the success of bulk inserts in application logic.

Operations

Support for Logging to Syslog See the SERVER-2957[1603] case and the documentation of the `syslogFacility` run-time option or the *`mongod --syslog`* and *`mongos --syslog`* command line-options.

`touch` Command Added the `touch` command to read the data and/or indexes from a collection into memory. See: SERVER-2023[1604] and `touch` for more information.

`indexCounters` No Longer Report Sampled Data `indexCounters` now report actual counters that reflect index use and state. In previous versions, these data were sampled. See SERVER-5784[1605] and `indexCounters` for more information.

Padding Specifiable on `compact` Command See the documentation of the `compact` and the SERVER-4018[1606] issue for more information.

Added Build Flag to Use System Libraries The Boost library, version 1.49, is now embedded in the MongoDB code base.

If you want to build MongoDB binaries using system Boost libraries, you can pass `scons` using the `--use-system-boost` flag, as follows:

[1597] https://jira.mongodb.org/browse/SERVER-4312
[1598] https://jira.mongodb.org/browse/SERVER-3470
[1599] https://jira.mongodb.org/browse/SERVER-3998
[1600] https://jira.mongodb.org/browse/SERVER-1651
[1601] https://jira.mongodb.org/browse/SERVER-3819
[1602] https://jira.mongodb.org/browse/SERVER-2395
[1603] https://jira.mongodb.org/browse/SERVER-2957
[1604] https://jira.mongodb.org/browse/SERVER-2023
[1605] https://jira.mongodb.org/browse/SERVER-5784
[1606] https://jira.mongodb.org/browse/SERVER-4018

```
scons --use-system-boost
```

When building MongoDB, you can also pass `scons` a flag to compile MongoDB using only system libraries rather than the included versions of the libraries. For example:

```
scons --use-system-all
```

See the SERVER-3829[1607] and SERVER-5172[1608] issues for more information.

Memory Allocator Changed to TCMalloc To improve performance, MongoDB 2.2 uses the TCMalloc memory allocator from Google Perftools. For more information about this change see the SERVER-188[1609] and SERVER-4683[1610]. For more information about TCMalloc, see the documentation of TCMalloc[1611] itself.

Replication

Improved Logging for Replica Set Lag When *secondary* members of a replica set fall behind in replication, `mongod` now provides better reporting in the log. This makes it possible to track replication in general and identify what process may produce errors or halt replication. See SERVER-3575[1612] for more information.

Replica Set Members can Sync from Specific Members The new `replSetSyncFrom` command and new `rs.syncFrom()` helper in the `mongo` shell make it possible for you to manually configure from which member of the set a replica will poll *oplog* entries. Use these commands to override the default selection logic if needed. Always exercise caution with `replSetSyncFrom` when overriding the default behavior.

Replica Set Members will not Sync from Members Without Indexes Unless `buildIndexes: false` To prevent inconsistency between members of replica sets, if the member of a replica set has `buildIndexes` set to `true`, other members of the replica set will *not* sync from this member, unless they also have `buildIndexes` set to `true`. See SERVER-4160[1613] for more information.

New Option To Configure Index Pre-Fetching during Replication By default, when replicating options, *secondaries* will pre-fetch *Indexes* (page 515) associated with a query to improve replication throughput in most cases. The `replication.secondaryIndexPrefetch` setting and `--replIndexPrefetch` option allow administrators to disable this feature or allow the `mongod` to pre-fetch only the index on the `_id` field. See SERVER-6718[1614] for more information.

Map Reduce Improvements

In 2.2 Map Reduce received the following improvements:

- Improved support for sharded MapReduce[1615], and
- MapReduce will retry jobs following a config error[1616].

[1607] https://jira.mongodb.org/browse/SERVER-3829
[1608] https://jira.mongodb.org/browse/SERVER-5172
[1609] https://jira.mongodb.org/browse/SERVER-188
[1610] https://jira.mongodb.org/browse/SERVER-4683
[1611] http://goog-perftools.sourceforge.net/doc/tcmalloc.html
[1612] https://jira.mongodb.org/browse/SERVER-3575
[1613] https://jira.mongodb.org/browse/SERVER-4160
[1614] https://jira.mongodb.org/browse/SERVER-6718
[1615] https://jira.mongodb.org/browse/SERVER-4521
[1616] https://jira.mongodb.org/browse/SERVER-4158

Sharding Improvements

Index on Shard Keys Can Now Be a Compound Index If your shard key uses the prefix of an existing index, then you do not need to maintain a separate index for your shard key in addition to your existing index. This index, however, cannot be a multi-key index. See the *Shard Key Indexes* (page 763) documentation and SERVER-1506[1617] for more information.

Migration Thresholds Modified The *migration thresholds* (page 759) have changed in 2.2 to permit more even distribution of *chunks* in collections that have smaller quantities of data. See the *Migration Thresholds* (page 759) documentation for more information.

Licensing Changes

Added License notice for Google Perftools (TCMalloc Utility). See the License Notice[1618] and the SERVER-4683[1619] for more information.

Resources

- MongoDB Downloads[1620].
- All JIRA issues resolved in 2.2[1621].
- All backwards incompatible changes[1622].
- All third party license notices[1623].
- What's New in MongoDB 2.2 Online Conference[1624].

15.2.5 Release Notes for MongoDB 2.0

On this page

- Upgrading (page 1054)
- Changes (page 1056)
- Resources (page 1060)

Upgrading

Although the major version number has changed, MongoDB 2.0 is a standard, incremental production release and works as a drop-in replacement for MongoDB 1.8.

[1617] https://jira.mongodb.org/browse/SERVER-1506
[1618] https://github.com/mongodb/mongo/blob/v2.2/distsrc/THIRD-PARTY-NOTICES#L231
[1619] https://jira.mongodb.org/browse/SERVER-4683
[1620] http://mongodb.org/downloads
[1621] https://jira.mongodb.org/secure/IssueNavigator.jspa?reset=true&jqlQuery=project+%3D+SERVER+AND+fixVersion+in+%28%222.1.0%22%2C+%222.1.1%22%2
rc0%22%2C+%222.2.0-rc1%22%2C+%222.2.0-rc2%22%29+ORDER+BY+component+ASC%2C+key+DESC
[1622] https://jira.mongodb.org/issues/?filter=11225&jql=project%20%3D%20SERVER%20AND%20fixVersion%20in%20(10483%2C%2010893%2C%2010894%2C%2
[1623] https://github.com/mongodb/mongo/blob/v2.2/distsrc/THIRD-PARTY-NOTICES
[1624] http://www.mongodb.com/events/webinar/mongodb-online-conference-sept

Preparation

Read through all release notes before upgrading, and ensure that no changes will affect your deployment.

If you create new indexes in 2.0, then downgrading to 1.8 is possible but you must reindex the new collections.

`mongoimport` and `mongoexport` now correctly adhere to the CSV spec for handling CSV input/output. This may break existing import/export workflows that relied on the previous behavior. For more information see SERVER-1097[1625].

Journaling (page 606) is **enabled by default** in 2.0 for 64-bit builds. If you still prefer to run without journaling, start `mongod` with the `--nojournal` run-time option. Otherwise, MongoDB creates journal files during startup. The first time you start `mongod` with journaling, you will see a delay as `mongod` creates new files. In addition, you may see reduced write throughput.

2.0 `mongod` instances are interoperable with 1.8 `mongod` instances; however, for best results, upgrade your deployments using the following procedures:

Upgrading a Standalone `mongod`

1. Download the v2.0.x binaries from the MongoDB Download Page[1626].
2. Shutdown your `mongod` instance. Replace the existing binary with the 2.0.x `mongod` binary and restart MongoDB.

Upgrading a Replica Set

1. Upgrade the *secondary* members of the set one at a time by shutting down the `mongod` and replacing the 1.8 binary with the 2.0.x binary from the MongoDB Download Page[1627].
2. To avoid losing the last few updates on failover you can temporarily halt your application (failover should take less than 10 seconds), or you can set *write concern* (page 179) in your application code to confirm that each update reaches multiple servers.
3. Use the `rs.stepDown()` to step down the primary to allow the normal *failover* (page 644) procedure.

   `rs.stepDown()` and `replSetStepDown` provide for shorter and more consistent failover procedures than simply shutting down the primary directly.

   When the primary has stepped down, shut down its instance and upgrade by replacing the `mongod` binary with the 2.0.x binary.

Upgrading a Sharded Cluster

1. Upgrade all *config server* instances *first*, in any order. Since config servers use two-phase commit, *shard* configuration metadata updates will halt until all are up and running.
2. Upgrade `mongos` routers in any order.

[1625] https://jira.mongodb.org/browse/SERVER-1097
[1626] http://downloads.mongodb.org/
[1627] http://downloads.mongodb.org/

Changes

Compact Command

A `compact` command is now available for compacting a single collection and its indexes. Previously, the only way to compact was to repair the entire database.

Concurrency Improvements

When going to disk, the server will yield the write lock when writing data that is not likely to be in memory. The initial implementation of this feature now exists:

See SERVER-2563[1628] for more information.

The specific operations yield in 2.0 are:

- Updates by `_id`
- Removes
- Long cursor iterations

Default Stack Size

MongoDB 2.0 reduces the default stack size. This change can reduce total memory usage when there are many (e.g., 1000+) client connections, as there is a thread per connection. While portions of a thread's stack can be swapped out if unused, some operating systems do this slowly enough that it might be an issue. The default stack size is lesser of the system setting or 1MB.

Index Performance Enhancements

v2.0 includes significant improvements to the index[1629]. Indexes are often 25% smaller and 25% faster (depends on the use case). When upgrading from previous versions, the benefits of the new index type are realized only if you create a new index or re-index an old one.

Dates are now signed, and the max index key size has increased slightly from 819 to 1024 bytes.

All operations that create a new index will result in a 2.0 index by default. For example:

- Reindexing results on an older-version index results in a 2.0 index. However, reindexing on a secondary does *not* work in versions prior to 2.0. Do not reindex on a secondary. For a workaround, see SERVER-3866[1630].
- The `repairDatabase` command converts indexes to a 2.0 indexes.

To convert all indexes for a given collection to the *2.0 type* (page 1056), invoke the `compact` command.

Once you create new indexes, downgrading to 1.8.x will require a re-index of any indexes created using 2.0. See /tutorial/roll-back-to-v1.8-index[1631].

Sharding Authentication

Applications can now use authentication with *sharded clusters*.

[1628] https://jira.mongodb.org/browse/SERVER-2563
[1629] https://docs.mongodb.org/v2.2/tutorial/roll-back-to-v1.8-index
[1630] https://jira.mongodb.org/browse/SERVER-3866
[1631] https://docs.mongodb.org/v2.2/tutorial/roll-back-to-v1.8-index

Replica Sets

Hidden Nodes in Sharded Clusters In 2.0, `mongos` instances can now determine when a member of a replica set becomes "hidden" without requiring a restart. In 1.8, `mongos` if you reconfigured a member as hidden, you *had* to restart `mongos` to prevent queries from reaching the hidden member.

Priorities Each *replica set* member can now have a priority value consisting of a floating-point from 0 to 1000, inclusive. Priorities let you control which member of the set you prefer to have as *primary* the member with the highest priority that can see a majority of the set will be elected primary.

For example, suppose you have a replica set with three members, `A`, `B`, and `C`, and suppose that their priorities are set as follows:

- `A`'s priority is `2`.
- `B`'s priority is `3`.
- `C`'s priority is `1`.

During normal operation, the set will always chose `B` as primary. If `B` becomes unavailable, the set will elect `A` as primary.

For more information, see the `priority` documentation.

Data-Center Awareness You can now "tag" *replica set* members to indicate their location. You can use these tags to design custom *write rules* (page 179) across data centers, racks, specific servers, or any other architecture choice.

For example, an administrator can define rules such as "very important write" or `customerData` or "audit-trail" to replicate to certain servers, racks, data centers, etc. Then in the application code, the developer would say:

```
db.foo.insert(doc, {w : "very important write"})
```

which would succeed if it fulfilled the conditions the DBA defined for "very important write".

For more information, see *Data Center Awareness* (page 308).

Drivers may also support tag-aware reads. Instead of specifying `slaveOk`, you specify `slaveOk` with tags indicating which data-centers to read from. For details, see the Drivers[1632] documentation.

`w : majority` You can also set `w` to `majority` to ensure that the write propagates to a majority of nodes, effectively committing it. The value for "majority" will automatically adjust as you add or remove nodes from the set.

For more information, see *Write Concern* (page 179).

Reconfiguration with a Minority Up If the majority of servers in a set has been permanently lost, you can now force a reconfiguration of the set to bring it back online.

For more information see *Reconfigure a Replica Set with Unavailable Members* (page 704).

Primary Checks for a Caught up Secondary before Stepping Down To minimize time without a *primary*, the `rs.stepDown()` method will now fail if the primary does not see a *secondary* within 10 seconds of its latest optime. You can force the primary to step down anyway, but by default it will return an error message.

See also *Force a Member to Become Primary* (page 697).

[1632] https://docs.mongodb.org/ecosystem/drivers

Extended Shutdown on the Primary to Minimize Interruption When you call the `shutdown` command, the *primary* will refuse to shut down unless there is a *secondary* whose optime is within 10 seconds of the primary. If such a secondary isn't available, the primary will step down and wait up to a minute for the secondary to be fully caught up before shutting down.

Note that to get this behavior, you must issue the `shutdown` command explicitly; sending a signal to the process will not trigger this behavior.

You can also force the primary to shut down, even without an up-to-date secondary available.

Maintenance Mode When `repair` or `compact` runs on a *secondary*, the secondary will automatically drop into "recovering" mode until the operation finishes. This prevents clients from trying to read from it while it's busy.

Geospatial Features

Multi-Location Documents Indexing is now supported on documents which have multiple location objects, embedded either inline or in embedded documents. Additional command options are also supported, allowing results to return with not only distance but the location used to generate the distance.

For more information, see *Multi-location Documents for 2d Indexes* (page 561).

Polygon searches Polygonal `$within` queries are also now supported for simple polygon shapes. For details, see the `$within` operator documentation.

Journaling Enhancements

- Journaling is now enabled by default for 64-bit platforms. Use the `--nojournal` command line option to disable it.
- The journal is now compressed for faster commits to disk.
- A new `--journalCommitInterval` run-time option exists for specifying your own group commit interval. The default settings do not change.
- A new `{ getLastError: { j: true } }` option is available to wait for the group commit. The group commit will happen sooner when a client is waiting on `{j: true}`. If journaling is disabled, `{j: true}` is a no-op.

New `ContinueOnError` Option for Bulk Insert

Set the `continueOnError` option for bulk inserts, in the `driver`, so that bulk insert will continue to insert any remaining documents even if an insert fails, as is the case with duplicate key exceptions or network interruptions. The `getLastError` command will report whether any inserts have failed, not just the last one. If multiple errors occur, the client will only receive the most recent `getLastError` results.

Note: For bulk inserts on sharded clusters, the `getLastError` command alone is insufficient to verify success. Applications should must verify the success of bulk inserts in application logic.

Map Reduce

Output to a Sharded Collection Using the new `sharded` flag, it is possible to send the result of a map/reduce to a sharded collection. Combined with the `reduce` or `merge` flags, it is possible to keep adding data to very large collections from map/reduce jobs.

For more information, see *Map-Reduce* (page 214) and the `mapReduce` reference.

Performance Improvements Map/reduce performance will benefit from the following:

- Larger in-memory buffer sizes, reducing the amount of disk I/O needed during a job
- Larger javascript heap size, allowing for larger objects and less GC
- Supports pure JavaScript execution with the `jsMode` flag. See the `mapReduce` reference.

New Querying Features

Additional regex options: `s` Allows the dot (`.`) to match all characters including new lines. This is in addition to the currently supported `i`, `m` and `x`. See `$regex`.

$and A special boolean `$and` query operator is now available.

Command Output Changes

The output of the `validate` command and the documents in the `system.profile` collection have both been enhanced to return information as BSON objects with keys for each value rather than as free-form strings.

Shell Features

Custom Prompt You can define a custom prompt for the `mongo` shell. You can change the prompt at any time by setting the prompt variable to a string or a custom JavaScript function returning a string. For examples, see *Customize the Prompt* (page 80).

Default Shell Init Script On startup, the shell will check for a `.mongorc.js` file in the user's home directory. The shell will execute this file after connecting to the database and before displaying the prompt.

If you would like the shell not to run the `.mongorc.js` file automatically, start the shell with *`--norc`*.

For more information, see the `mongo` reference.

Most Commands Require Authentication

In 2.0, when running with authentication (e.g. `authorization`) *all* database commands require authentication, *except* the following commands.

- `isMaster`
- `authenticate`
- `getnonce`
- `buildInfo`

- `ping`
- `isdbgrid`

Resources

- MongoDB Downloads[1633]
- All JIRA Issues resolved in 2.0[1634]
- All Backward Incompatible Changes[1635]

15.2.6 Release Notes for MongoDB 1.8

On this page

- Upgrading (page 1060)
- Changes (page 1063)
- Resources (page 1066)

Upgrading

MongoDB 1.8 is a standard, incremental production release and works as a drop-in replacement for MongoDB 1.6, except:

- *Replica set* members should be upgraded in a particular order, as described in *Upgrading a Replica Set* (page 1061).
- The `mapReduce` command has changed in 1.8, causing incompatibility with previous releases. `mapReduce` no longer generates temporary collections (thus, `keepTemp` has been removed). Now, you must always supply a value for `out`. See the `out` field options in the `mapReduce` document. If you use MapReduce, this also likely means you need a recent version of your client driver.

Preparation

Read through all release notes before upgrading and ensure that no changes will affect your deployment.

Upgrading a Standalone `mongod`

1. Download the v1.8.x binaries from the MongoDB Download Page[1636].
2. Shutdown your `mongod` instance.
3. Replace the existing binary with the 1.8.x `mongod` binary.
4. Restart MongoDB.

[1633] http://mongodb.org/downloads
[1634] https://jira.mongodb.org/secure/IssueNavigator.jspa?mode=hide&requestId=11002
[1635] https://jira.mongodb.org/issues/?filter=11023&jql=project%20%3D%20SERVER%20AND%20fixVersion%20in%20(10889%2C%2010886%2C%2010784%2C%2
[1636] http://downloads.mongodb.org/

Upgrading a Replica Set

1.8.x *secondaries* **can** replicate from 1.6.x *primaries*.

1.6.x secondaries **cannot** replicate from 1.8.x primaries.

Thus, to upgrade a *replica set* you must replace all of your secondaries first, then the primary.

For example, suppose you have a replica set with a primary, an *arbiter* and several secondaries. To upgrade the set, do the following:

1. For the arbiter:
   (a) Shut down the arbiter.
   (b) Restart it with the 1.8.x binary from the MongoDB Download Page[1637].
2. Change your config (optional) to prevent election of a new primary.

   It is possible that, when you start shutting down members of the set, a new primary will be elected. To prevent this, you can give all of the secondaries a priority of `0` before upgrading, and then change them back afterwards. To do so:
   (a) Record your current config. Run `rs.config()` and paste the results into a text file.
   (b) Update your config so that all secondaries have priority `0`. For example:

```
config = rs.conf()
{
      "_id" : "foo",
      "version" : 3,
      "members" : [
              {
                      "_id" : 0,
                      "host" : "ubuntu:27017"
              },
              {
                      "_id" : 1,
                      "host" : "ubuntu:27018"
              },
              {
                      "_id" : 2,
                      "host" : "ubuntu:27019",
                      "arbiterOnly" : true
              }
              {
                      "_id" : 3,
                      "host" : "ubuntu:27020"
              },
              {
                      "_id" : 4,
                      "host" : "ubuntu:27021"
              },
      ]
}
config.version++
3
rs.isMaster()
{
      "setName" : "foo",
```

[1637] http://downloads.mongodb.org/

```
        "ismaster" : false,
        "secondary" : true,
        "hosts" : [
                "ubuntu:27017",
                "ubuntu:27018"
        ],
        "arbiters" : [
                "ubuntu:27019"
        ],
        "primary" : "ubuntu:27018",
        "ok" : 1
}
// for each secondary
config.members[0].priority = 0
config.members[3].priority = 0
config.members[4].priority = 0
rs.reconfig(config)
```

3. For each secondary:
   (a) Shut down the secondary.
   (b) Restart it with the 1.8.x binary from the MongoDB Download Page[1638].
4. If you changed the config, change it back to its original state:

```
config = rs.conf()
config.version++
config.members[0].priority = 1
config.members[3].priority = 1
config.members[4].priority = 1
rs.reconfig(config)
```

5. Shut down the primary (the final 1.6 server), and then restart it with the 1.8.x binary from the MongoDB Download Page[1639].

Upgrading a Sharded Cluster

1. Turn off the balancer:

```
mongo <a_mongos_hostname>
use config
db.settings.update({_id:"balancer"},{$set : {stopped:true}}, true)
```

2. For each *shard*:
   - If the shard is a *replica set*, follow the directions above for *Upgrading a Replica Set* (page 1061).
   - If the shard is a single `mongod` process, shut it down and then restart it with the 1.8.x binary from the MongoDB Download Page[1640].
3. For each `mongos`:
   (a) Shut down the `mongos` process.
   (b) Restart it with the 1.8.x binary from the MongoDB Download Page[1641].

[1638] http://downloads.mongodb.org/
[1639] http://downloads.mongodb.org/
[1640] http://downloads.mongodb.org/
[1641] http://downloads.mongodb.org/

4. For each config server:
   (a) Shut down the config server process.
   (b) Restart it with the 1.8.x binary from the MongoDB Download Page[1642].
5. Turn on the balancer:

```
use config
db.settings.update({_id:"balancer"},{$set : {stopped:false}})
```

Returning to 1.6

If for any reason you must move back to 1.6, follow the steps above in reverse. Please be careful that you have not inserted any documents larger than 4MB while running on 1.8 (where the max size has increased to 16MB). If you have you will get errors when the server tries to read those documents.

Journaling Returning to 1.6 after using 1.8 *Journaling* (page 606) works fine, as journaling does not change anything about the data file format. Suppose you are running 1.8.x with journaling enabled and you decide to switch back to 1.6. There are two scenarios:

- If you shut down cleanly with 1.8.x, just restart with the 1.6 mongod binary.
- If 1.8.x shut down uncleanly, start 1.8.x up again and let the journal files run to fix any damage (incomplete writes) that may have existed at the crash. Then shut down 1.8.x cleanly and restart with the 1.6 mongod binary.

Changes

Journaling

MongoDB now supports write-ahead *Journaling* (page 606) to facilitate fast crash recovery and durability in the storage engine. With journaling enabled, a `mongod` can be quickly restarted following a crash without needing to repair the *collections*. The aggregation framework makes it possible to do aggregation

Sparse and Covered Indexes

Sparse Indexes (page 574) are indexes that only include documents that contain the fields specified in the index. Documents missing the field will not appear in the index at all. This can significantly reduce index size for indexes of fields that contain only a subset of documents within a *collection*.

Covered Indexes (page 106) enable MongoDB to answer queries entirely from the index when the query only selects fields that the index contains.

Incremental MapReduce Support

The `mapReduce` command supports new options that enable incrementally updating existing *collections*. Previously, a MapReduce job could output either to a temporary collection or to a named permanent collection, which it would overwrite with new data.

You now have several options for the output of your MapReduce jobs:

[1642] http://downloads.mongodb.org/

- You can merge MapReduce output into an existing collection. Output from the Reduce phase will replace existing keys in the output collection if it already exists. Other keys will remain in the collection.
- You can now re-reduce your output with the contents of an existing collection. Each key output by the reduce phase will be reduced with the existing document in the output collection.
- You can replace the existing output collection with the new results of the MapReduce job (equivalent to setting a permanent output collection in previous releases)
- You can compute MapReduce inline and return results to the caller without persisting the results of the job. This is similar to the temporary collections generated in previous releases, except results are limited to 8MB.

For more information, see the `out` field options in the `mapReduce` document.

Additional Changes and Enhancements

1.8.1

- Sharding migrate fix when moving larger chunks.
- Durability fix with background indexing.
- Fixed mongos concurrency issue with many incoming connections.

1.8.0

- All changes from 1.7.x series.

1.7.6

- Bug fixes.

1.7.5

- *Journaling* (page 606).
- Extent allocation improvements.
- Improved *replica set* connectivity for `mongos`.
- `getLastError` improvements for *sharding*.

1.7.4

- `mongos` routes `slaveOk` queries to *secondaries* in *replica sets*.
- New `mapReduce` output options.
- *Sparse Indexes* (page 574).

1.7.3

- Initial *covered index* (page 106) support.
- Distinct can use data from indexes when possible.
- `mapReduce` can merge or reduce results into an existing collection.
- `mongod` tracks and `mongostat` displays network usage. See *mongostat*.

- Sharding stability improvements.

1.7.2

- `$rename` operator allows renaming of fields in a document.
- `db.eval()` not to block.
- Geo queries with sharding.
- *`mongostat --discover`* option
- Chunk splitting enhancements.
- Replica sets network enhancements for servers behind a nat.

1.7.1

- Many sharding performance enhancements.
- Better support for `$elemMatch` on primitives in embedded arrays.
- Query optimizer enhancements on range queries.
- Window service enhancements.
- Replica set setup improvements.
- `$pull` works on primitives in arrays.

1.7.0

- Sharding performance improvements for heavy insert loads.
- Slave delay support for replica sets.
- `getLastErrorDefaults` for replica sets.
- Auto completion in the shell.
- Spherical distance for geo search.
- All fixes from 1.6.1 and 1.6.2.

Release Announcement Forum Pages

- 1.8.1[1643], 1.8.0[1644]
- 1.7.6[1645], 1.7.5[1646], 1.7.4[1647], 1.7.3[1648], 1.7.2[1649], 1.7.1[1650], 1.7.0[1651]

[1643] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/v09MbhEm62Y
[1644] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/JeHQOnam6Qk
[1645] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/3t6GNZ1qGYc
[1646] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/S5R0Tx9wkEg
[1647] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/9Om3Vuw-y9c
[1648] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/DfNUrdbmflI
[1649] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/df7mwK6Xixo
[1650] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/HUR9zYtTpA8
[1651] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/TUnJCg9161A

Resources

- MongoDB Downloads[1652]
- All JIRA Issues resolved in 1.8[1653]

15.2.7 Release Notes for MongoDB 1.6

On this page

- Upgrading (page 1066)
- Sharding (page 1066)
- Replica Sets (page 1066)
- Other Improvements (page 1067)
- Installation (page 1067)
- 1.6.x Release Notes (page 1067)
- 1.5.x Release Notes (page 1067)

Upgrading

MongoDB 1.6 is a drop-in replacement for 1.4. To upgrade, simply shutdown `mongod` then restart with the new binaries.

Please note that you should upgrade to the latest version of whichever driver you're using. Certain drivers, including the Ruby driver, will require the upgrade, and all the drivers will provide extra features for connecting to replica sets.

Sharding

Sharding (page 733) is now production-ready, making MongoDB horizontally scalable, with no single point of failure. A single instance of `mongod` can now be upgraded to a distributed cluster with zero downtime when the need arises.

- *Sharding* (page 733)
- *Deploy a Sharded Cluster* (page 765)
- *Convert a Replica Set to a Sharded Cluster* (page 775)

Replica Sets

Replica sets (page 623), which provide automated failover among a cluster of `n` nodes, are also now available.

Please note that replica pairs are now deprecated; we strongly recommend that replica pair users upgrade to replica sets.

- *Replication* (page 623)
- *Deploy a Replica Set* (page 667)
- *Convert a Standalone to a Replica Set* (page 678)

[1652] http://mongodb.org/downloads
[1653] https://jira.mongodb.org/secure/IssueNavigator.jspa?mode=hide&requestId=10172

Other Improvements

- The `w` option (and `wtimeout`) forces writes to be propagated to `n` servers before returning success (this works especially well with replica sets)
- `$or queries`
- Improved concurrency
- `$slice` operator for returning subsets of arrays
- 64 indexes per collection (formerly 40 indexes per collection)
- 64-bit integers can now be represented in the shell using NumberLong
- The `findAndModify` command now supports upserts. It also allows you to specify fields to return
- $showDiskLoc option to see disk location of a document
- Support for IPv6 and UNIX domain sockets

Installation

- Windows service improvements
- The C++ client is a separate tarball from the binaries

1.6.x Release Notes

- 1.6.5[1654]

1.5.x Release Notes

- 1.5.8[1655]
- 1.5.7[1656]
- 1.5.6[1657]
- 1.5.5[1658]
- 1.5.4[1659]
- 1.5.3[1660]
- 1.5.2[1661]
- 1.5.1[1662]
- 1.5.0[1663]

[1654] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/06_QCC05Fpk
[1655] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/uJfF1QN6Thk
[1656] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/OYvz40RWs90
[1657] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/4l0N2U_H0cQ
[1658] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/oO749nvTARY
[1659] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/380V_Ec_q1c
[1660] https://groups.google.com/forum/?hl=en&fromgroups=#!topic/mongodb-user/hsUQL9CxTQw
[1661] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/94EE3HVidAA
[1662] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/7SBPQ2RSfdM
[1663] https://groups.google.com/forum/?fromgroups=#!topic/mongodb-user/VAhJcjDGTy0

You can see a full list of all changes on JIRA[1664].

Thank you everyone for your support and suggestions!

15.2.8 Release Notes for MongoDB 1.4

On this page

- Upgrading (page 1068)
- Core Server Enhancements (page 1068)
- Replication and Sharding (page 1068)
- Deployment and Production (page 1068)
- Query Language Improvements (page 1069)
- Geo (page 1069)

Upgrading

We're pleased to announce the 1.4 release of MongoDB. 1.4 is a drop-in replacement for 1.2. To upgrade you just need to shutdown `mongod`, then restart with the new binaries. (Users upgrading from release 1.0 should review the *1.2 release notes* (page 1069), in particular the instructions for upgrading the DB format.)

Release 1.4 includes the following improvements over release 1.2:

Core Server Enhancements

- *concurrency* (page 835) improvements
- indexing memory improvements
- *background index creation* (page 577)
- better detection of regular expressions so the index can be used in more cases

Replication and Sharding

- better handling for restarting slaves offline for a while
- fast new slaves from snapshots (`--fastsync`)
- configurable slave delay (`--slavedelay`)
- replication handles clock skew on master
- `$inc` replication fixes
- sharding alpha 3 - notably 2-phase commit on config servers

Deployment and Production

- *configure "slow threshold" for profiling* (page 327)
- ability to do `fsync + lock` for backing up raw files

[1664] https://jira.mongodb.org/secure/IssueNavigator.jspa?mode=hide&requestId=10107

- option for separate directory per db (`--directoryperdb`)
- `http://localhost:28017/_status` to get serverStatus via http
- REST interface is off by default for security (`--rest` to enable)
- can rotate logs with a db command, `logRotate`
- enhancements to `serverStatus` command (db.serverStatus()) - counters and *replication lag* (page 712) stats
- new `mongostat` tool

Query Language Improvements

- `$all` with regex
- `$not`
- partial matching of array elements `$elemMatch`
- $ operator for updating arrays
- `$addToSet`
- `$unset`
- `$pull` supports object matching
- `$set` with array indexes

Geo

- *2d geospatial search* (page 561)
- geo `$center` and `$box` searches

15.2.9 Release Notes for MongoDB 1.2.x

On this page

- New Features (page 1069)
- DB Upgrade Required (page 1070)
- Replication Changes (page 1070)
- mongoimport (page 1070)
- field filter changing (page 1070)

New Features

- More indexes per collection
- Faster index creation
- Map/Reduce
- Stored JavaScript functions
- Configurable fsync time
- Several small features and fixes

DB Upgrade Required

There are some changes that will require doing an upgrade if your previous version is <= 1.0.x. If you're already using a version >= 1.1.x then these changes aren't required. There are 2 ways to do it:

- `--upgrade`
  - stop your `mongod` process
  - run `./mongod --upgrade`
  - start `mongod` again
- use a slave
  - start a slave on a different port and data directory
  - when its synced, shut down the master, and start the new slave on the regular port.

Ask in the forums or IRC for more help.

Replication Changes

- There have been minor changes in replication. If you are upgrading a master/slave setup from <= 1.1.2 you have to update the slave first.

mongoimport

- `mongoimportjson` has been removed and is replaced with `mongoimport` that can do json/csv/tsv

field filter changing

- We've changed the semantics of the field filter a little bit. Previously only objects with those fields would be returned. Now the field filter only changes the output, not which objects are returned. If you need that behavior, you can use `$exists`

15.3 MongoDB Version Numbers

For MongoDB `2.4.1`, `2.4` refers to the release series and `.1` refers to the revision. The second component of the release series (e.g. `4` in `2.4.1`) describes the type of release series. Release series ending with even numbers (e.g. `4` above) are *stable* and ready for production, while odd numbers are for *development* and testing only.

Generally, changes in the release series (e.g. `2.2` to `2.4`) mark the introduction of new features that may break backwards compatibility. Changes to the revision number mark the release bug fixes and backwards-compatible changes.

Important: Always upgrade to the latest stable revision of your release series.

The version numbering system for MongoDB differs from the system used for the MongoDB drivers. Drivers use only the first number to indicate a major version. For details, see *drivers-version-numbers*.

Example

Version numbers

- 2.0.0 : Stable release.

- 2.0.1 : Revision.
- 2.1.0 : Development release *for testing only*. Includes new features and changes for testing. Interfaces and stability may not be compatible in development releases.
- 2.2.0 : Stable release. This is a culmination of the 2.1.x development series.

CHAPTER 16

About MongoDB Documentation

On this page

- License (page 1073)
- Editions (page 1073)
- Version and Revisions (page 1074)
- Report an Issue or Make a Change Request (page 1074)
- Contribute to the Documentation (page 1075)

The MongoDB Manual[1] contains comprehensive documentation on MongoDB. This page describes the manual's licensing, editions, and versions, and describes how to make a change request and how to contribute to the manual.

16.1 License

16.2 Editions

In addition to the MongoDB Manual[3], you can also access this content in the following editions:

- PDF Format[4] (without reference).
- HTML tar.gz[5]
- ePub Format[6]

You also can access PDF files that contain subsets of the MongoDB Manual:

- MongoDB Reference Manual[7]

[1] http://docs.mongodb.org/manual/#
[2] http://creativecommons.org/licenses/by-nc-sa/3.0/us/
[3] http://docs.mongodb.org/manual/#
[4] http://docs.mongodb.org/master/MongoDB-manual.pdf
[5] http://docs.mongodb.org/master/manual.tar.gz
[6] http://docs.mongodb.org/master/MongoDB-manual.epub
[7] http://docs.mongodb.org/master/MongoDB-reference-manual.pdf

- MongoDB CRUD Operations[8]
- Data Models for MongoDB[9]
- MongoDB Data Aggregation[10]
- Replication and MongoDB[11]
- Sharding and MongoDB[12]
- MongoDB Administration[13]
- MongoDB Security[14]

MongoDB Reference documentation is also available as part of dash[15]. You can also access the MongoDB Man Pages[16] which are also distributed with the official MongoDB Packages.

16.3 Version and Revisions

This version of the manual reflects version 3.2 of MongoDB.

See the MongoDB Documentation Project Page[17] for an overview of all editions and output formats of the MongoDB Manual. You can see the full revision history and track ongoing improvements and additions for all versions of the manual from its GitHub repository[18].

This edition reflects "`master`" branch of the documentation as of the "`ed9746b481b9ab0a337e8e3f8aef56854d51eca1`" revision. This branch is explicitly accessible via "https://docs.mongodb.org/master" and you can always reference the commit of the current manual in the release.txt[19] file.

The most up-to-date, current, and stable version of the manual is always available at "http://docs.mongodb.org/manual/".

16.4 Report an Issue or Make a Change Request

To report an issue with this manual or to make a change request, file a ticket at the MongoDB DOCS Project on Jira[20].

[8] http://docs.mongodb.org/master/MongoDB-crud-guide.pdf
[9] http://docs.mongodb.org/master/MongoDB-data-models-guide.pdf
[10] http://docs.mongodb.org/master/MongoDB-aggregation-guide.pdf
[11] http://docs.mongodb.org/master/MongoDB-replication-guide.pdf
[12] http://docs.mongodb.org/master/MongoDB-sharding-guide.pdf
[13] http://docs.mongodb.org/master/MongoDB-administration-guide.pdf
[14] http://docs.mongodb.org/master/MongoDB-security-guide.pdf
[15] http://kapeli.com/dash
[16] http://docs.mongodb.org/master/manpages.tar.gz
[17] http://docs.mongodb.org
[18] https://github.com/mongodb/docs
[19] http://docs.mongodb.org/master/release.txt
[20] https://jira.mongodb.org/browse/DOCS

16.5 Contribute to the Documentation

16.5.1 MongoDB Manual Translation

The original language of all MongoDB documentation is American English. However it is of critical importance to the documentation project to ensure that speakers of other languages can read and understand the documentation.

To this end, the MongoDB Documentation Project is preparing to launch a translation effort to allow the community to help bring the documentation to speakers of other languages.

If you would like to express interest in helping to translate the MongoDB documentation once this project is opened to the public, please:

- complete the MongoDB Contributor Agreement[21], and
- join the mongodb-translators[22] user group.

The mongodb-translators[23] user group exists to facilitate collaboration between translators and the documentation team at large. You can join the group without signing the Contributor Agreement, but you will not be allowed to contribute translations.

See also:

- *Contribute to the Documentation* (page 1075)
- *Style Guide and Documentation Conventions* (page 1076)
- *MongoDB Manual Organization* (page 1085)
- *MongoDB Documentation Practices and Processes* (page 1082)
- *MongoDB Documentation Build System* (page 1086)

The entire documentation source for this manual is available in the mongodb/docs repository[24], which is one of the MongoDB project repositories on GitHub[25].

To contribute to the documentation, you can open a GitHub account[26], fork the mongodb/docs repository[27], make a change, and issue a pull request.

In order for the documentation team to accept your change, you must complete the MongoDB Contributor Agreement[28].

You can clone the repository by issuing the following command at your system shell:

```
git clone git://github.com/mongodb/docs.git
```

16.5.2 About the Documentation Process

The MongoDB Manual uses Sphinx[29], a sophisticated documentation engine built upon Python Docutils[30]. The original reStructured Text[31] files, as well as all necessary Sphinx extensions and build tools, are available in the same

[21] http://www.mongodb.com/legal/contributor-agreement
[22] http://groups.google.com/group/mongodb-translators
[23] http://groups.google.com/group/mongodb-translators
[24] https://github.com/mongodb/docs
[25] http://github.com/mongodb
[26] https://github.com/
[27] https://github.com/mongodb/docs
[28] http://www.mongodb.com/contributor
[29] http://sphinx-doc.org//
[30] http://docutils.sourceforge.net/
[31] http://docutils.sourceforge.net/rst.html

repository as the documentation.

For more information on the MongoDB documentation process, see:

Style Guide and Documentation Conventions

This document provides an overview of the style for the MongoDB documentation stored in this repository. The overarching goal of this style guide is to provide an accessible base style to ensure that our documentation is easy to read, simple to use, and straightforward to maintain.

For information regarding the MongoDB Manual organization, see *MongoDB Manual Organization* (page 1085).

Document History

2011-09-27: Document created with a (very) rough list of style guidelines, conventions, and questions.

2012-01-12: Document revised based on slight shifts in practice, and as part of an effort of making it easier for people outside of the documentation team to contribute to documentation.

2012-03-21: Merged in content from the Jargon, and cleaned up style in light of recent experiences.

2012-08-10: Addition to the "Referencing" section.

2013-02-07: Migrated this document to the manual. Added "map-reduce" terminology convention. Other edits.

2013-11-15: Added new table of preferred terms.

2016-01-05: Standardizing on 'embedded document'

Naming Conventions

This section contains guidelines on naming files, sections, documents and other document elements.

- File naming Convention:
  - For Sphinx, all files should have a `.txt` extension.
  - Separate words in file names with hyphens (i.e. `-`.)
  - For most documents, file names should have a terse one or two word name that describes the material covered in the document. Allow the path of the file within the document tree to add some of the required context/categorization. For example it's acceptable to have `https://docs.mongodb.org/manual/core/sharding.rst` and `https://docs.mongodb.org/manual/administration/sharding.rst`.
  - For tutorials, the full title of the document should be in the file name. For example, `https://docs.mongodb.org/manual/tutorial/replace-one-configuration-server-in-a-shar`
- Phrase headlines and titles so users can determine what questions the text will answer, and material that will be addressed, without needing them to read the content. This shortens the amount of time that people spend looking for answers, and improvise search/scanning, and possibly "SEO."
- Prefer titles and headers in the form of "Using foo" over "How to Foo."
- When using target references (i.e. `:ref:` references in documents), use names that include enough context to be intelligible through all documentation. For example, use "`replica-set-secondary-only-node`" as opposed to "`secondary-only-node`". This makes the source more usable and easier to maintain.

Style Guide

This includes the local typesetting, English, grammatical, conventions and preferences that all documents in the manual should use. The goal here is to choose good standards, that are clear, and have a stylistic minimalism that does not interfere with or distract from the content. A uniform style will improve user experience and minimize the effect of a multi-authored document.

Spelling Use American spelling.

Punctuation

- Use the Oxford comma.

  Oxford commas are the commas in a list of things (e.g. "something, something else, and another thing") before the conjunction (e.g. "and" or "or.").

- Do not add two spaces after terminal punctuation, such as periods.
- Place commas and periods inside quotation marks.

Headings Use title case for headings and document titles. Title case capitalizes the first letter of the first, last, and all significant words.

Verbs Verb tense and mood preferences, with examples:

- **Avoid** the first person. For example do not say, "We will begin the backup process by locking the database," or "I begin the backup process by locking my database instance."
- **Use** the second person. "If you need to back up your database, start by locking the database first." In practice, however, it's more concise to imply second person using the imperative, as in "Before initiating a backup, lock the database."
- When indicated, use the imperative mood. For example: "Back up your databases often" and "To prevent data loss, back up your databases."
- The future perfect is also useful in some cases. For example, "Creating disk snapshots without locking the database will lead to an invalid state."
- Avoid helper verbs, as possible, to increase clarity and concision. For example, attempt to avoid "this does foo" and "this will do foo" when possible. Use "does foo" over "will do foo" in situations where "this foos" is unacceptable.

Referencing

- To refer to future or planned functionality in MongoDB or a driver, *always* link to the Jira case. The Manual's `conf.py` provides an `:issue:` role that links directly to a Jira case (e.g. `:issue:\'SERVER-9001\'`).
- For non-object references (i.e. functions, operators, methods, database commands, settings) always reference only the first occurrence of the reference in a section. You should *always* reference objects, except in section headings.
- Structure references with the *why* first; the link second.

  For example, instead of this:

  Use the *Convert a Replica Set to a Sharded Cluster* (page 775) procedure if you have an existing replica set.

  Type this:

To deploy a sharded cluster for an existing replica set, see *Convert a Replica Set to a Sharded Cluster* (page 775).

General Formulations

- Contractions are acceptable insofar as they are necessary to increase readability and flow. Avoid otherwise.
- Make lists grammatically correct.
  - Do not use a period after every item unless the list item completes the unfinished sentence before the list.
  - Use appropriate commas and conjunctions in the list items.
  - Typically begin a bulleted list with an introductory sentence or clause, with a colon or comma.
- The following terms are one word:
  - standalone
  - workflow
- Use "unavailable," "offline," or "unreachable" to refer to a `mongod` instance that cannot be accessed. Do not use the colloquialism "down."
- Always write out units (e.g. "megabytes") rather than using abbreviations (e.g. "MB".)

Structural Formulations

- There should be at least two headings at every nesting level. Within an "h2" block, there should be either: no "h3" blocks, 2 "h3" blocks, or more than 2 "h3" blocks.
- Section headers are in title case (capitalize first, last, and all important words) and should effectively describe the contents of the section. In a single document you should strive to have section titles that are not redundant and grammatically consistent with each other.
- Use paragraphs and paragraph breaks to increase clarity and flow. Avoid burying critical information in the middle of long paragraphs. Err on the side of shorter paragraphs.
- Prefer shorter sentences to longer sentences. Use complex formations only as a last resort, if at all (e.g. compound complex structures that require semi-colons).
- Avoid paragraphs that consist of single sentences as they often represent a sentence that has unintentionally become too complex or incomplete. However, sometimes such paragraphs are useful for emphasis, summary, or introductions.

  As a corollary, most sections should have multiple paragraphs.
- For longer lists and more complex lists, use bulleted items rather than integrating them inline into a sentence.
- Do not expect that the content of any example (inline or blocked) will be self explanatory. Even when it feels redundant, make sure that the function and use of every example is clearly described.

ReStructured Text and Typesetting

- Place spaces between nested parentheticals and elements in JavaScript examples. For example, prefer `{ [ a, a, a ] }` over `{[a,a,a]}`.
- For underlines associated with headers in RST, use:
  - = for heading level 1 or h1s. Use underlines and overlines for document titles.
  - − for heading level 2 or h2s.
  - ~ for heading level 3 or h3s.

  - ` for heading level 4 or h4s.

- Use hyphens (-) to indicate items of an ordered list.
- Place footnotes and other references, if you use them, at the end of a section rather than the end of a file.

  Use the footnote format that includes automatic numbering and a target name for ease of use. For instance a footnote tag may look like: `[#note]_` with the corresponding directive holding the body of the footnote that resembles the following: `.. [#note]`.

  Do **not** include `.. code-block:: [language]` in footnotes.
- As it makes sense, use the `.. code-block:: [language]` form to insert literal blocks into the text. While the double colon, `::`, is functional, the `.. code-block:: [language]` form makes the source easier to read and understand.
- For all mentions of referenced types (i.e. commands, operators, expressions, functions, statuses, etc.) use the reference types to ensure uniform formatting and cross-referencing.

Paths and Hostnames

- Use angle brackets to denote areas that users should input the relevant path, as in `--dbpath <path>`.
- When including sample hostnames, use `example.com`, `example.net`, or `example.org`, which are reserved for documentation purposes. See RFC2606[32] and RFC6761[33] for more information.

[32] http://tools.ietf.org/html/rfc2606
[33] http://tools.ietf.org/html/rfc6761

Jargon and Common Terms

Preferred Term	Concept	Dispreferred Alternatives	Notes
document	A single, top-level object/record in a MongoDB collection.	record, object, row	Prefer document over object because of concerns about cross-driver language handling of objects. Reserve record for "allocation" of storage. Avoid "row," as possible.
database	A group of collections. Refers to a group of data files. This is the "logical" sense of the term "database."		Avoid genericizing "database." Avoid using database to refer to a server process or a data set. This applies both to the datastoring contexts as well as other (related) operational contexts (command context, authentication/authorization context.)
instance	A daemon process. (e.g. **mongos** or **mongod**)	process (acceptable sometimes), node (never acceptable), server.	Avoid using instance, unless it modifies something specifically. Having a descriptor for a process/instance makes it possible to avoid needing to make mongod or mongos plural. Server and node are both vague and contextually difficult to disambiguate with regards to application servers, and underlying hardware.
field name	The identifier of a value in a document.	key, column	Avoid introducing unrelated terms for a single field. In the documentation we've rarely had to discuss the identifier of a field, so the extra word here isn't burdensome.
field/value	The name/value pair that describes a unit of data in MongoDB.	key, slot, attribute	Use to emphasize the difference between the name of a field and its value For example, "_id" is the field and the default value is an ObjectId.
value	The data content of a field.	data	
MongoDB	A group of processes, or deployment that implement the MongoDB interface.	mongo, mongodb, cluster	Stylistic preference, mostly. In some cases it's useful to be able to refer generically to instances (that may be either **mongod** or **mongos**.)
embedded document	An embedded or nested document within a document or an array.	nested document	
mapreduce	An operation performed by the mapReduce command.	mapReduce, map reduce, map/reduce	Avoid confusion with the command, shell helper, and driver interfaces. Makes it possible to discuss the operation generally.
cluster	A sharded cluster.	grid, shard cluster, set, deployment	Cluster is a great word for a group of processes; however, it's important to avoid letting the term become generic. Do not use for any group of MongoDB processes or deployments.
sharded cluster	A *sharded cluster*.	shard cluster, cluster, sharded system	
replica set	A deployment of replicating **mongod** programs that provide redundancy and automatic failover.	set, replication deployment	
deployment	A group of MongoDB processes, or a standalone **mongod** instance.	cluster, system	Typically in the form MongoDB deployment. Includes standalones, replica sets and sharded clusters.
data	The collection of physical	database, data	Important to keep the distinction between the

Database Systems and Processes

- To indicate the entire database system, use "MongoDB," not mongo or Mongo.
- To indicate the database process or a server instance, use `mongod` or `mongos`. Refer to these as "processes" or "instances." Reserve "database" for referring to a database structure, i.e., the structure that holds collections and refers to a group of files on disk.

Distributed System Terms

- Refer to partitioned systems as "sharded clusters." Do not use shard clusters or sharded systems.
- Refer to configurations that run with replication as "replica sets" (or "master/slave deployments") rather than "clusters" or other variants.

Data Structure Terms

- "document" refers to "rows" or "records" in a MongoDB database. Potential confusion with "JSON Documents."

  Do not refer to documents as "objects," because drivers (and MongoDB) do not preserve the order of fields when fetching data. If the order of objects matter, use an array.
- "field" refers to a "key" or "identifier" of data within a MongoDB document.
- "value" refers to the contents of a "field".
- "embedded document" describes a nested document.

Other Terms

- Use `example.net` (and `.org` or `.com` if needed) for all examples and samples.
- Hyphenate "map-reduce" in order to avoid ambiguous reference to the command name. Do not camel-case.

Notes on Specific Features

- Geo-Location
  1. While MongoDB *is capable* of storing coordinates in embedded documents, in practice, users should only store coordinates in arrays. (See: DOCS-41[34].)

MongoDB Documentation Practices and Processes

This document provides an overview of the practices and processes.

Commits

When relevant, include a Jira case identifier in a commit message. Reference documentation cases when applicable, but feel free to reference other cases from jira.mongodb.org[35].

Err on the side of creating a larger number of discrete commits rather than bundling large set of changes into one commit.

[34] https://jira.mongodb.org/browse/DOCS-41
[35] http://jira.mongodb.org/

For the sake of consistency, remove trailing whitespaces in the source file.

"Hard wrap" files to between 72 and 80 characters per-line.

Standards and Practices

- At least two people should vet all non-trivial changes to the documentation before publication. One of the reviewers should have significant technical experience with the material covered in the documentation.
- All development and editorial work should transpire on GitHub branches or forks that editors can then merge into the publication branches.

Collaboration

To propose a change to the documentation, do either of the following:

- Open a ticket in the documentation project[36] proposing the change. Someone on the documentation team will make the change and be in contact with you so that you can review the change.
- Using GitHub[37], fork the mongodb/docs repository[38], commit your changes, and issue a pull request. Someone on the documentation team will review and incorporate your change into the documentation.

Builds

Building the documentation is useful because Sphinx[39] and docutils can catch numerous errors in the format and syntax of the documentation. Additionally, having access to an example documentation as it *will* appear to the users is useful for providing more effective basis for the review process. Besides Sphinx, Pygments, and Python-Docutils, the documentation repository contains all requirements for building the documentation resource.

Talk to someone on the documentation team if you are having problems running builds yourself.

Publication

The makefile for this repository contains targets that automate the publication process. Use `make html` to publish a test build of the documentation in the `build/` directory of your repository. Use `make publish` to build the full contents of the manual from the current branch in the `../public-docs/` directory relative the docs repository.

Other targets include:

- `man` - builds UNIX Manual pages for all Mongodb utilities.
- `push` - builds and deploys the contents of the `../public-docs/`.
- `pdfs` - builds a PDF version of the manual (requires LaTeX dependencies.)

Branches

This section provides an overview of the git branches in the MongoDB documentation repository and their use.

[36] https://jira.mongodb.org/browse/DOCS
[37] https://github.com/
[38] https://github.com/mongodb/docs
[39] http://sphinx.pocoo.org/

At the present time, future work transpires in the `master`, with the main publication being `current`. As the documentation stabilizes, the documentation team will begin to maintain branches of the documentation for specific MongoDB releases.

Migration from Legacy Documentation

The MongoDB.org Wiki contains a wealth of information. As the transition to the Manual (i.e. this project and resource) continues, it's *critical* that no information disappears or goes missing. The following process outlines *how* to migrate a wiki page to the manual:

1. Read the relevant sections of the Manual, and see what the new documentation has to offer on a specific topic.

   In this process you should follow cross references and gain an understanding of both the underlying information and how the parts of the new content relates its constituent parts.

2. Read the wiki page you wish to redirect, and take note of all of the factual assertions, examples presented by the wiki page.

3. Test the factual assertions of the wiki page to the greatest extent possible. Ensure that example output is accurate. In the case of commands and reference material, make sure that documented options are accurate.

4. Make corrections to the manual page or pages to reflect any missing pieces of information.

   The target of the redirect need *not* contain every piece of information on the wiki page, **if** the manual as a whole does, and relevant section(s) with the information from the wiki page are accessible from the target of the redirection.

5. As necessary, get these changes reviewed by another writer and/or someone familiar with the area of the information in question.

   At this point, update the relevant Jira case with the target that you've chosen for the redirect, and make the ticket unassigned.

6. When someone has reviewed the changes and published those changes to Manual, you, or preferably someone else on the team, should make a final pass at both pages with fresh eyes and then make the redirect.

   Steps 1-5 should ensure that no information is lost in the migration, and that the final review in step 6 should be trivial to complete.

Review Process

Types of Review The content in the Manual undergoes many types of review, including the following:

Initial Technical Review Review by an engineer familiar with MongoDB and the topic area of the documentation. This review focuses on technical content, and correctness of the procedures and facts presented, but can improve any aspect of the documentation that may still be lacking. When both the initial technical review and the content review are complete, the piece may be "published."

Content Review Textual review by another writer to ensure stylistic consistency with the rest of the manual. Depending on the content, this may precede or follow the initial technical review. When both the initial technical review and the content review are complete, the piece may be "published."

Consistency Review This occurs post-publication and is content focused. The goals of consistency reviews are to increase the internal consistency of the documentation as a whole. Insert relevant cross-references, update the style as needed, and provide background fact-checking.

When possible, consistency reviews should be as systematic as possible and we should avoid encouraging stylistic and information drift by editing only small sections at a time.

Subsequent Technical Review If the documentation needs to be updated following a change in functionality of the server or following the resolution of a user issue, changes may be significant enough to warrant additional technical review. These reviews follow the same form as the "initial technical review," but is often less involved and covers a smaller area.

Review Methods If you're not a usual contributor to the documentation and would like to review something, you can submit reviews in any of the following methods:

- If you're reviewing an open pull request in GitHub, the best way to comment is on the "overview diff," which you can find by clicking on the "diff" button in the upper left portion of the screen. You can also use the following URL to reach this interface:

  ```
  https://github.com/mongodb/docs/pull/[pull-request-id]/files
  ```

  Replace `[pull-request-id]` with the identifier of the pull request. Make all comments inline, using GitHub's comment system.

  You may also provide comments directly on commits, or on the pull request itself but these commit-comments are archived in less coherent ways and generate less useful emails, while comments on the pull request lead to less specific changes to the document.

- Leave feedback on Jira cases in the DOCS[40] project. These are better for more general changes that aren't necessarily tied to a specific line, or affect multiple files.

- Create a fork of the repository in your GitHub account, make any required changes and then create a pull request with your changes.

  If you insert lines that begin with any of the following annotations:

  ```
  .. TODO:
  TODO:
  .. TODO
  TODO
  ```

  followed by your comments, it will be easier for the original writer to locate your comments. The two dots `..` format is a comment in reStructured Text, which will hide your comments from Sphinx and publication if you're worried about that.

  This format is often easier for reviewers with larger portions of content to review.

MongoDB Manual Organization

This document provides an overview of the global organization of the documentation resource. Refer to the notes below if you are having trouble understanding the reasoning behind a file's current location, or if you want to add new documentation but aren't sure how to integrate it into the existing resource.

If you have questions, don't hesitate to open a ticket in the Documentation Jira Project[41] or contact the documentation team[42].

[40] http://jira.mongodb.org/browse/DOCS
[41] https://jira.mongodb.org/browse/DOCS
[42] docs@mongodb.com

Global Organization

Indexes and Experience The documentation project has two "index files": `https://docs.mongodb.org/manual/contents.txt` and `https://docs.mongodb.org/manual/index.txt`. The "contents" file provides the documentation's tree structure, which Sphinx uses to create the left-pane navigational structure, to power the "Next" and "Previous" page functionality, and to provide all overarching outlines of the resource. The "index" file is not included in the "contents" file (and thus builds will produce a warning here) and is the page that users first land on when visiting the resource.

Having separate "contents" and "index" files provides a bit more flexibility with the organization of the resource while also making it possible to customize the primary user experience.

Topical Organization The placement of files in the repository depends on the *type* of documentation rather than the *topic* of the content. Like the difference between `contents.txt` and `index.txt`, by decoupling the organization of the files from the organization of the information the documentation can be more flexible and can more adequately address changes in the product and in users' needs.

Files in the `source/` directory represent the tip of a logical tree of documents, while *directories* are containers of types of content. The `administration` and `applications` directories, however, are legacy artifacts and with a few exceptions contain sub-navigation pages.

With several exceptions in the `reference/` directory, there is only one level of sub-directories in the `source/` directory.

Tools

The organization of the site, like all Sphinx sites derives from the `toctree` structure. However, in order to annotate the table of contents and provide additional flexibility, the MongoDB documentation generates `toctree` structures using data from YAML files stored in the `source/includes/` directory. These files start with `ref-toc` or `toc` and generate output in the `source/includes/toc/` directory. Briefly this system has the following behavior:

- files that start with `ref-toc` refer to the documentation of API objects (i.e. commands, operators and methods), and the build system generates files that hold `toctree` directives as well as files that hold *tables* that list objects and a brief description.
- files that start with `toc` refer to all other documentation and the build system generates files that hold `toctree` directives as well as files that hold *definition lists* that contain links to the documents and short descriptions the content.
- file names that have `spec` following `toc` or `ref-toc` will generate aggregated tables or definition lists and allow ad-hoc combinations of documents for landing pages and quick reference guides.

MongoDB Documentation Build System

This document contains more direct instructions for building the MongoDB documentation.

Getting Started

Install Dependencies The MongoDB Documentation project depends on the following tools:

- Python
- Git
- Inkscape (Image generation.)

- LaTeX/PDF LaTeX (typically texlive; for building PDFs)
- Giza[43]

OS X Install Sphinx, Docutils, and their dependencies with `easy_install` the following command:

```
easy_install giza
```

Feel free to use `pip` rather than `easy_install` to install python packages.

To generate the images used in the documentation, download and install Inkscape[44].

Optional

To generate PDFs for the full production build, install a TeX distribution (for building the PDF.) If you do not have a LaTeX installation, use MacTeX[45]. This is **only** required to build PDFs.

Arch Linux Install packages from the system repositories with the following command:

```
pacman -S inkscape python2-pip
```

Then install the following Python packages:

```
pip2 install giza
```

Optional

To generate PDFs for the full production build, install the following packages from the system repository:

```
pacman -S texlive-bin texlive-core texlive-latexextra
```

Debian/Ubuntu Install the required system packages with the following command:

```
apt-get install inkscape python-pip
```

Then install the following Python packages:

```
pip install giza
```

Optional

To generate PDFs for the full production build, install the following packages from the system repository:

```
apt-get install texlive-latex-recommended texlive-latex-recommended
```

Setup and Configuration Clone the repository:

```
git clone git://github.com/mongodb/docs.git
```

[43] https://pypi.python.org/pypi/giza
[44] http://inkscape.org/download/
[45] http://www.tug.org/mactex/2011/

Building the Documentation

The MongoDB documentation build system is entirely accessible via `make` targets. For example, to build an HTML version of the documentation issue the following command:

```
make html
```

You can find the build output in `build/<branch>/html`, where `<branch>` is the name of the current branch.

In addition to the `html` target, the build system provides the following targets:

`publish` Builds and integrates all output for the production build. Build output is in `build/public/<branch>/`. When you run `publish` in the `master`, the build will generate some output in `build/public/`.

`push`; `stage` Uploads the production build to the production or staging web servers. Depends on `publish`. Requires access production or staging environment.

`push-all`; `stage-all` Uploads the entire content of `build/public/` to the web servers. Depends on `publish`. Not used in common practice.

`push-with-delete`; `stage-with-delete` Modifies the action of `push` and `stage` to remove remote file that don't exist in the local build. Use with caution.

`html`; `latex`; `dirhtml`; `epub`; `texinfo`; `man`; `json` These are standard targets derived from the default Sphinx Makefile, with adjusted dependencies. Additionally, for all of these targets you can append `-nitpick` to increase Sphinx's verbosity, or `-clean` to remove all Sphinx build artifacts.

`latex` performs several additional post-processing steps on `.tex` output generated by Sphinx. This target will also compile PDFs using `pdflatex`.

`html` and `man` also generates a `.tar.gz` file of the build outputs for inclusion in the final releases.

If you have any questions, please feel free to open a Jira Case[46].

[46] https://jira.mongodb.org/browse/DOCS