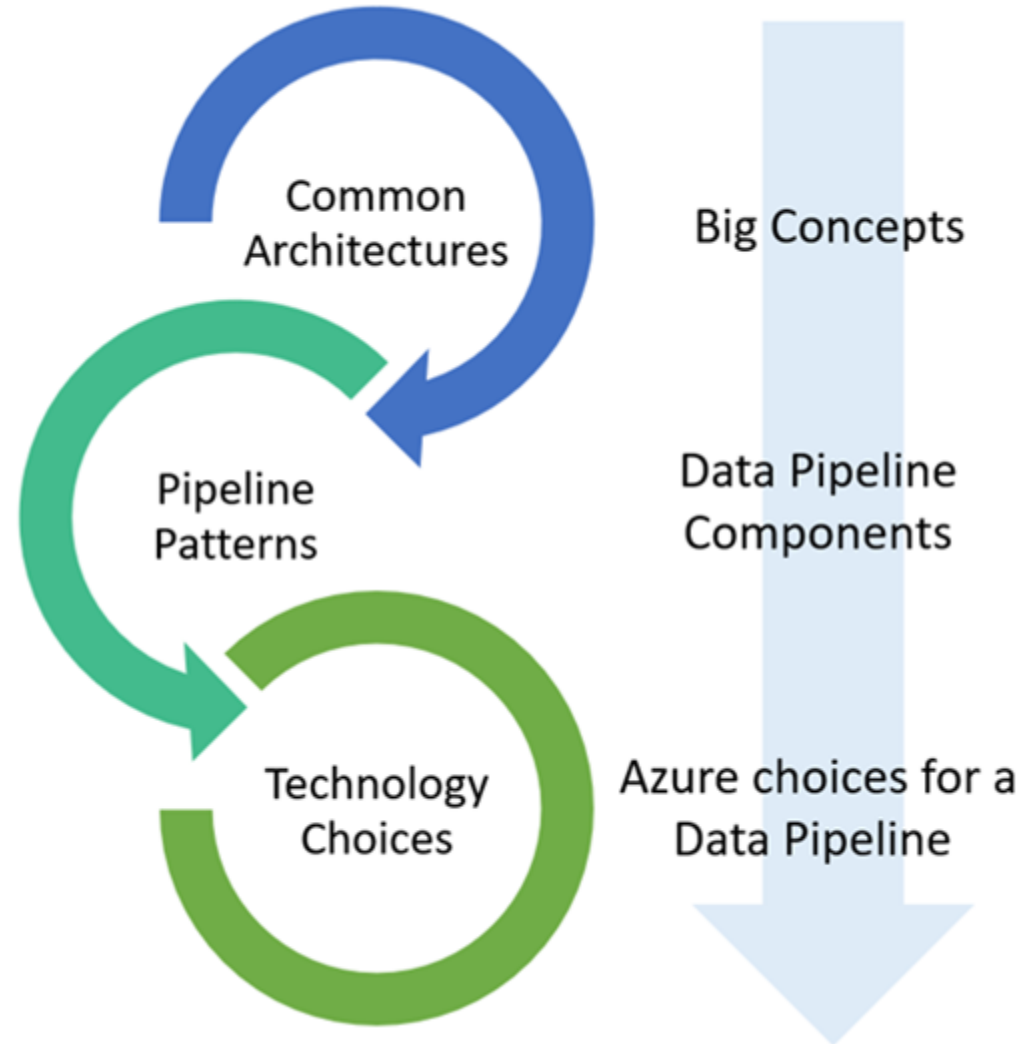
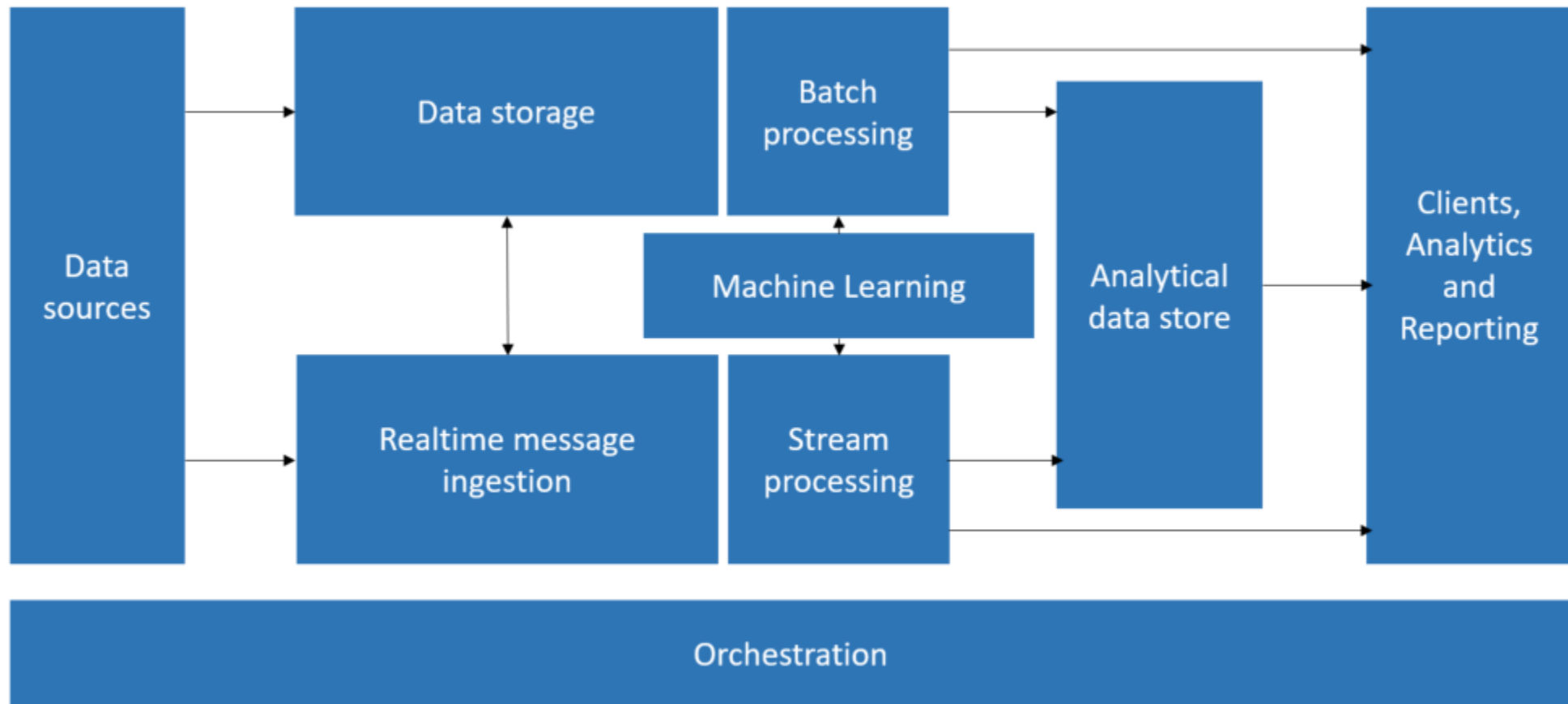


Azure Data Architecture Guide

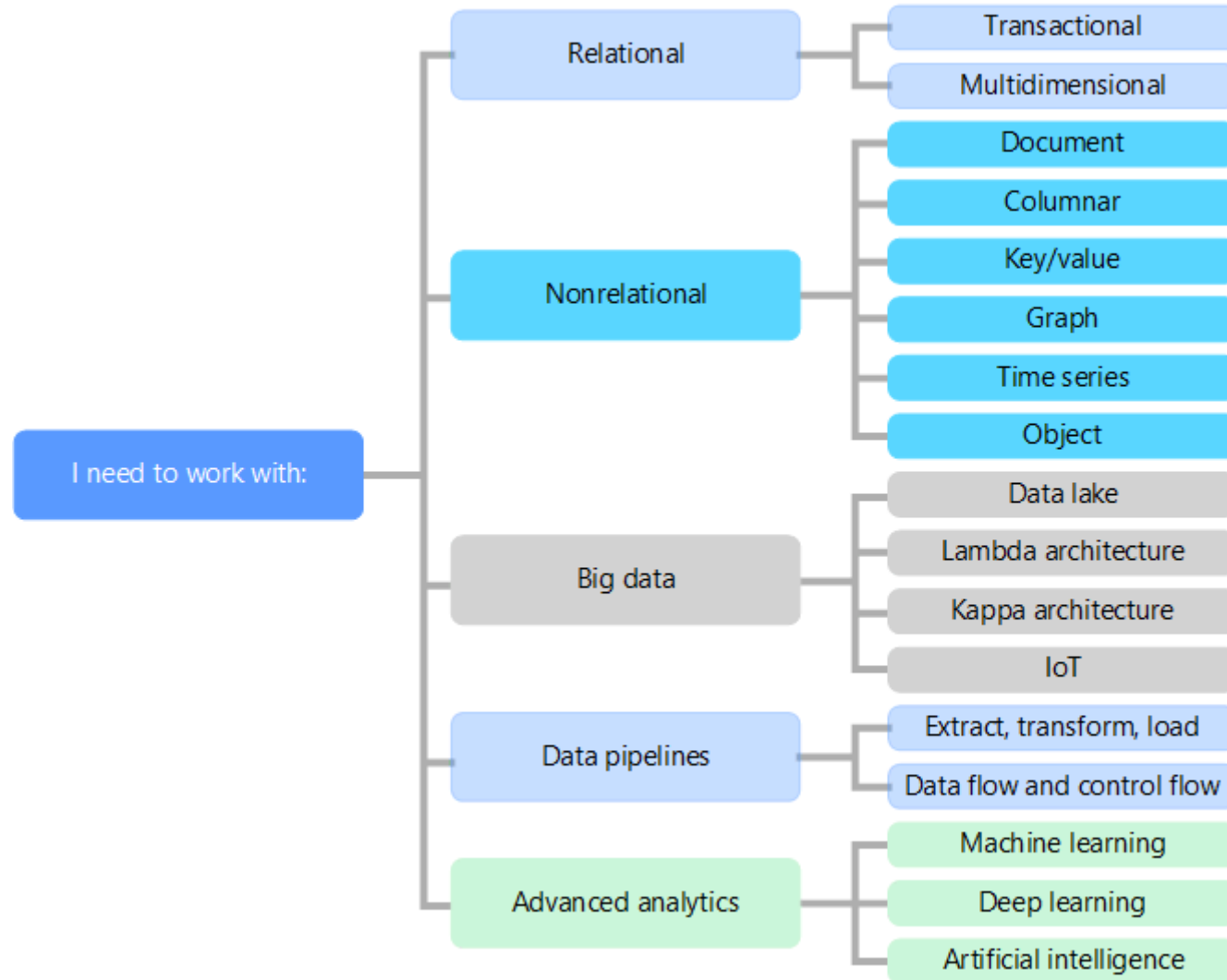
Agenda



Subway Map

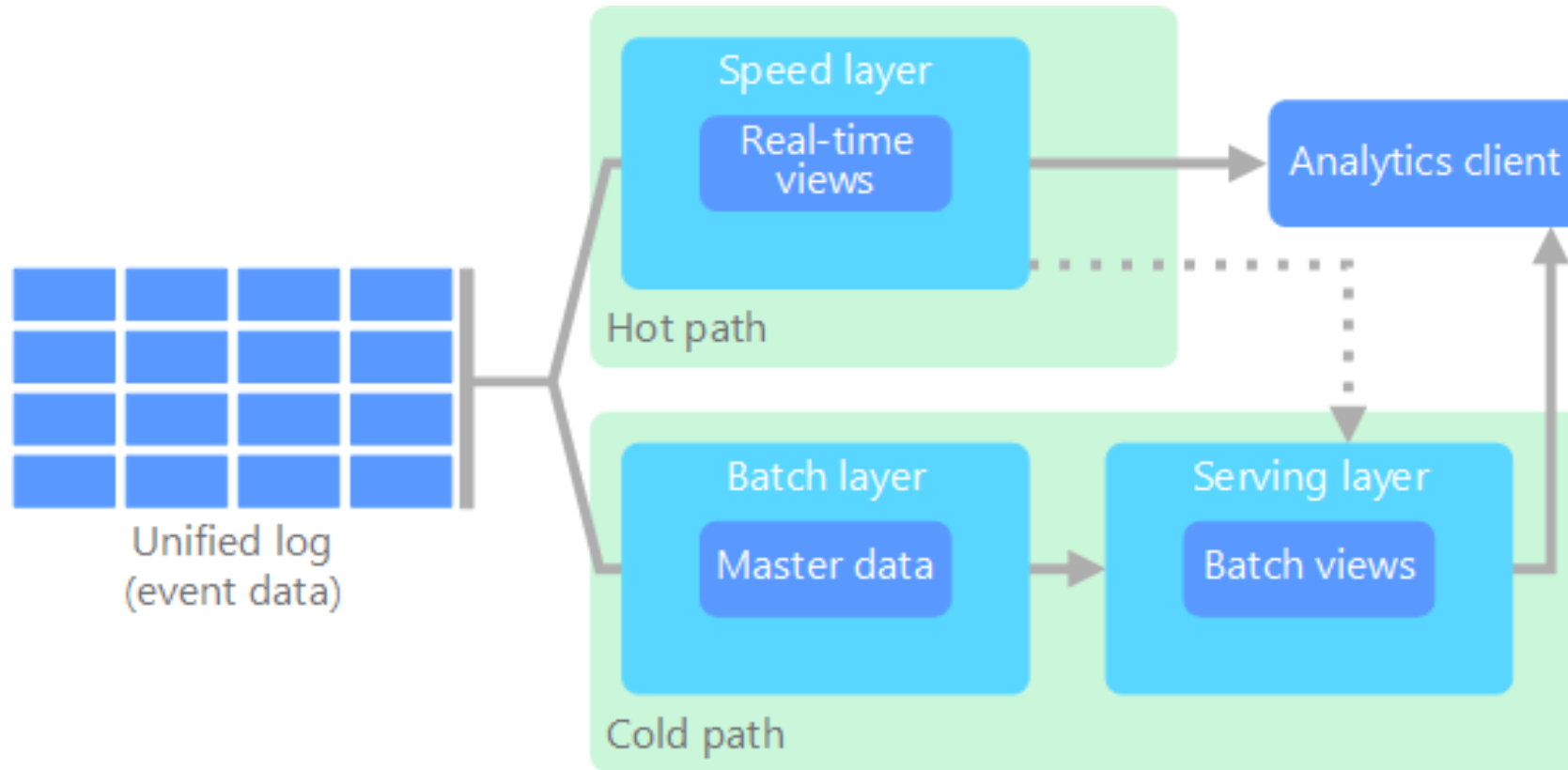


Types of Data



Real-Time Streaming Architecture

Lambda Architecture



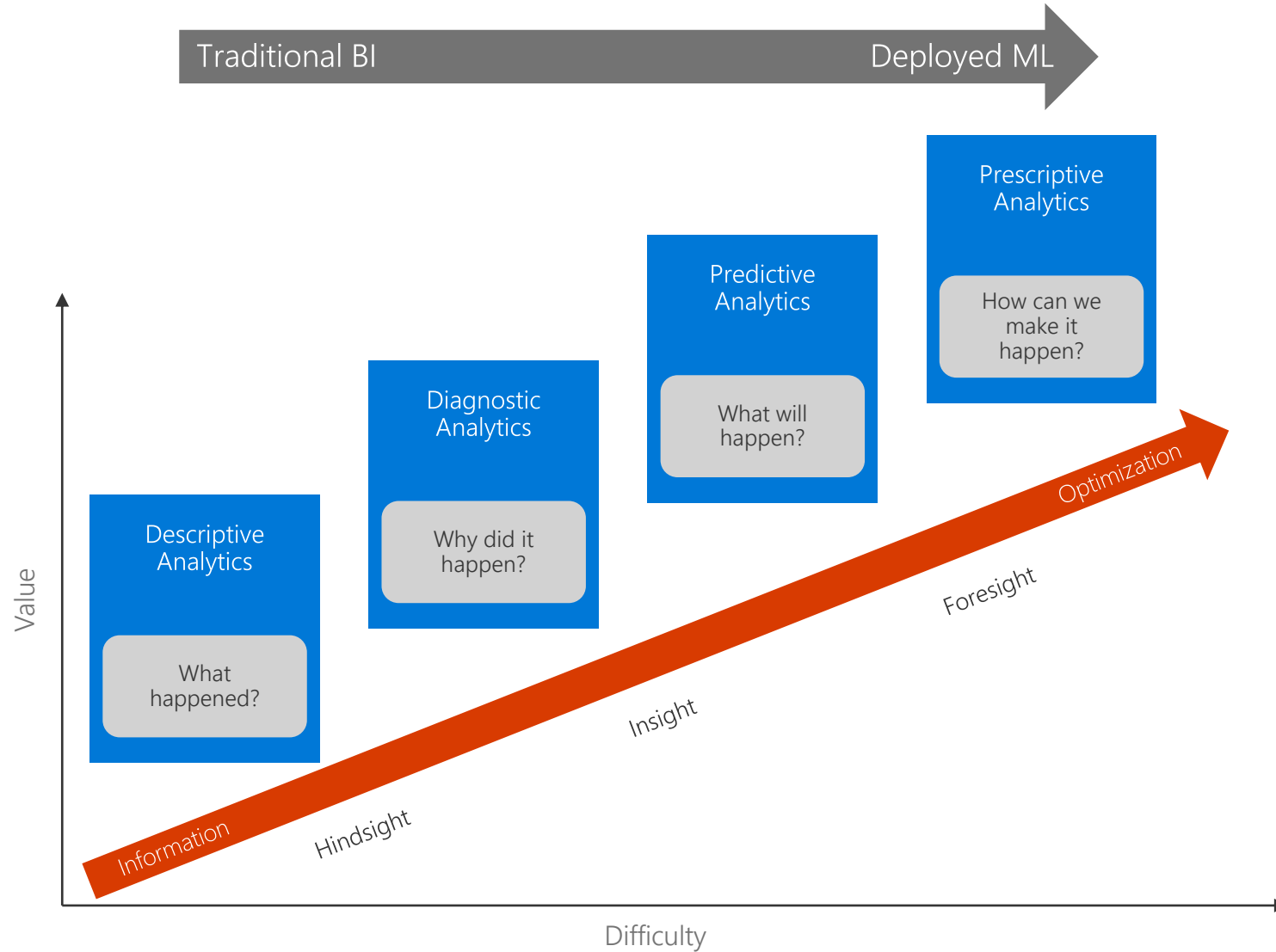
Hot Path

- View Actionable Data Quickly
- Likely Looking at an Aggregate
- Requires Professional Presentation
- Better if Ad-Hoc Querying is Supported

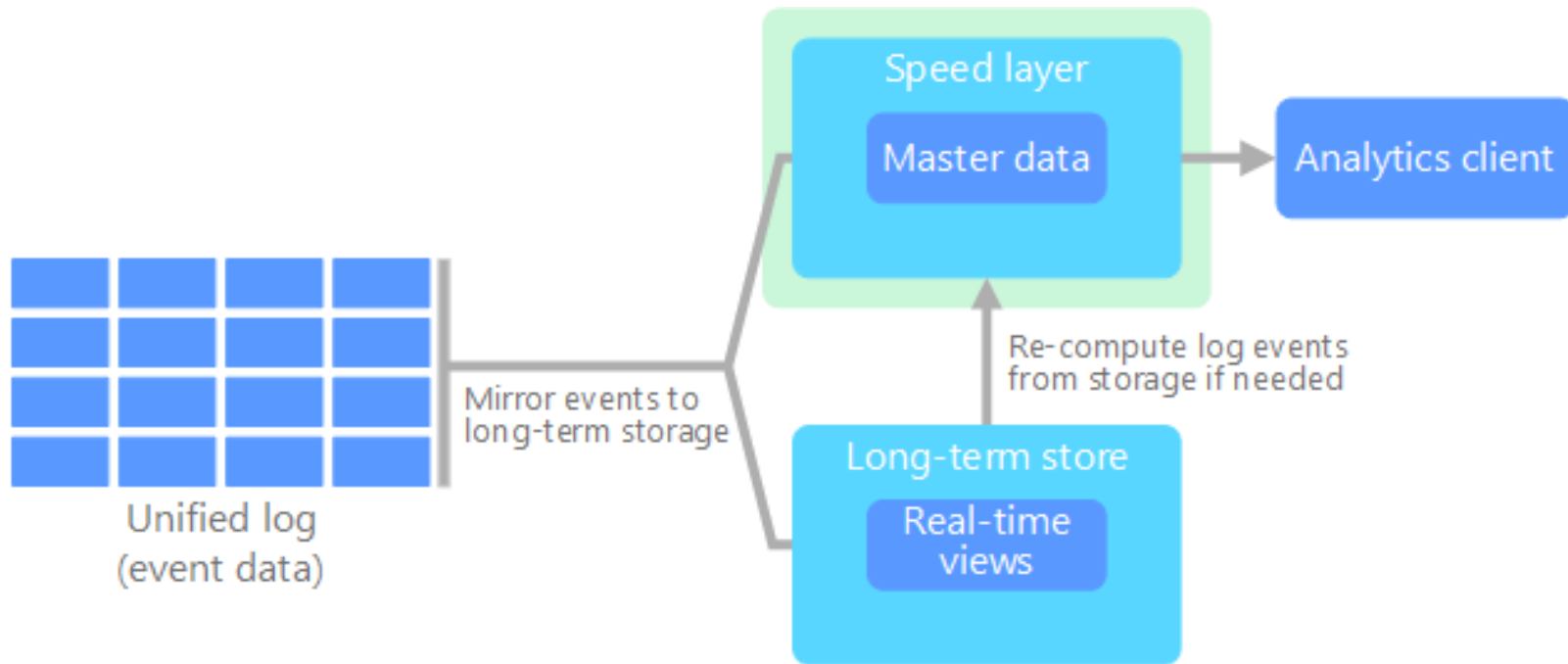
Cold Path

- Long-term Storage
- Must be cheap to query and store
- May be processed multiple times
 - Analysis
 - Prediction
 - Learning

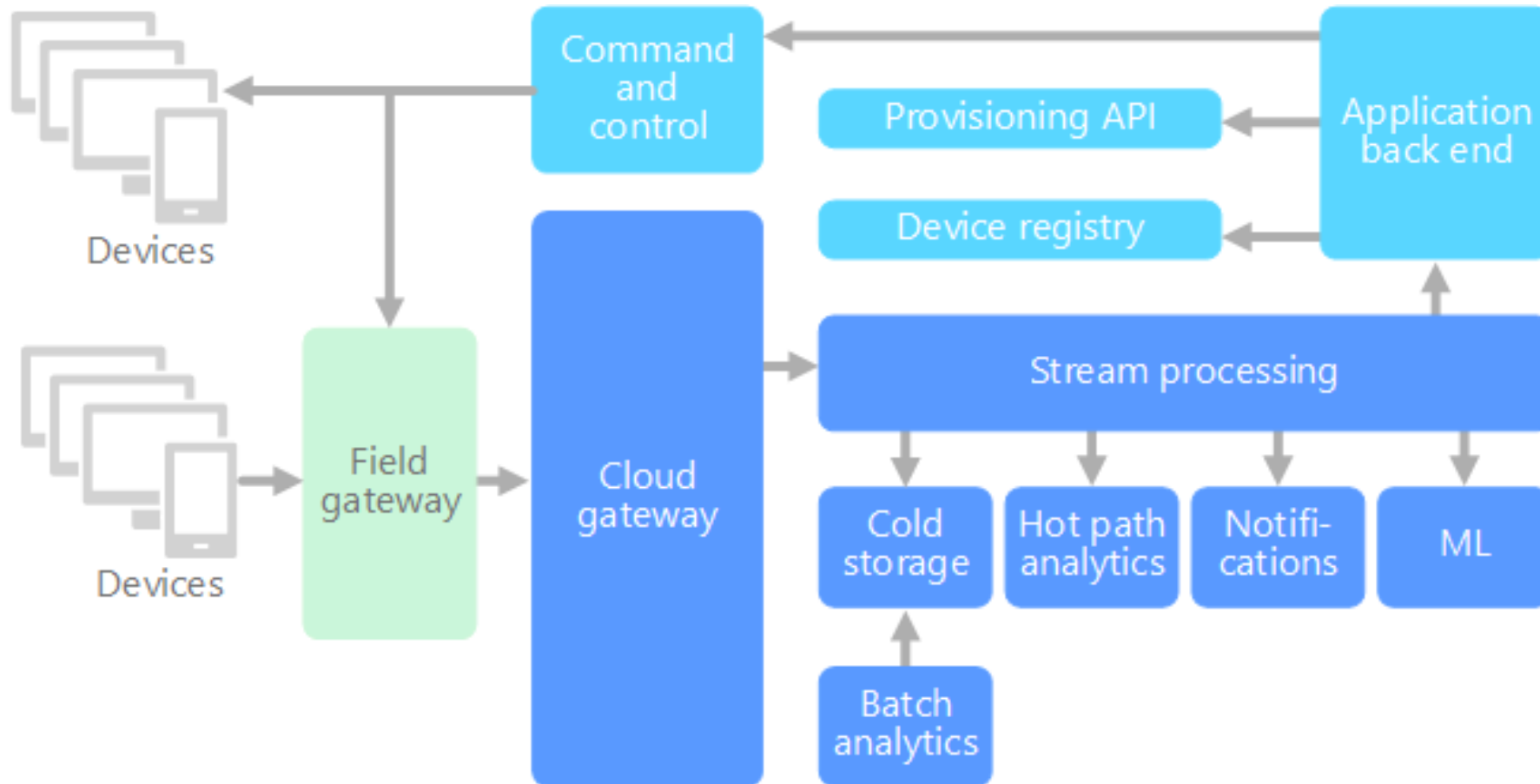
Hot & Cold Path Analysis



Kappa Architecture

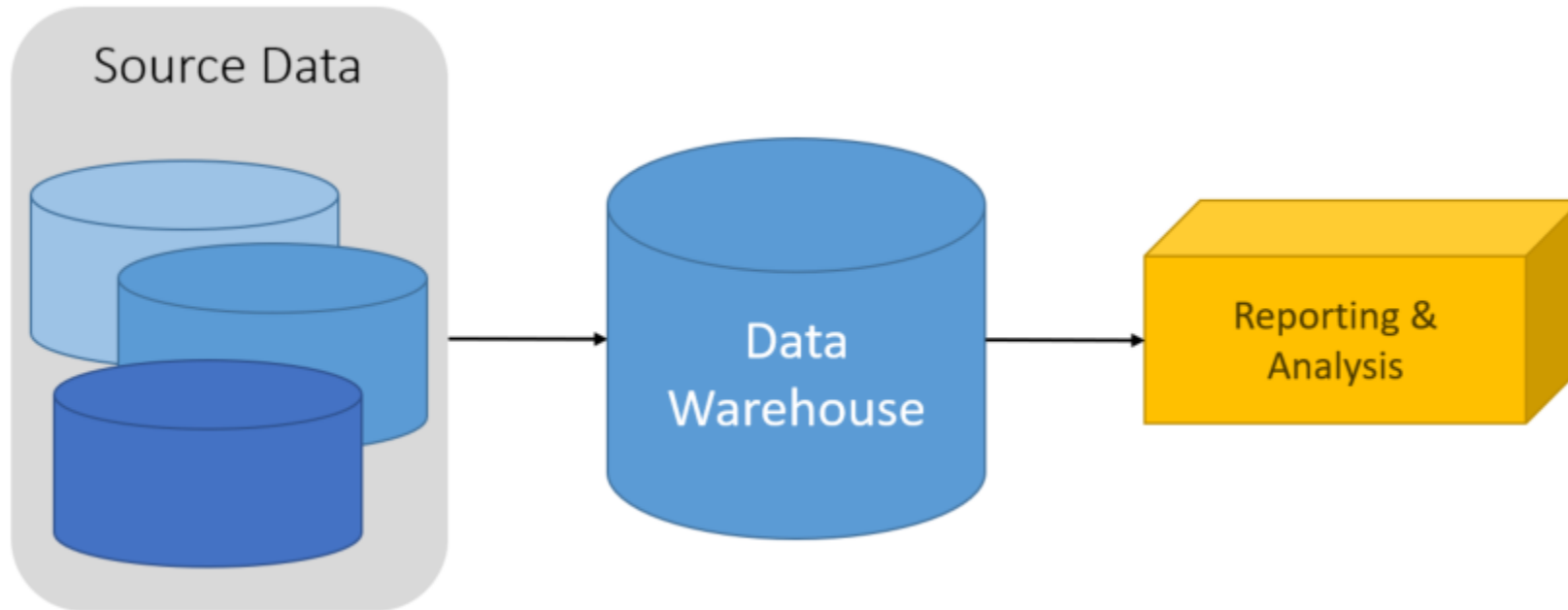


IoT Message Processing

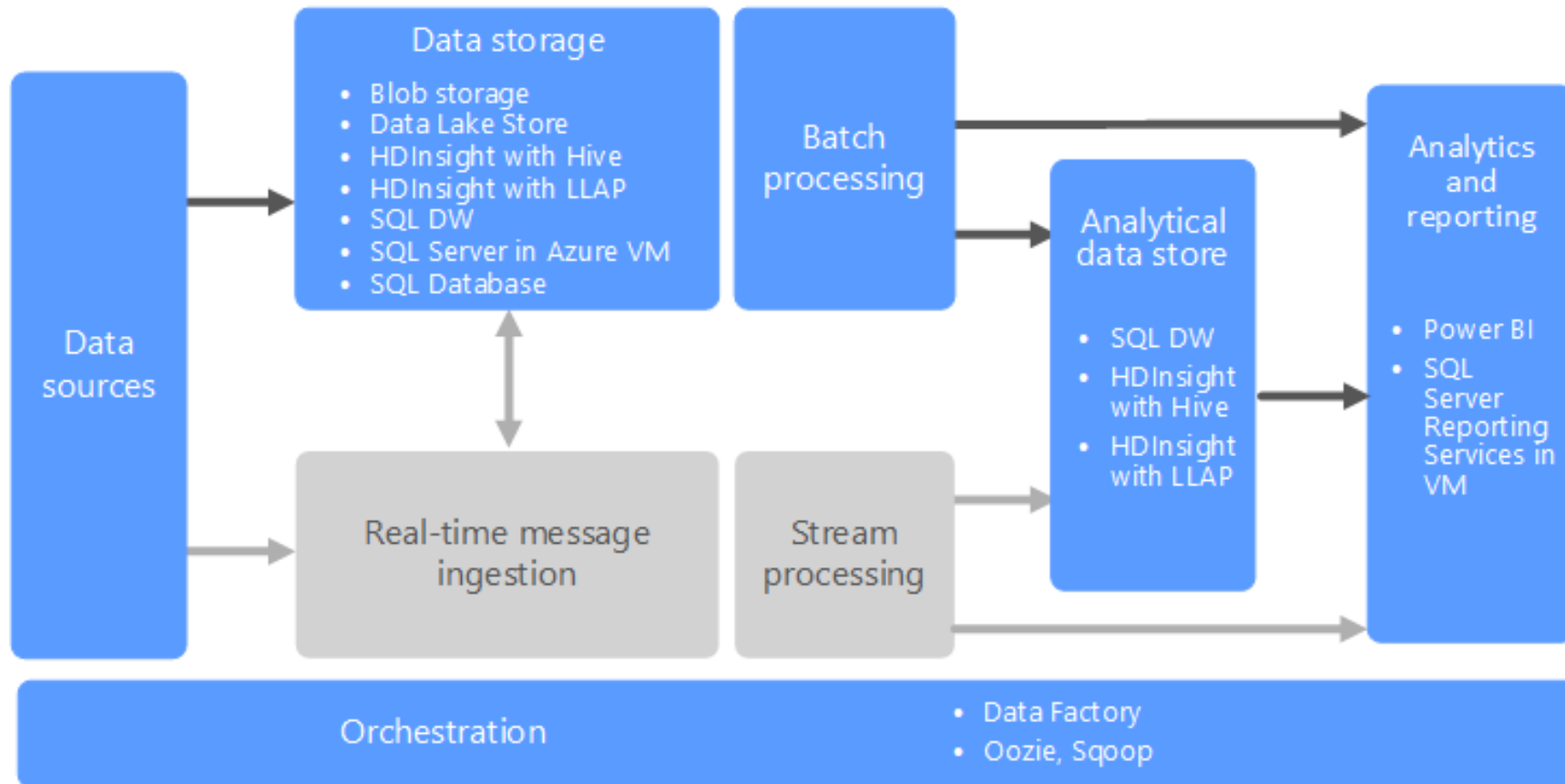


Data Warehousing

Data Warehousing



Data Warehouse Pattern



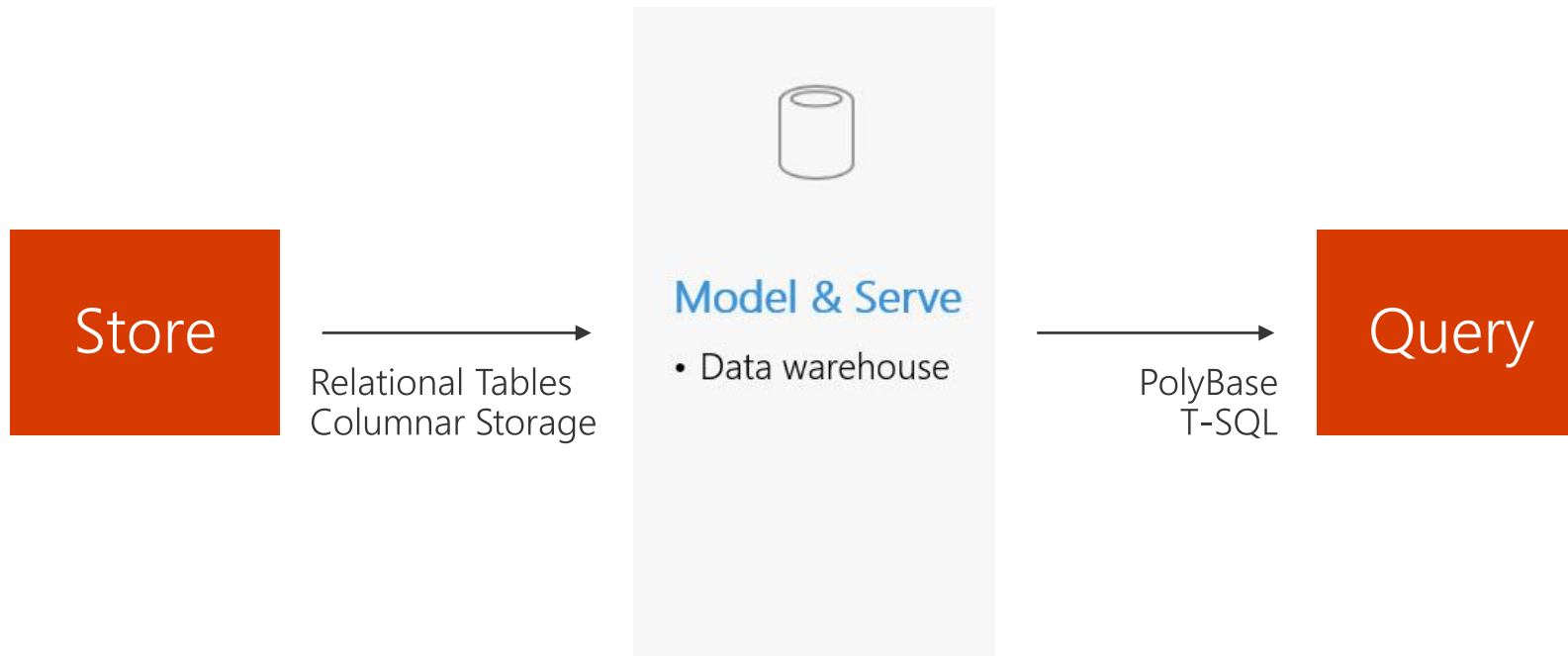
Data Warehouse Technology Choices

- SMP (small/medium data)
 - Azure SQL Database
 - SQL Server in a virtual machine
 - Azure SQL Database managed instance

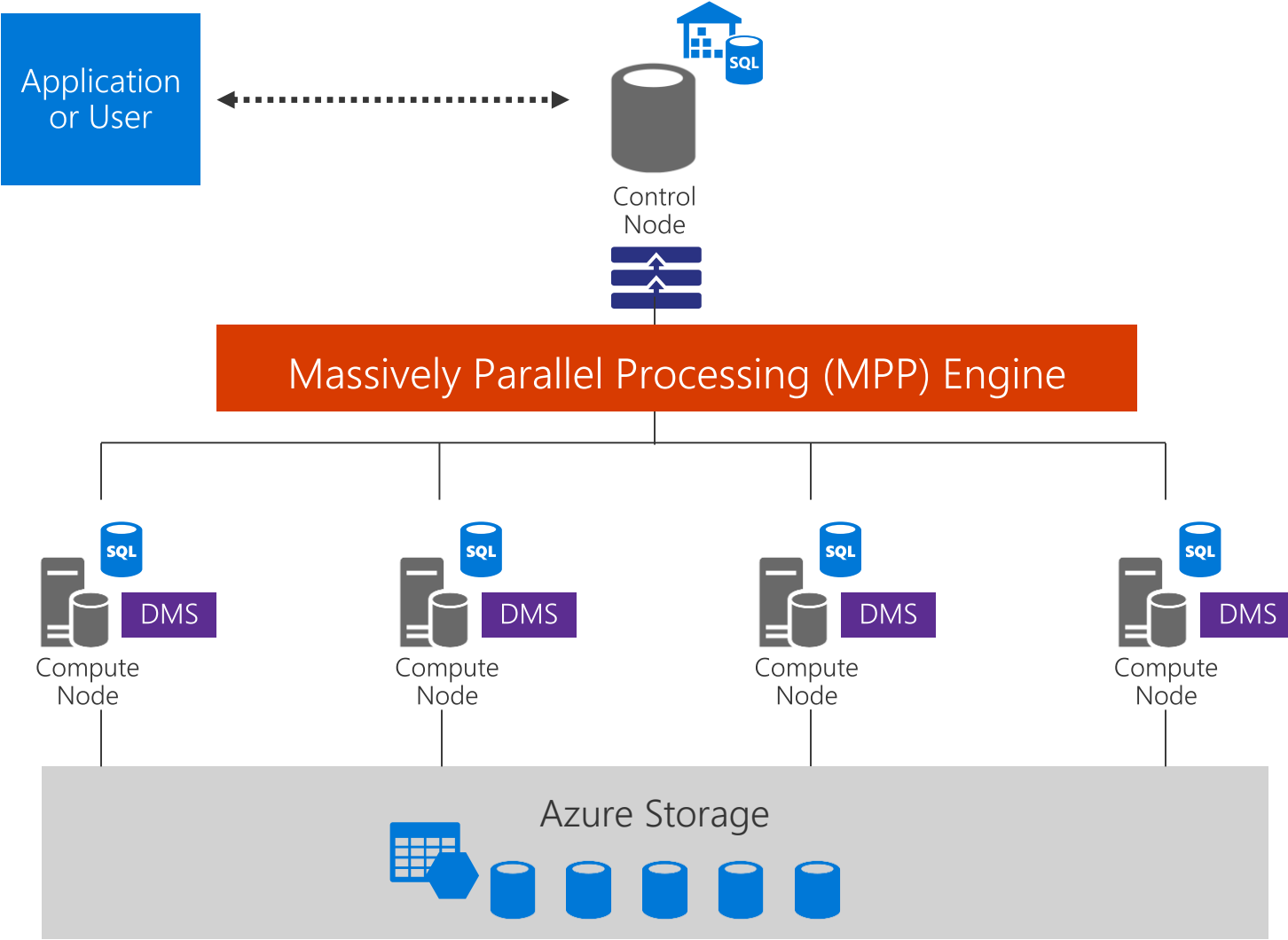
Data Warehouse Technology Choices

- MPP (big data)
 - Azure Data Warehouse
 - Apache Hive on HDInsight
 - Interactive Query (Hive LLAP) on HDInsight

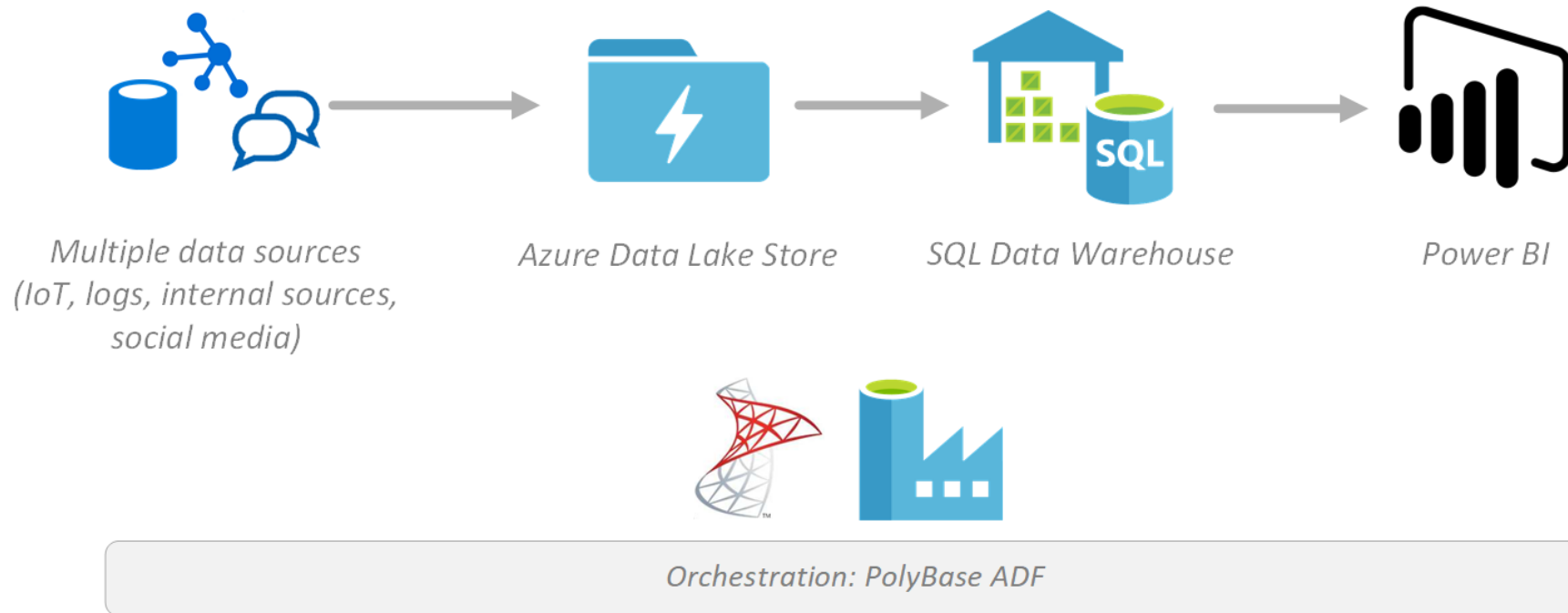
SQL Data Warehouse



SQL Data Warehouse



Data Warehousing



No-SQL


Document Data Stores

Key	Document
1001	<pre>{ "CustomerID": 99, "OrderItems": [{ "ProductID": 2010, "Quantity": 2, "Cost": 520 }, { "ProductID": 4365, "Quantity": 1, "Cost": 18 }], "OrderDate": "04/01/2017" }</pre>
1002	<pre>{ "CustomerID": 220, "OrderItems": [{ "ProductID": 1285, "Quantity": 1, "Cost": 120 }], "OrderDate": "05/08/2017" }</pre>

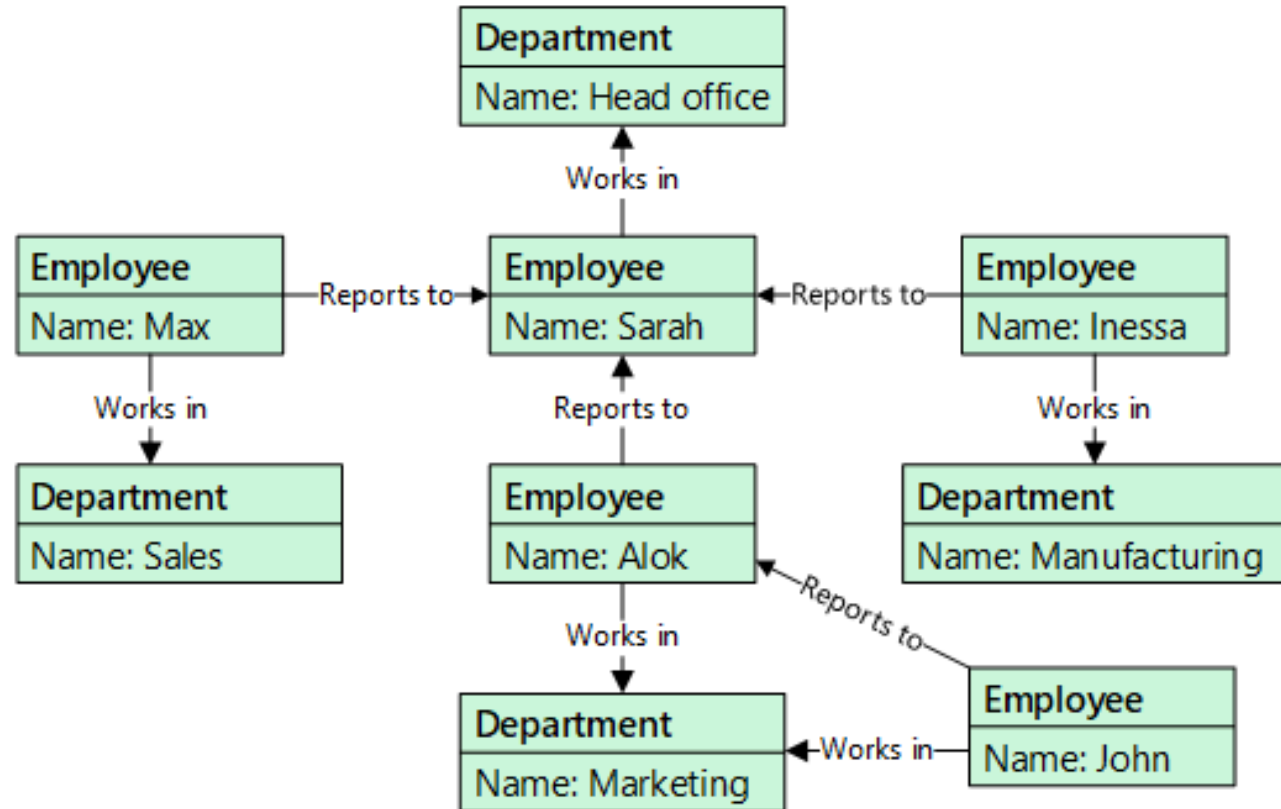
Key-Value Data Stores

Key	Value
AAAAA	11010011110100100011001...
AABAB	00000101011100110110010...
DFA766	00000001111011101110101...
FABCC4	11110011001101011101001...

Not visible
to data store



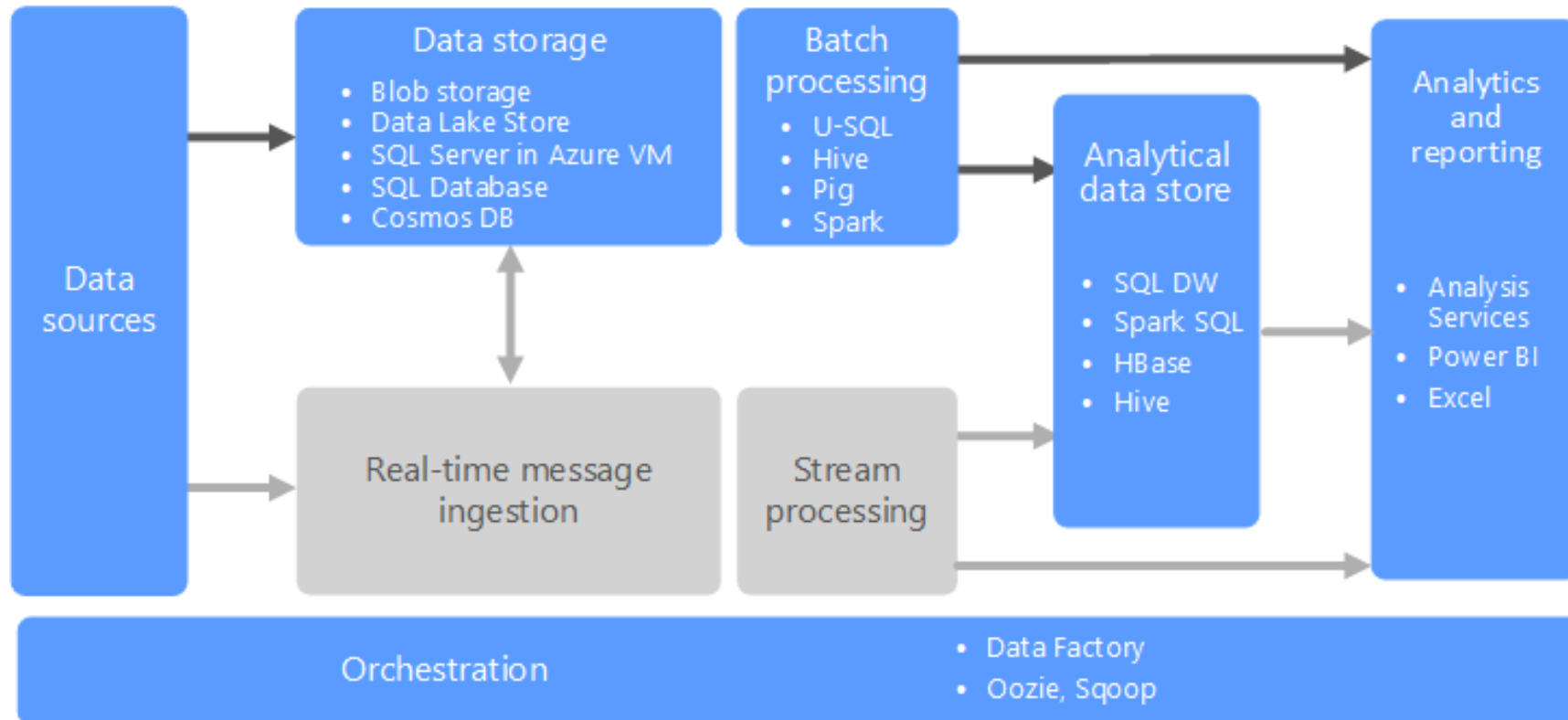
Graph Data Stores



And Then There's More...

- Columnar
- Object
- Time Series
- External Index

Processing JSON



No-SQL Technology Choices

- Azure Data Factory
- Azure Logic Apps
- Azure Functions
- App Service
- Azure Data Lake Analytics
- Azure HDInsight

No-SQL Technology Choices

- Spark SQL
- HBase
- Hive
- SQL Data Warehouse
- Azure Machine Learning Workbench
- SQL SSIS

Azure Cosmos DB

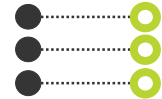
Core Database

Column-family

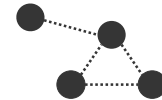


cassandra

Key-Value

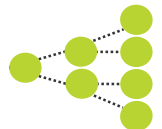


Graph



Gremlin
 $G = (V, E)$

Documents



mongoDB

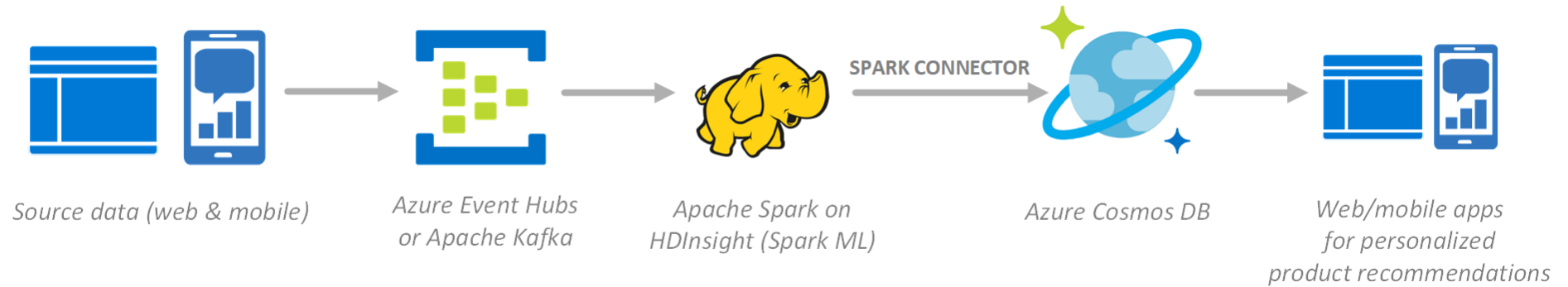


SQL



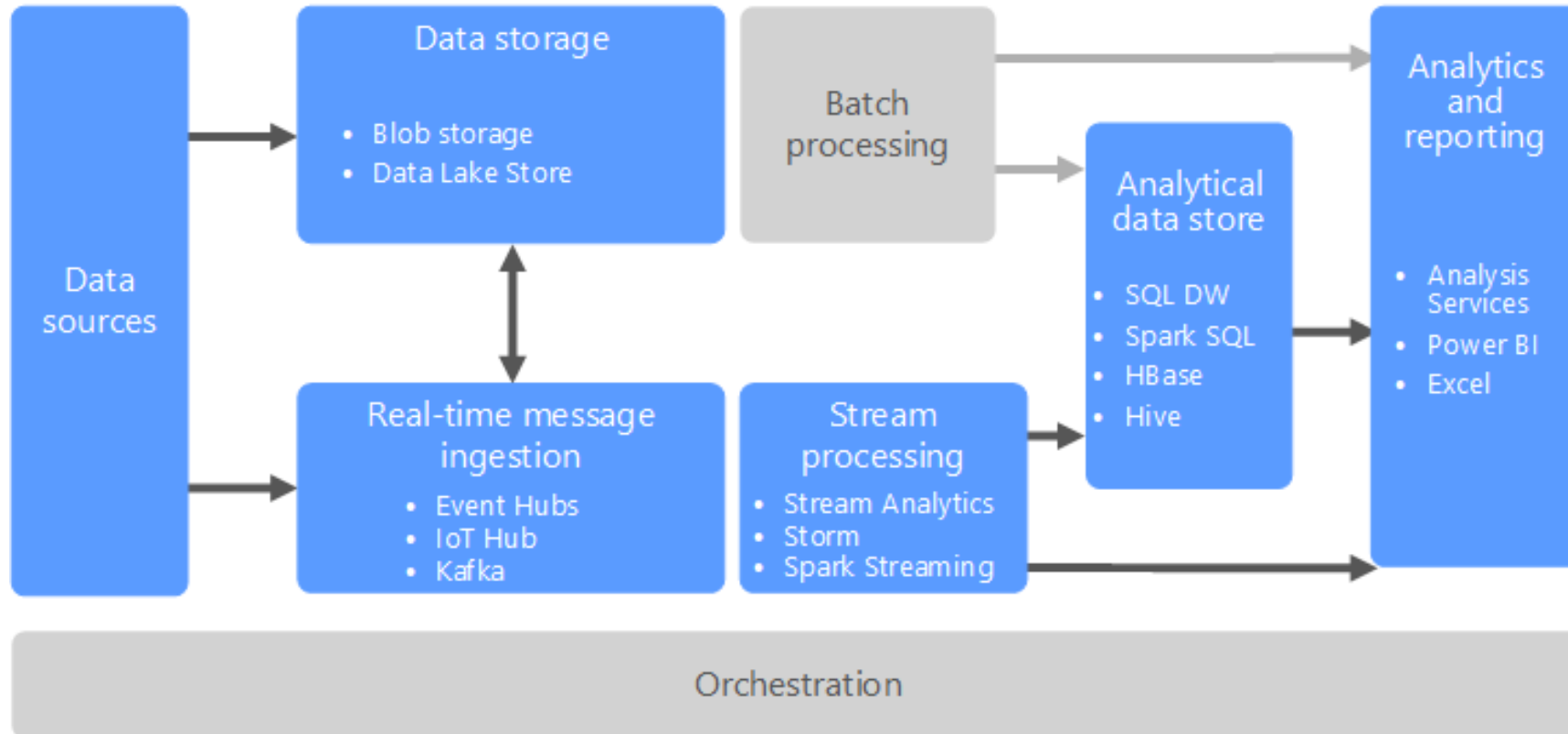
JavaScript

Clickstream Analysis



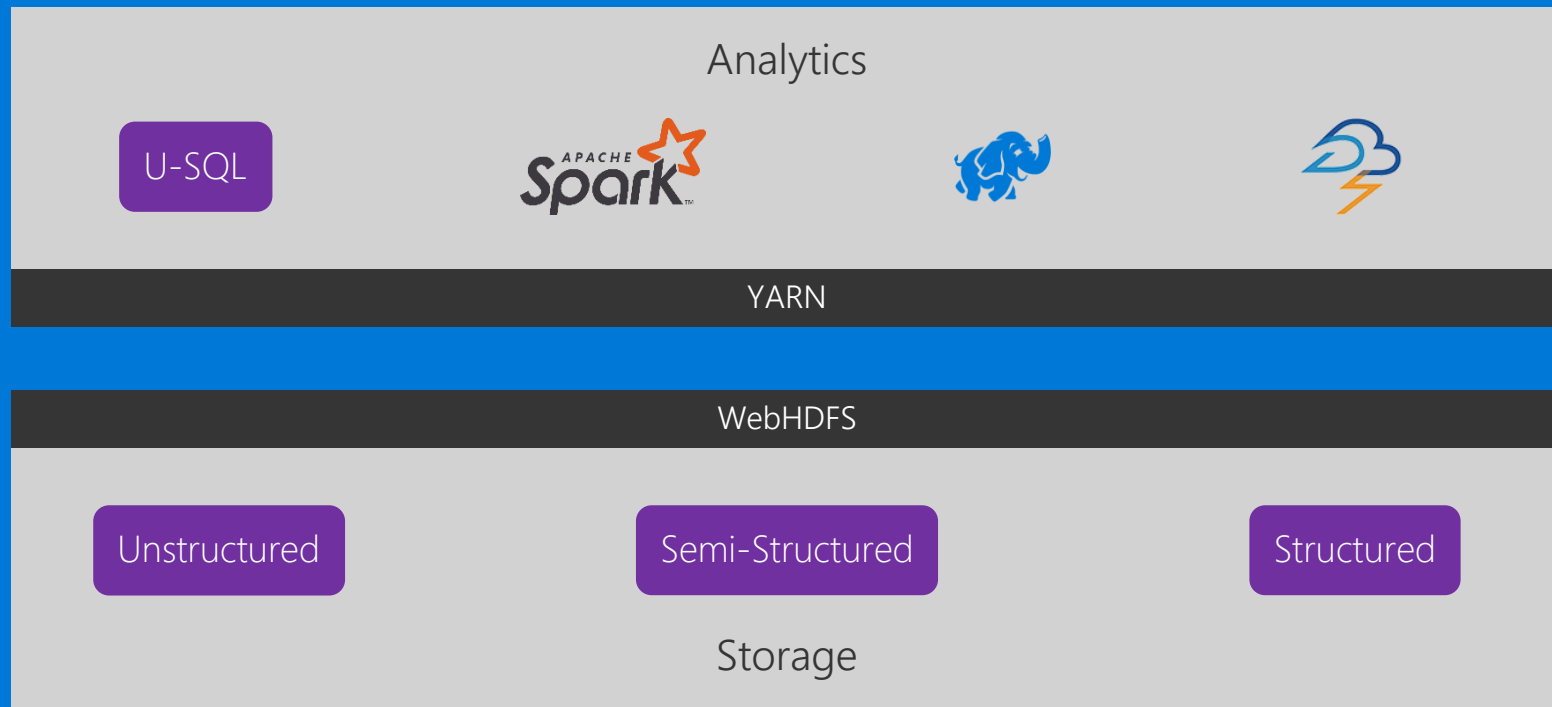
Storing Relational and No-SQL Data

Processing JSON in Real-Time

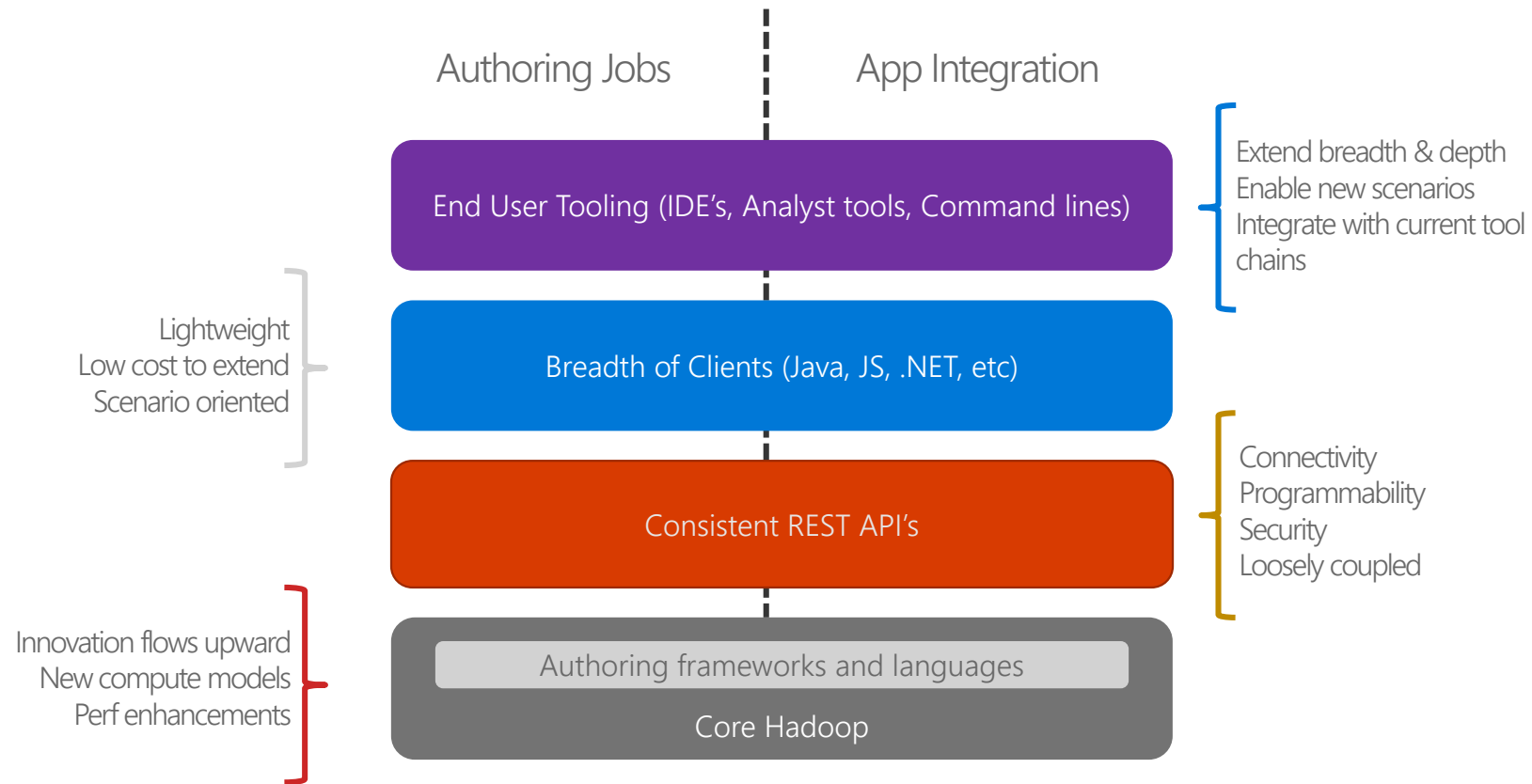


Data Lake

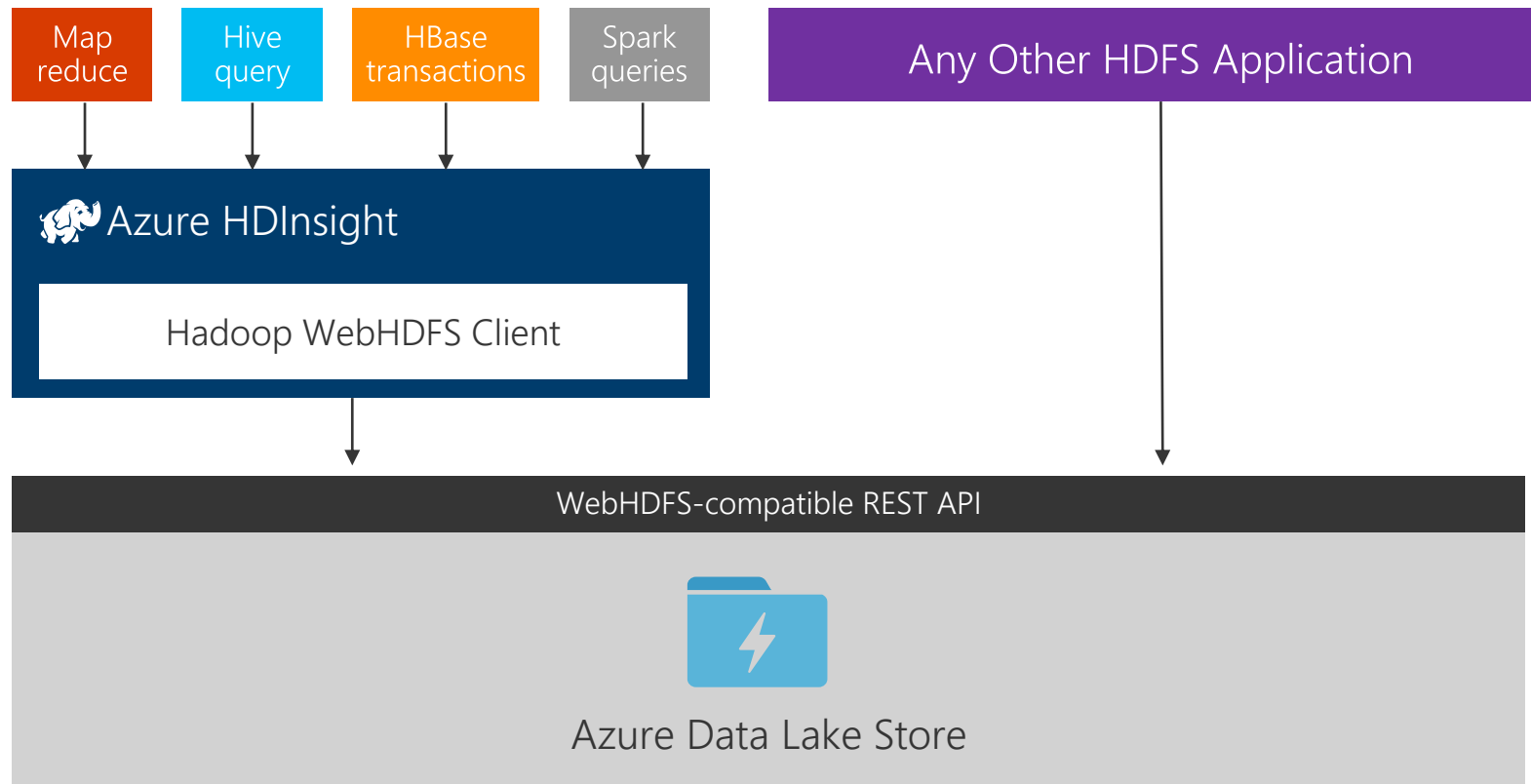
Azure Data Lake



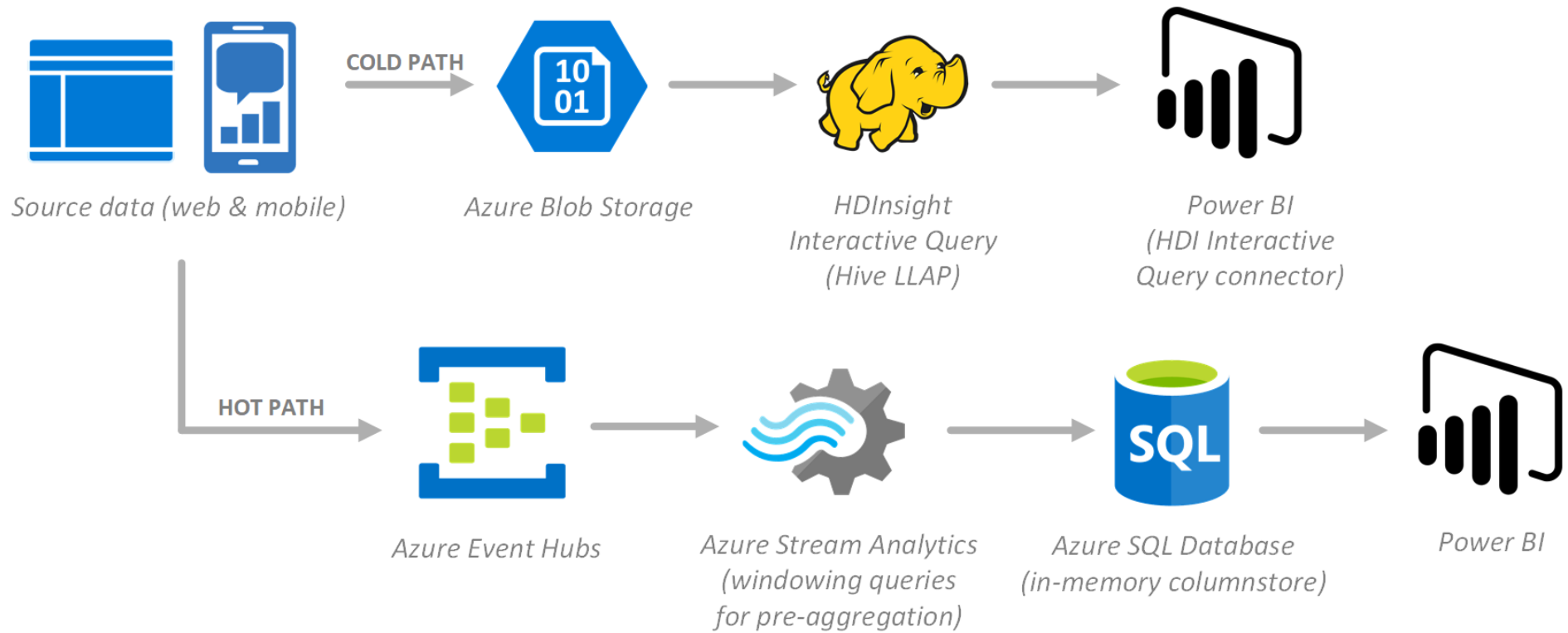
HDInsight



HDInsight & Data Lake

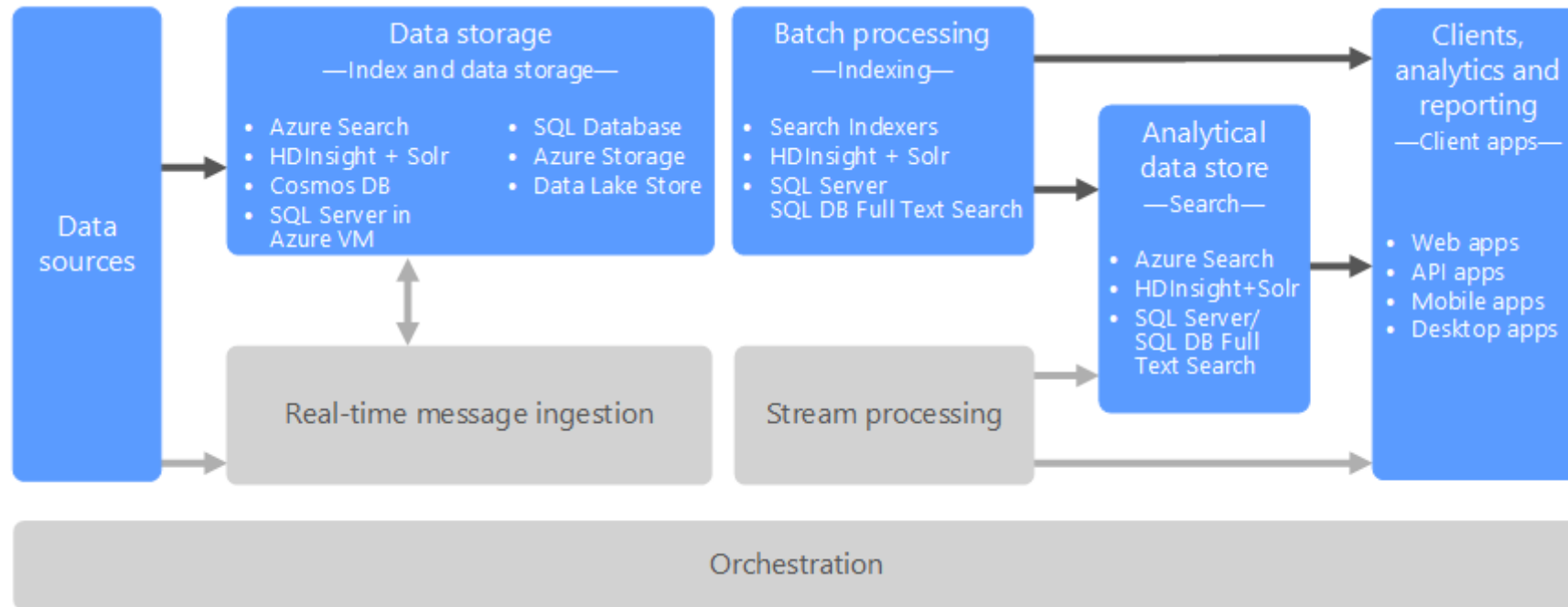


On-Demand Big Data Analytics



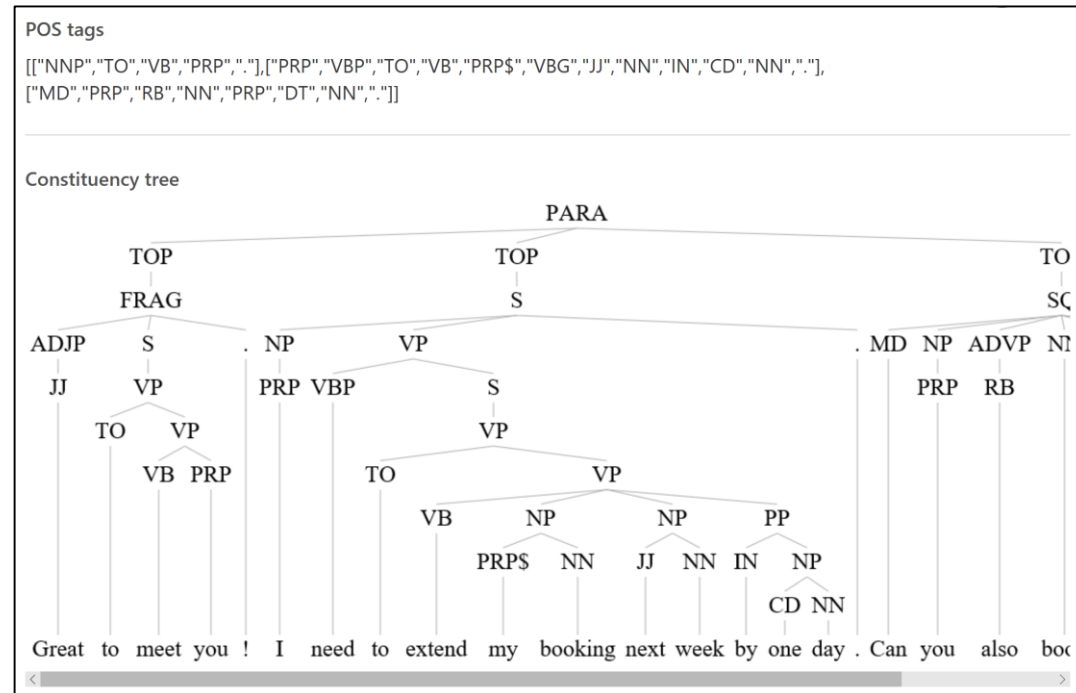
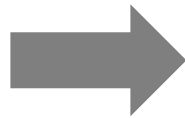
Natural Language Processing

Processing Free-Form Text

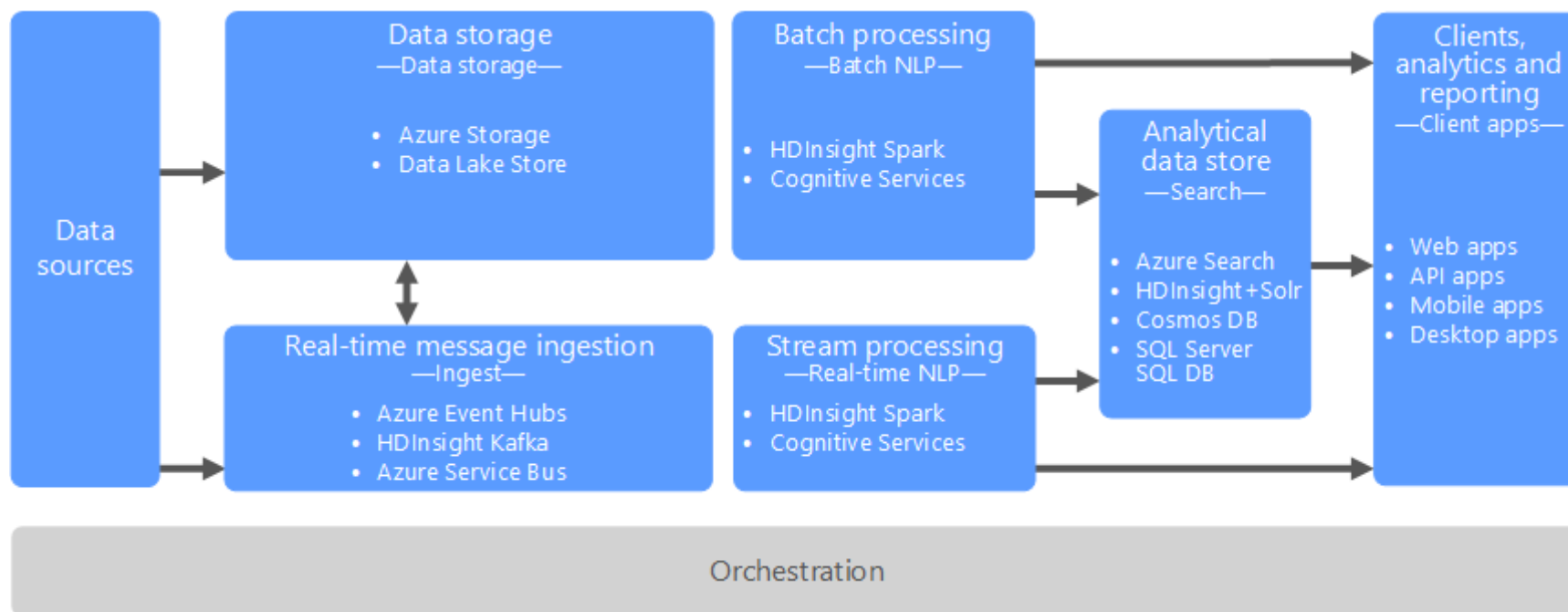


Natural Language Processing (NLP)

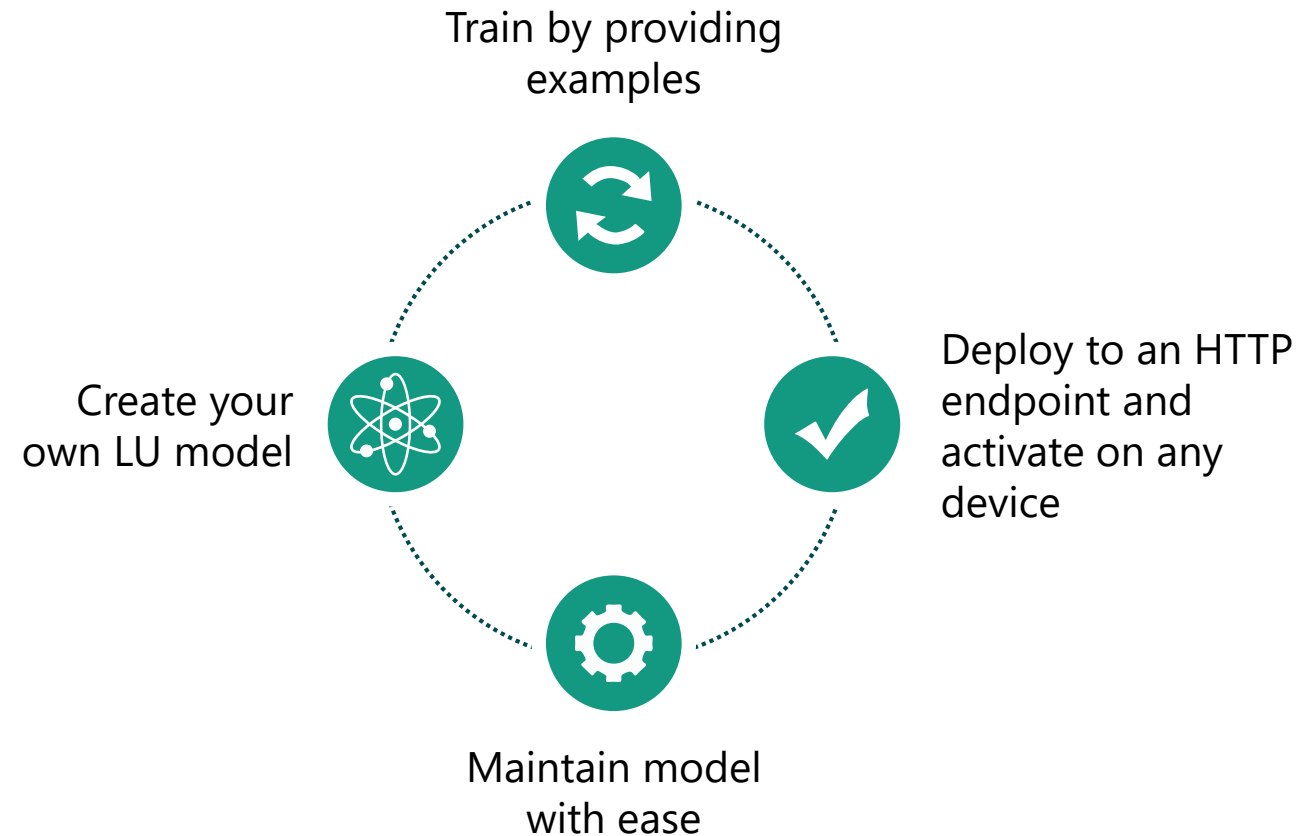
*“Great to meet you! I
need to extend my
booking next week by one
day. Can you also book
me a car?”*



Processing Free-Form Text using NLP



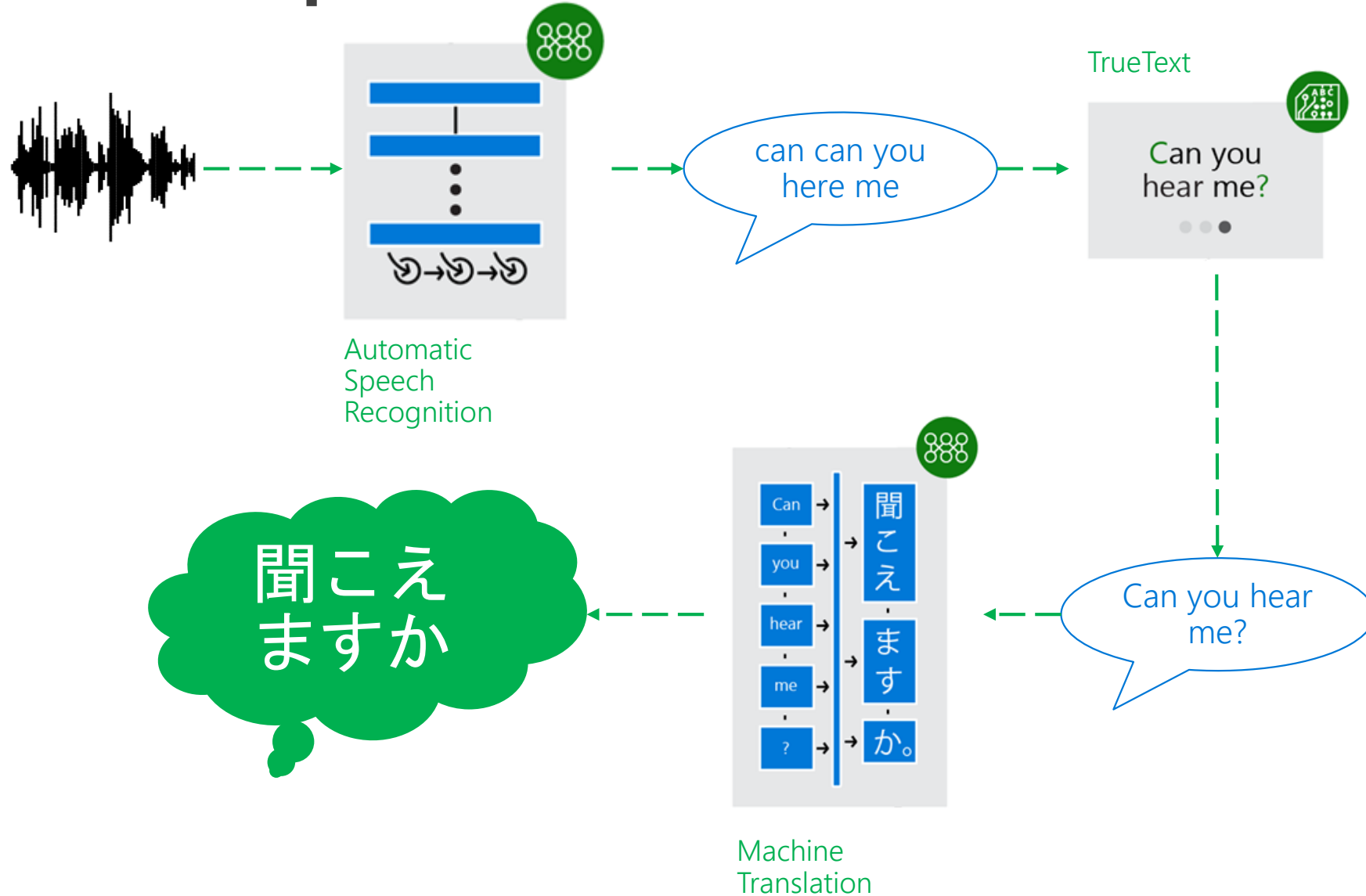
Language Understanding (LUIS)



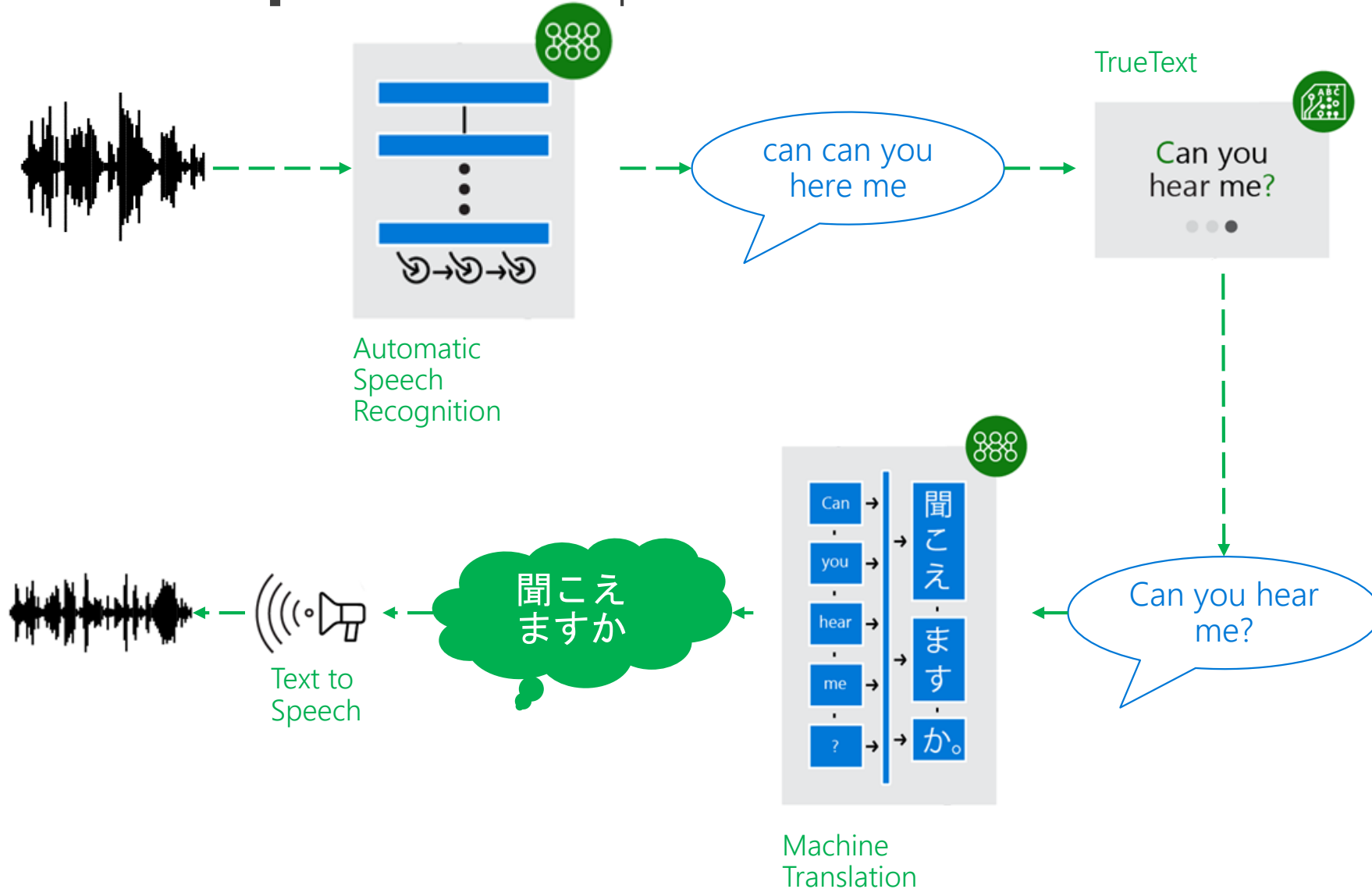
Natural Language Processing Technology Choices

- Azure HDInsight
 - with Spark and Spark MLlib
- Microsoft Cognitive Services
 - LUIS
 - Bing Search APIs

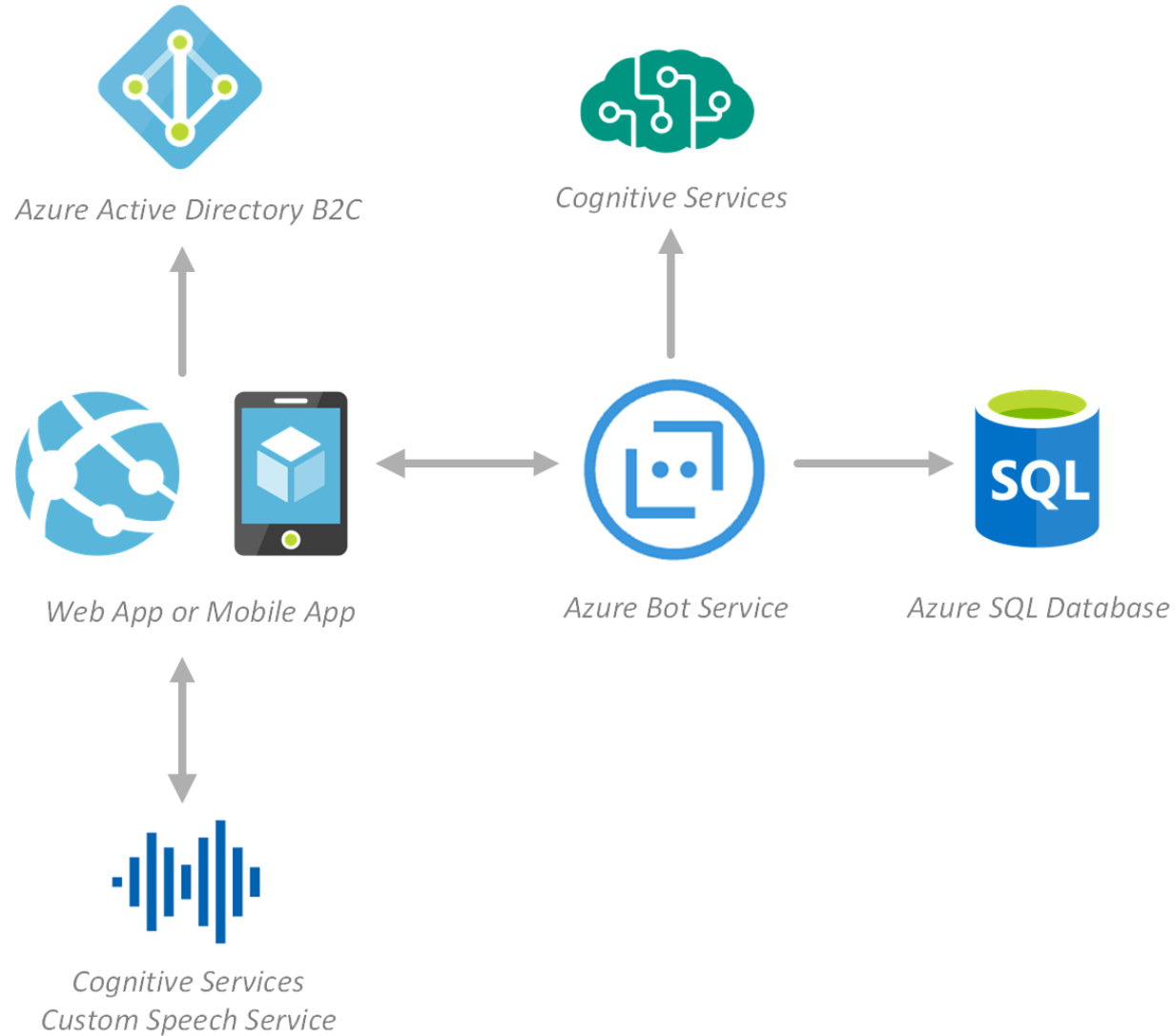
Speech-to-Text Translation



Speech-to-Speech Translation

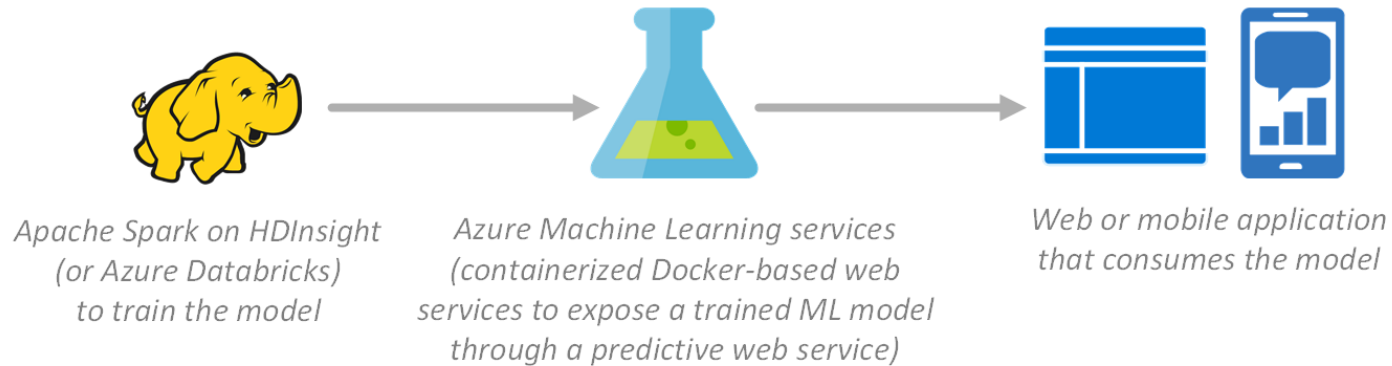


Intelligent Applications

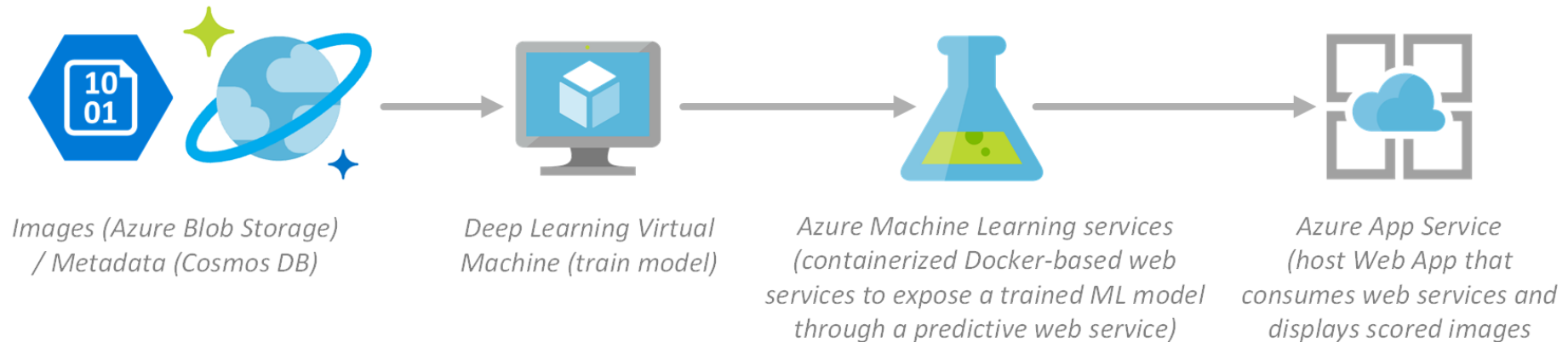


Advanced Analytics & Deep Learning

ADVANCED ANALYTICS

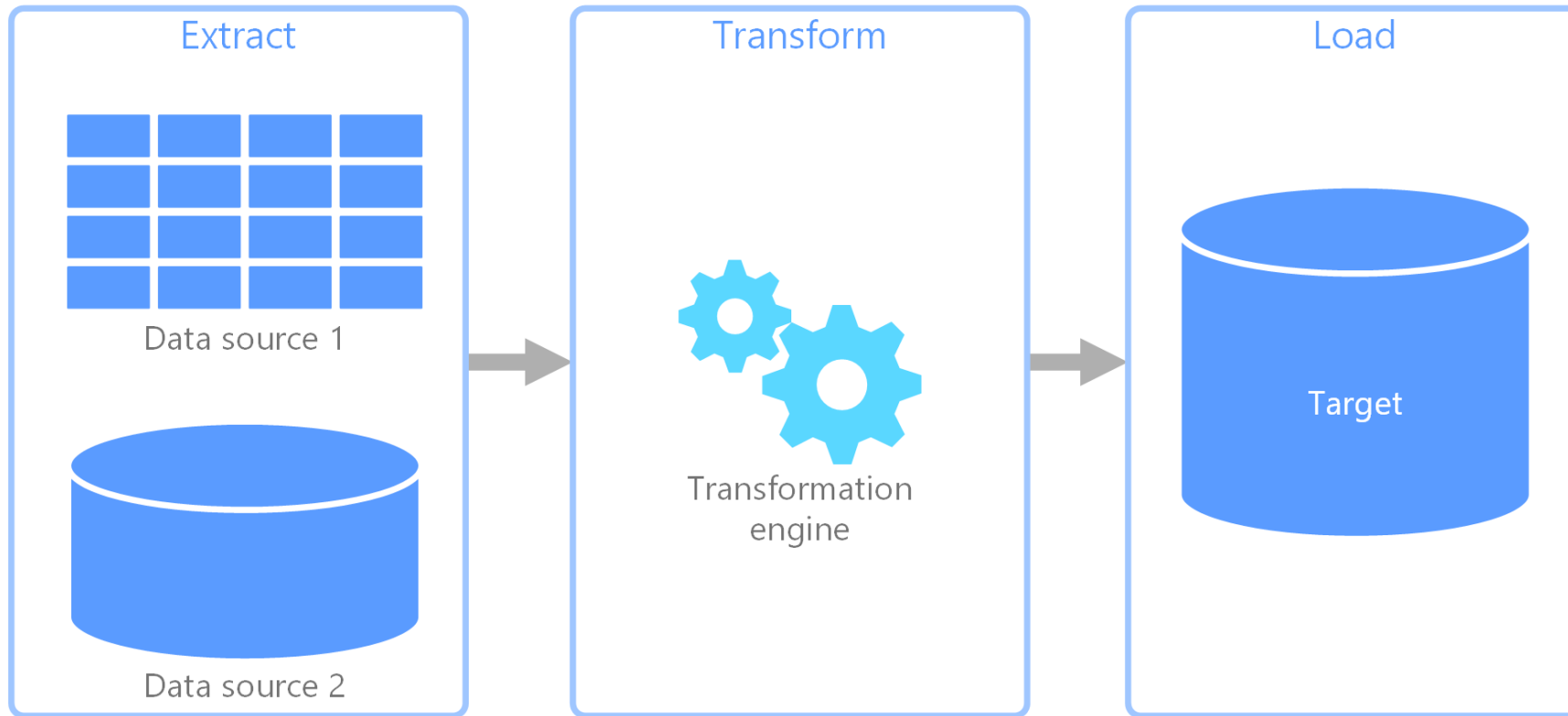


DEEP LEARNING

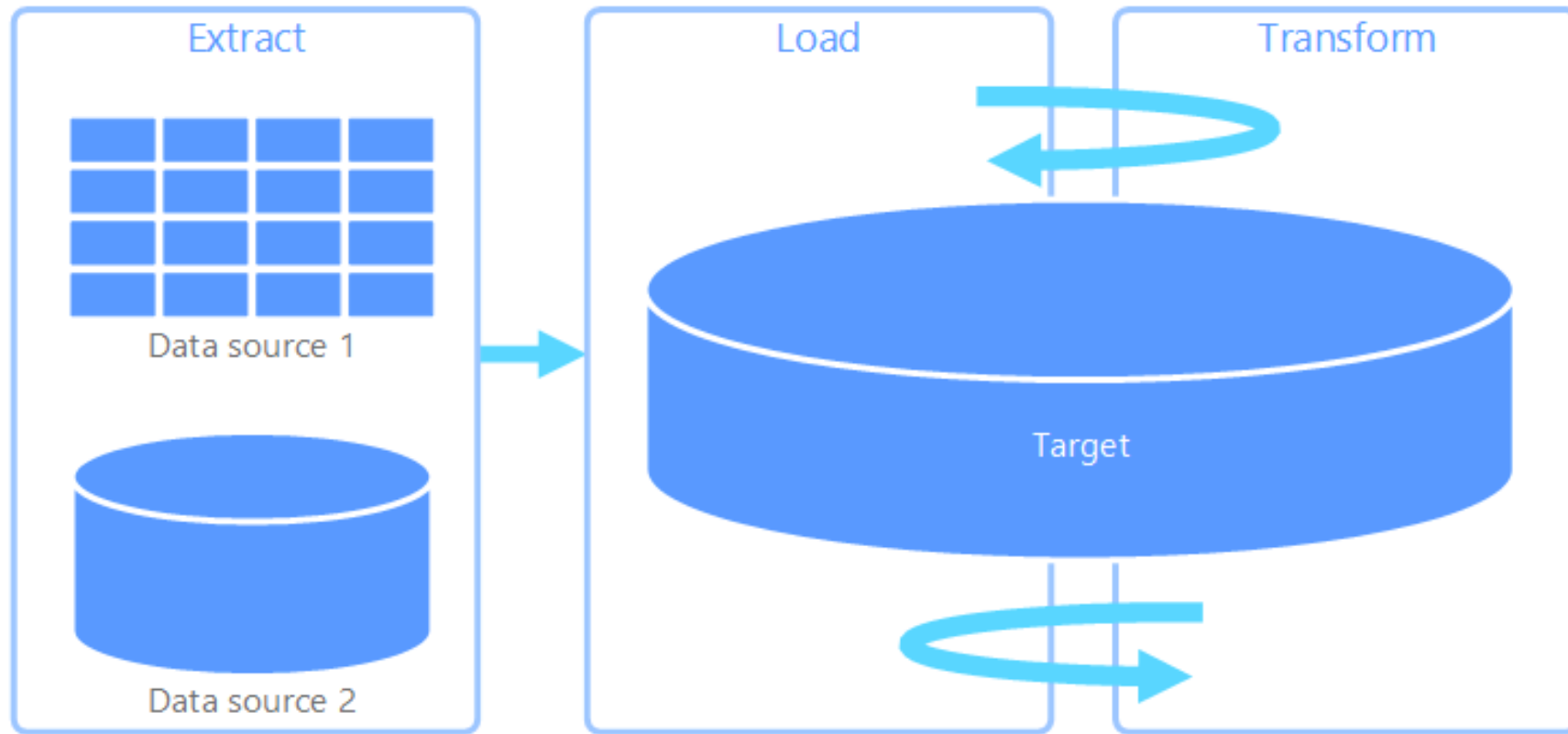


Data Pipeline

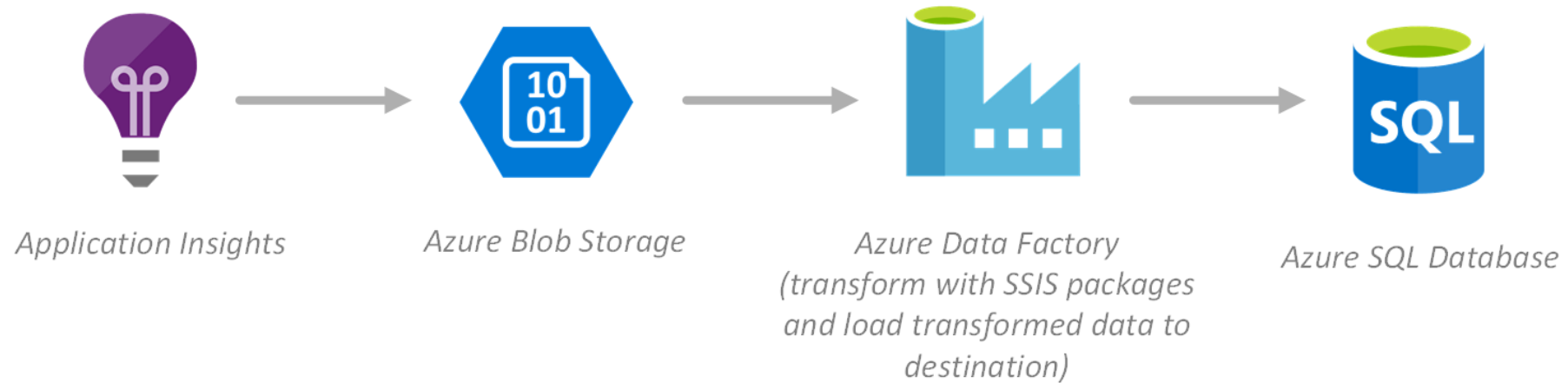
Extract, transform, and load



Extract, load, and transform

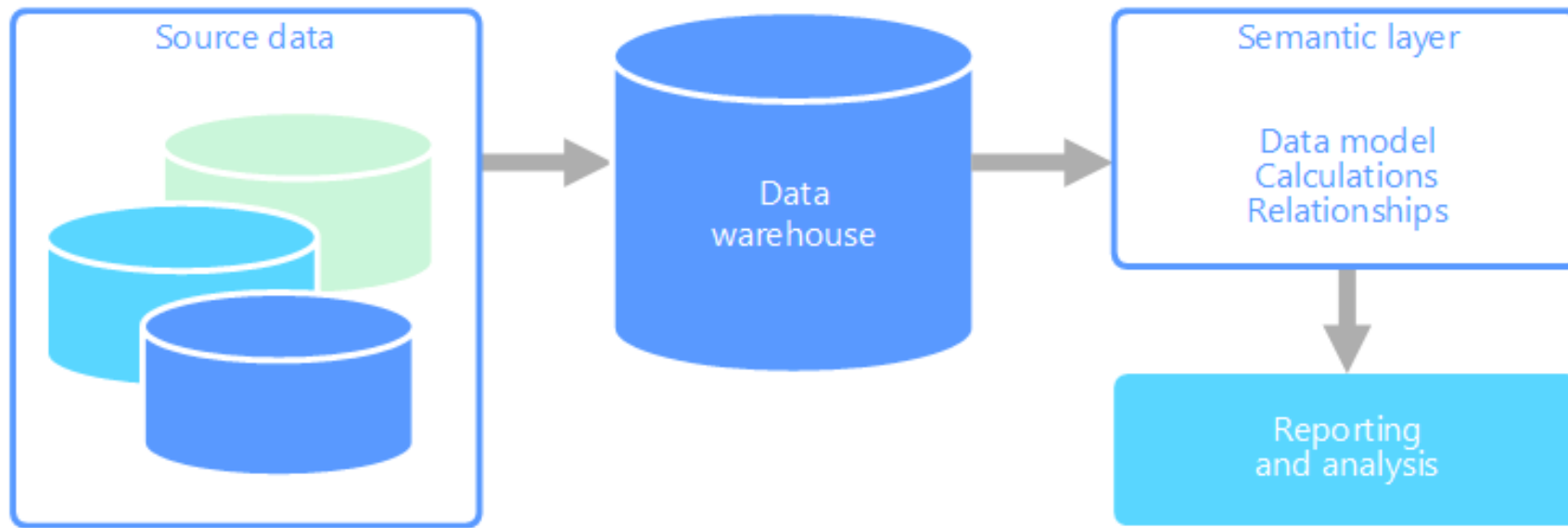


Extract, Transform, Load (ETL)

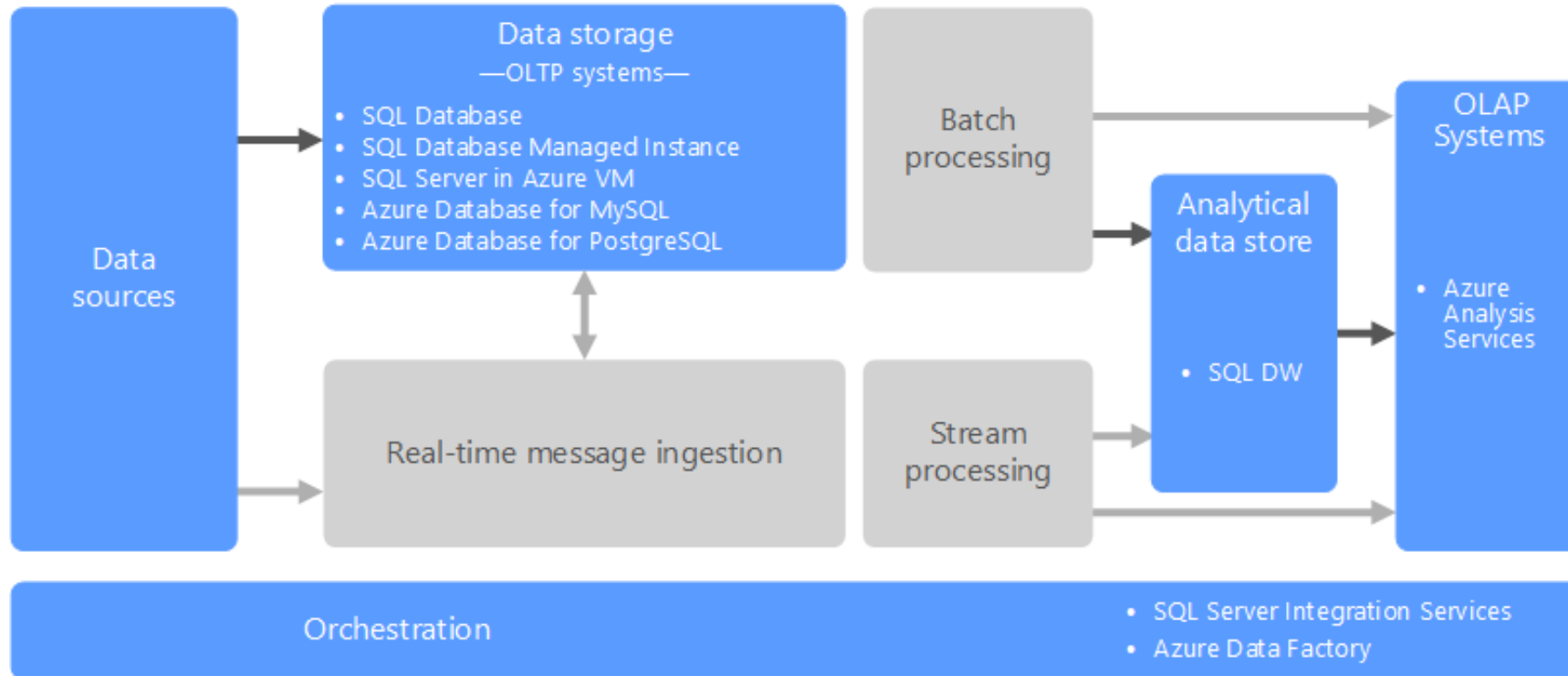


Semantic Modeling

Semantic Modeling



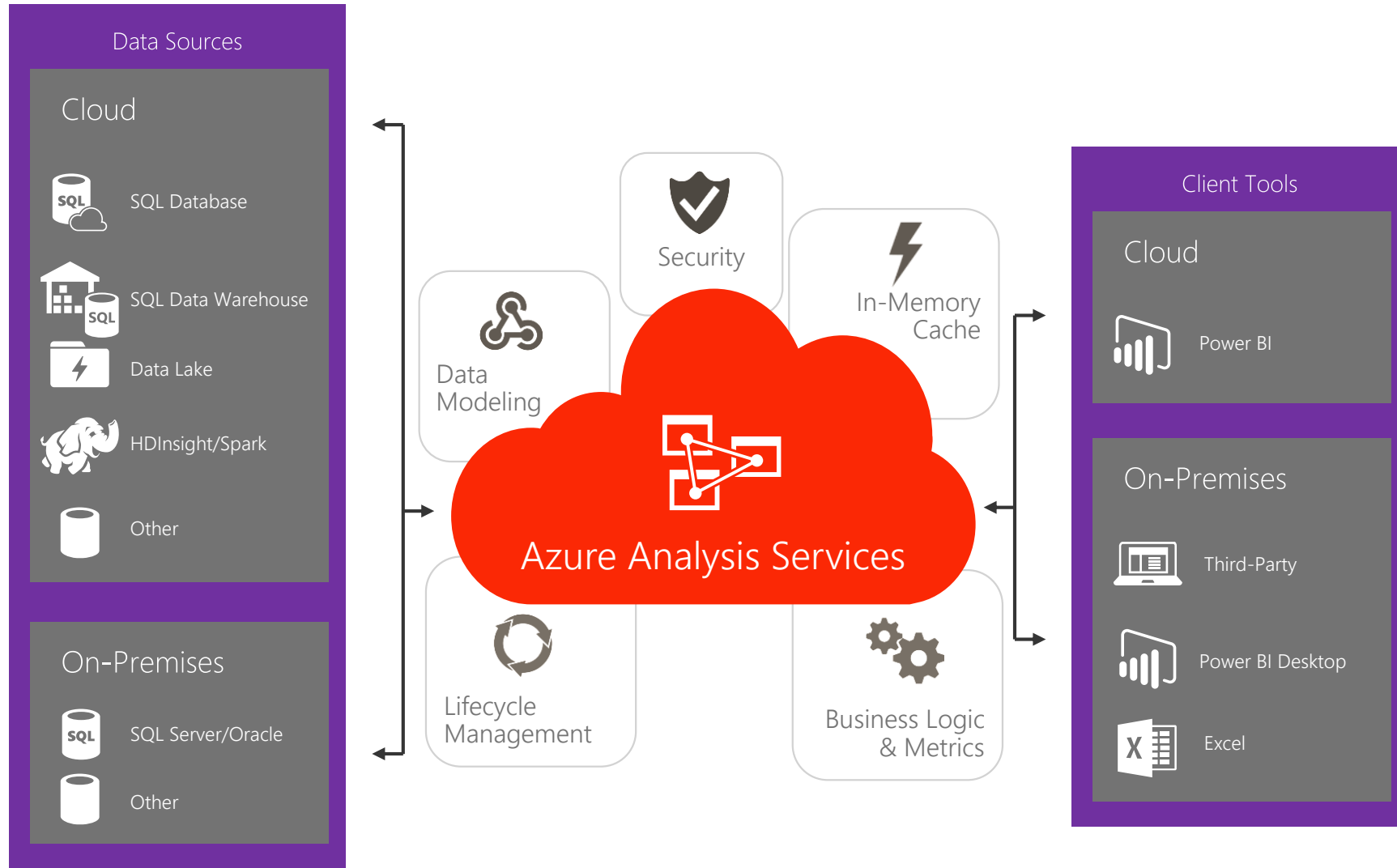
Online Analytical Processing Pattern (OLAP)



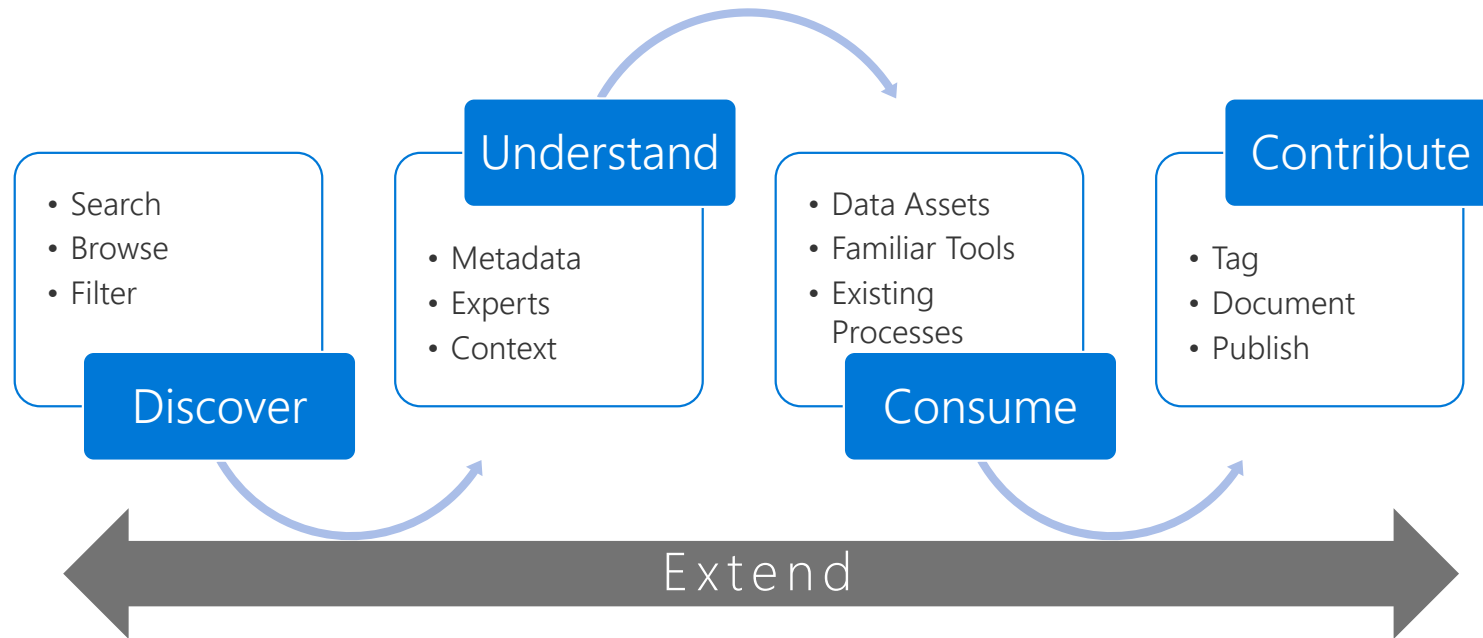
OLAP Technology Choices

- SQL Server with Columnstore indexes
- Azure Analysis Services
- SQL Server Analysis Services (SSAS)

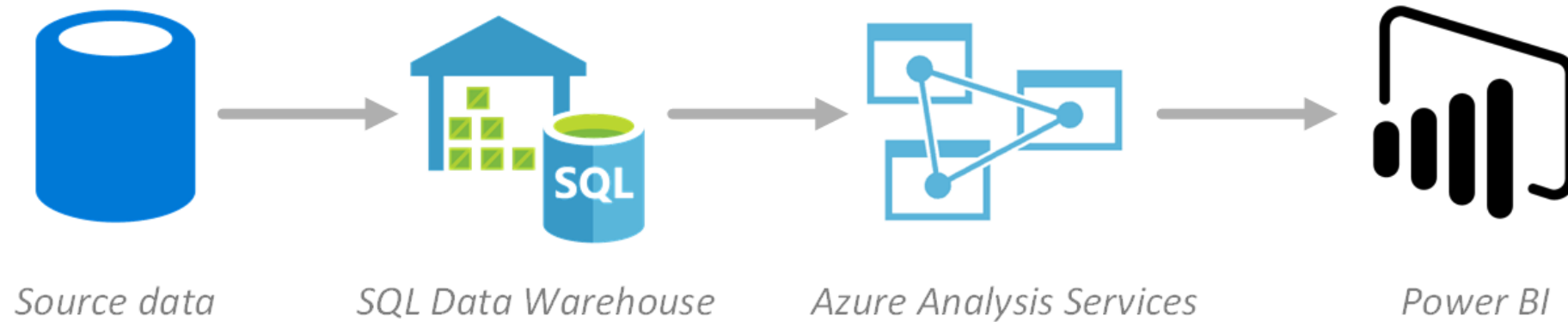
Azure Analysis Services



Data Mart



Business Intelligence



Azure Data Factory drives the process

Relational Data

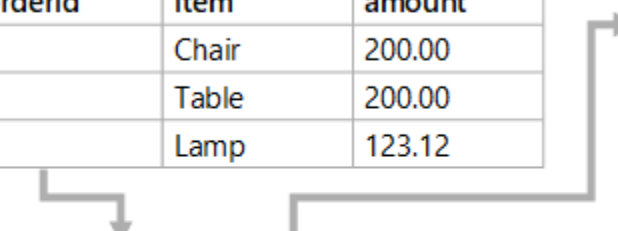
Relational Data

OrderItems table

itemid	orderid	item	amount
5	1	Chair	200.00
6	1	Table	200.00
7	1	Lamp	123.12

Customers table

customerid	name	email
4	Dan Drayton	dan@adatum.com
3	Kari Tran	kari@cpandl.com
1	Margo Ayers	margo@cpandl.com



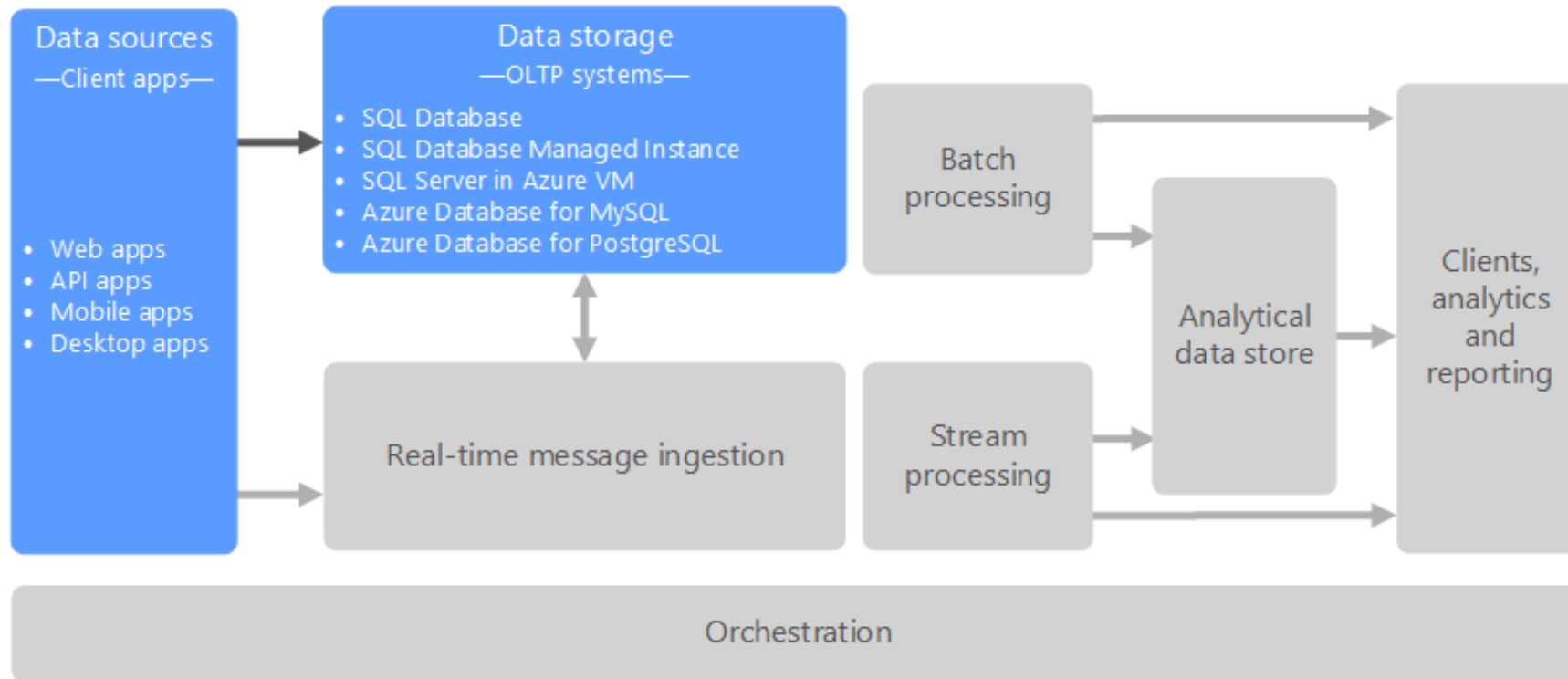
orderid	customerid	date	amount
1	4	11/1/17	523.12
2	3	11/15/17	32.99
3	1	11/21/17	23.99

Orders table

Transactional Data



Online Transaction Processing (OLTP) Pattern



OLTP Technology Choices

- Azure SQL Database
- Azure SQL Database Managed Instance
- SQL Server in an Azure Virtual Machine
- Azure Database for MySQL
- Azure Database for PostgreSQL

Thank You!

E-mail: **sidney@seesharprun.net**

Twitter: **[@sidney_andrews](https://twitter.com/sidney_andrews)**