Microsoft

SQL Server Technical Article

Writer: Jonathan Foster

Technical Reviewer: Karthik Pinnamaneni; Andre Ciabattari

Published: November, 2013

Applies to: SQL Server 2012

Summary: This paper examines the performance of an XtremIO All Flash array in an I/O intensive BI environment.

Copyright

Introduction

The Customer Service & Support BI group at Microsoft (CSSBI) was offered the opportunity to test BI-related SQL loads on EMC's XtremIO storage unit for ultra-fast IO rates. The group decided to create a load test that simulated peak activity in their data acquisition platform (DAP). The DAP is a single SQL Instance containing 100's of various size databases hosted on a very powerful server. These databases are updated w/ new data at various frequencies while at the same time queried by managed systems, reports, and ad hoc users. This paper presents the disk I/O performance of the XtremIO unit in a DAP test environment and also provides disk I/O performance metrics from a standard SAN to be used as a reference point. The purpose of this paper is to provide BI architects and DBA's with viable performance information for helping select viable hardware configurations to meet specific performance needs.

Background

As mentioned previously, the CSSBI DAP server is subjected to competitive IO requests throughout each day with the heaviest load occurring in the early morning business hours. Many of these IO transfers involve large datasets like clickstream or user event data and/or large-object data types like online forum chats and XML data. A portion of these large data flows are repetitive meaning the same query types occur multiple times each day. Near-real time (NRT) subsets of these data flows are also occurring simultaneously. CSSBI's value to stakeholders is its ability to provide consistently good performance in an easily accessible symmetric multiprocessing (SMP) environment. The project's success has resulted in exponential growth of both data and users which has forced the IT owners to regularly upgrade the underlying infrastructure & platform to ensure optimal performance.

Purpose

When EMC shared the technical details and future integration plans of the XtremIO storage unit, the CSSBI IT team wanted to evaluate whether the product might solve a specific need for expanding the NRT or real-time (RT) capability of the DAP for larger datasets as described previously. Most often in these scenarios the solution is one that involves more expensive massive parallel processing (MPP) capability. The CSSBI IT team's goal is to try and meet the business requirements with a less expensive SMP capability for as long as possible.

Test Environment

Hardware	
Platform	Cisco B230-Base-M2 Blade
CPU (Clock Speed, Cache, Max TDP)	2 x 10 core (hyper threading) Intel Xeon E72860 @ 2.27GHz, 24MB L3 Cache, 130W (40 logical processors)
Memory	28 x 8GB DDR3 (224GB total)
HDDs (Capacity, Rotational Speed, Interface, Form Factor)	Local Storage 2 x 64GB SSD (Raid 1) (C: / D:)
XtremIO Storage	1 XBrick @ 7.5TB usable presented as 5 separate volumes of 1TB (1) & 1.5TB(4)
SAN Storage	EMC VNX 5700 having 1 auto-tiered pool composed of SAS 10k (95%) & SSD (5%)
Controller	Cisco N20-AC0002 Con adapter (dual 8GB FC paths)
Power Supply	(Power supplies in chassis not blades)
Network	Single Dual Port Embedded Cisco NIC @ 10Gbps

Storage Configuration	
XtremIO X-Brick	XtremIO Data Protection methodology (XDP)
SAN Pool	RAID 5

Software	
OS	Microsoft Windows Server 2012 Datacenter
SQL	Microsoft SQL Server 2012 Developer SP1 (X64)

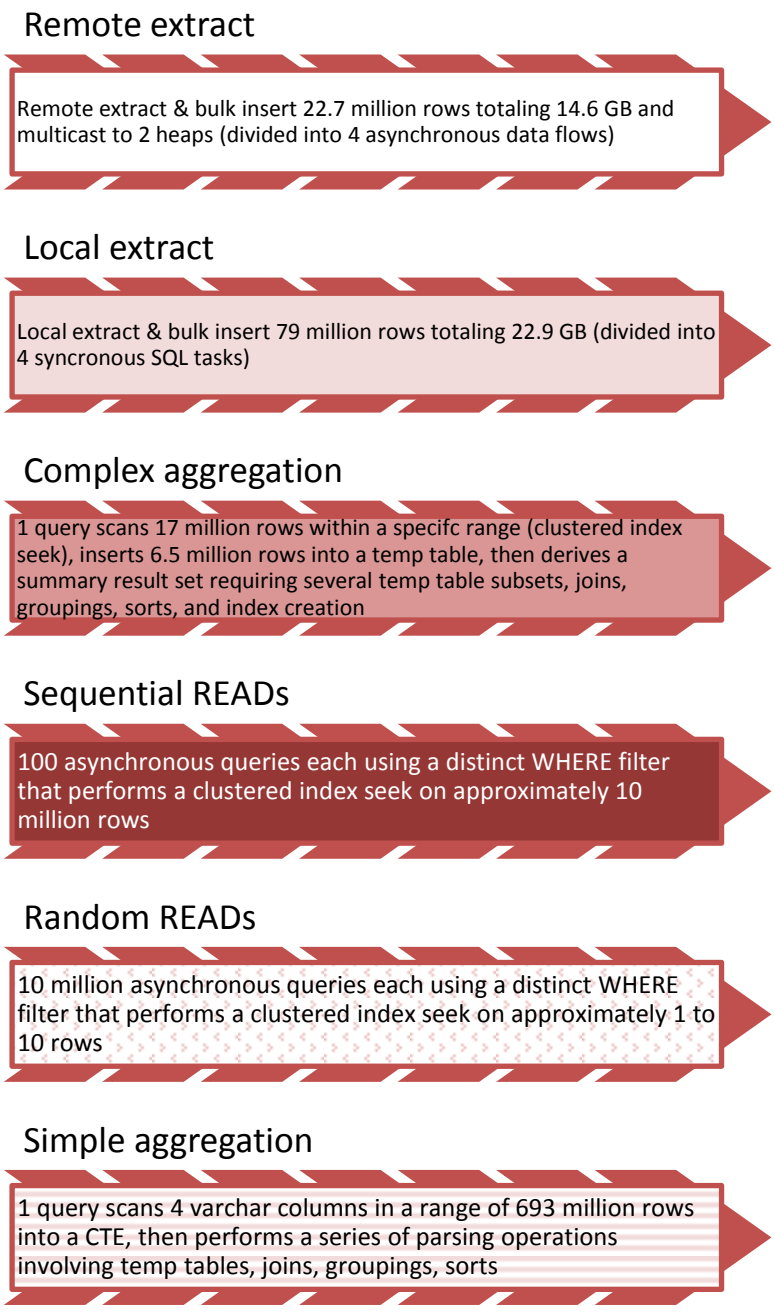
Test

The test involved a variety of DDL & DML operations having distinctively different I/O patterns organized to run predominately synchronously in order to force longer disk queues. The test was designed and executed within SSIS with the beginning and ending of each step logged to audit tables. The relational landscape consisted of 9 databases containing a mixture of clickstream, incident management, and chat data with a storage footprint of approximately 3.5TB. Other important environment details to call out are:

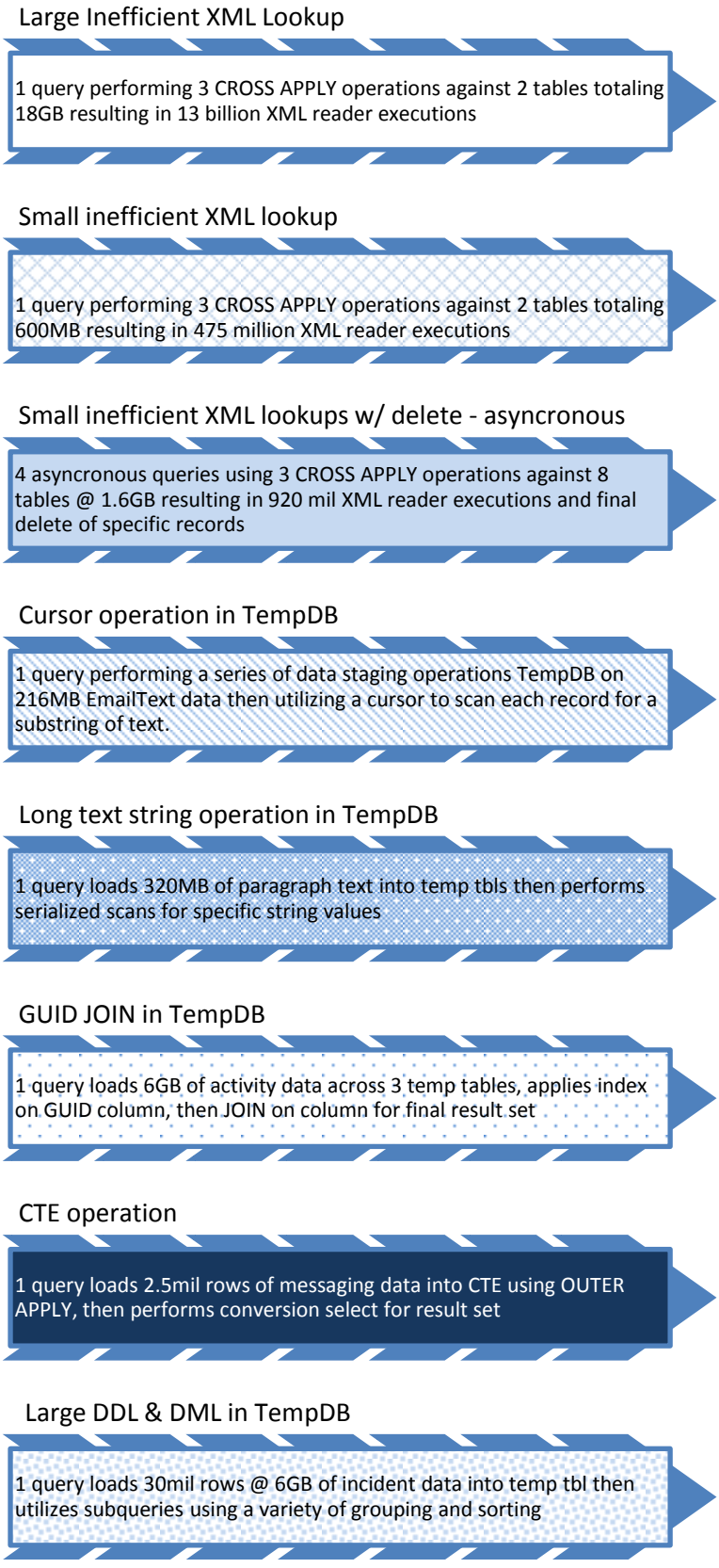
- The source tables all contained clustered indexes having various levels of fragmentation.
- All SQL engine, system and user db files were intentionally placed on volumes hosted by 1 storage device. For example, during the XtremIO testing all SQL and db files were placed on volumes presented by the XtremIO device. Conversely, during the SAN comparison test all SQL and db files were placed on volumes presented by the SAN.
- SQL Server memory was intentionally fixed @ 190GB for min and max settings to eliminate memory allocation swaps.
- SQL system cache & buffer pool were cleansed prior to each single test run
- All data flows used OLE DB connections with Rows Per Batch set to 10k & Maximum Insert Commit Size set to 100k
- The clickstream database totaled 1.7TB and was composed of 3 monthly partitions
- The incident & chat databases which contained the bulk of LOB data totaled 1.9TB w/ each having their largest tables spread in monthly partitions

The following pages summarize each individual test component in a graphical format. The color and pattern of each graphic will correspond to individual line items in the disk latency chart presented later in the results.

Click stream test flow



LOB test flow (temporary object focus)



LOB test flow (persistent object focus)

Large single remote extract

1 dataflow imports 30mil rows of incident data @ 6GB from remote server to physical table

Small single remote extract

1 dataflow imports 90k rows rows of xml data @ 1.3GB from remote server to physical table

Random READ

1k asynchronus queries selecting a random date range of records from 2 tbls totaling 600MB

Large DML operation

1 query loads 21mil rows @ 10GB of event detail data into physical table then performs distinct select on non-LOB values

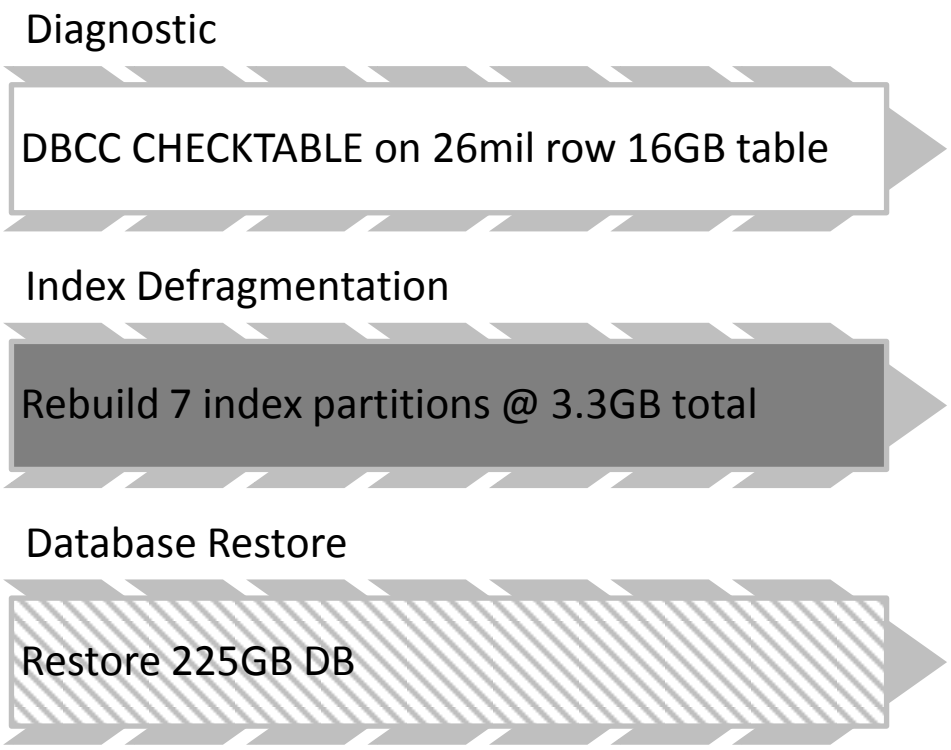
Small multiple remote extract & complex agg

17 dataflows load 2GB of data from remote server into physical tables followed by 1 stored procedure performs several summary aggregations of 2GB incident topic data using physical tables, then returns grouped & sorted result set

Small inefficient XML lookups - syncronous

3 syncronous queries using 3 CROSS APPLY operations against 2 tables @ 1.3GB resulting in 890 mil XML reader executions

Maintenance Tasks

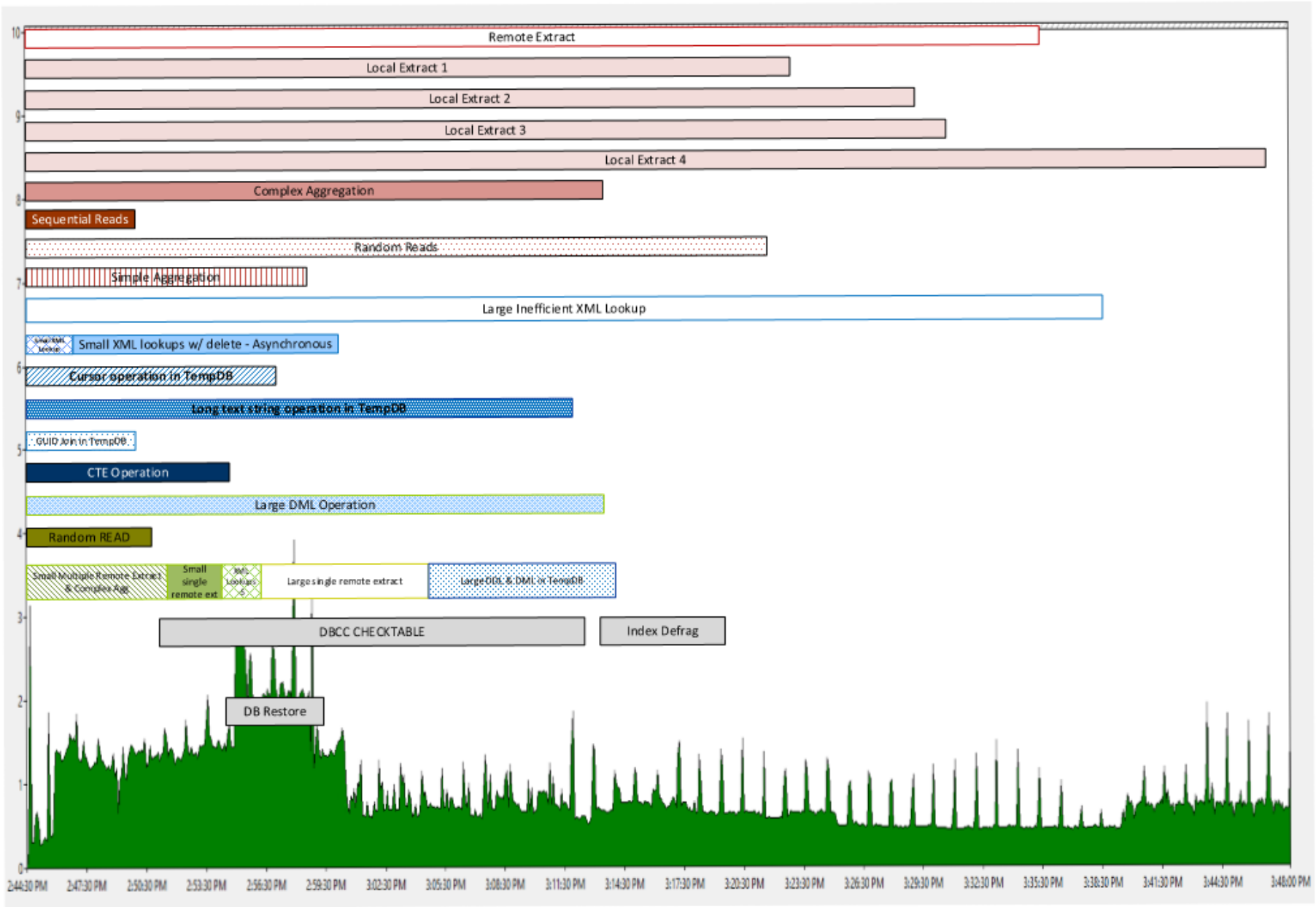


Success Criteria

Our success criteria was very simple - disk latency avg across all the XtremIO volumes should never exceeded 10 milliseconds and should be sustained below 5 milliseconds. Since it is common for most storage systems to experience short spikes in latency as cache is swapped, we allowed the transgression of the 5ms sustainment for activities shorter than 15 seconds.

Test Results

The avg test duration was 70 mins and during that time approx. 104k IOPS were generated from approx. 58GB of data movement. Disk latency across the XtremIO volumes never exceeded 4ms including the temporary spikes mentioned previously. The avg disk latency for the first phase of the test which was also the most I/O intensive was < 2ms. The avg disk latency reached its highest when the database restore executed causing sustained latency to bounce between 2ms & 3ms. Once the restore was completed, avg sustained latency remained < 1ms with isolated spikes at 1.75ms.


Remote Extract
Local Extract 1
Local Extract 2
Local Extract 3
Local Extract 4
Complex Aggregation
Sequential Reads
Random Reads
Simple Aggregation
Large Inefficient XML Lookup
Small XML lookups w/ delete - Asynchronous
Cursor operation in TempDB
Long text string operation in TempDB
GUID Join in TempDB
CTE Operation
Large DML Operation
Random READ
Small multiple Remote Extract & Complex Agg
Small single remote ext
XML Lookups <5
Large single remote extract
Large DDL & DML in TempDB
DBCC CHECKTABLE
Index Defrag
DB Restore
10
9
8
7
6
5
4
3
2
1
0
2:44:30 PM
2:47:30 PM
2:50:30 PM
2:53:30 PM
2:56:30 PM
2:59:30 PM
3:02:30 PM
3:05:30 PM
3:08:30 PM
3:11:30 PM
3:14:30 PM
3:17:30 PM
3:20:30 PM
3:23:30 PM
3:26:30 PM
3:29:30 PM
3:32:30 PM
3:35:30 PM
3:38:30 PM
3:41:30 PM
3:44:30 PM
3:48:00 PM

Summary & Conclusion

The success criteria outlined for the test was entirely met with final results actually exceeding our most optimistic expectations. Sustained latency rates were well below the recommended 10ms threshold for OLTP applications, therefore, to exceed this threshold in an I/O contentious BI environment is highly commendable.

As stated previously, the test package incorporated all the most I/O intensive operations that the CSSBI team will typically observe on the stand-alone DAP instance during peak usage periods. What's more, it incorporated maintenance operations that typically are schedule during non-peak usage periods. When considering how well the unit performed in an unlikely "worst case scenario" it is feasible that the I/O could be substantially increased further before any storage performance bottleneck would be encountered.

In addition, the fact that the XtremIO storage unit was able to easily handle the load with no care given to the actual placement of files on the volumes, or the segregation of files based upon READ/WRITE balance (ex. hosting TempDB files on the same volumes hosting LOB and clickstream data files) has the potential to greatly reduce the time and effort required of DBA's to organize the storage volumes and the files they host. SAN's achieve this to some degree with pools and block tiering, however, the setup of the pools and tiers still requires a SAN administrator. What's more, in most BI environments the rate of change and growth is exponential which requires DBA's to frequently work with the SAN administrator on storage optimization. However, the XtremIO storage product has the ability to be a "plug and play" solution where the optimization efforts of the DBA and SAN administrator are greatly reduced.

Despite the very large storage footprint of most BI systems, it's common that less than 20% of the data is actively used by the BI end-users. It's feasible that an XtremIO unit could be provisioned to host the most active data files, thereby allowing the storage of the remaining 80% on less expensive and highly redundant SATA pools either in DAS or SAN. Such a setup could ensure that storage performance was highly optimized without the higher associated expense of the hardware and the support personnel.

For more information:

http://www.microsoft.com/sqlserver/: SQL Server Web site

http://technet.microsoft.com/en-us/sqlserver/: SQL Server TechCenter

http://msdn.microsoft.com/en-us/sqlserver/: SQL Server DevCenter

http://www.xtremio.com/ EMC XtremIO

Did this paper help you? Please give us your feedback. Tell us on a scale of 1 (poor) to 5 (excellent), how would you rate this paper and why have you given it this rating? For example:

- Are you rating it high due to having good examples, excellent screen shots, clear writing, or another reason?
- Are you rating it low due to poor examples, fuzzy screen shots, or unclear writing?

This feedback will help us improve the quality of white papers we release.

Send feedback.