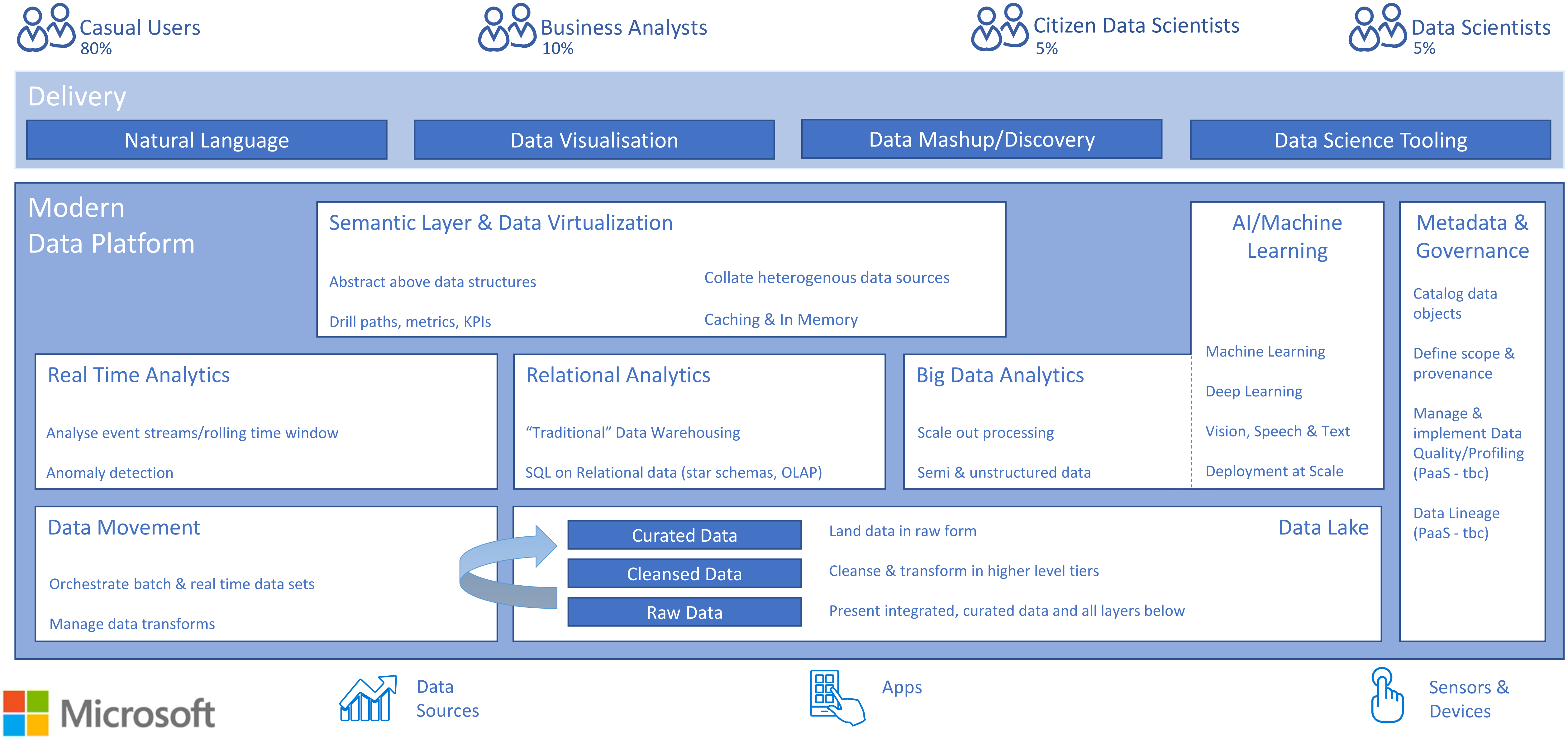
Microsoft

Modern Data Platform

Straight Through File Processing

Azure Global Bootcamp 2019 - Neil Millington

Capability Architecture



Casual Users
80%

Business Analysts
10%

Citizen Data Scientists
5%

Data Scientists
5%

Delivery

- Natural Language
- Data Visualisation
- Data Mashup/Discovery
- Data Science Tooling

Modern Data Platform

Semantic Layer & Data Virtualization

Abstract above data structures

Collate heterogenous data sources

Drill paths, metrics, KPIs

Caching & In Memory

AI/Machine Learning

Machine Learning

Deep Learning

Vision, Speech & Text

Deployment at Scale

Metadata & Governance

Catalog data objects

Define scope & provenance

Manage & implement Data Quality/Profiling (PaaS - tbc)

Data Lineage (PaaS - tbc)

Real Time Analytics

Analyse event streams/rolling time window

Anomaly detection

Relational Analytics

“Traditional” Data Warehousing

SQL on Relational data (star schemas, OLAP)

Big Data Analytics

Scale out processing

Semi & unstructured data

Data Movement

Orchestrate batch & real time data sets

Manage data transforms

Data Lake

Curated Data

Cleansed Data

Raw Data

Land data in raw form

Cleanse & transform in higher level tiers

Present integrated, curated data and all layers below

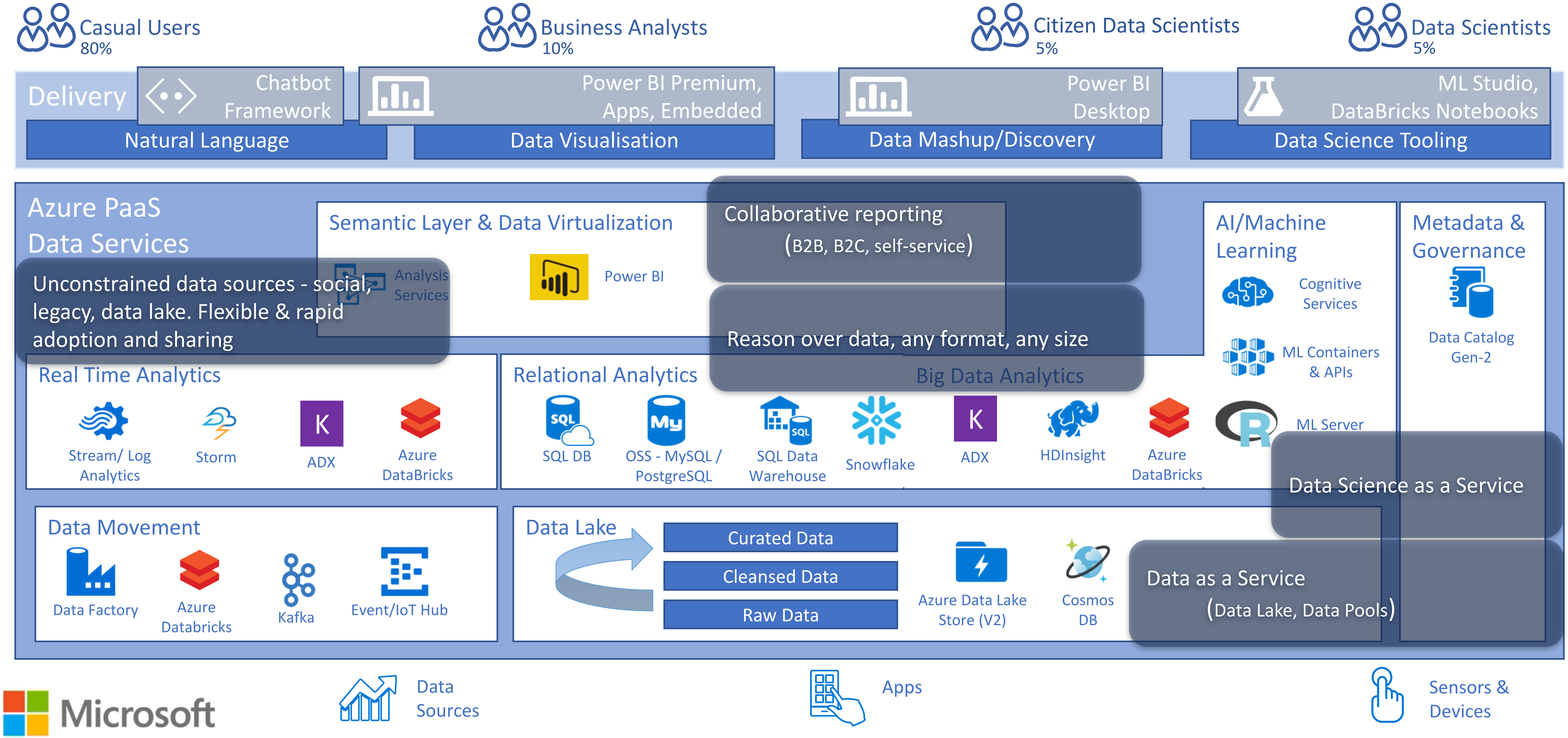
Data Sources

Apps

Sensors & Devices

Microsoft

Modern Data Platform – wants & needs



Casual Users 80%

Business Analysts 10%

Citizen Data Scientists 5%

Data Scientists 5%

Delivery

Chatbot Framework

Natural Language

Power BI Premium, Apps, Embedded

Data Visualisation

Power BI Desktop

Data Mashup/Discovery

ML Studio, DataBricks Notebooks

Data Science Tooling

Azure PaaS Data Services

Unconstrained data sources - social, legacy, data lake. Flexible & rapid adoption and sharing

Semantic Layer & Data Virtualization

Analysis Services

Power BI

Collaborative reporting (B2B, B2C, self-service)

Reason over data, any format, any size

Big Data Analytics

AI/Machine Learning

Cognitive Services

ML Containers & APIs

ML Server

Metadata & Governance

Data Catalog Gen-2

Real Time Analytics

Stream/ Log Analytics

Storm

ADX

Azure DataBricks

Relational Analytics

SQL DB

OSS - MySQL / PostgreSQL

SQL Data Warehouse

Snowflake

ADX

HDInsight

Azure DataBricks

Data Science as a Service

Data Movement

Data Factory

Azure Databricks

Kafka

Event/IoT Hub

Data Lake

Curated Data

Cleansed Data

Raw Data

Azure Data Lake Store (V2)

Cosmos DB

Data as a Service (Data Lake, Data Pools)

Data Sources

Apps

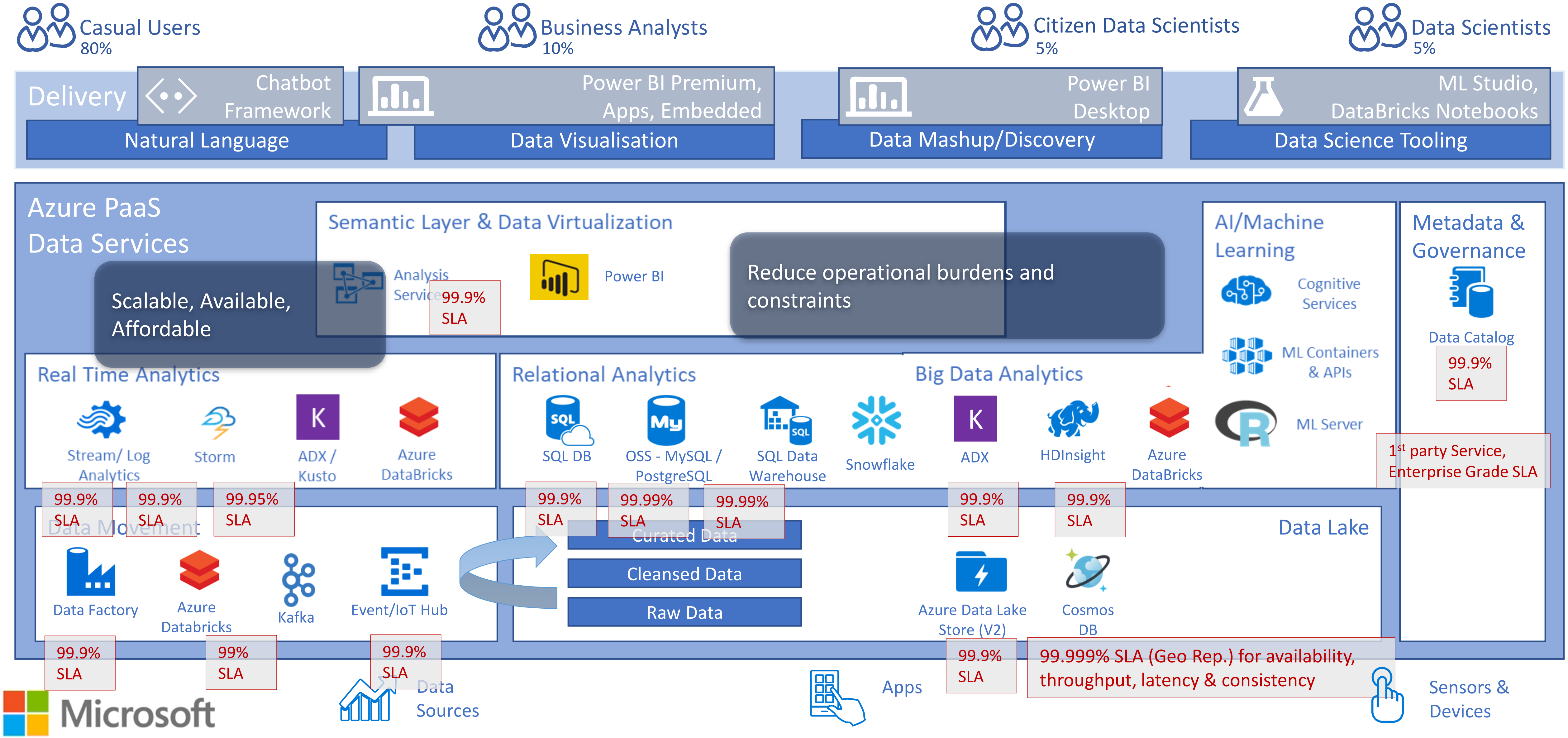
Sensors & Devices

On Demand Scalability…..

- Wide range of scale on demand services
- Sandpit capability, scale as needed
- Adapt to peaks and troughs as needed

Azure Analysis Services	Scale up: 400GB compressed Scale out: up to 8 instances	Pause & Resize
SQL DW	Scale out: 100 to 18000 DWUs	Pause & Resize
ADX / Kusto	Cluster creation and teardown	Pause & Resize
Databricks	Lightweight cluster creation and teardown	
Cosmos DB	Enterprise Scale NoSQL as a Service	
Azure Data Lake Store	Near infinite scale, no limit on object size	
Azure Data Factory	Data Movement as a Service	Up to 1GB/s

…and availability



Casual Users 80%

Business Analysts 10%

Citizen Data Scientists 5%

Data Scientists 5%

Delivery

Chatbot Framework

Natural Language

Power BI Premium, Apps, Embedded

Data Visualisation

Power BI Desktop

Data Mashup/Discovery

ML Studio, DataBricks Notebooks

Data Science Tooling

Azure PaaS Data Services

Scalable, Available, Affordable

Semantic Layer & Data Virtualization

Analysis Services 99.9% SLA

Power BI

Reduce operational burdens and constraints

AI/Machine Learning

Cognitive Services

ML Containers & APIs

ML Server

Metadata & Governance

Data Catalog 99.9% SLA

1st party Service, Enterprise Grade SLA

Real Time Analytics

Stream/ Log Analytics 99.9% SLA

Storm 99.9% SLA

ADX / Kusto 99.95% SLA

Azure DataBricks

Relational Analytics

SQL DB 99.9% SLA

OSS - MySQL / PostgreSQL 99.99% SLA

SQL Data Warehouse 99.99% SLA

Snowflake

Big Data Analytics

ADX 99.9% SLA

HDInsight 99.9% SLA

Azure DataBricks

Data Movement

Data Factory 99.9% SLA

Azure Databricks 99% SLA

Kafka

Event/IoT Hub 99.9% SLA

Data Lake

Curated Data

Cleansed Data

Raw Data

Azure Data Lake Store (V2) 99.9% SLA

Cosmos DB

99.999% SLA (Geo Rep.) for availability, throughput, latency & consistency

Data Sources

Apps

Sensors & Devices

Warehousing on Azure

Data Factory
- Data Movement as a Service
- Host SSIS packages in pipelines
- Visual Tooling

Power BI
- Data discovery for everyone
- Class leading Data Shaping

Azure Analysis Services
- Enterprise wide semantic layer
- Stunning response times
- Scale up/Scale out

Data Catalog
- Catalog Enterprise
- GDPR initiatives

Analytics
- Scalable
- Lambda (Hot, Warm, Cool data)

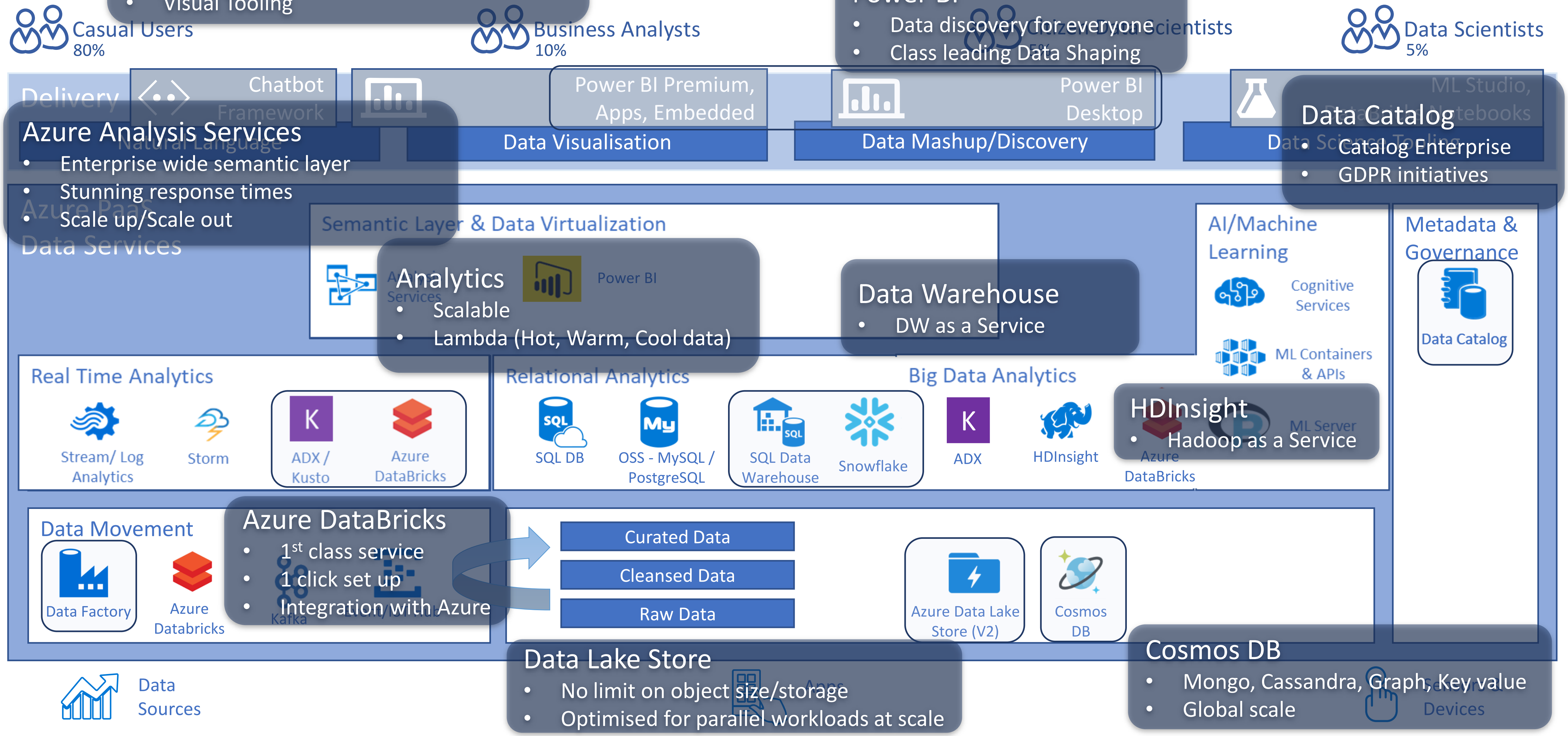
Data Warehouse
- DW as a Service

HDInsight
- Hadoop as a Service

Azure DataBricks
- 1^{st} class service
- 1 click set up
- Integration with Azure

Data Lake Store
- No limit on object size/storage
- Optimised for parallel workloads at scale

Cosmos DB
- Mongo, Cassandra, Graph, Key value
- Global scale



Casual Users 80%

Business Analysts 10%

Citizen Data Scientists 5%

Data Scientists 5%

Delivery

Chatbot Framework

Natural Language

Power BI Premium, Apps, Embedded

Data Visualisation

Power BI Desktop

Data Mashup/Discovery

ML Studio, Data Science Notebooks

Data Science Tooling

Azure PaaS

Data Services

Semantic Layer & Data Virtualization

Analysis Services

Power BI

AI/Machine Learning

Cognitive Services

ML Containers & APIs

ML Server

Metadata & Governance

Data Catalog

Real Time Analytics

Stream/ Log Analytics

Storm

ADX / Kusto

Azure DataBricks

Relational Analytics

SQL DB

OSS - MySQL / PostgreSQL

SQL Data Warehouse

Snowflake

Big Data Analytics

ADX

HDInsight

Azure DataBricks

Data Movement

Data Factory

Azure Databricks

Kafka

Event Hub/IoT Hub

Curated Data

Cleansed Data

Raw Data

Azure Data Lake Store (V2)

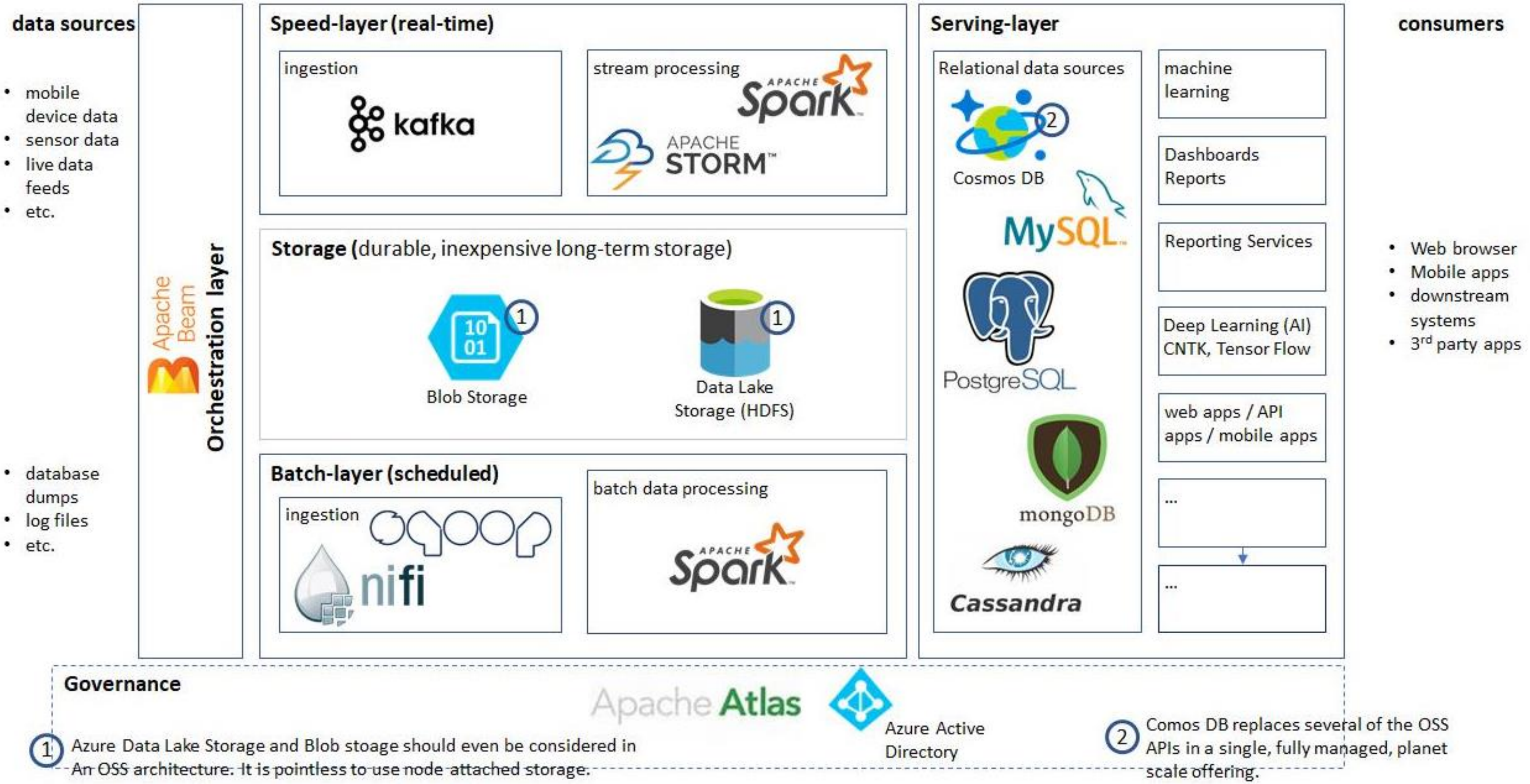
Cosmos DB

Data Sources

Apps

Sensors & Devices

An OSS viewpoint



data sources

- mobile device data
- sensor data
- live data feeds
- etc.

- database dumps
- log files
- etc.

Apache Beam

Orchestration layer

Speed-layer (real-time)

ingestion

kafka

stream processing

APACHE Spark

APACHE STORM

Storage (durable, inexpensive long-term storage)

Blob Storage ①

Data Lake Storage (HDFS) ①

Batch-layer (scheduled)

ingestion

sqoop

nifi

batch data processing

APACHE Spark

Serving-layer

Relational data sources

Cosmos DB ②

MySQL

PostgreSQL

mongoDB

Cassandra

machine learning

Dashboards Reports

Reporting Services

Deep Learning (AI) CNTK, Tensor Flow

web apps / API apps / mobile apps

...

...

consumers

- Web browser
- Mobile apps
- downstream systems
- 3rd party apps

Governance

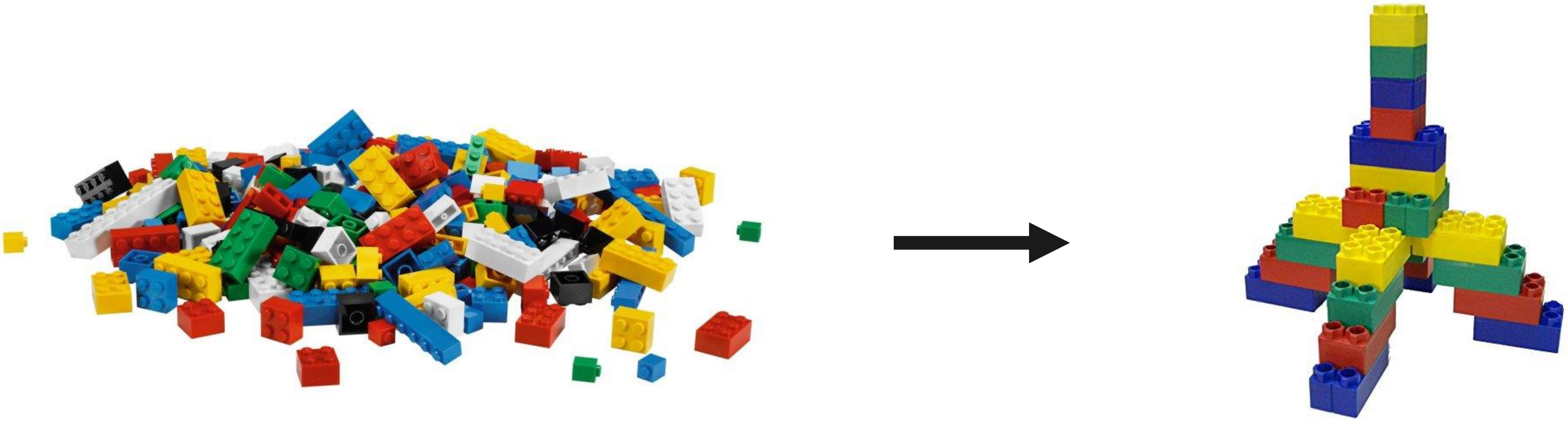
Apache Atlas

Azure Active Directory

① Azure Data Lake Storage and Blob stoage should even be considered in An OSS architecture. It is pointless to use node attached storage.

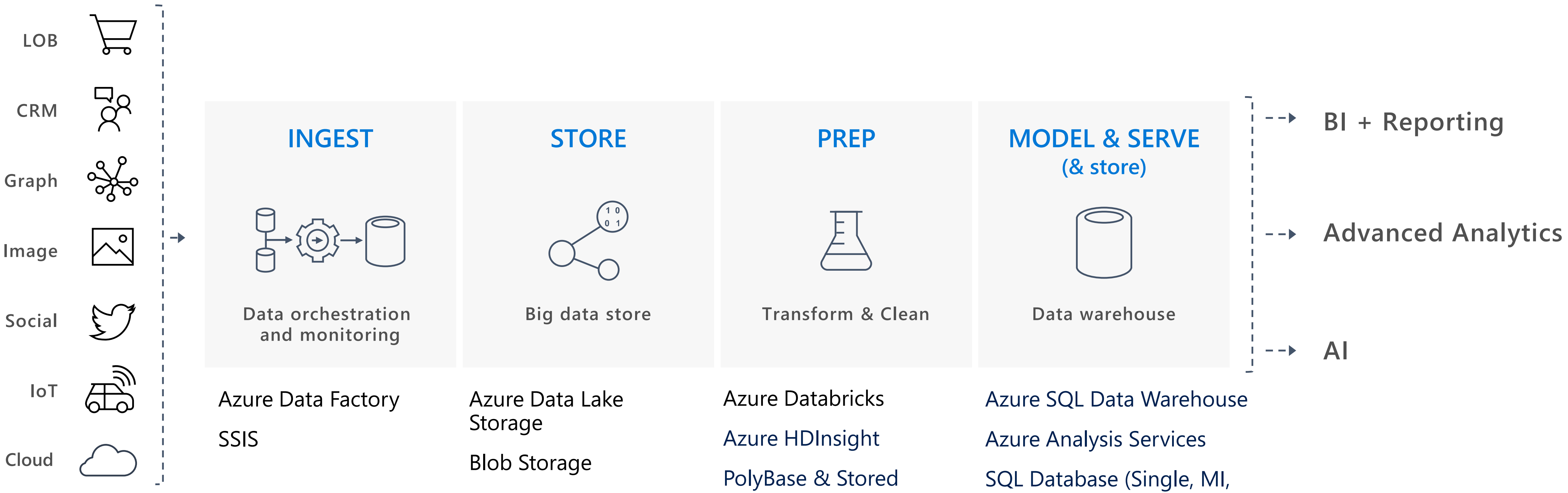
② Comos DB replaces several of the OSS APIs in a single, fully managed, planet scale offering.

Our job



Plug n Play Products

Modern Data Platform (**commoditized patterns**)



LOB

CRM

Graph

Image

Social

IoT

Cloud

INGEST

Data orchestration and monitoring

STORE

Big data store

PREP

Transform & Clean

MODEL & SERVE
(& store)

Data warehouse

BI + Reporting

Advanced Analytics

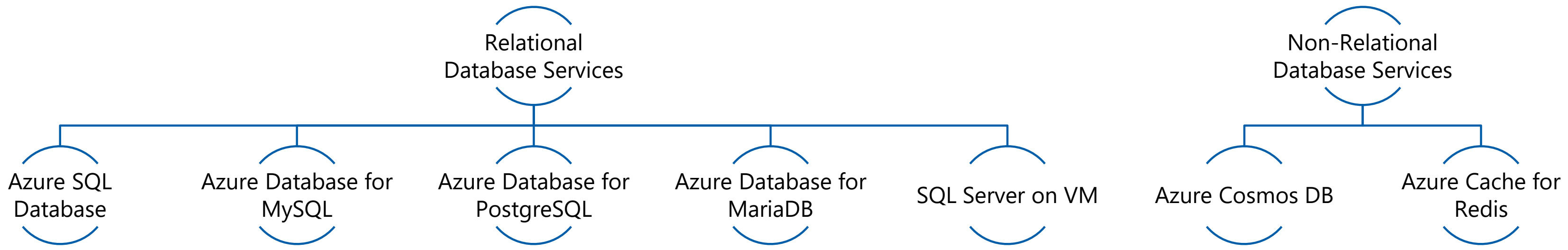
AI

INGEST	STORE	PREP	MODEL & SERVE
Azure Data Factory SSIS	Azure Data Lake Storage Blob Storage SQL Server OSS (PostgreSQL, MySQL, MariaDB)	Azure Databricks Azure HDInsight PolyBase & Stored Procedures Power BI Dataflow Azure Data Lake Analytics	Azure SQL Data Warehouse Azure Analysis Services SQL Database (Single, MI, HyperScale, Serverless) SQL Server in a VM Cosmos DB Power BI Aggregations

Microsoft

Credit James Serra, Microsoft

Azure Database Platform – Microsoft & OSS redefined



Relational Database Services

Azure SQL Database

Azure Database for MySQL

Azure Database for PostgreSQL

Azure Database for MariaDB

SQL Server on VM

Non-Relational Database Services

Azure Cosmos DB

Azure Cache for Redis

Fully-Managed

Flexible

Enterprise Scale & Performance

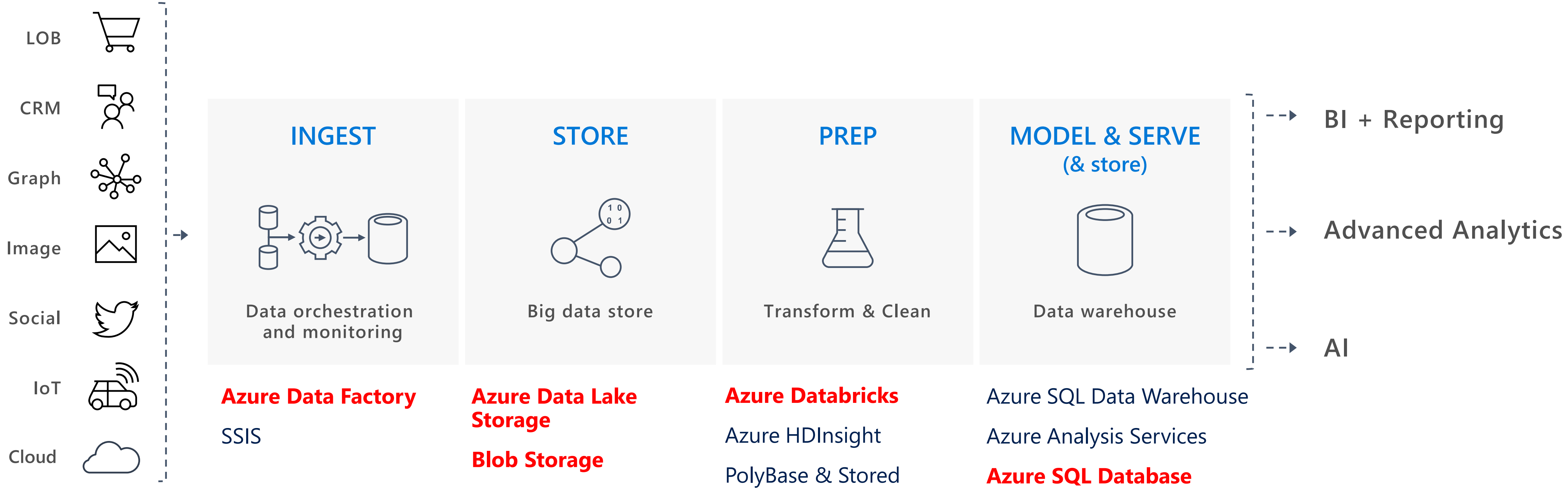
Security & Compliance

Products not covered

- Streaming (i.e. Iot Hub, Event Hub, Azure Stream Analytics)
- On-prem products (i.e. APS)
- Open source (i.e. Kafka, PostgreSQL, MySQL, MariaDB, Storm, Spark, Hbase, Redis)
- Machine learning/AI tools (i.e. Azure ML, Machine Learning Services, Cognitive Services)
- Reporting tools (i.e. Power BI)
- 3rd-party products (i.e. Informatica, Profisee)
- Competitor products (AWS, Google)
- Snowflake
- Azure Data Catalog (ADC) Gen2
- Azure Data Explorer
- Azure Database Migration Service
- Data Box, Data Box Disk
- Azure Search / Cognitive Search / Knowledge Mining
- Master Data Services (MDS)
- Azure Functions (for Prep) / Azure Logic Apps (for Ingest)
- OSS !!

Credit James Serra, Microsoft

Modern Data Platform (**possible products by four areas**)



Azure Data Factory

SSIS

Azure Data Lake Storage

Blob Storage

SQL Server

OSS (PostgreSQL, MySQL, MariaDB

Azure Databricks

Azure HDInsight

PolyBase & Stored Procedures

Power BI Dataflow

Azure SQL Data Warehouse

Azure Analysis Services

Azure SQL Database (Single, MI, HyperScale) = PaaS

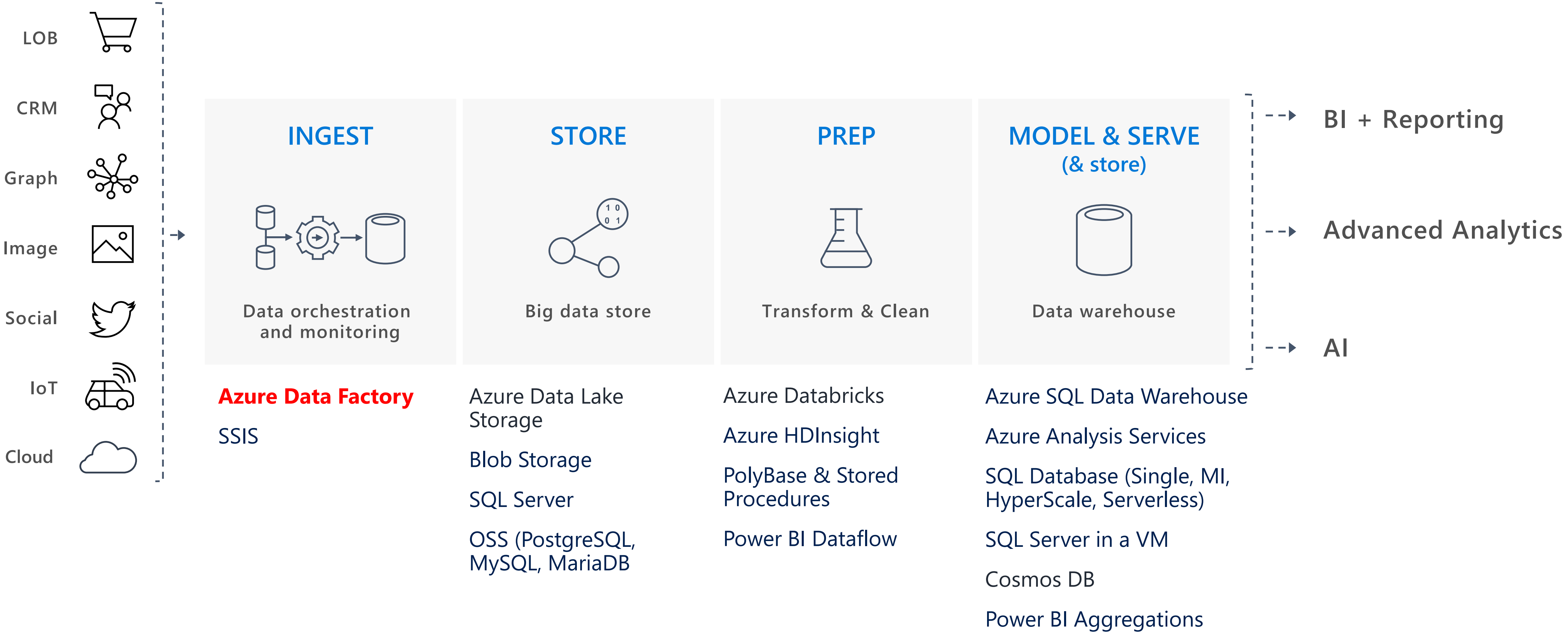
SQL Server in a VM = IaaS

Cosmos DB

Power BI Aggregations

Note: Those products that span more than one area are listed in their primary area

Modern Data Platform (commoditized patterns)



LOB

CRM

Graph

Image

Social

IoT

Cloud

INGEST

Data orchestration and monitoring

Azure Data Factory

SSIS

STORE

Big data store

Azure Data Lake Storage

Blob Storage

SQL Server

OSS (PostgreSQL, MySQL, MariaDB

PREP

Transform & Clean

Azure Databricks

Azure HDInsight

PolyBase & Stored Procedures

Power BI Dataflow

MODEL & SERVE (& store)

Data warehouse

Azure SQL Data Warehouse

Azure Analysis Services

SQL Database (Single, MI, HyperScale, Serverless)

SQL Server in a VM

Cosmos DB

Power BI Aggregations

BI + Reporting

Advanced Analytics

AI

AZURE DATA FACTORY

Ingest – Data Orchestration and Monitoring

Product: Azure Data Factory (ADF)

Overview:

With Mapping Data Flow, can now transform data, so ETL/ELT tool. Copy Data tool to easily copy from source to destination.

Use cases:

Any new usecase for data movement and curation. SSIS packages migration.

How to use:

PaaS based service – *Hybrid configuration, Hybrid operation*

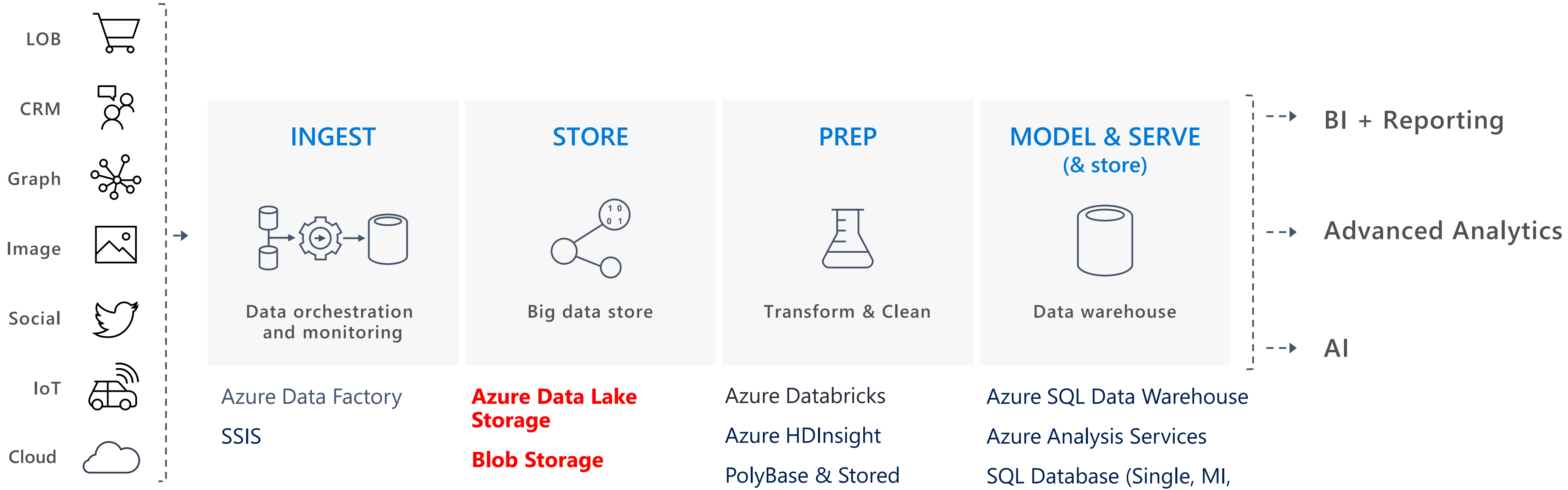
Area also used for:

Preparation of data (as part of a pipeline)

Microsoft

Credit James Serra, Microsoft

Modern Data Platform (commoditized patterns)



Azure Data Factory

SSIS

Azure Data Lake Storage

Blob Storage

SQL Server

OSS (PostgreSQL, MySQL, MariaDB

Azure Databricks

Azure HDInsight

PolyBase & Stored Procedures

Power BI Dataflow

Azure SQL Data Warehouse

Azure Analysis Services

SQL Database (Single, MI, HyperScale, Serverless)

SQL Server in a VM

Cosmos DB

Power BI Aggregations

Store – Data Lake / Big Data

Product: **Azure Data Lake Storage Gen2 (ADLS Gen2)**

Overview:

Combines features of blob storage and ADLS Gen1. ADLS Gen2 adds a high performance HDFS Endpoint to Azure Blob Storage and inherits the rich feature set of Azure Blob Storage

Use cases:

Any new project. Convert Blob and Gen1 over time

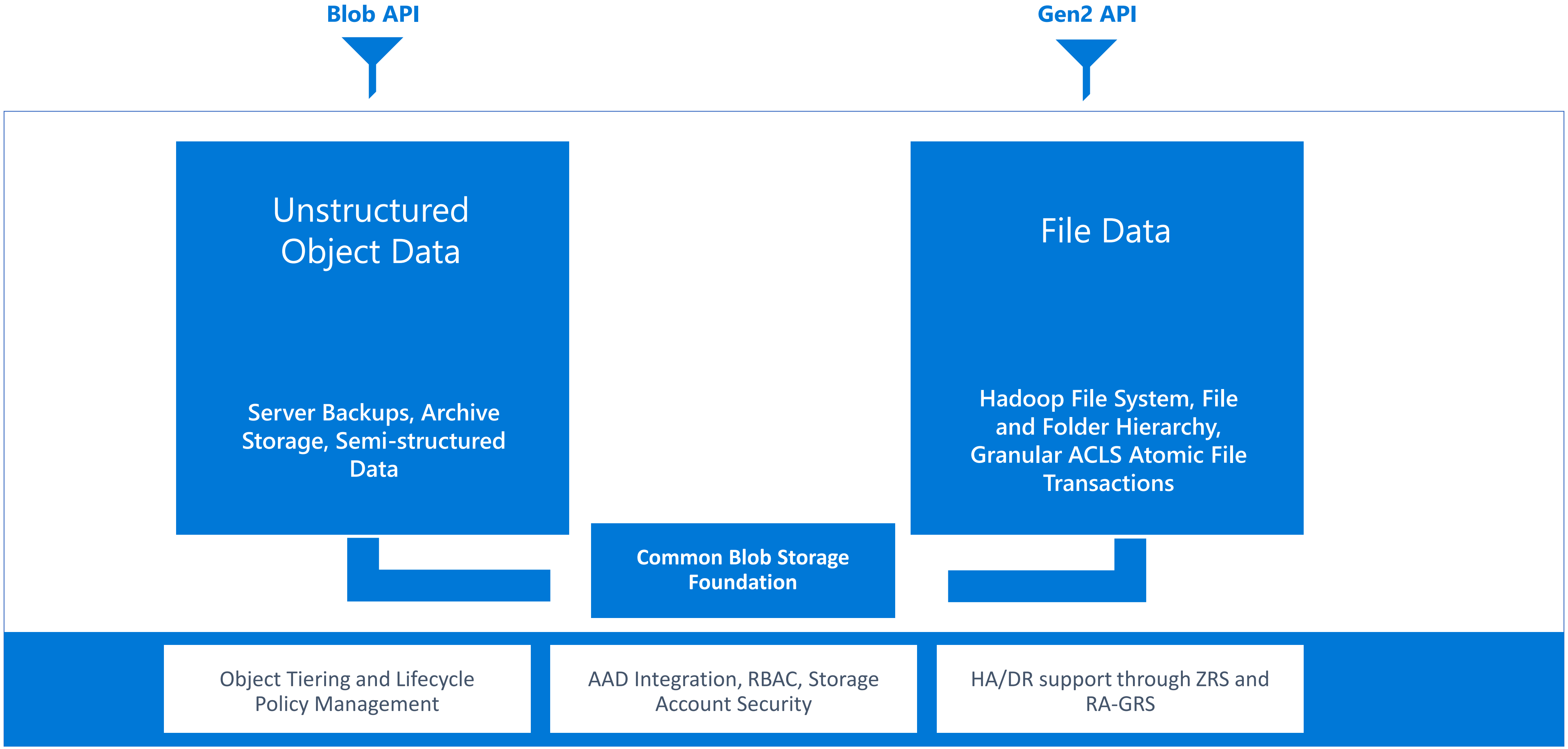
How to use:

PaaS

Area also used for:

n/a

Credit James Serra, Microsoft



Blob API

Gen2 API

Unstructured Object Data

Server Backups, Archive Storage, Semi-structured Data

File Data

Hadoop File System, File and Folder Hierarchy, Granular ACLS Atomic File Transactions

Common Blob Storage Foundation

Object Tiering and Lifecycle Policy Management

AAD Integration, RBAC, Storage Account Security

HA/DR support through ZRS and RA-GRS

Microsoft

10
01
Blob

Store – Blob / Big Data

Product: Blob Storage

Overview:

Original storage, foundational to Azure

Use cases:

non-analytical use cases that only need object storage rather than hierarchical storage (i.e. video, images, backup files).

ADLS Gen2 – no need to use if current data does not need features of ADLS Gen2

How to use:

PaaS

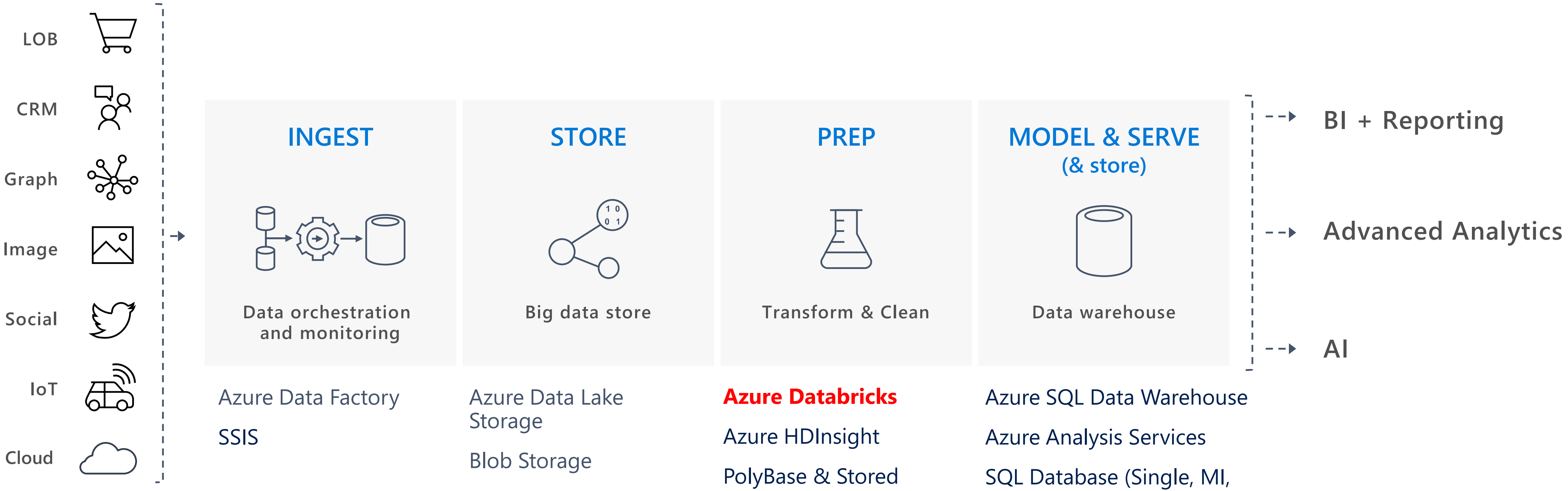
Area also used for:

n/a

Microsoft

Credit James Serra, Microsoft

Modern Data Platform (commoditized patterns)



LOB

CRM

Graph

Image

Social

IoT

Cloud

INGEST

Data orchestration and monitoring

STORE

Big data store

PREP

Transform & Clean

MODEL & SERVE
(& store)

Data warehouse

BI + Reporting

Advanced Analytics

AI

Azure Data Factory

SSIS

Azure Data Lake Storage

Blob Storage

SQL Server

OSS (PostgreSQL, MySQL, MariaDB

Azure Databricks

Azure HDInsight

PolyBase & Stored Procedures

Power BI Dataflow

Azure SQL Data Warehouse

Azure Analysis Services

SQL Database (Single, MI, HyperScale, Serverless)

SQL Server in a VM

Cosmos DB

Power BI Aggregations

Prep – Big Data

Product: Azure Databricks

Overview:

Tool for curating and processing massive amounts of data and developing, training and deploying models on that data, and managing the whole workflow process throughout the project. SPARK.

Use cases:

Comfortable with Spark and notebooks, integration with ADLS, SQL DW, PBI, etc, need auto-scaling and auto-termination, need fast Spark

How to use:

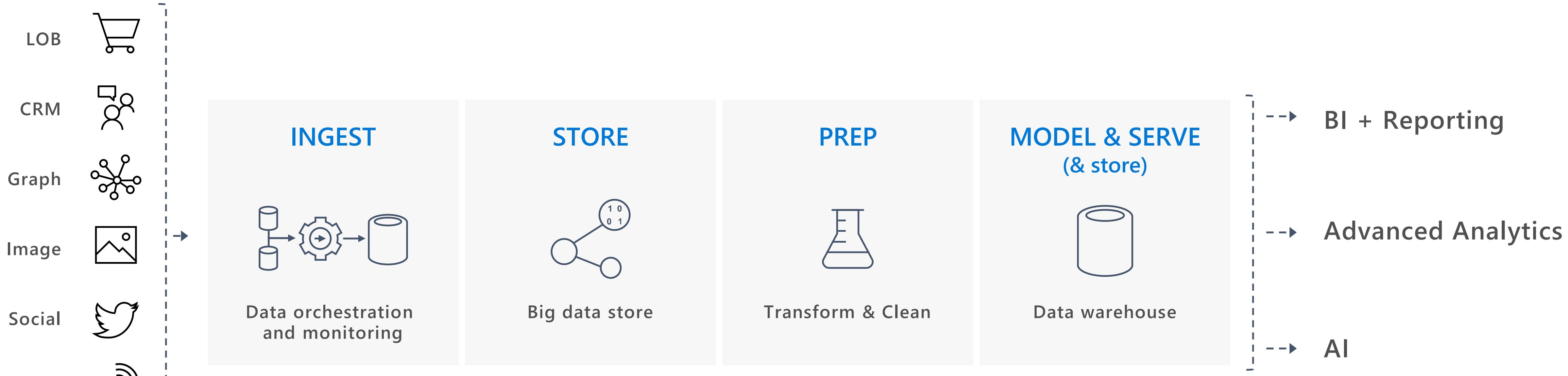
PaaS

Area also used for:

ETL/ELT - Ingest, Model & Serve

Credit James Serra, Microsoft

Modern Data Platform (commoditized patterns)



LOB

CRM

Graph

Image

Social

IoT

Cloud

INGEST

Data orchestration and monitoring

STORE

Big data store

PREP

Transform & Clean

MODEL & SERVE

(& store)

Data warehouse

BI + Reporting

Advanced Analytics

AI

INGEST

- Azure Data Factory
- SSIS

STORE

- Azure Data Lake Storage
- Blob Storage
- SQL Server
- OSS (PostgreSQL, MySQL, MariaDB

PREP

- Azure Databricks
- Azure HDInsight
- PolyBase & Stored Procedures
- Power BI Dataflow

MODEL & SERVE

- Azure SQL Data Warehouse
- Azure Analysis Services
- **SQL Database (Single, MI, HyperScale, Serverless)**
- SQL Server in a VM
- **Cosmos DB**
- Power BI Aggregations

Serve - Big Data

Product: Cosmos DB

Overview:

A globally distributed, multi-model (key-value, graph, and document) database service. It fits into the NoSQL camp by having a non-relational model (supporting schema-on-read and JSON documents)

Use cases:

Works really well for large-scale OLTP solutions. Spark to Cosmos DB connector for DW aggregations. Use for data lake to have one datastore for both operational and analytical queries

How to use:

PaaS

Area also used for:

Store, Prep

Credit James Serra, Microsoft

Azure Cosmos DB

A globally distributed, fully managed, massively scalable, multi-model database service

SQL

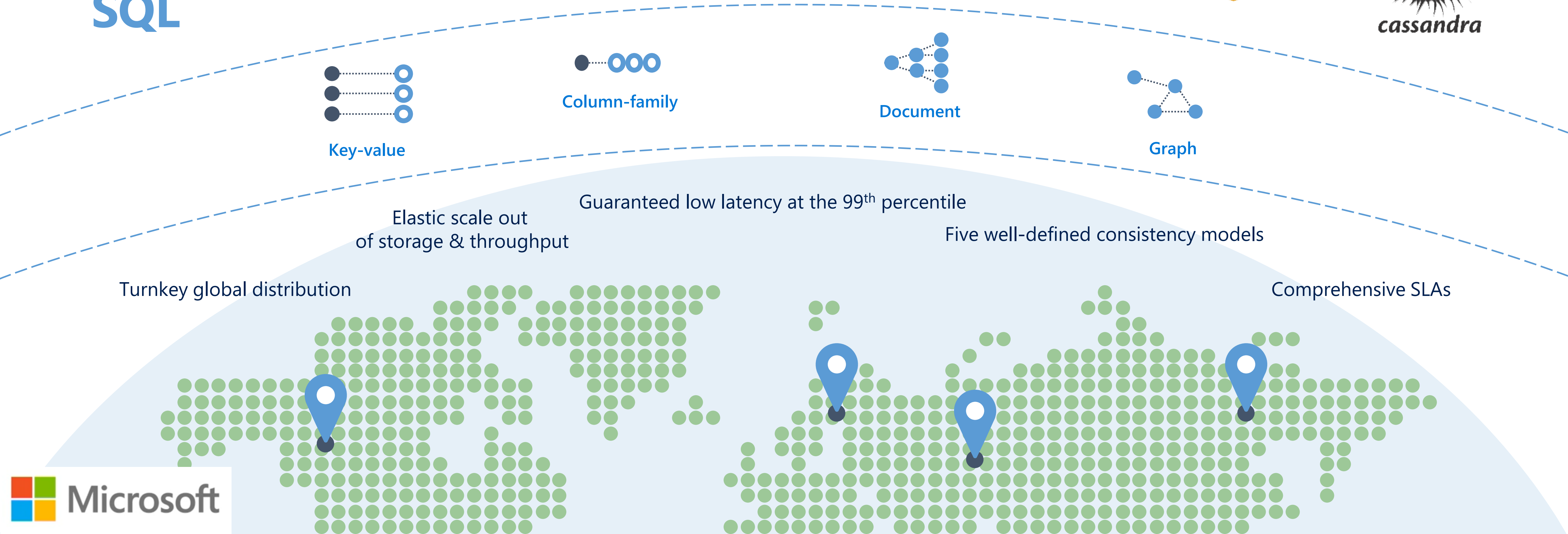
MongoDB

Table API

Gremlin $G = (V, E)$

JS

cassandra



Key-value

Column-family

Document

Graph

Guaranteed low latency at the 99th percentile

Elastic scale out of storage & throughput

Five well-defined consistency models

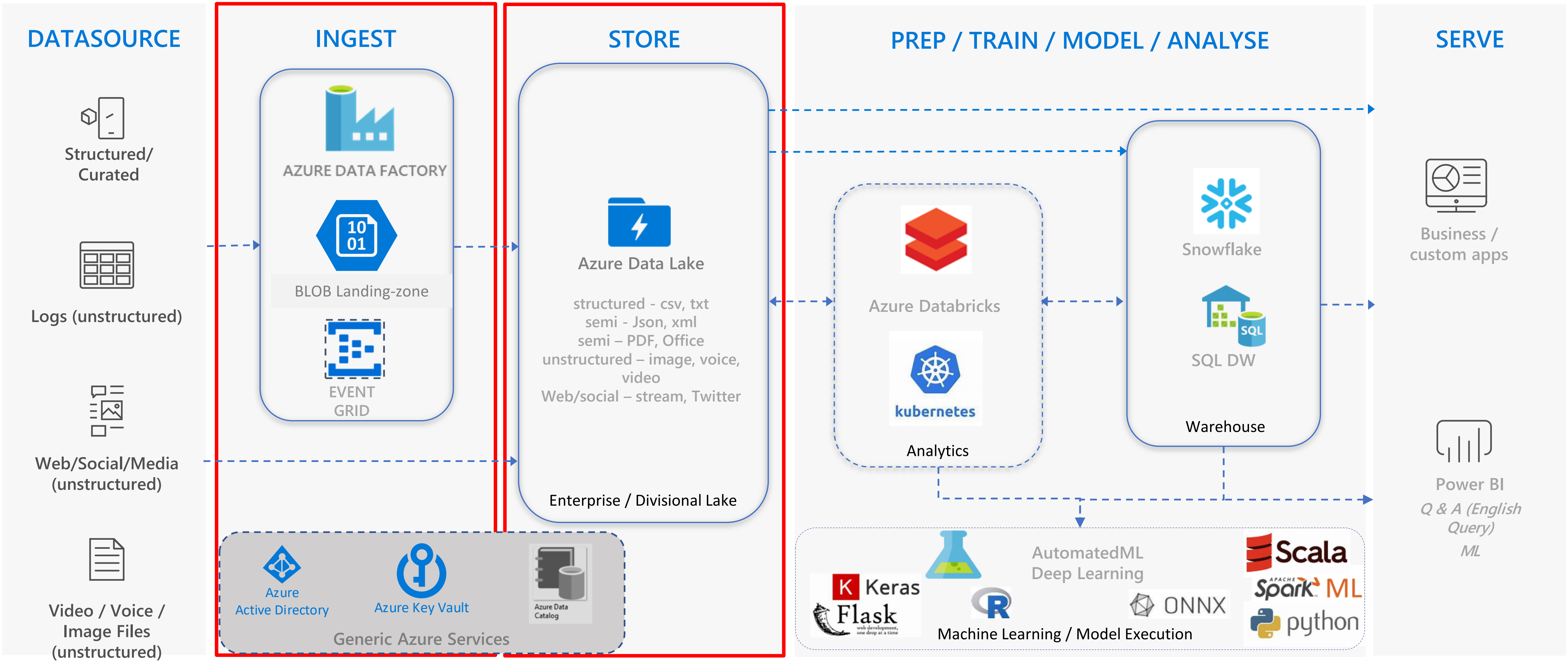
Turnkey global distribution

Comprehensive SLAs

Microsoft

Ok, but why ?

The Data Estate



DATASOURCE

Structured/ Curated

Logs (unstructured)

Web/Social/Media (unstructured)

Video / Voice / Image Files (unstructured)

INGEST

AZURE DATA FACTORY

BLOB Landing-zone

EVENT GRID

STORE

Azure Data Lake

structured - csv, txt
semi - Json, xml
semi – PDF, Office
unstructured – image, voice, video
Web/social – stream, Twitter

Enterprise / Divisional Lake

PREP / TRAIN / MODEL / ANALYSE

Azure Databricks

kubernetes

Analytics

Snowflake

SQL DW

Warehouse

AutomatedML
Deep Learning

Keras

Flask

ONNX

Scala

Spark ML

python

Machine Learning / Model Execution

SERVE

Business / custom apps

Power BI

Q & A (English Query)

ML

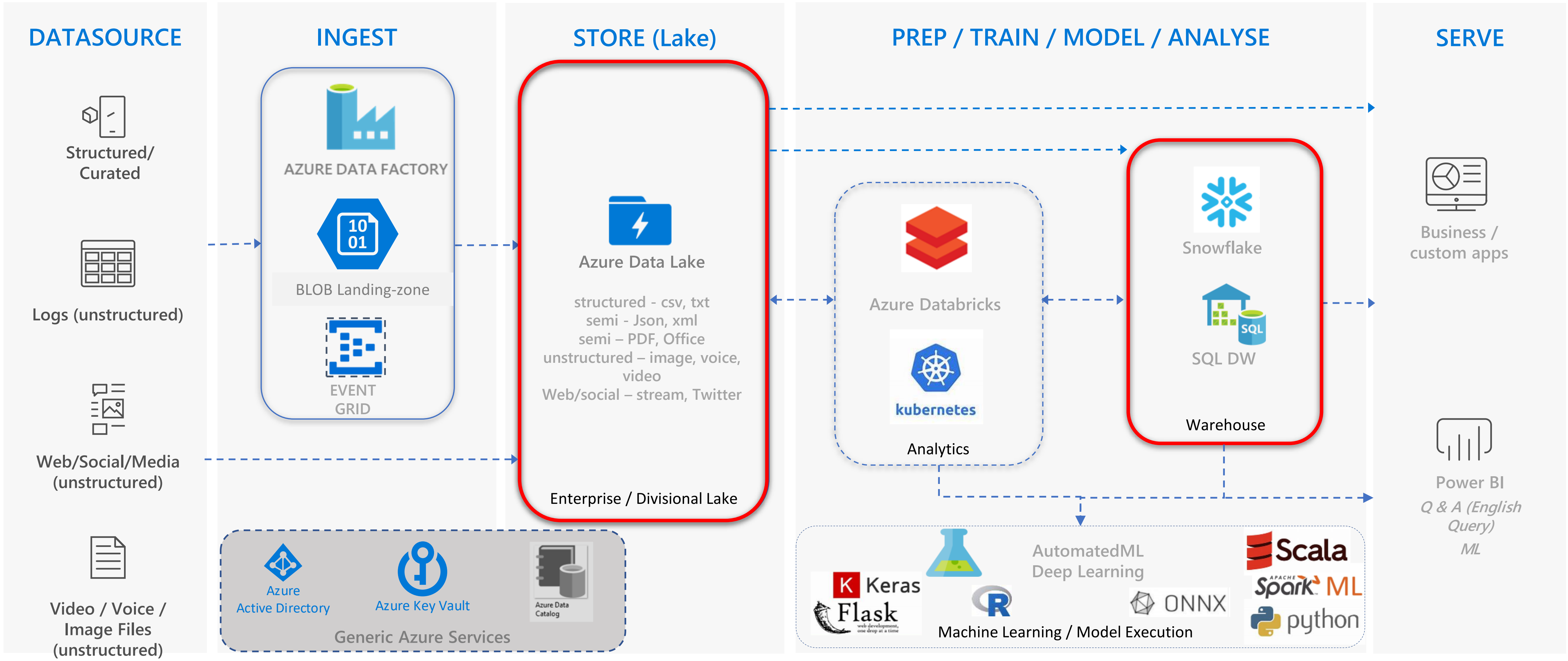
Azure Active Directory

Azure Key Vault

Azure Data Catalog

Generic Azure Services

Data as a Service - Theoretical



DATASOURCE

Structured/ Curated

Logs (unstructured)

Web/Social/Media (unstructured)

Video / Voice / Image Files (unstructured)

INGEST

AZURE DATA FACTORY

BLOB Landing-zone

EVENT GRID

Azure Active Directory

Azure Key Vault

Azure Data Catalog

Generic Azure Services

STORE (Lake)

Azure Data Lake

structured - csv, txt
semi - Json, xml
semi – PDF, Office
unstructured – image, voice, video
Web/social – stream, Twitter

Enterprise / Divisional Lake

PREP / TRAIN / MODEL / ANALYSE

Azure Databricks

kubernetes

Analytics

Snowflake

SQL DW

Warehouse

AutomatedML
Deep Learning

Keras

Flask

R

ONNX

Scala

Spark ML

python

Machine Learning / Model Execution

SERVE

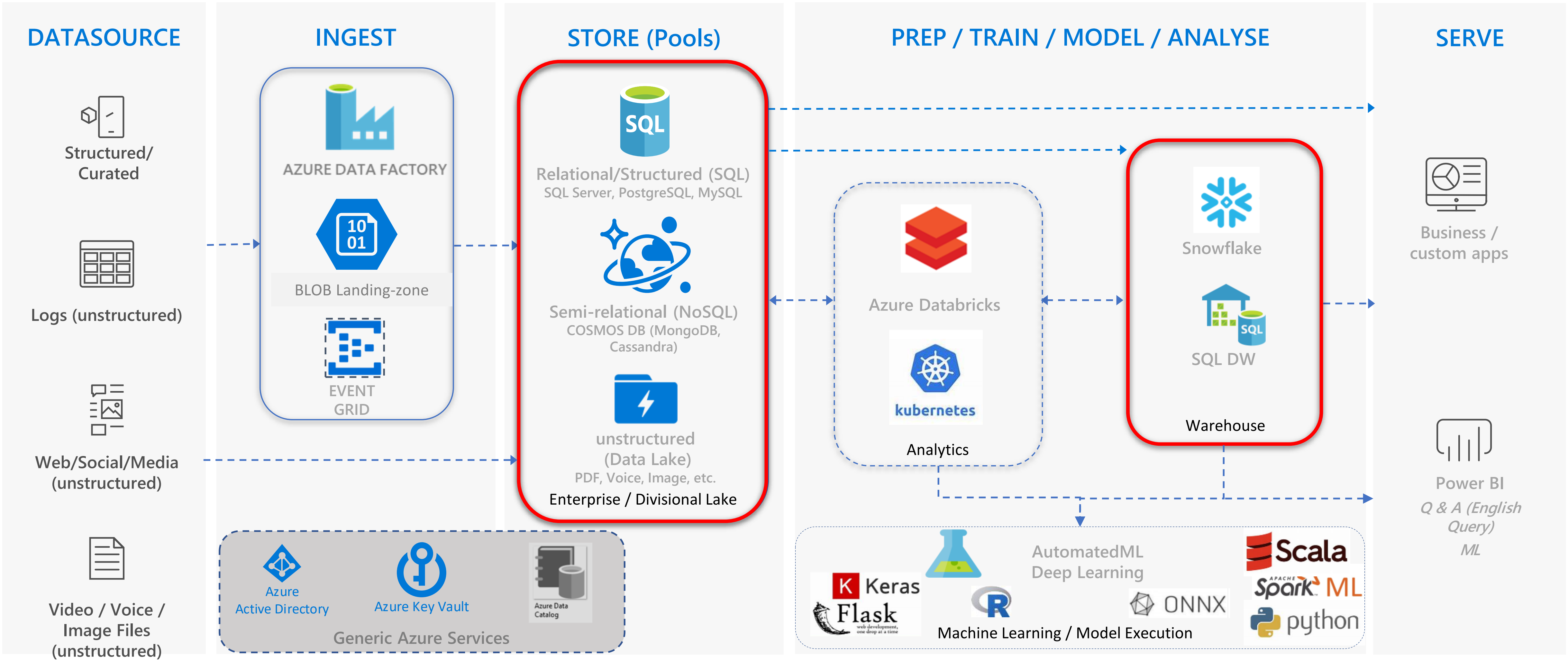
Business / custom apps

Power BI

Q & A (English Query)

ML

Data as a Service - Reality



DATASOURCE

Structured/ Curated

Logs (unstructured)

Web/Social/Media (unstructured)

Video / Voice / Image Files (unstructured)

INGEST

AZURE DATA FACTORY

BLOB Landing-zone

EVENT GRID

STORE (Pools)

Relational/Structured (SQL)
SQL Server, PostgreSQL, MySQL

Semi-relational (NoSQL)
COSMOS DB (MongoDB, Cassandra)

unstructured (Data Lake)
PDF, Voice, Image, etc.

Enterprise / Divisional Lake

PREP / TRAIN / MODEL / ANALYSE

Azure Databricks

kubernetes

Analytics

Snowflake

SQL DW

Warehouse

AutomatedML
Deep Learning

Keras

Flask
web development, one drop at a time

R

ONNX

Scala

Spark ML

python

Machine Learning / Model Execution

SERVE

Business / custom apps

Power BI

Q & A (English Query)

ML

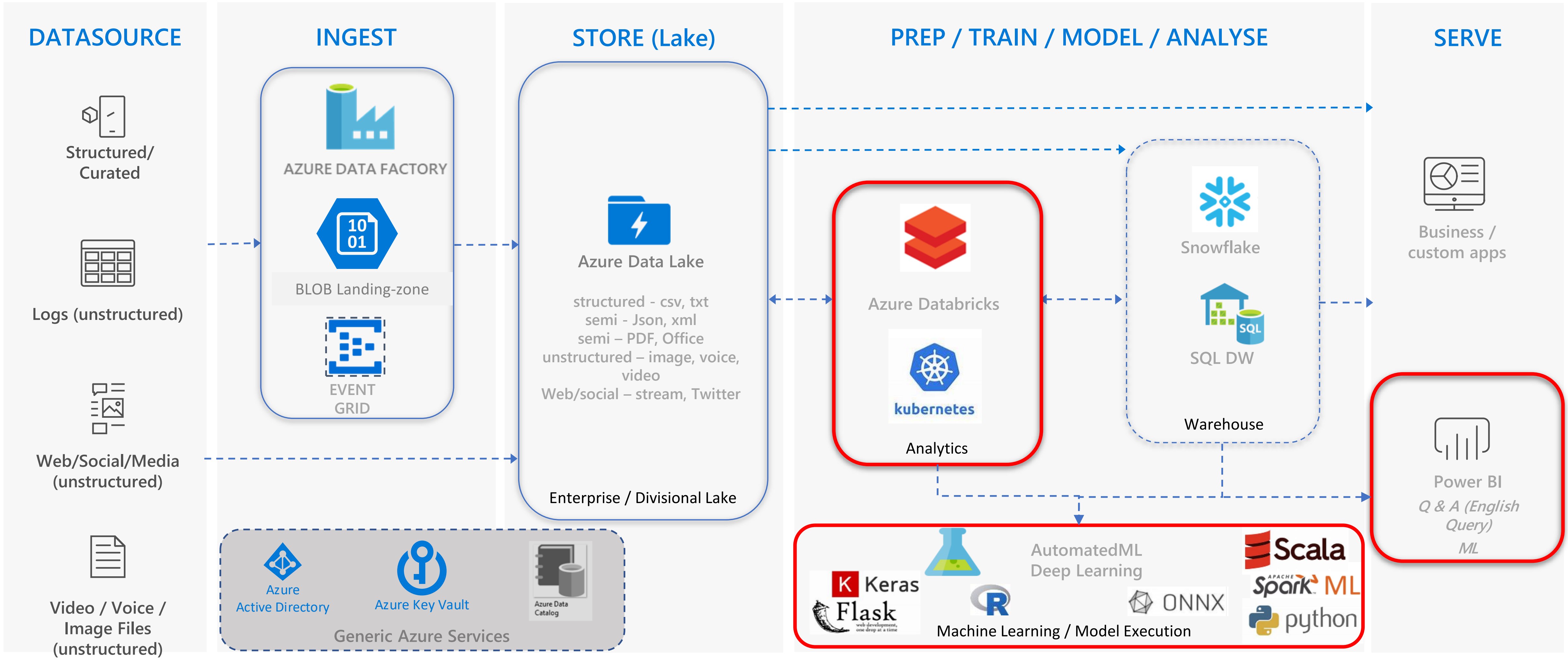
Azure Active Directory

Azure Key Vault

Azure Data Catalog

Generic Azure Services

Data Science / Analytics as a Service



DATASOURCE

Structured/ Curated

Logs (unstructured)

Web/Social/Media (unstructured)

Video / Voice / Image Files (unstructured)

INGEST

AZURE DATA FACTORY

BLOB Landing-zone

EVENT GRID

STORE (Lake)

Azure Data Lake

structured - csv, txt
semi - Json, xml
semi – PDF, Office
unstructured – image, voice, video
Web/social – stream, Twitter

Enterprise / Divisional Lake

Azure Active Directory

Azure Key Vault

Azure Data Catalog

Generic Azure Services

PREP / TRAIN / MODEL / ANALYSE

Azure Databricks

kubernetes

Analytics

Snowflake

SQL DW

Warehouse

AutomatedML
Deep Learning

Keras

Flask
web development, one drop at a time

R

ONNX

Scala

Spark ML

python

Machine Learning / Model Execution

SERVE

Business / custom apps

Power BI

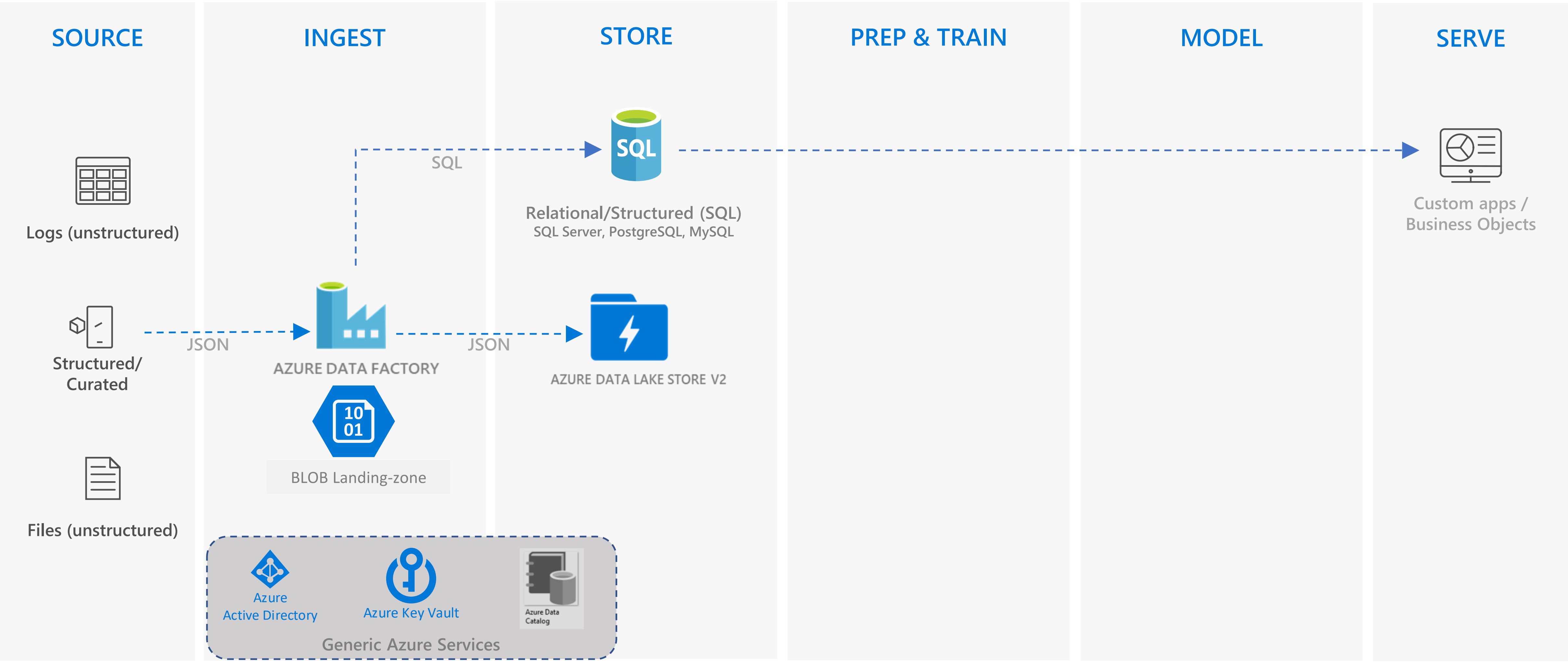
Q & A (English Query)

ML

Microsoft

Demos: Straight Through Processing (the real world)

Real World – Pattern 1



SOURCE

INGEST

STORE

PREP & TRAIN

MODEL

SERVE

Logs (unstructured)

Structured/ Curated

Files (unstructured)

JSON

AZURE DATA FACTORY

BLOB Landing-zone

SQL

JSON

SQL

Relational/Structured (SQL)
SQL Server, PostgreSQL, MySQL

AZURE DATA LAKE STORE V2

Custom apps / Business Objects

Azure Active Directory

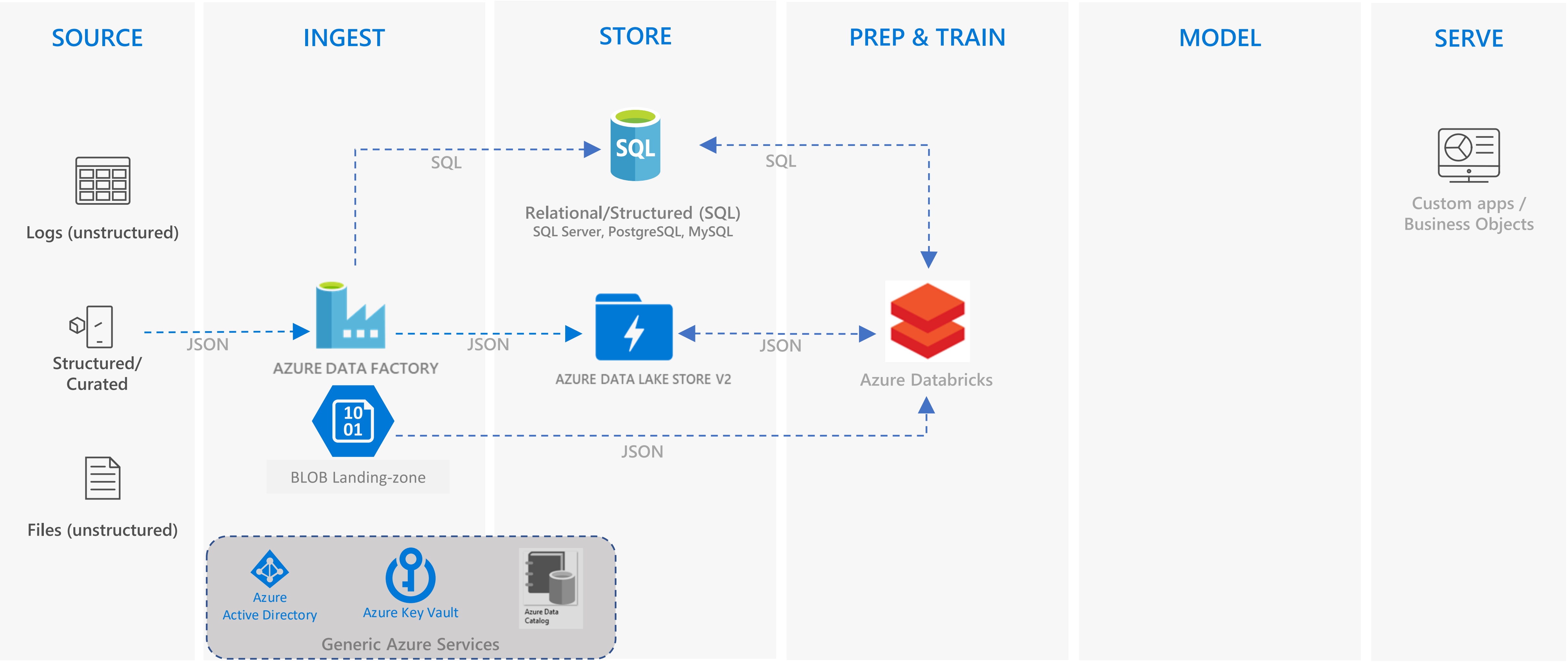
Azure Key Vault

Azure Data Catalog

Generic Azure Services

Microsoft

Real World – Pattern 2



SOURCE

INGEST

STORE

PREP & TRAIN

MODEL

SERVE

Logs (unstructured)

Structured/ Curated

Files (unstructured)

JSON

AZURE DATA FACTORY

SQL

BLOB Landing-zone

Relational/Structured (SQL)
SQL Server, PostgreSQL, MySQL

JSON

AZURE DATA LAKE STORE V2

SQL

JSON

Azure Databricks

JSON

Custom apps / Business Objects

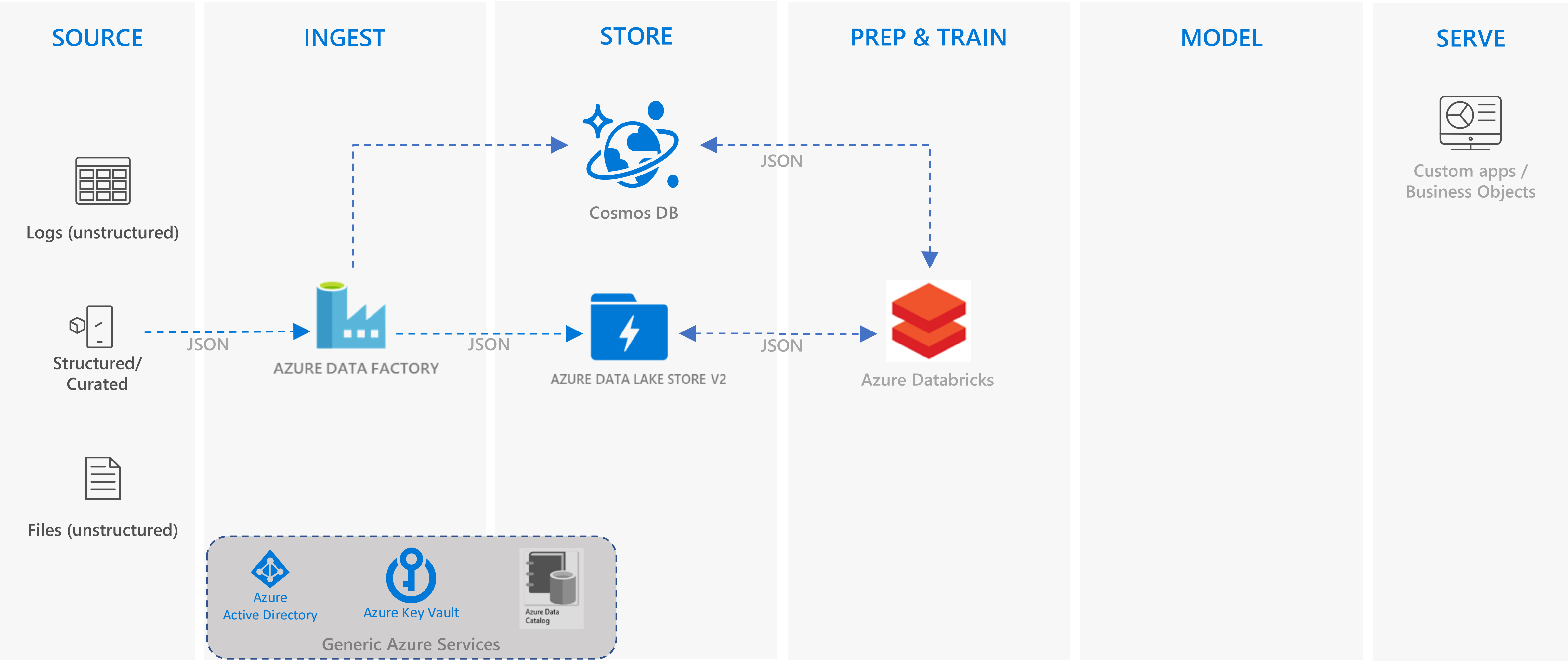
Azure Active Directory

Azure Key Vault

Azure Data Catalog

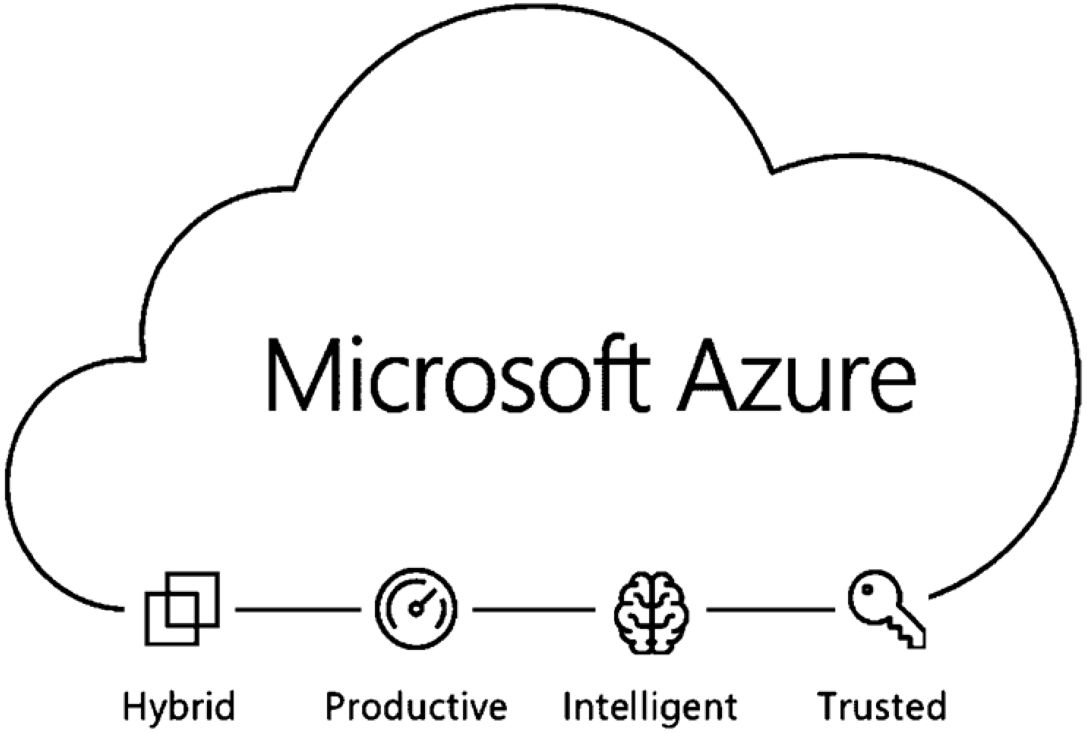
Generic Azure Services

Real World – Pattern 3



Microsoft

Q & A



Appendices

Use-case: data

Ability to query across multiple entity types with a single network request.

For example, we have two types of documents: cat and person.

```
{
   "id": "Andrew",
   "type": "Person",
   "familyId": "Liu",
   "worksOn": "Azure Cosmos DB"
}
```

```
{
   "id": "Ralph",
   "type": "Cat",
   "familyId": "Liu",
   "fur": {
         "length": "short",
         "color": "brown"
   }
}
```

Ability to query across multiple entity types with a single network request.

For example, we have two types of documents: cat and person.

```
{
   "id": "Andrew",
   "type": "Person",
   "familyId": "Liu",
   "worksOn": "Azure Cosmos DB"
}
```

```
{
   "id": "Ralph",
   "type": "Cat",
   "familyId": "Liu",
   "fur": {
         "length": "short",
         "color": "brown"
   }
}
```

We can query both types of documents without needing a JOIN simply by running a query without a filter on type:

```
SELECT * FROM c WHERE c.familyId = "Liu"
```

Ability to query across multiple entity types with a single network request.

For example, we have two types of documents: cat and person.

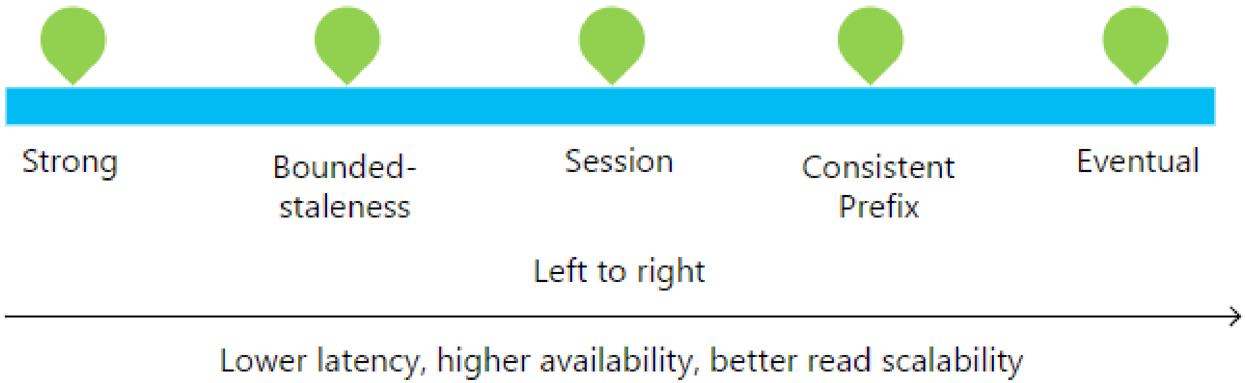
```
{
   "id": "Andrew",
   "type": "Person",
   "familyId": "Liu",
   "worksOn": "Azure Cosmos DB"
}
```

```
{
   "id": "Ralph",
   "type": "Cat",
   "familyId": "Liu",
   "fur": {
         "length": "short",
         "color": "brown"
   }
}
```

If we wanted to filter on type = “Person”, we can simply add a filter on type to our query:

SELECT * FROM c WHERE c.familyId = "Liu" **AND c.type = "Person"**

Use-case: application & services



Strong

Bounded-staleness

Session

Consistent Prefix

Eventual

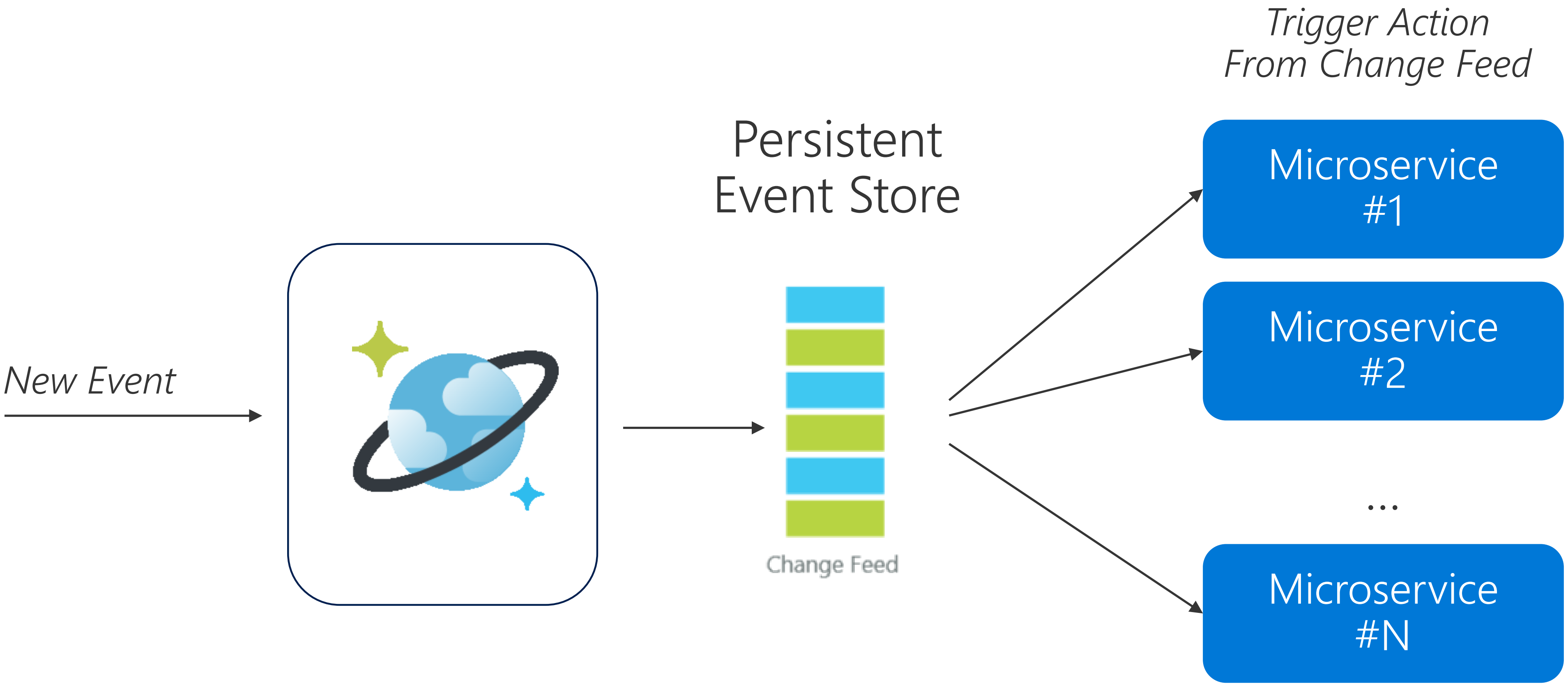
Left to right

Lower latency, higher availability, better read scalability

Consistency Level	**Guarantees**
Strong	Linearizability (once operation is complete, it will be visible to all)
Bounded Staleness	Consistent Prefix. Reads lag behind writes by at most k prefixes or t interval Similar properties to strong consistency (except within staleness window), while preserving 99.99% availability and low latency.
Session	Consistent Prefix. Within a session: monotonic reads, monotonic writes, read-your-writes, write-follows-reads Predictable consistency for a session, high read throughput + low latency
Consistent Prefix	Reads will never see out of order writes (no gaps).
Eventual	Potential for out of order reads. Lowest cost for reads of all consistency levels.

Event Sourcing for Microservices

sqlbits



Trigger Action From Change Feed

Persistent Event Store

New Event

Microservice #1

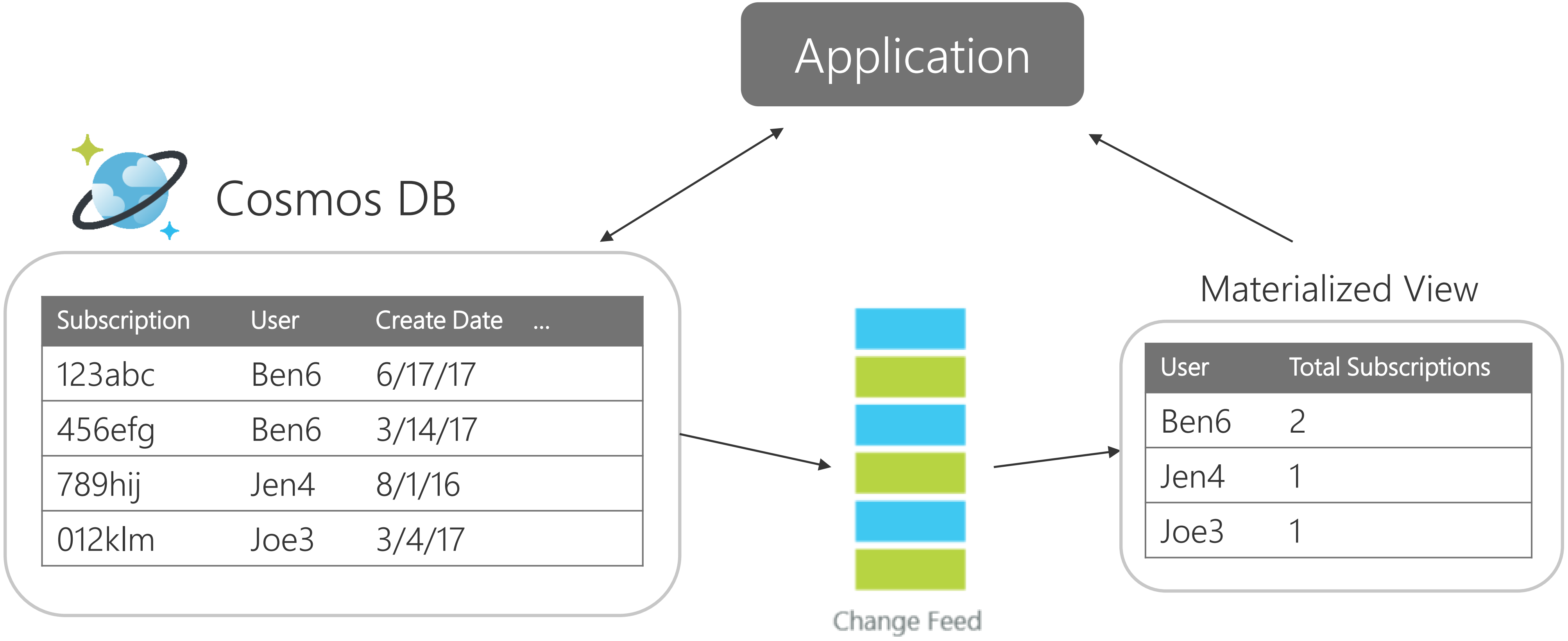
Microservice #2

...

Microservice #N

Microsoft

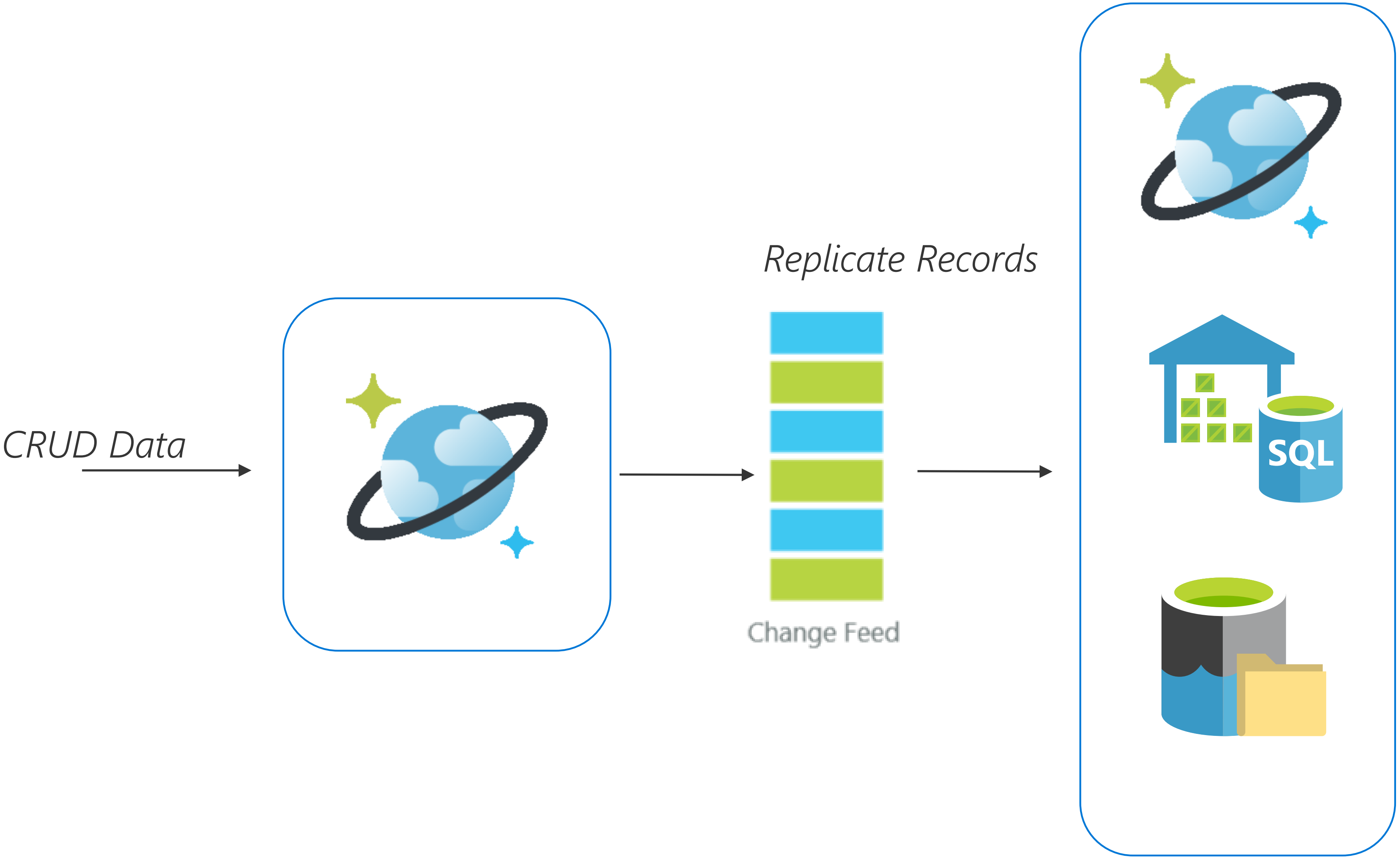
Materializing Views



Subscription	User	Create Date	...
123abc	Ben6	6/17/17	
456efg	Ben6	3/14/17	
789hij	Jen4	8/1/16	
012klm	Joe3	3/4/17	

User	Total Subscriptions
Ben6	2
Jen4	1
Joe3	1

Replicating Data



CRUD Data

Replicate Records

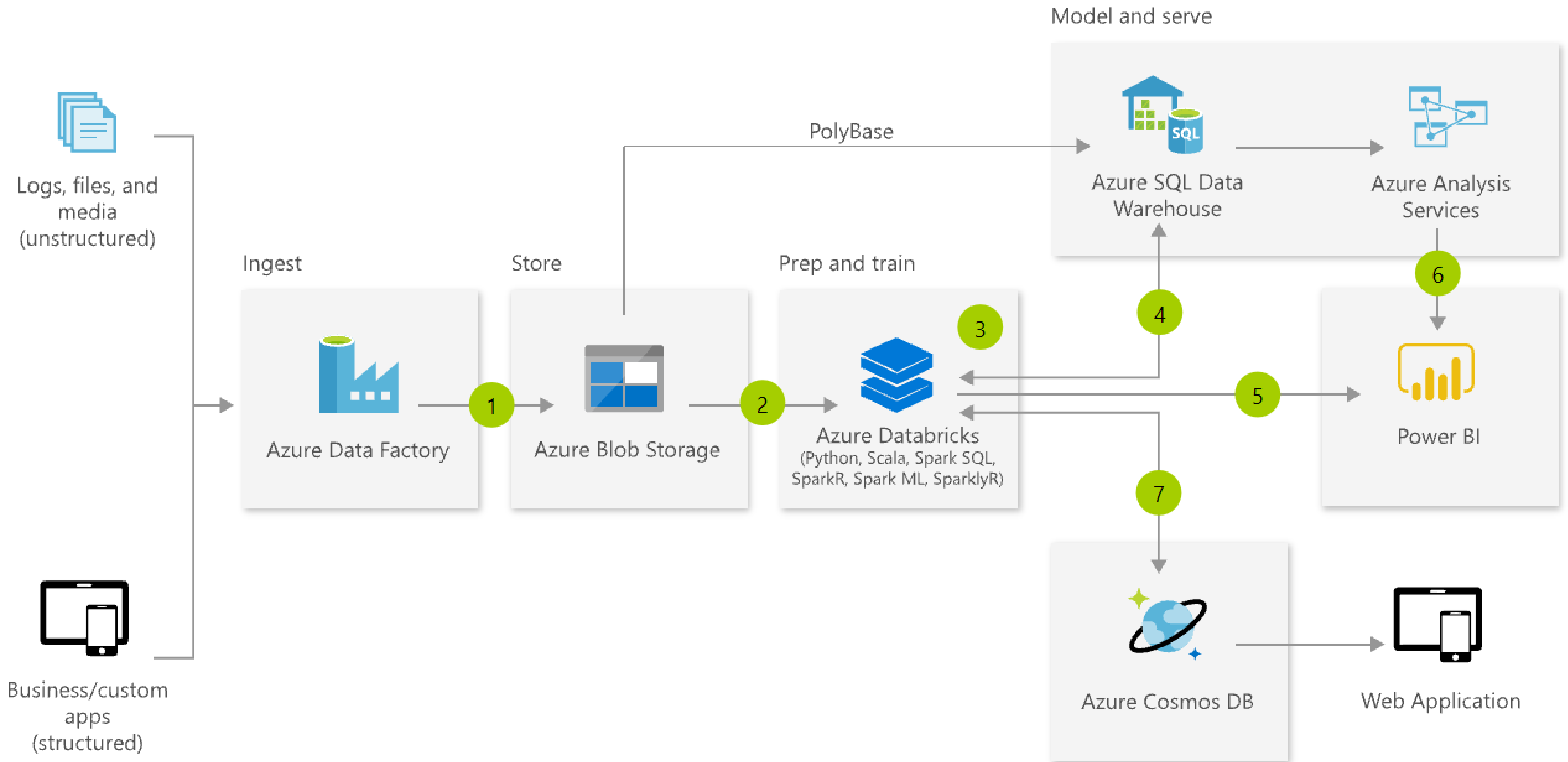
Change Feed

Secondary Datastore (e.g. archive)

Common architectures

Modern DW for big data

sqlbits

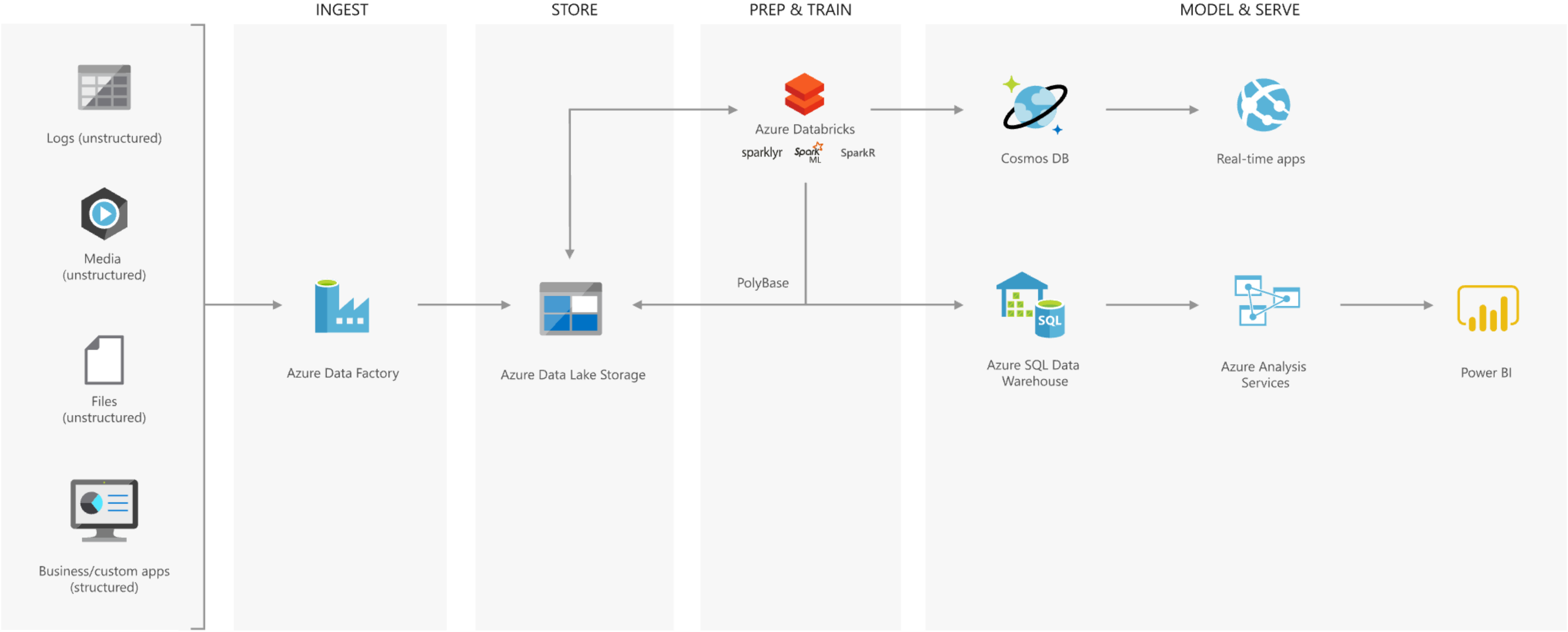


AZURE DATA FACTORY ORCHESTRATES DATA PIPELINE ACTIVITY WORKFLOW & SCHEDULING

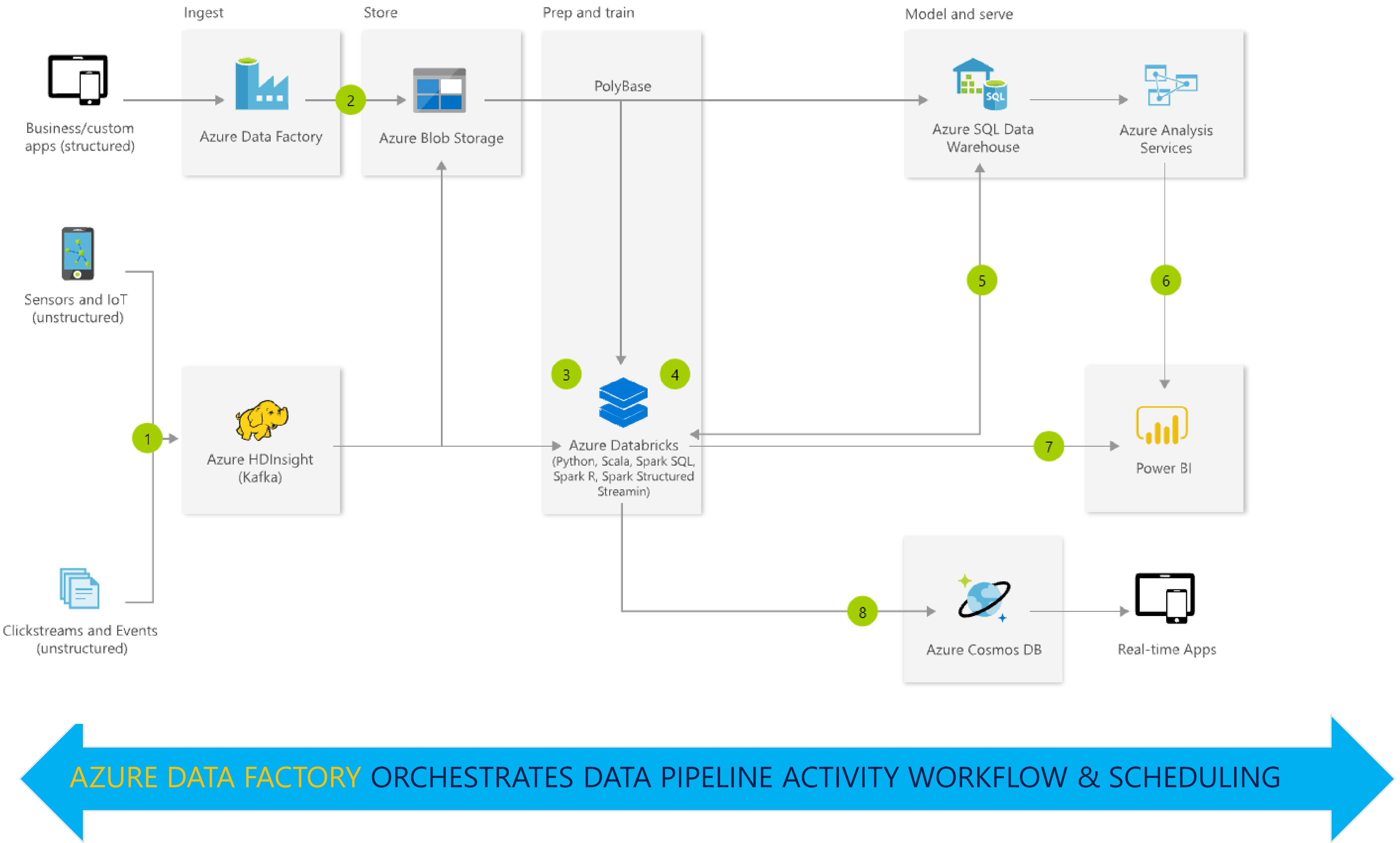
Microsoft

Modern DW for big data continued

Advanced analytics on big data



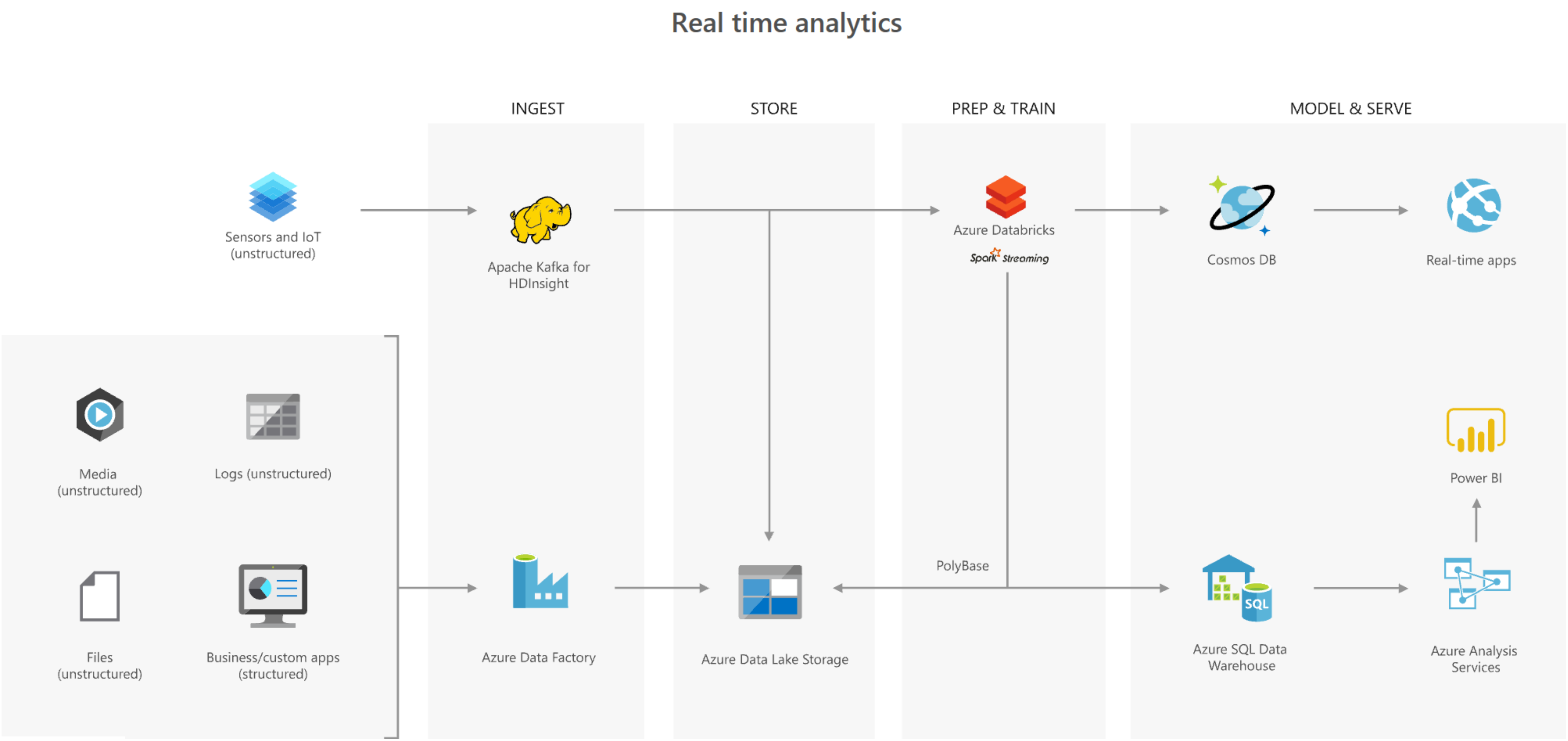
Real-time analytics



Ingest
Store
Prep and train
Model and serve
Business/custom apps (structured)
Azure Data Factory
2
Azure Blob Storage
PolyBase
Azure SQL Data Warehouse
Azure Analysis Services
Sensors and IoT (unstructured)
5
6
1
Azure HDInsight (Kafka)
3
4
Azure Databricks (Python, Scala, Spark SQL, Spark R, Spark Structured Streamin)
7
Power BI
Clickstreams and Events (unstructured)
8
Azure Cosmos DB
Real-time Apps
AZURE DATA FACTORY ORCHESTRATES DATA PIPELINE ACTIVITY WORKFLOW & SCHEDULING

Real-time analytics continued

sqlbits



Microsoft