Azure Data Platform Overview

James Serra
Big Data Evangelist
Microsoft
jamesserra3@gmail.com
Blog: JamesSerra.com

About Me

- Microsoft, Big Data Evangelist
- In IT for 30 years, worked on many BI and DW projects
- Worked as desktop/web/database developer, DBA, BI and DW architect and developer, MDM architect, PDW/APS developer
- Been perm employee, contractor, consultant, business owner
- Presenter at PASS Business Analytics Conference, PASS Summit, Enterprise Data World conference
- Certifications: MCSE: Data Platform, Business Intelligence; MS: Architecting Microsoft Azure Solutions, Design and Implement Big Data Analytics Solutions, Design and Implement Cloud Data Platform Solutions
- Blog at JamesSerra.com
- Former SQL Server MVP
- Author of book “Reporting with Microsoft SQL Server 2012”

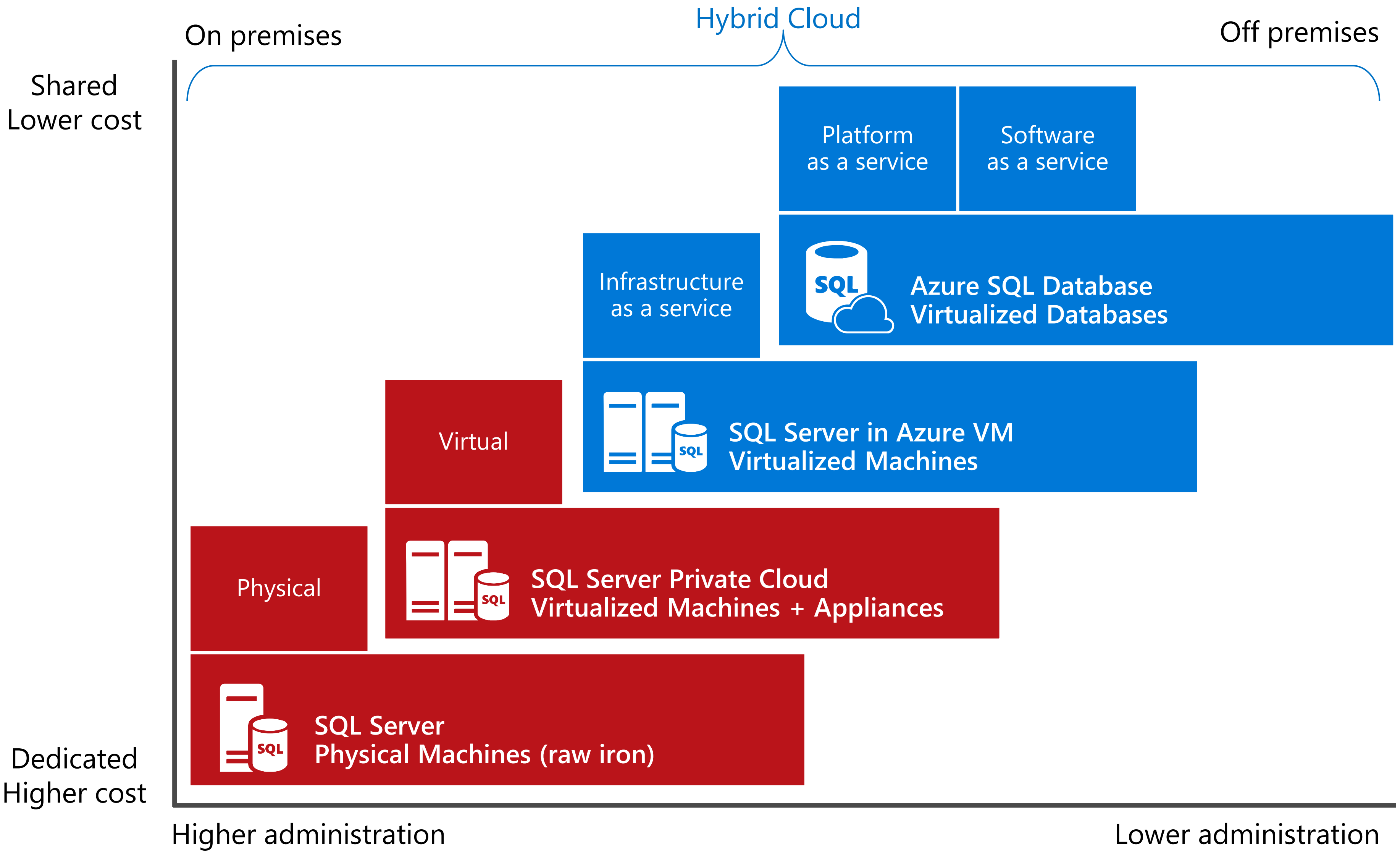
MVP Microsoft Most Valuable Professional

I tried to understand the Microsoft data platform on my own…

And felt like I was body slammed by Randy Savage:

Let’s prevent that from happening…

Data platform continuum



Hybrid Cloud

On premises

Off premises

Shared
Lower cost

Dedicated
Higher cost

Higher administration

Lower administration

Platform
as a service

Software
as a service

Infrastructure
as a service

Azure SQL Database
Virtualized Databases

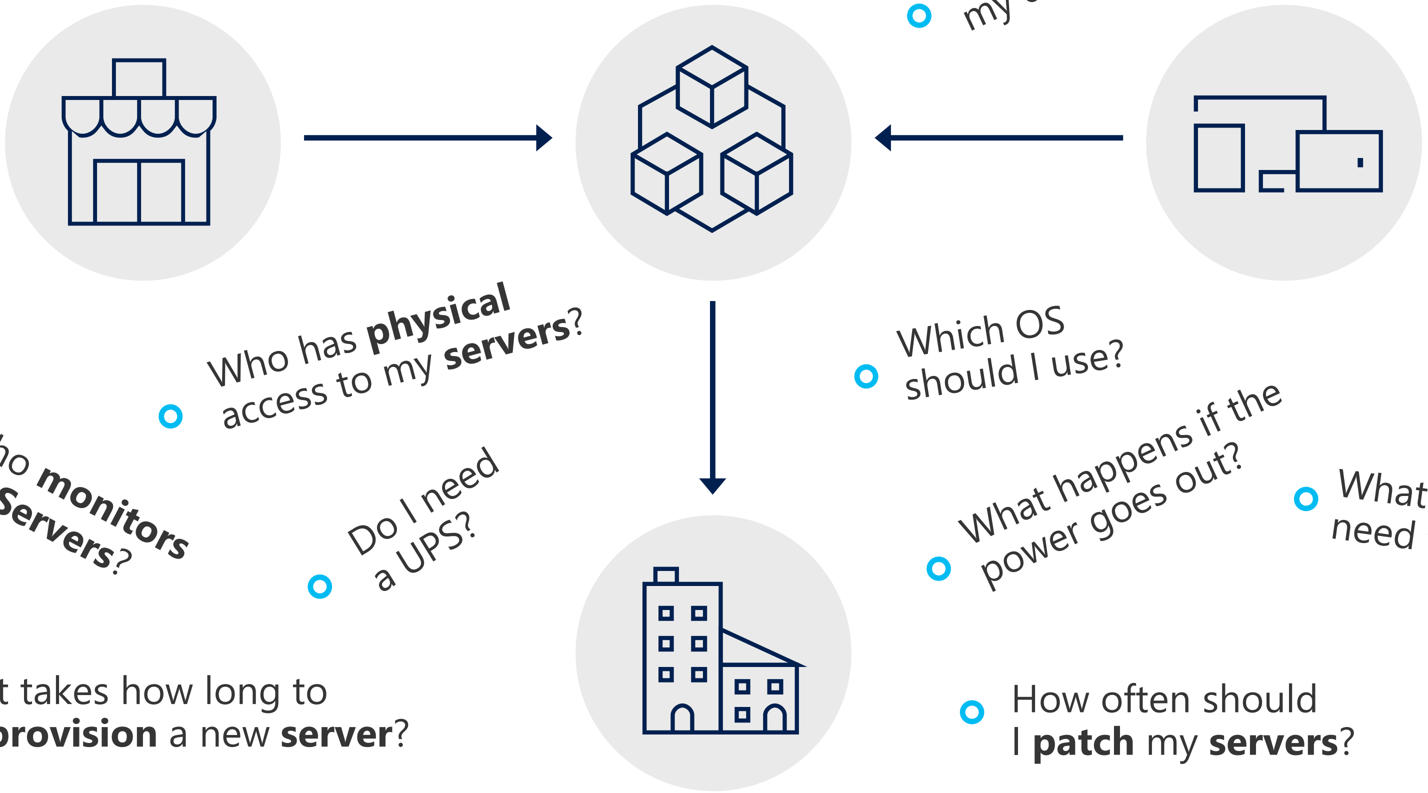
Virtual

SQL Server in Azure VM
Virtualized Machines

Physical

SQL Server Private Cloud
Virtualized Machines + Appliances

SQL Server
Physical Machines (raw iron)



- What media should I use to keep backup?
- What is the right **size** of **servers** for my business needs?
- What happens in case of **server hardware** failure?
- Which packages should be on my **server**?
- What size of **servers** should I **buy**?
- Who **monitors** my **database**?
- How often should I backup my **server**?
- How can I **scale** my database?
- How can I increase **server** utilization?
- Who has **physical** access to my **servers**?
- Which OS should I use?
- Are my **servers** in a secure location?
- Do I need secondary network connection?
- Who **monitors** my **Servers**?
- Do I need a UPS?
- What happens if the power goes out?
- What storage I need to use?
- How many **servers** do I need?
- It takes how long to **provision** a new **server**?
- How often should I **patch** my **servers**?

On-Premises

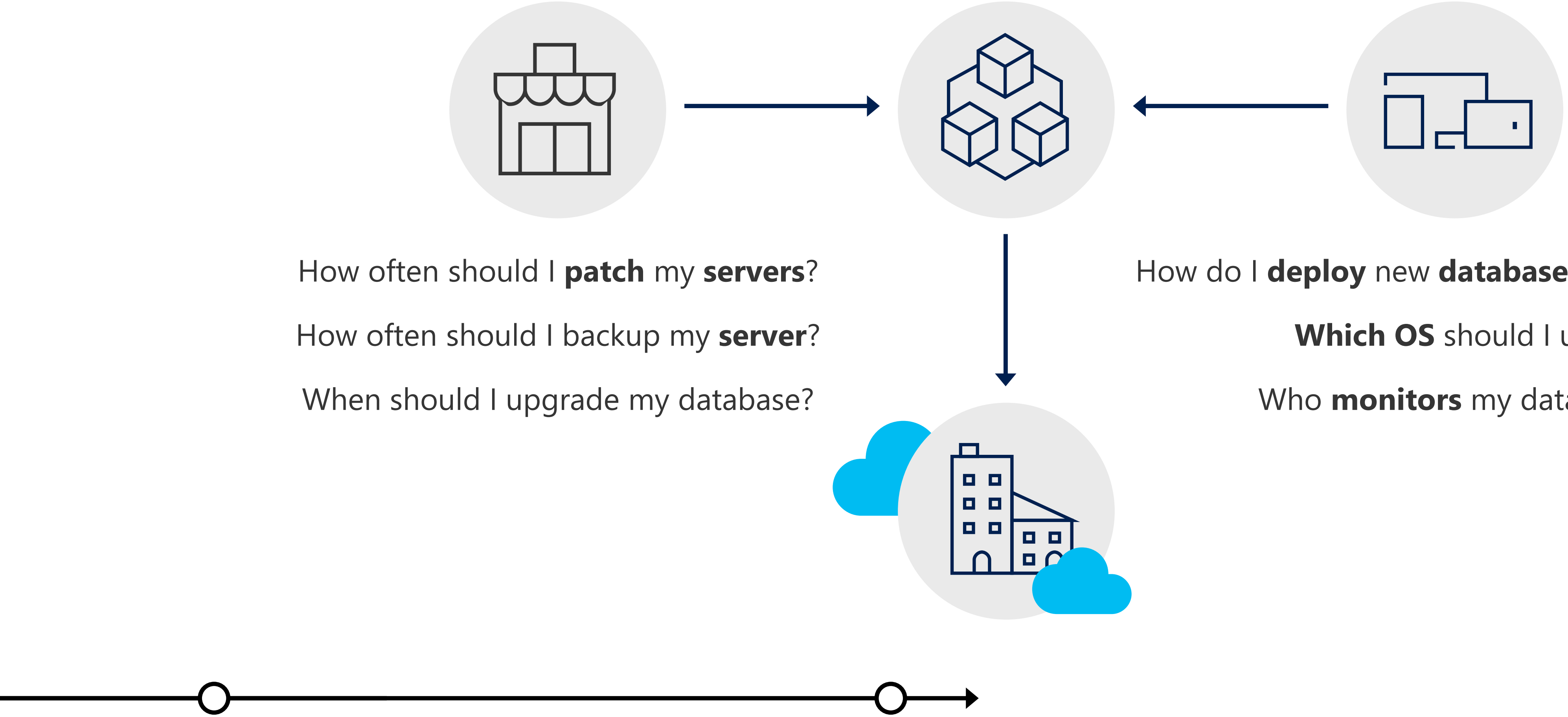
The “evolution” of data platforms

What is the right **size** of **servers** for my business needs?

How can I increase **server** utilization?

How many **servers** do I need?

How can I **scale** my database?



How often should I **patch** my **servers**?

How often should I backup my **server**?

When should I upgrade my database?

How do I **deploy** new **databases** to my **server**?

Which OS should I use?

Who **monitors** my database?

On-Premises

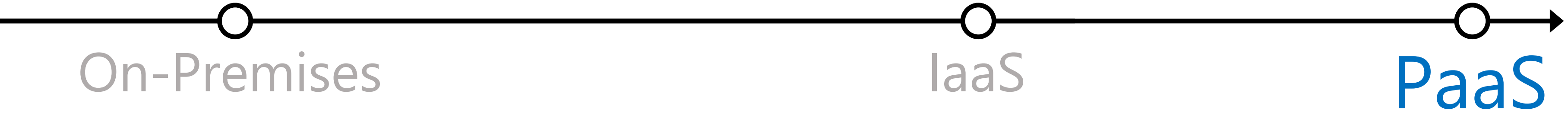
IaaS

The “evolution” of data platforms

What is the right **size** of **“servers”** for my business needs?

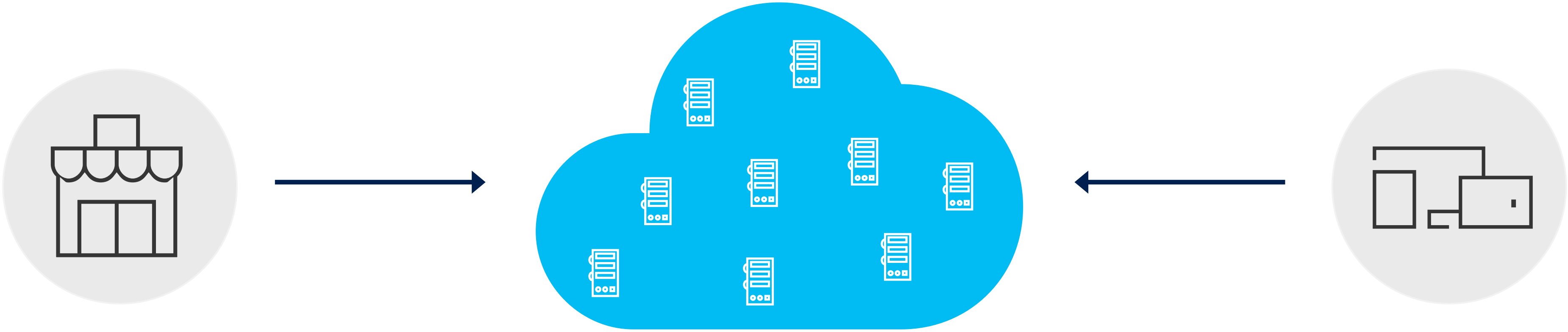
How can I increase **“server”** utilization?

How can I **scale** my database?



The “evolution” of data platforms

How do I **architect** my database?



On-Premises

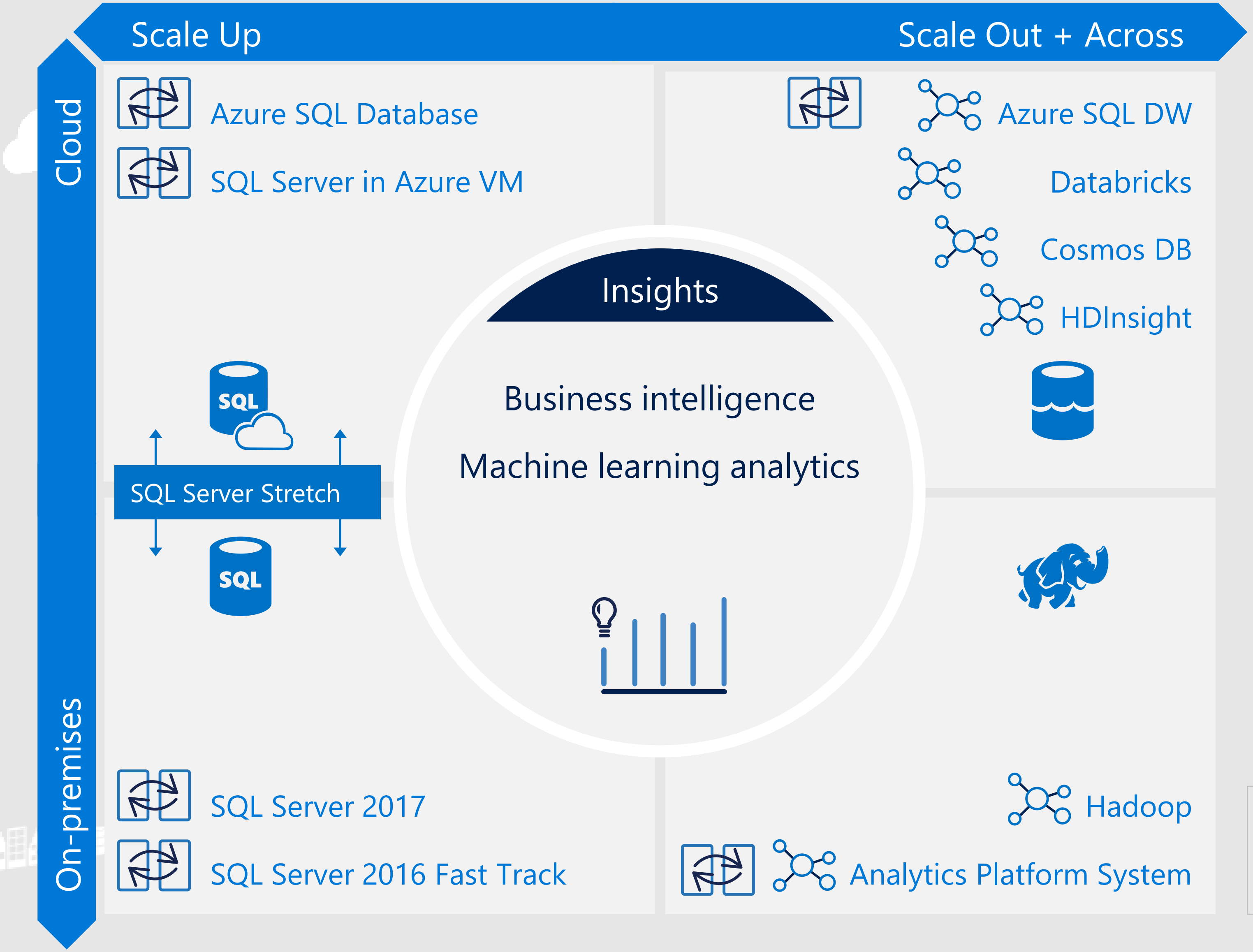
IaaS

PaaS

Pay per query

The "evolution" of data platforms

Microsoft Big Data Portfolio



Scale Up

Scale Out + Across

Cloud

On-premises

Azure SQL Database

SQL Server in Azure VM

SQL Server Stretch

SQL

SQL

Azure SQL DW

Databricks

Cosmos DB

HDInsight

Insights

Business intelligence

Machine learning analytics

SQL Server 2017

SQL Server 2016 Fast Track

Hadoop

Analytics Platform System

Microsoft has solutions covering and connecting all four quadrants – that’s why SQL Server is one of the most utilized databases in the world

Key

Relational

Non-relational

Microsoft Azure VMs

- VM hosted on Microsoft Azure Infrastructure (“IaaS”)
  - From Microsoft images (gallery) or your own images (custom)

    SQL 2008R2 / 2012 / 2014 / 2016 / 2017 Web / Standard / Enterprise

    Images refreshed with latest version, SP, CU
  - Windows Server 2008 R2 / 2012 R2 / 2016, Linux RHEL / Ubuntu
  - Fast provisioning (~10 minutes). Provision groups of servers with resource templates
  - Accessible via RDP and Powershell
  - Full compatibility with SQL Server “Box” software

- Pay per use
  - Per minute (only when running)
  - Cost depends on size and licensing
    - EA customers can use existing SQL licenses (BYOL)
  - Network: only outgoing (not incoming)
  - Storage: only used (not allocated)

- Elasticity
  - 1 core / 2 GB mem / 1 TB ← → 128 cores / 3.5 TB mem / 256 TB

Azure SQL Database

A relational **database-as-a-service (“PaaS”)**, fully managed by Microsoft.

For cloud-designed apps when **near-zero administration** and **enterprise-grade** capabilities are key.

Perfect for organizations looking to dramatically **increase the DB:IT ratio**.

Elastic scale & performance

Predictable performance levels

Programmatic scale-out

Dashboard views of DB metrics

Business continuity & data protection

Self-service restore

Disaster recovery

Compliance-enabled

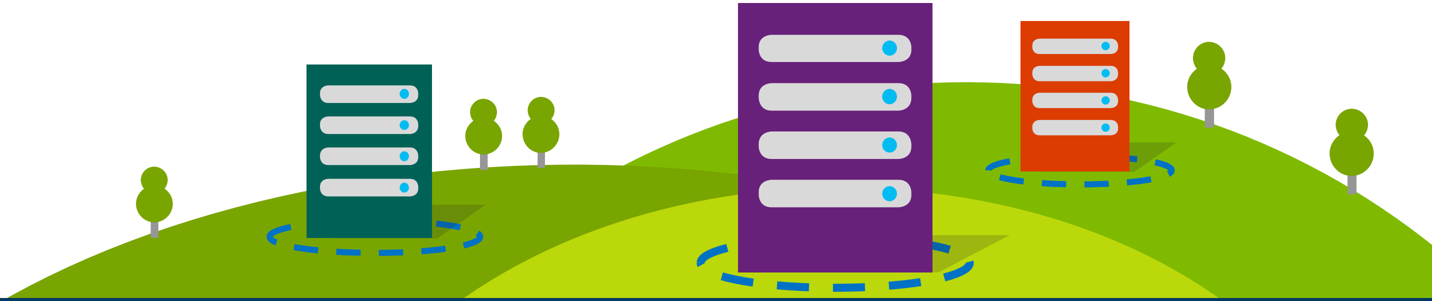
Familiar & self-managed

Familiar & compatible

Programmatic

Self-managed

Note: New features will be in SQL Database before SQL Server!



Azure SQL Database Managed Instance

Managed Instance
Instance scoped programming model with high compatibility to on-premises databases

Single
Standalone managed database best for predictable and stable workloads

Elastic pool
Shared resource model best for greater efficiency through multi-tenancy

Best for modernization at scale with low cost and effort

Easy migration: nearly 100% like SQL Server

Data migration

- Native backup/restore
- Log shipping (DMS)

Security

- TDE
- SQL Audit
- Row level security
- Always Encrypted

Programmability

- Global temp tables
- Cross-database queries and transactions
- Linked servers
- CLR modules

Operational

- DMVs & XEvents
- Query Store
- SQL Agent
- DB Mail (external SMTP)

Scenario enablers

- Service Broker
- Change Data Capture
- Transactional Replication

Supports compatibility modes (SQL Server 2005+), Instance sizes up to 8TB

Azure SQL Database Hyperscale

Adapts on-demand to your workload's needs, auto-scaling up to 100TB per database.

Reliable and available

- Multiple levels of redundancy
- No single points of failure
- 99.99% availability

Scalable

- Auto-scales quickly up to 100TB
- Data size and cores scale independently
- No size of data operations

High performance

- Low latency, high throughput for large databases
- Snapshot-based backups – no impact on query performance
- Rapid database restore

Best for VLDB workloads with highly scalable storage and read-scale requirements, optimized for OLTP and HTAP workloads.

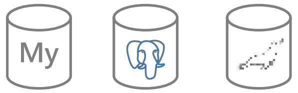
SQL Database vCore options

Programming Model	General Purpose	Business Critical	Hyperscale	Elastic Pools
Instance (MI)	GA, 8TB	GA, 4TB	Private Preview, 100TB	April private preview
Database (Single)	GA, 4TB	GA, 4TB	Public Preview, 100TB	GA

AZURE DATABASE SERVICES FOR MYSQL, POSTGRESQL, AND MARIADB

More choices and full integration into Azure's ecosystem and services

Managed community MySQL, PostgreSQL, and MariaDB



Languages and frameworks of your choice

Scale in seconds with built-in high availability

Secure and compliant

Industry-leading global reach

← Easy Lift and Shift →

← Enterprise Ready →

SMP vs MPP

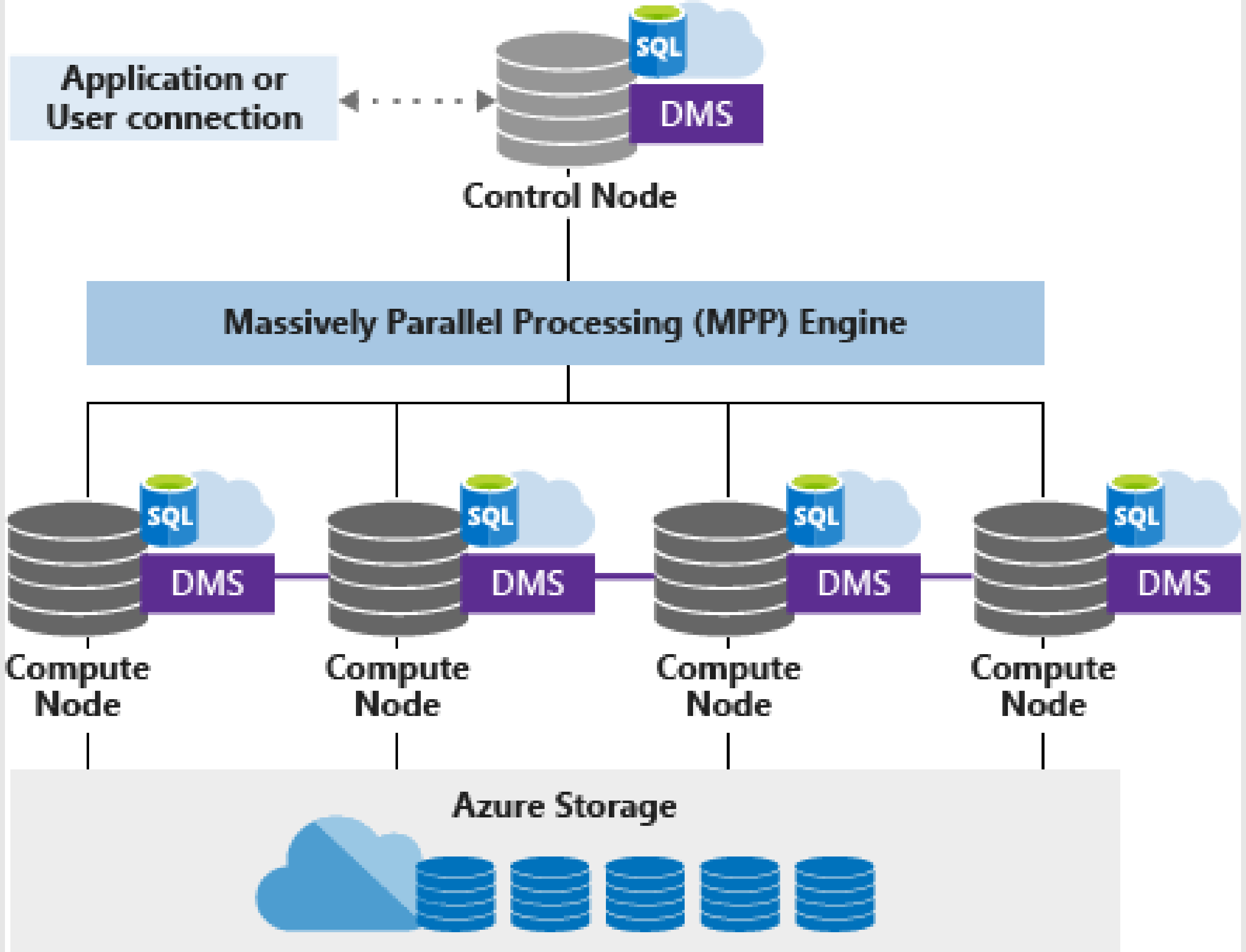
SMP - Symmetric Multiprocessing

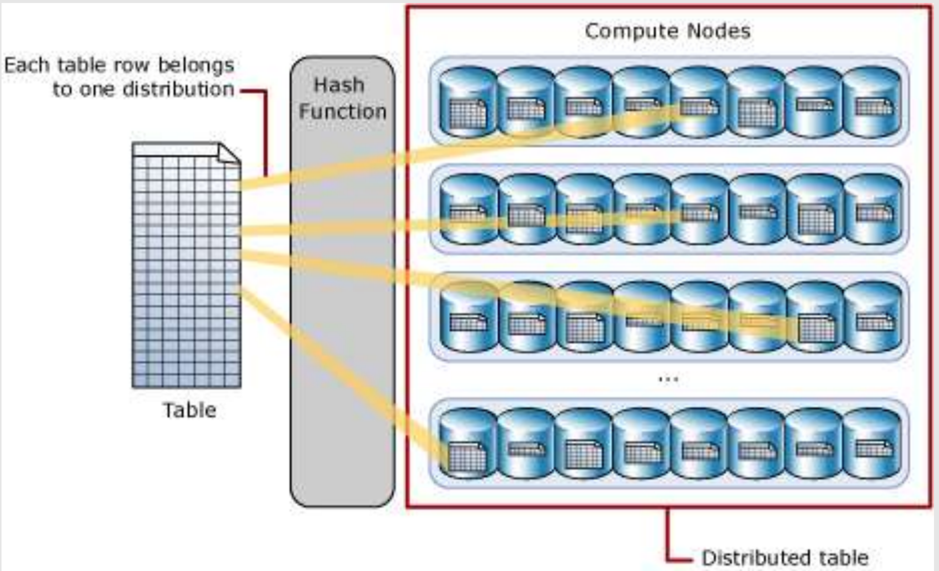
- Multiple CPUs used to complete individual processes simultaneously
- All CPUs share the same memory, disks, and network controllers (scale-up)
- SQL Server implementations traditionally have been SMP
- Mostly, the solution is housed on a shared storage

MPP - Massively Parallel Processing

- Uses many separate CPUs running in parallel to execute a single program
- Shared Nothing: Each CPU has its own memory and disk (scale-out)
- Segments communicate using high-speed network between nodes

SMP vs MPP


Application or User connection
SQL
DMS
Control Node
Massively Parallel Processing (MPP) Engine
SQL
DMS
Compute Node
SQL
DMS
Compute Node
SQL
DMS
Compute Node
SQL
DMS
Compute Node
Azure Storage


Compute Nodes
Each table row belongs to one distribution
Hash Function
Table
...
Distributed table

Azure SQL Data Warehouse

A relational **data warehouse-as-a-service**, fully managed by Microsoft.

Industries first **elastic** cloud data warehouse with **enterprise-grade** capabilities.

Support your **smallest to your largest** data storage needs while handling queries up to **100x faster**.

Elastic scale & performance

Scales to petabytes of data

Massively Parallel Processing

Instant-on compute scales in seconds

Query Relational / Non-Relational

Powered by the Cloud

Get started in minutes

Integrated with Azure ML, PowerBI & ADF

Enterprise Ready

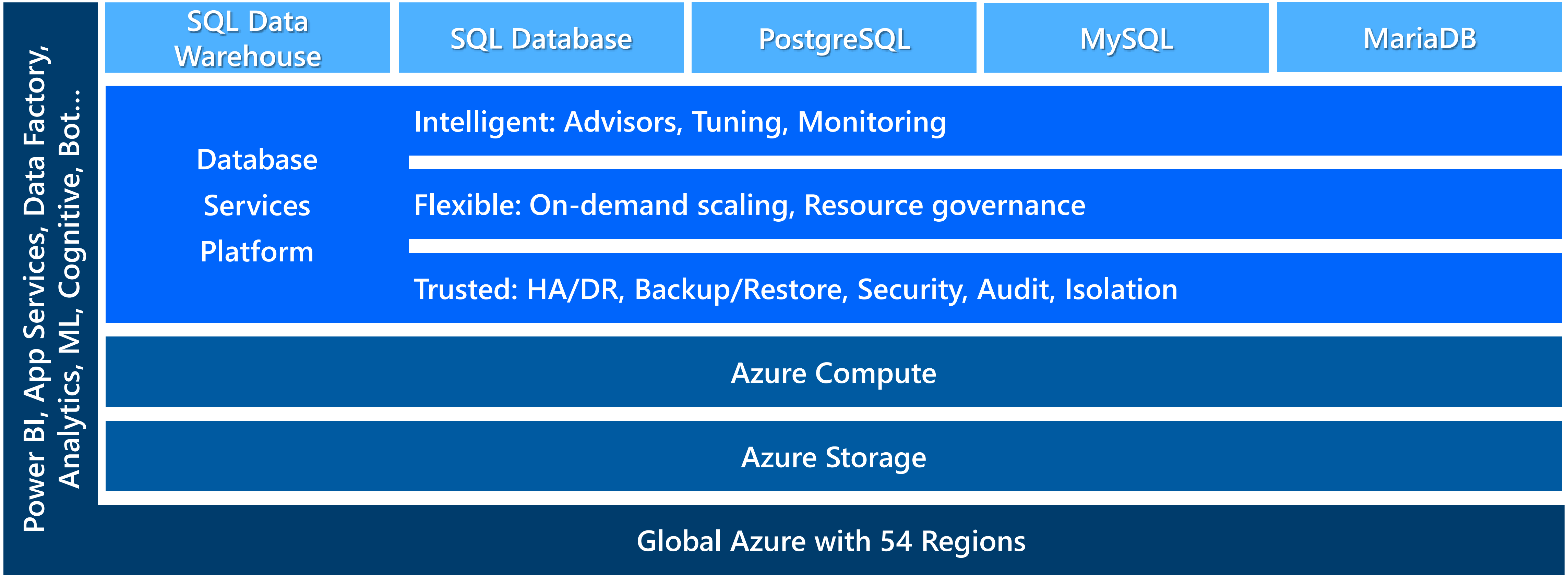
Market Leading Price & Performance

Simple billing compute & storage

Pay for what you need, when you need it with dynamic pause

Bring DW to the Cloud without rewriting

AZURE RELATIONAL DATABASE PLATFORM



SQL Data Warehouse

SQL Database

PostgreSQL

MySQL

MariaDB

Database Services Platform

Intelligent: Advisors, Tuning, Monitoring

Flexible: On-demand scaling, Resource governance

Trusted: HA/DR, Backup/Restore, Security, Audit, Isolation

Azure Compute

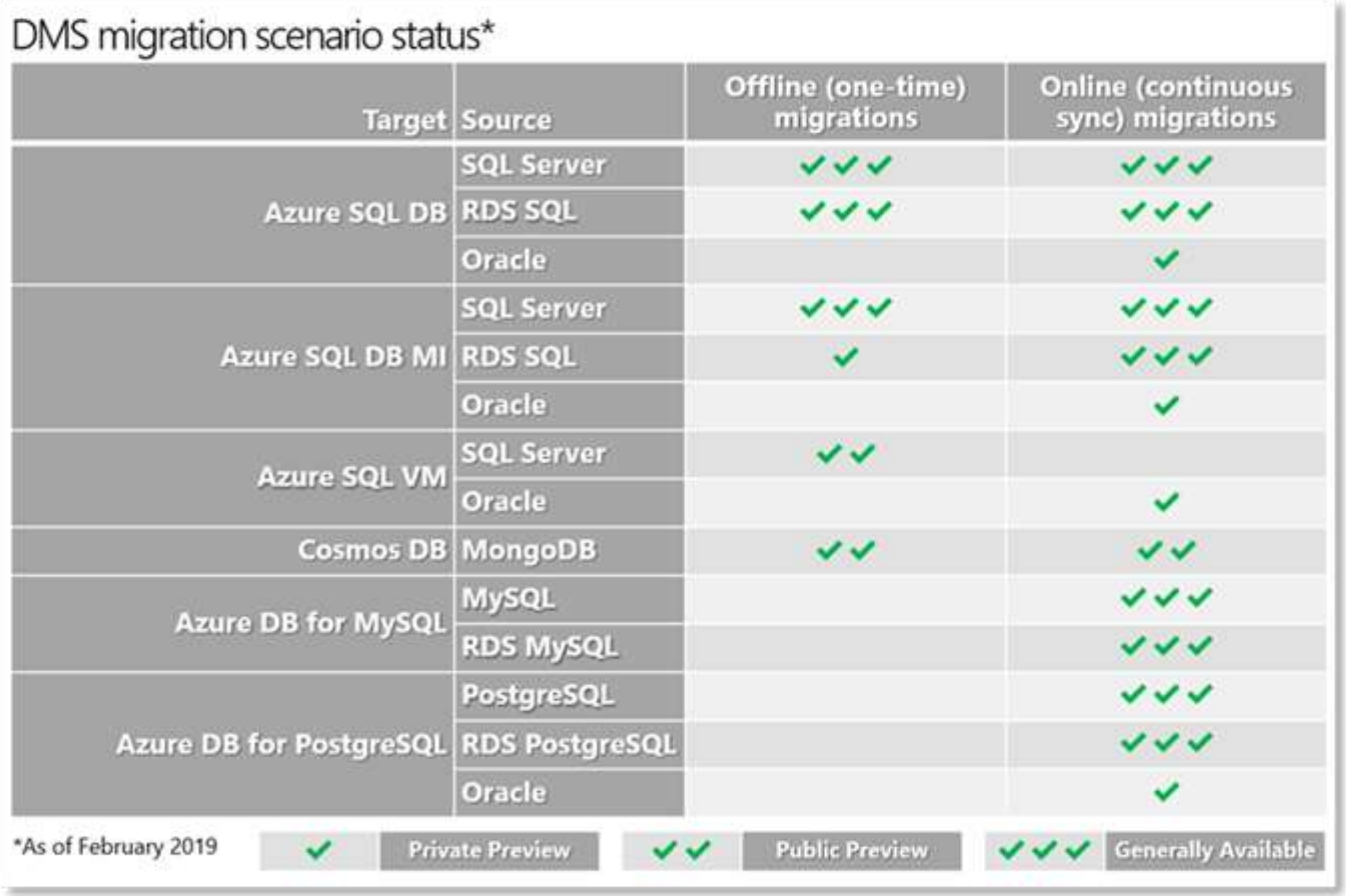
Azure Storage

Global Azure with 54 Regions

Power BI, App Services, Data Factory, Analytics, ML, Cognitive, Bot…

Azure Database Migration Service (DMS)

A seamless, end-to-end solution for moving on-premises SQL Server, Oracle, and other relational databases to the cloud.



DMS migration scenario status*

Target	Source	Offline (one-time) migrations	Online (continuous sync) migrations
Azure SQL DB	SQL Server	✓✓✓	✓✓✓
	RDS SQL	✓✓✓	✓✓✓
	Oracle		✓
Azure SQL DB MI	SQL Server	✓✓✓	✓✓✓
	RDS SQL	✓	✓✓✓
	Oracle		✓
Azure SQL VM	SQL Server	✓✓	
	Oracle		✓
Cosmos DB	MongoDB	✓✓	✓✓
Azure DB for MySQL	MySQL		✓✓✓
	RDS MySQL		✓✓✓
Azure DB for PostgreSQL	PostgreSQL		✓✓✓
	RDS PostgreSQL		✓✓✓
	Oracle		✓

*As of February 2019 ✓ Private Preview ✓✓ Public Preview ✓✓✓ Generally Available

Azure Database Migration Guide
https://datamigration.microsoft.com/

Relational and non-relational defined

Relational databases (RDBMS, SQL Databases)

- Example: Microsoft SQL Server, Oracle Database, IBM DB2
- Mostly used in large enterprise scenarios
- Analytical RDBMS (OLAP, MPP) solutions are SQL DW, Redshift, Teradata, Netezza

Non-relational databases (NoSQL databases)

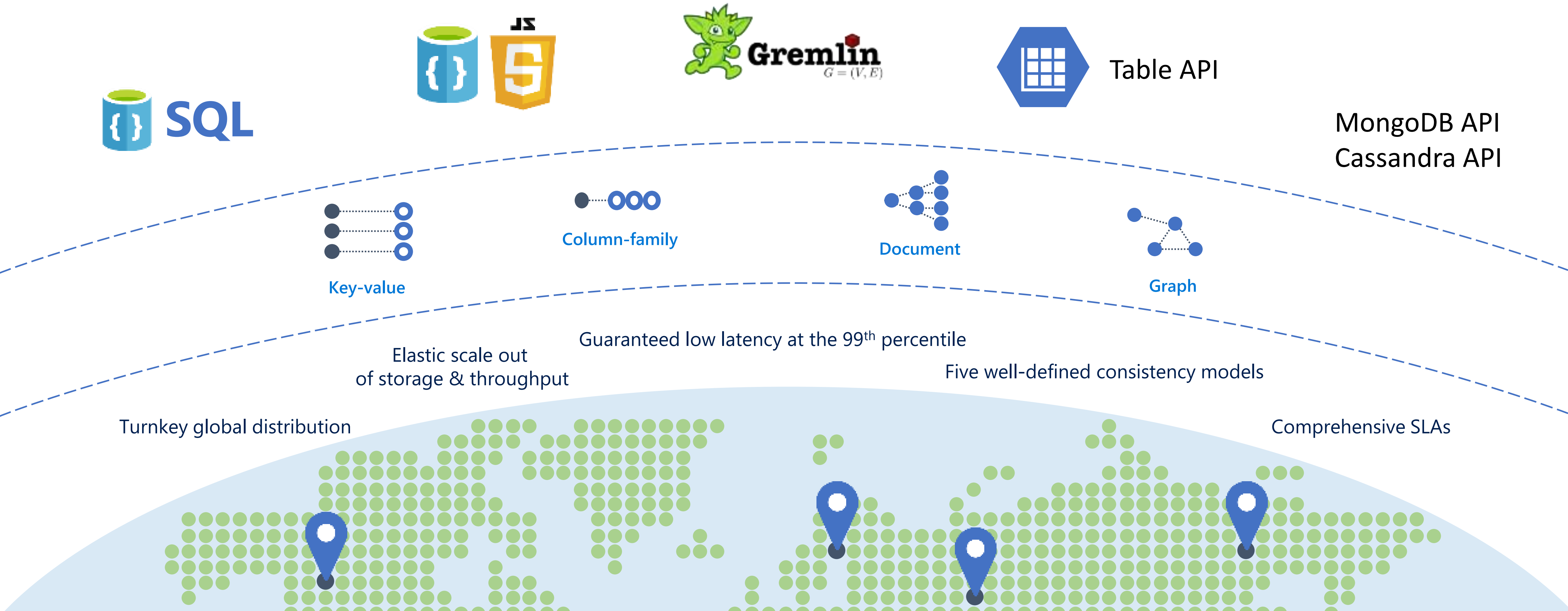
- Example: Azure Cosmos DB, MongoDB, Cassandra
- Four categories: Key-value stores, Wide-column stores, Document stores and Graph stores

Hadoop: Made up of Hadoop Distributed File System (HDFS), YARN and MapReduce (Ideal for data lake)

OLTP vs OLAP/DW
SMP vs MPP

Azure Cosmos DB

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



SQL

Table API

MongoDB API
Cassandra API

Key-value

Column-family

Document

Graph

Elastic scale out
of storage & throughput

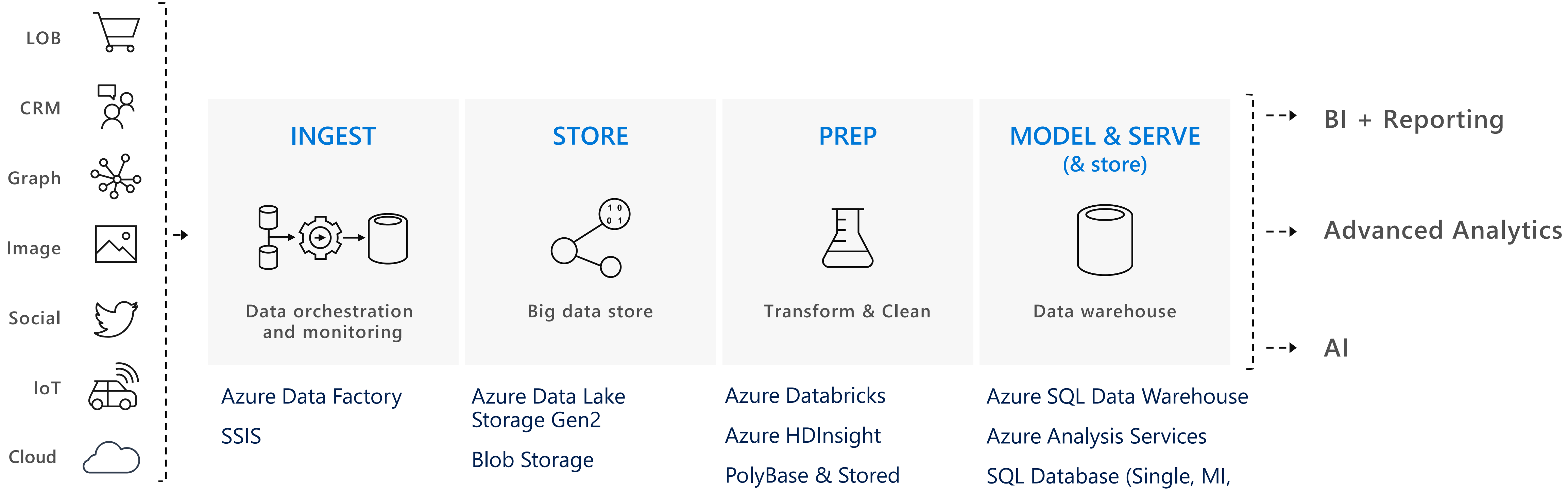
Guaranteed low latency at the 99th percentile

Five well-defined consistency models

Turnkey global distribution

Comprehensive SLAs

Modern Data Warehouse (possible products by four areas)



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 Gen2

Blob Storage

Azure Data Lake Storage Gen1

SQL Server 2019 Big Data Cluster

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)

SQL Server in a VM

Cosmos DB

Power BI Aggregations

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

ADLS Gen2: Convergence of two Storage Services

Blob Storage

General Purpose Object Storage

Large partner ecosystem
Global scale – All 50 regions
Durability options
Tiered - Hot/Cool/Archive
Cost Efficient

Data Lake Store

Optimized for Big Data analytics

Built for Hadoop
Hierarchical namespace
ACLs, AAD and RBAC
Performance tuned for big data
Very high scale capacity and throughput

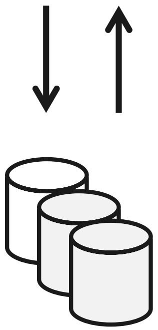
Azure Data Lake Storage Gen2

The best of Blobs and ADLS

Large partner ecosystem
Global scale – All 50 regions
Durability options
Tiered - Hot/Cool/Archive
Cost Efficient

Built for Hadoop
Hierarchical namespace
ACLs, AAD and RBAC
Performance tuned for big data
Very high scale capacity and throughput

Azure Storage Replication Options



LRS

Multiple replicas across a datacenter

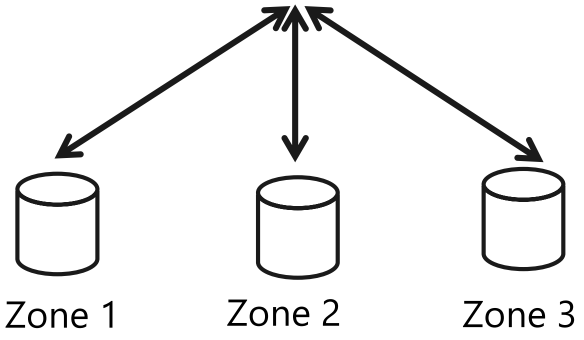
Protect against disk, node, rack failures

Write is ack'd when all replicas are committed

Superior to dual-parity RAID

11 9s of durability

SLA: 99.9%



ZRS

Replicas across 3 Zones

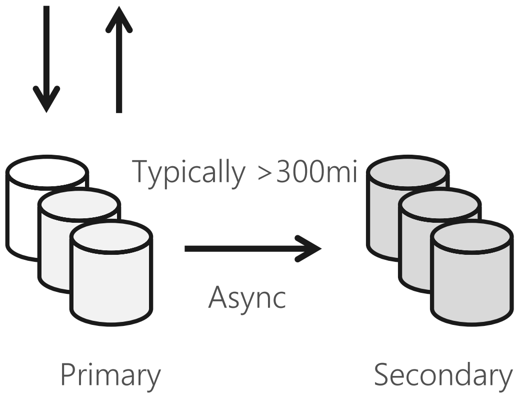
Protect against disk, node, rack and zone failures

Synchronous writes to all 3 zones

12 9s of durability

Available in 8 regions

SLA: 99.9%



GRS

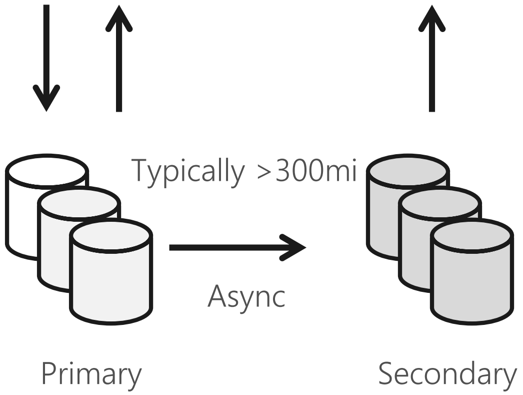
Multiple replicas across each of 2 regions

Protects against major regional disasters

Asynchronous to secondary

16 9s of durability

SLA: 99.9%



RA-GRS

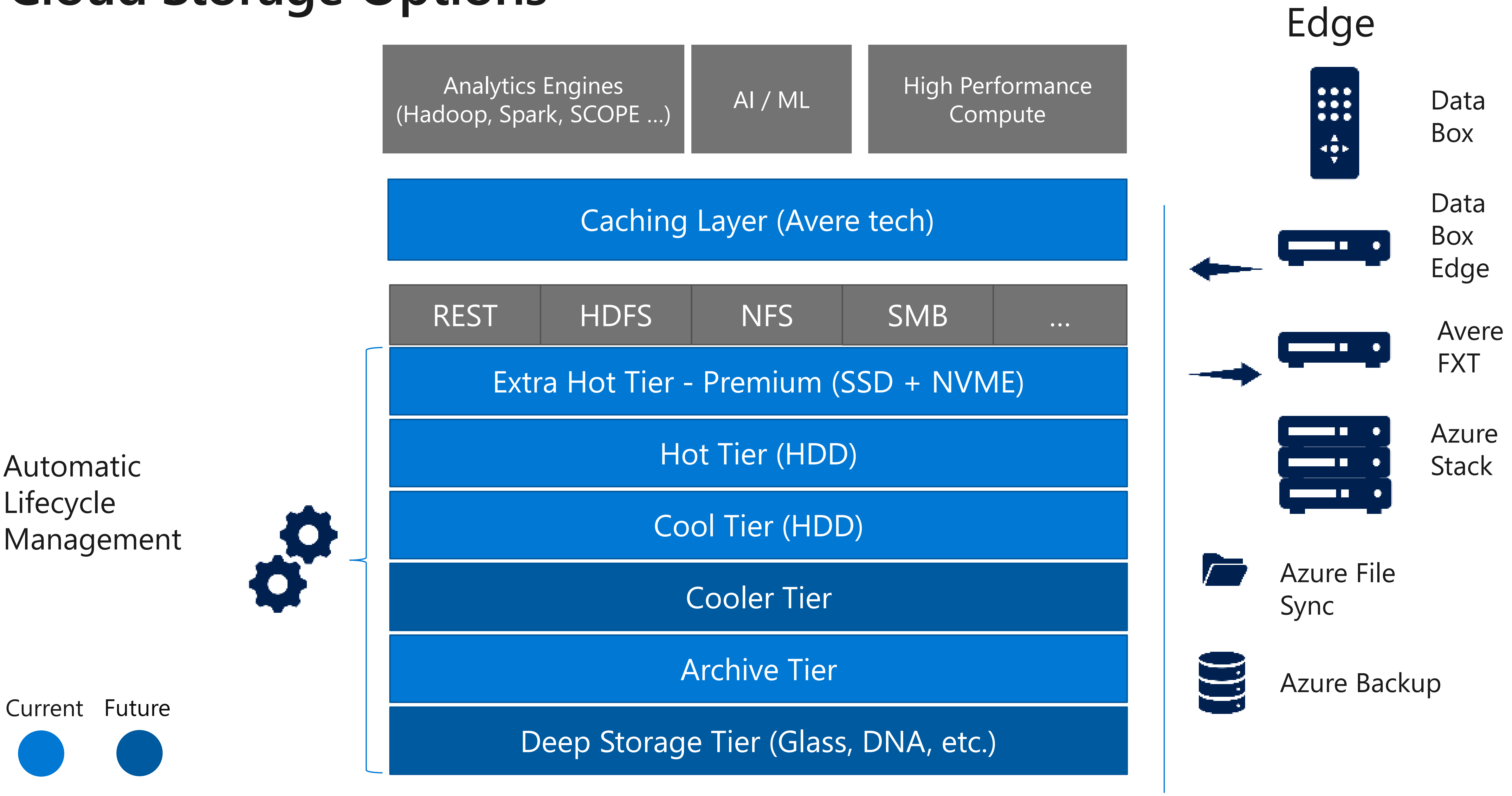
GRS + Read access to secondary

Separate secondary endpoint

RPO delay to secondary can be queried

SLA: 99.99% (read), 99.9% (write)

Cloud Storage Options



Analytics Engines (Hadoop, Spark, SCOPE ...)

AI / ML

High Performance Compute

Caching Layer (Avere tech)

REST | HDFS | NFS | SMB | ...

Extra Hot Tier - Premium (SSD + NVME)

Hot Tier (HDD)

Cool Tier (HDD)

Cooler Tier

Archive Tier

Deep Storage Tier (Glass, DNA, etc.)

Automatic Lifecycle Management

Current ● Future ●

Edge

Data Box

Data Box Edge

Avere FXT

Azure Stack

Azure File Sync

Azure Backup

Data Transport Methods

File Sync

- Windows Srv <-> Azure
- Local caching
- With offline (Databox) can 'sync' remainder

Fuse

- Mount blobs as local FS
- Commit on write
- Linux

Site Replication

- On premise & cloud
- Windows, Linux
- Physical, virtual
- Hyper-V, VMWare

Network Acceleration

- Aspera
- Signiant

AZCopy

- Throughput +30%
- S3 to Azure Blobs
- Sync to cloud
- Hi Latency 10-100%

NetApp

- CloudSync
- SnapMirror
- SnapVault

Data Factory

- On premise & cloud sources
- Structured & unstructured
- Over 60 connectors
- UI design data flow

Partners

- Peer Global File Service
- Talon FAST
- Zerto
- …

Offline

- Data Box
- Data Box Heavy
- Data Box Disk
- Disk Import / Export

Fast Data Transfer
microsoft.com/en-us/garage/profiles/fast-data-transfer/

Azure Data Box Family

Offline Data Transfer

Data Box

- Capacity: 100 TB
- Weight: ~50 lbs
- Secure, ruggedized appliance
- GA September 2018

- Data Box enables bulk migration to Azure when network isn't an option.

Data Box Disk PREVIEW

- Capacity: 8TB ea.; 40TB/order
- Secure, ruggedized USB drives orderable in packs of 5 (up to 40TB).
- Currently in Preview

- Perfect for projects that require a smaller form factor, e.g., autonomous vehicles.

Data Box Heavy PREVIEW

- Capacity: 1 PB
- Weight 500+ lbs
- Secure, ruggedized appliance
- Preview September 2018

- Same service as Data Box, but targeted to petabyte-sized datasets.

Online Data Transfer

Data Box Gateway PREVIEW

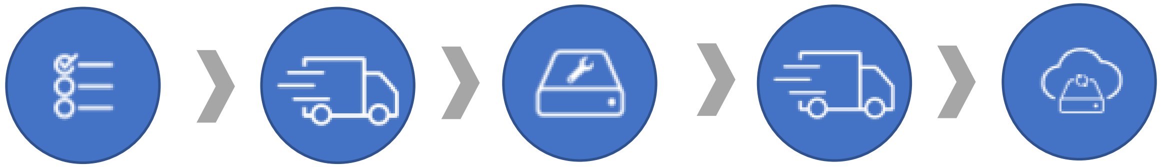
- Virtual device provisioned in your hypervisor
- Supports storage gateway, SMB, NFS, Azure blob, files
- Preview: September 2018

- Virtual network transfer appliance (VM), runs on your choice of hardware.

Data Box Edge PREVIEW

- Local Cache Capacity: ~12 TB
- Includes Data Box Gateway and Azure IoT Edge.
- Preview: September 2018

- Data Box Edge manages uploads to Azure and can pre-process data prior to upload.



Order › Send › Fill › Return › Upload

Network Data Transfer

Cloud to Edge

Edge to Cloud

Edge Compute

Pre-processing

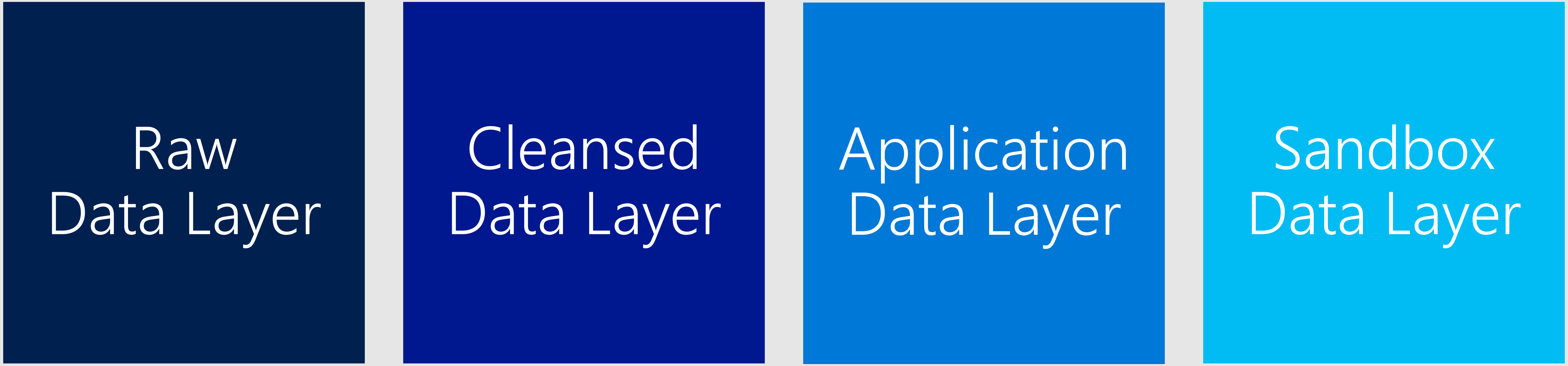
ML Inferencing

Exactly what is a data lake?

A storage repository, usually Hadoop, that holds a vast amount of raw data in its native format until it is needed.

- Inexpensively store unlimited data
- Collect all data “just in case”
- Store data with no modeling – “Schema on read”
- Complements EDW
- Frees up expensive EDW resources
- Quick user access to data
- ETL Hadoop tools
- Easily scalable
- Place to move older data (archive)
- Place to backup data to

Data Lake Layers



Raw Data Layer

Cleansed Data Layer

Application Data Layer

Sandbox Data Layer

Needs data governance so your data lake does not turn into a data swamp!

Organizing a Data Lake – Folder structure

Objectives

✓ Plan the structure based on optimal data retrieval

✓ Avoid a chaotic, unorganized data swamp

Common ways to organize the data:

Time Partitioning

Year/Month/Day/Hour/Minute

Subject Area

Security Boundaries

Department

Business unit

etc…

Downstream App/Purpose

Data Retention Policy

Temporary data

Permanent data

Applicable period (ex: project lifetime)

etc…

Business Impact / Criticality

High (HBI)

Medium (MBI)

Low (LBI)

etc…

Owner / Steward / SME

Probability of Data Access

Recent/current data

Historical data

etc…

Confidential Classification

Public information

Internal use only

Supplier/partner confidential

Personally identifiable information (PII)

Sensitive – financial

Sensitive – intellectual property

etc…

Data Lake with DW use cases

Data Lake

Staging & preparation

- Data scientists/Power users
- Batch processing
- Data refinement/cleaning
- ETL workloads
- Store older/backup data
- Sandbox for data exploration
- One-time reports
- Quick access to data
- Don’t know questions

Data Warehouse

Serving, Security & Compliance

- Business people
- Low latency
- Complex joins
- Interactive ad-hoc query
- High number of users
- Additional security
- Large support for tools
- Dashboards
- Easily create reports (Self-service BI)
- Know questions

What is Azure Databricks?

A fast, easy and collaborative Apache® Spark™ based analytics platform optimized for Azure

databricks

Best of Databricks

+

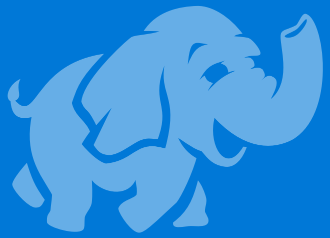
Microsoft

Best of Microsoft

- Designed in collaboration with the founders of Apache Spark
- One-click set up; streamlined workflows
- Interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.
- Native integration with Azure services (Power BI, SQL DW, Cosmos DB, Blob Storage)
- Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs)

Azure HDInsight

Hadoop and Spark as a Service on Azure



Fully-managed Hadoop and Spark for the cloud

100% Open Source Hortonworks data platform

Clusters up and **running in minutes**

Managed, monitored and supported by Microsoft with the **industry's best SLA**

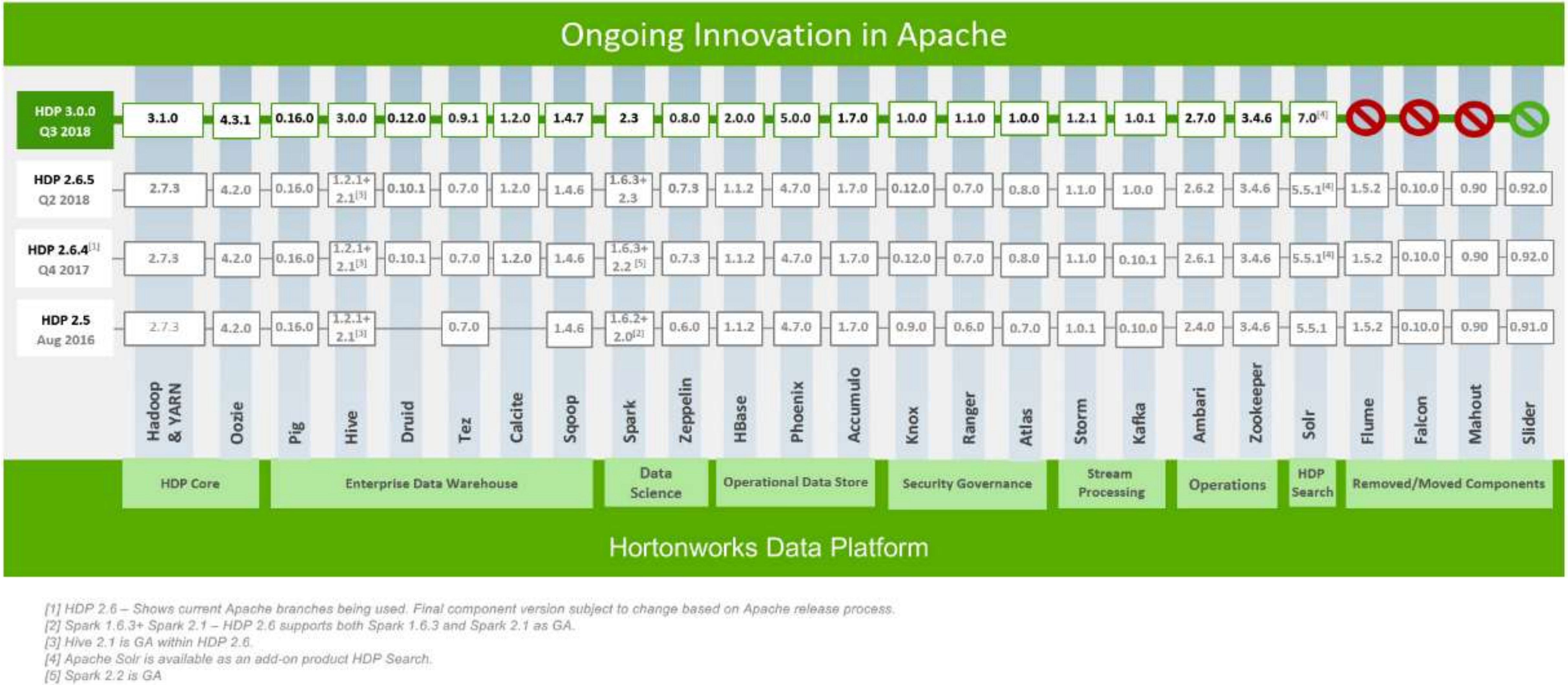
Familiar **BI tools for analysis**, or open source notebooks for **interactive data science**

63% lower TCO than deploy your own Hadoop on-premises*

*IDC study "The Business Value and TCO Advantage of Apache Hadoop in the Cloud with Microsoft Azure HDInsight"

Hortonworks Data Platform (HDP) 3.0

(under the covers of HDInsight 4.0 – public preview)



Ongoing Innovation in Apache

	Hadoop & YARN	Oozie	Pig	Hive	Druid	Tez	Calcite	Sqoop	Spark	Zeppelin	HBase	Phoenix	Accumulo	Knox	Ranger	Atlas	Storm	Kafka	Ambari	Zookeeper	Solr	Flume	Falcon	Mahout	Slider
HDP 3.0.0 Q3 2018	3.1.0	4.3.1	0.16.0	3.0.0	0.12.0	0.9.1	1.2.0	1.4.7	2.3	0.8.0	2.0.0	5.0.0	1.7.0	1.0.0	1.1.0	1.0.0	1.2.1	1.0.1	2.7.0	3.4.6	7.0[4]	🚫	🚫	🚫	⊘
HDP 2.6.5 Q2 2018	2.7.3	4.2.0	0.16.0	1.2.1+ 2.1[3]	0.10.1	0.7.0	1.2.0	1.4.6	1.6.3+ 2.3	0.7.3	1.1.2	4.7.0	1.7.0	0.12.0	0.7.0	0.8.0	1.1.0	1.0.0	2.6.2	3.4.6	5.5.1[4]	1.5.2	0.10.0	0.90	0.92.0
HDP 2.6.4[1] Q4 2017	2.7.3	4.2.0	0.16.0	1.2.1+ 2.1[3]	0.10.1	0.7.0	1.2.0	1.4.6	1.6.3+ 2.2[5]	0.7.3	1.1.2	4.7.0	1.7.0	0.12.0	0.7.0	0.8.0	1.1.0	0.10.1	2.6.1	3.4.6	5.5.1[4]	1.5.2	0.10.0	0.90	0.92.0
HDP 2.5 Aug 2016	2.7.3	4.2.0	0.16.0	1.2.1+ 2.1[3]		0.7.0		1.4.6	1.6.2+ 2.0[2]	0.6.0	1.1.2	4.7.0	1.7.0	0.9.0	0.6.0	0.7.0	1.0.1	0.10.0	2.4.0	3.4.6	5.5.1	1.5.2	0.10.0	0.90	0.91.0
	HDP Core		Enterprise Data Warehouse						Data Science		Operational Data Store			Security Governance			Stream Processing		Operations		HDP Search	Removed/Moved Components			

Hortonworks Data Platform

[1] HDP 2.6 – Shows current Apache branches being used. Final component version subject to change based on Apache release process.

[2] Spark 1.6.3+ Spark 2.1 – HDP 2.6 supports both Spark 1.6.3 and Spark 2.1 as GA.

[3] Hive 2.1 is GA within HDP 2.6.

[4] Apache Solr is available as an add-on product HDP Search.

[5] Spark 2.2 is GA

Simply put, Hortonworks ties all the open source products together (20)

Azure Analysis Services

Enterprise grade analytics engine as a service

Build rich semantic models

Transform complex data into business user friendly semantic models

Gain insights at the speed of thought

Gain instant insights with in-memory cache using your preferred visualization tools

Proven technology

Based on powerful, proven SQL Server Analysis Services

Provision and scale with ease

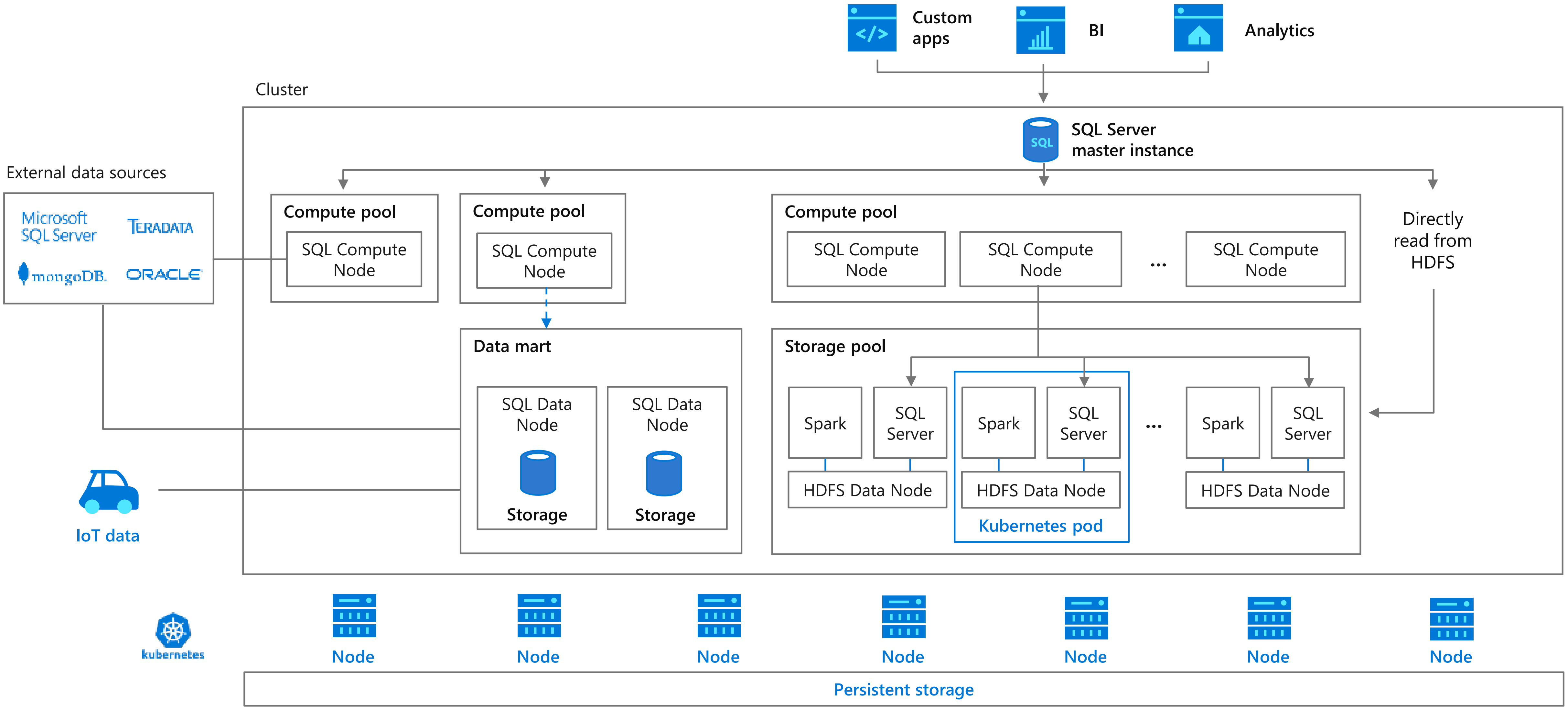
Easy to deploy, scale, and manage as platform-as-a-service

Azure Analysis Services Cubes

Reasons to report off cubes instead of the data warehouse:

- Semantic layer
- Handle many concurrent users
- Aggregating data for performance
- Multidimensional analysis
- No joins or relationships
- Hierarchies, KPI’s
- Row-level Security
- Advanced time-calculations
- Slowly Changing Dimensions (SCD)
- Required for some reporting tools

SQL Server 2019 big data clusters



Machine Learning and AI in the Microsoft Stack

1. Machine Learning
2. Deep Learning
3. Data Science Virtual Machine
4. Azure Databricks
5. Cognitive Services
6. Batch AI
7. Machine Learning Server
8. SQL Server
9. Power BI
10. Azure Machine Learning service

Machine learning and AI portfolio

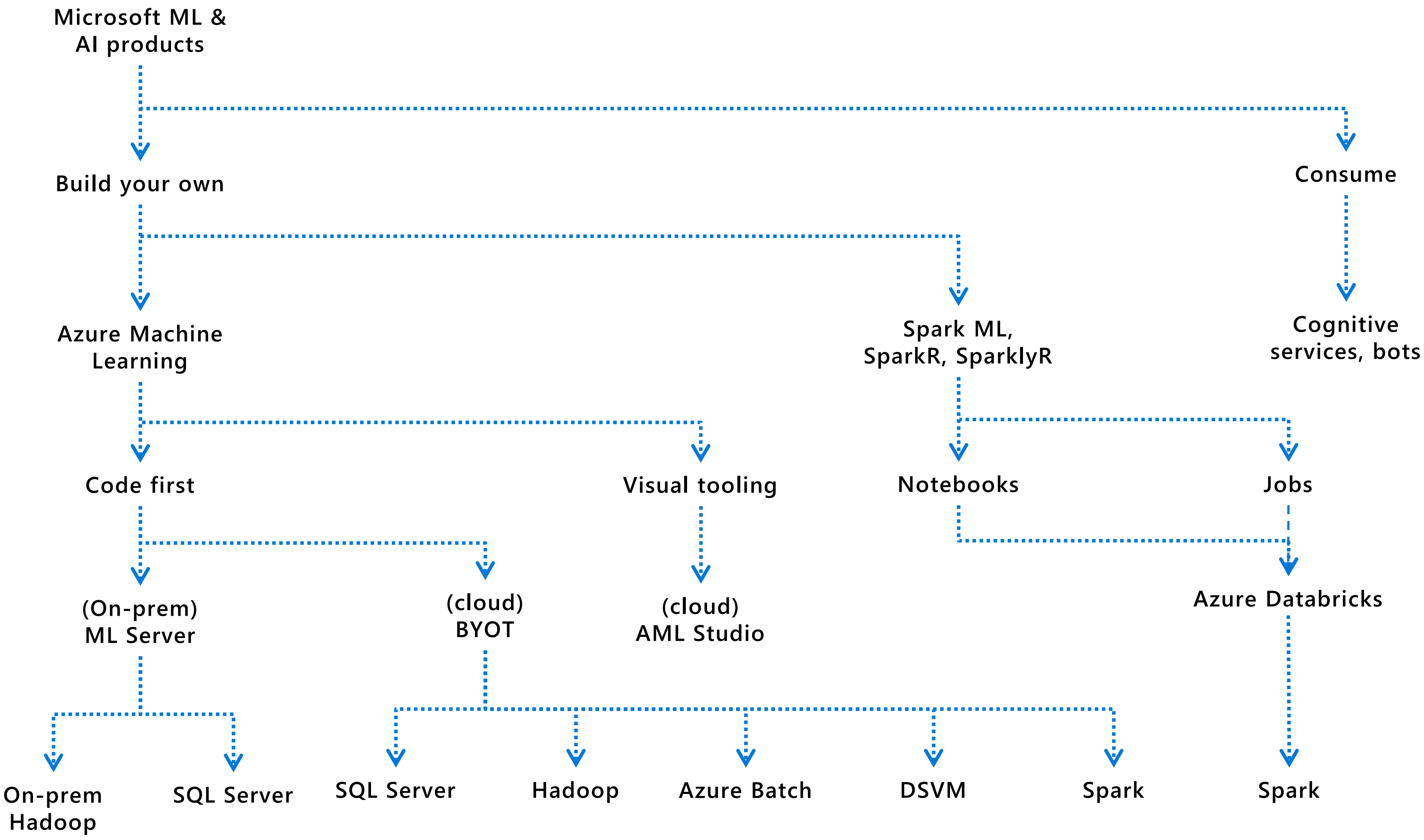
When to use what

Build your own or consume pre-trained models?

Which experience do you want?

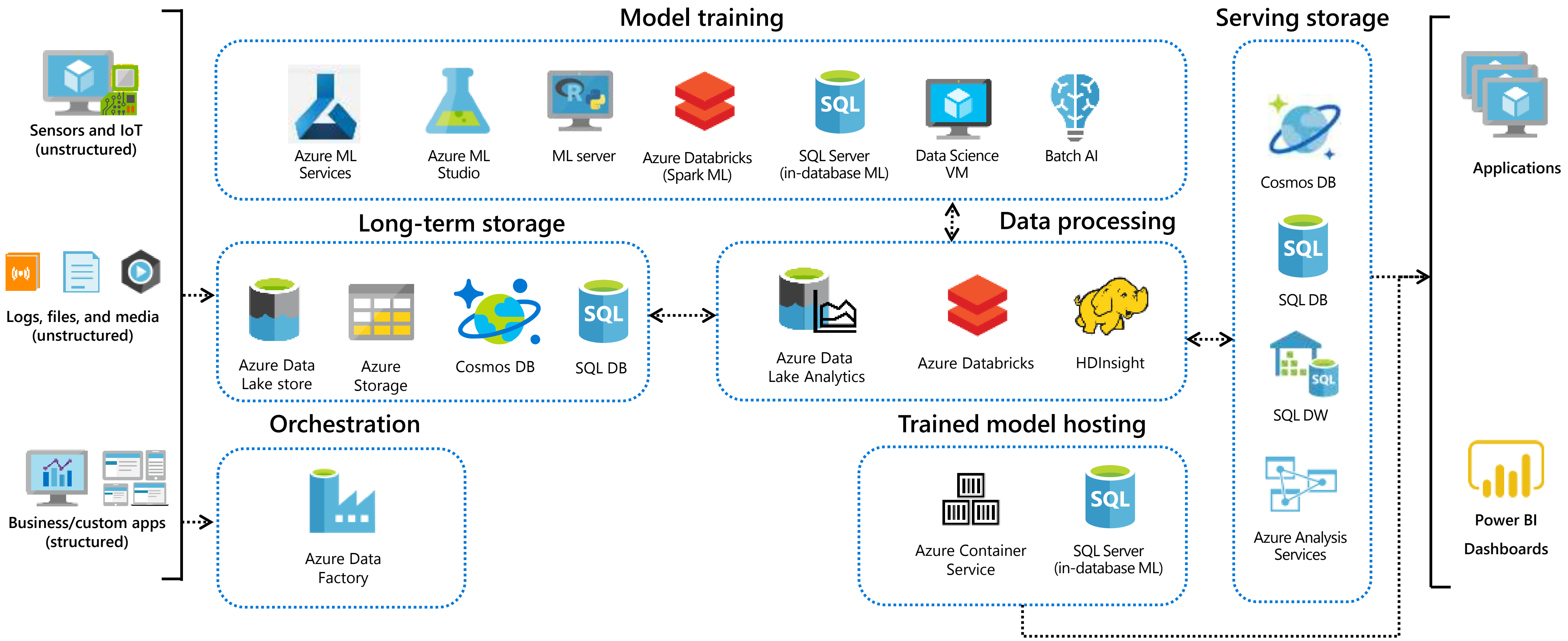
Deployment target

What engines do you want to use?



Advanced analytics pattern in Azure

Data collection and understanding, modeling, and deployment



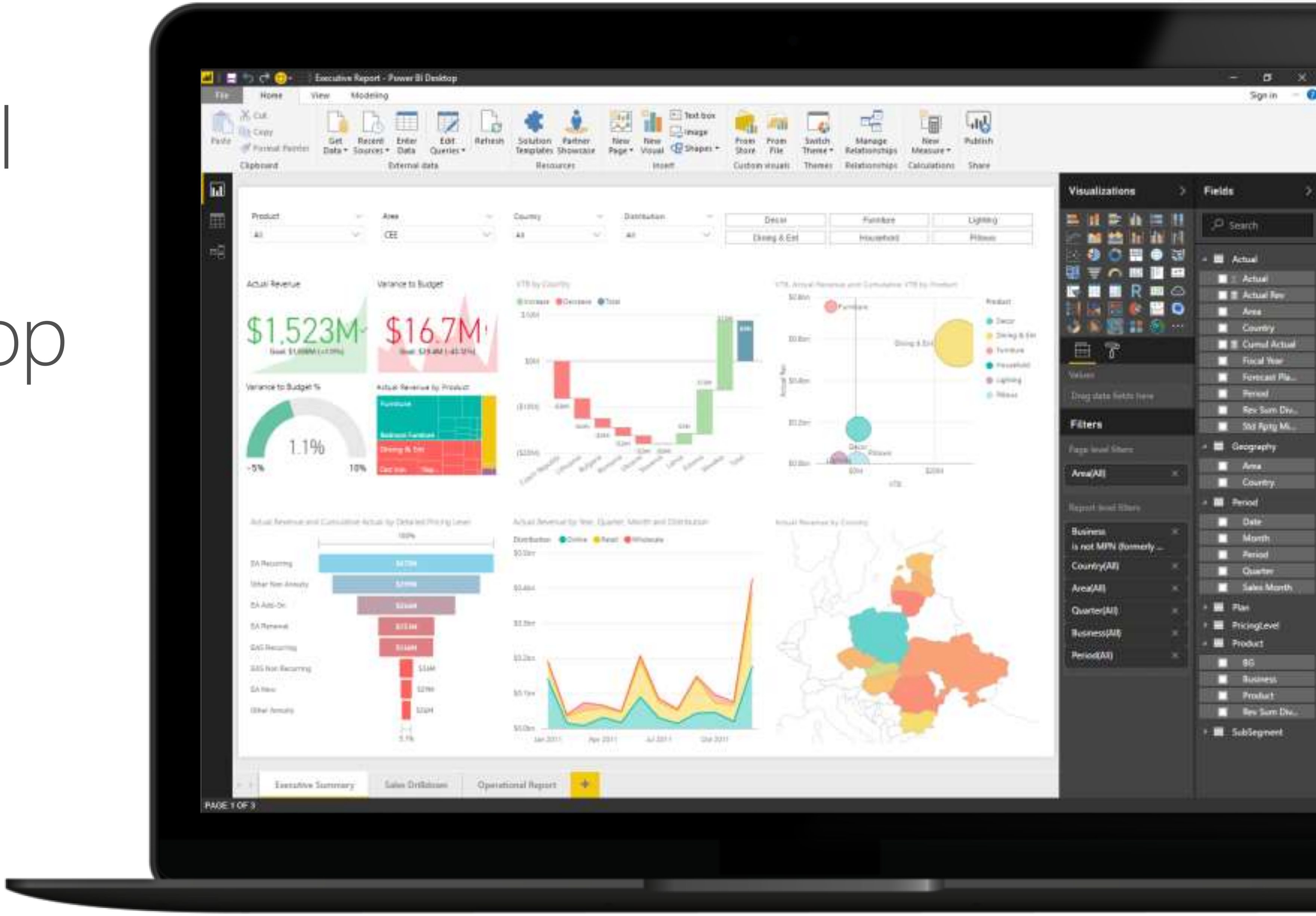
Create powerful reports with Power BI Desktop

Discovery & exploration

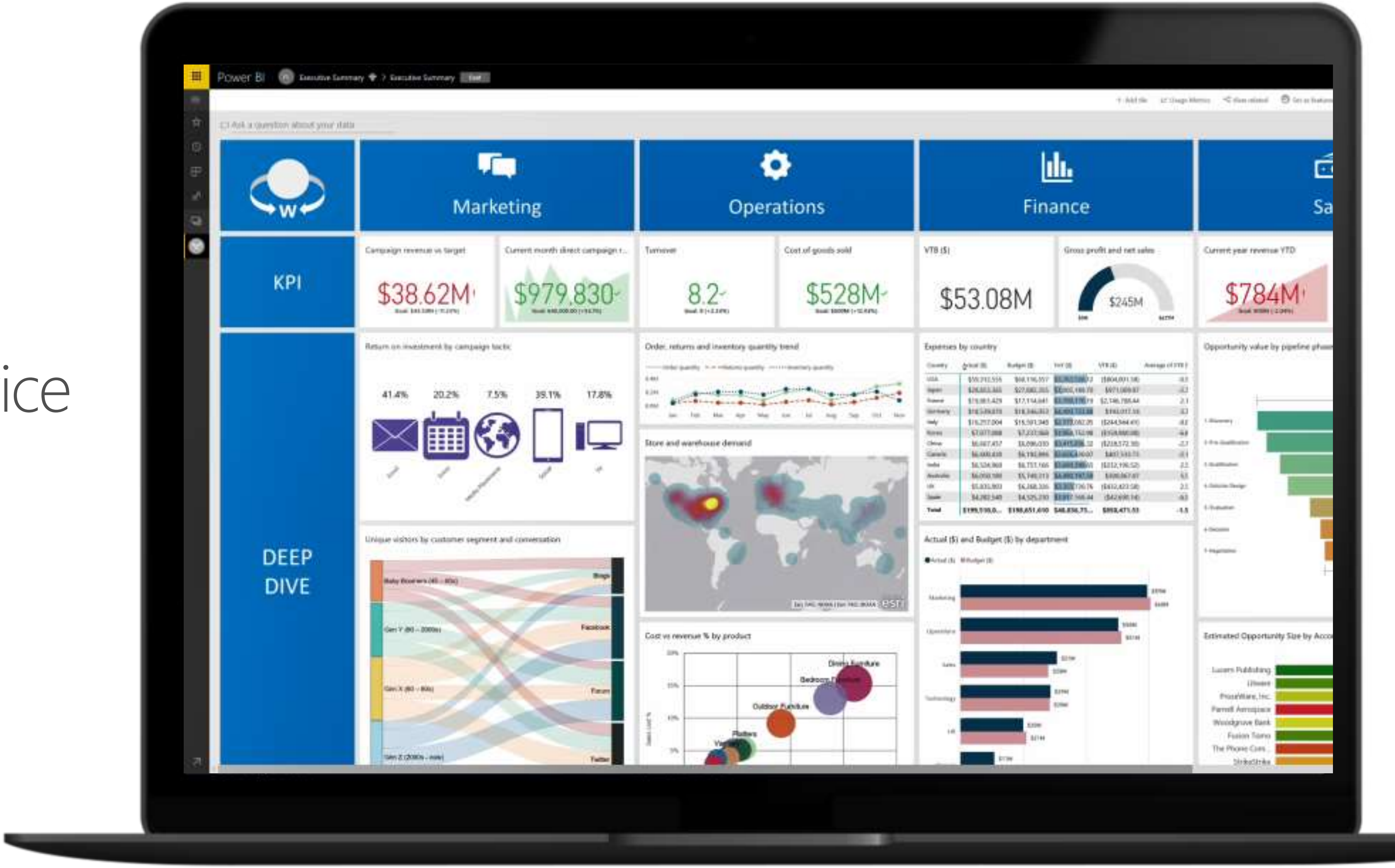
Easy report authoring

Custom visualizations

R & Python integration



Power BI Service Dashboard


Power BI
Executive Summary
Ask a question about your data
Marketing
Operations
Finance
Sa
KPI
DEEP DIVE
Campaign revenue vs target
$38.62M
Current month direct campaign r...
$979,830
Turnover
8.2
Cost of goods sold
$528M
VTB ($)
$53.08M
Gross profit and net sales
$245M
Current year revenue YTD
$784M
Return on investment by campaign tactic
41.4%
20.2%
7.5%
39.1%
17.8%
Order, returns and inventory quantity trend
Expenses by country
Opportunity value by pipeline phase
Store and warehouse demand
Unique visitors by customer segment and conversation
Actual ($) and Budget ($) by department
Cost vs revenue % by product
Dining Furniture
Bedroom Furniture
Outdoor Furniture
Estimated Opportunity Size by Account

Resources

- Ivan Kosyakov:
- Artificial Intelligence Decision Tree
- Big Data Decision Tree v4
- Business Intelligence Solutions Decision Tree

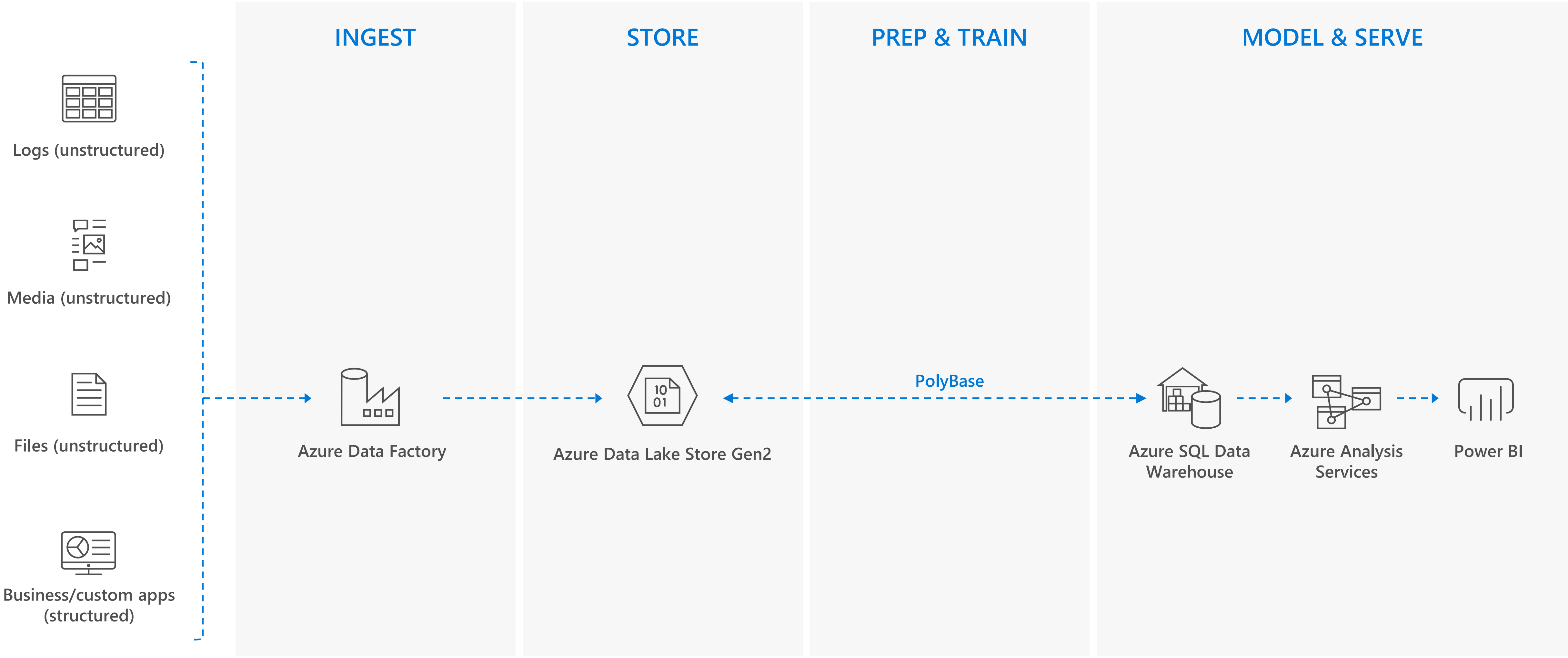
Q & A

James Serra, Big Data Evangelist
Email me at: jamesserra3@gmail.com
Follow me at: @JamesSerra
Link to me at: www.linkedin.com/in/JamesSerra
Visit my blog at: JamesSerra.com (where this slide deck is posted via the "Presentations" link on the top menu)

Appendix

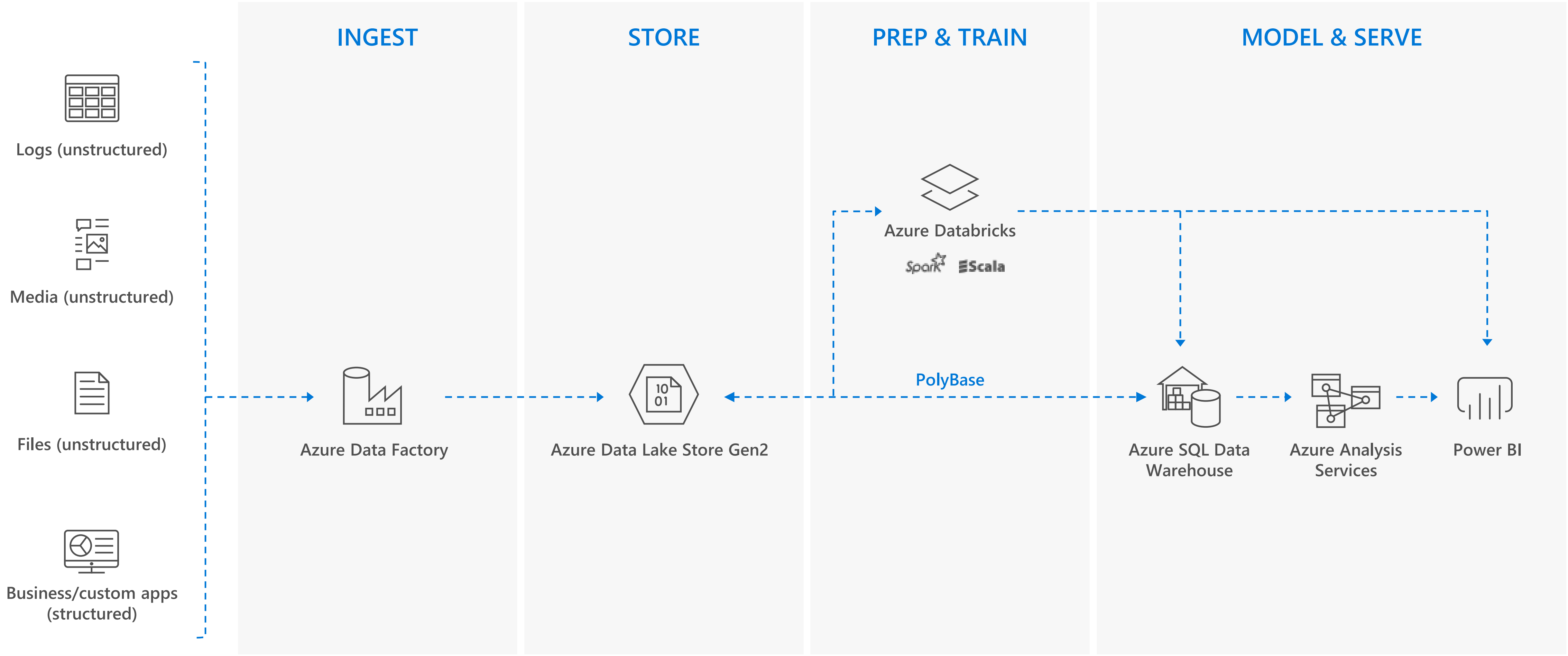
Architecture Patterns

CLOUD DATA WAREHOUSE



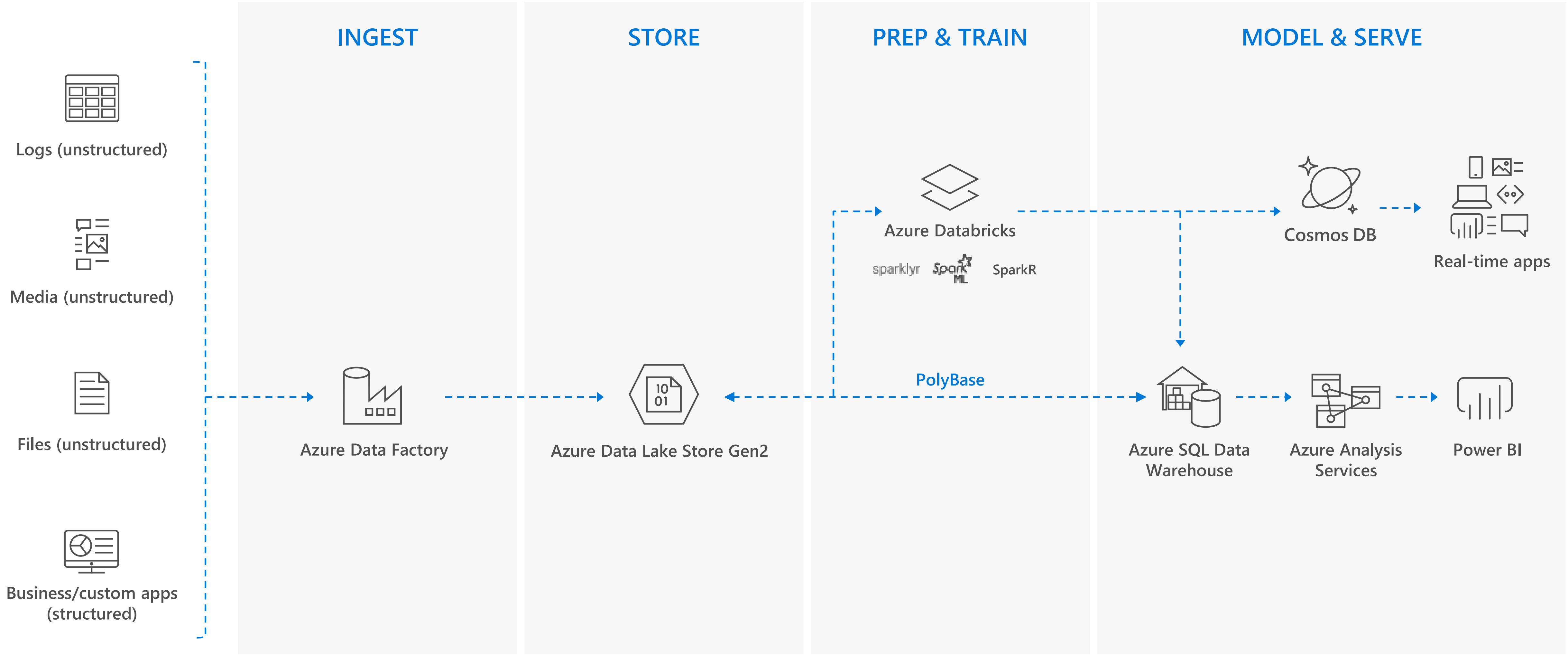
Microsoft Azure also supports other Big Data services like Azure HDInsight to allow customers to tailor the above architecture to meet their unique needs.

MODERN DATA WAREHOUSE



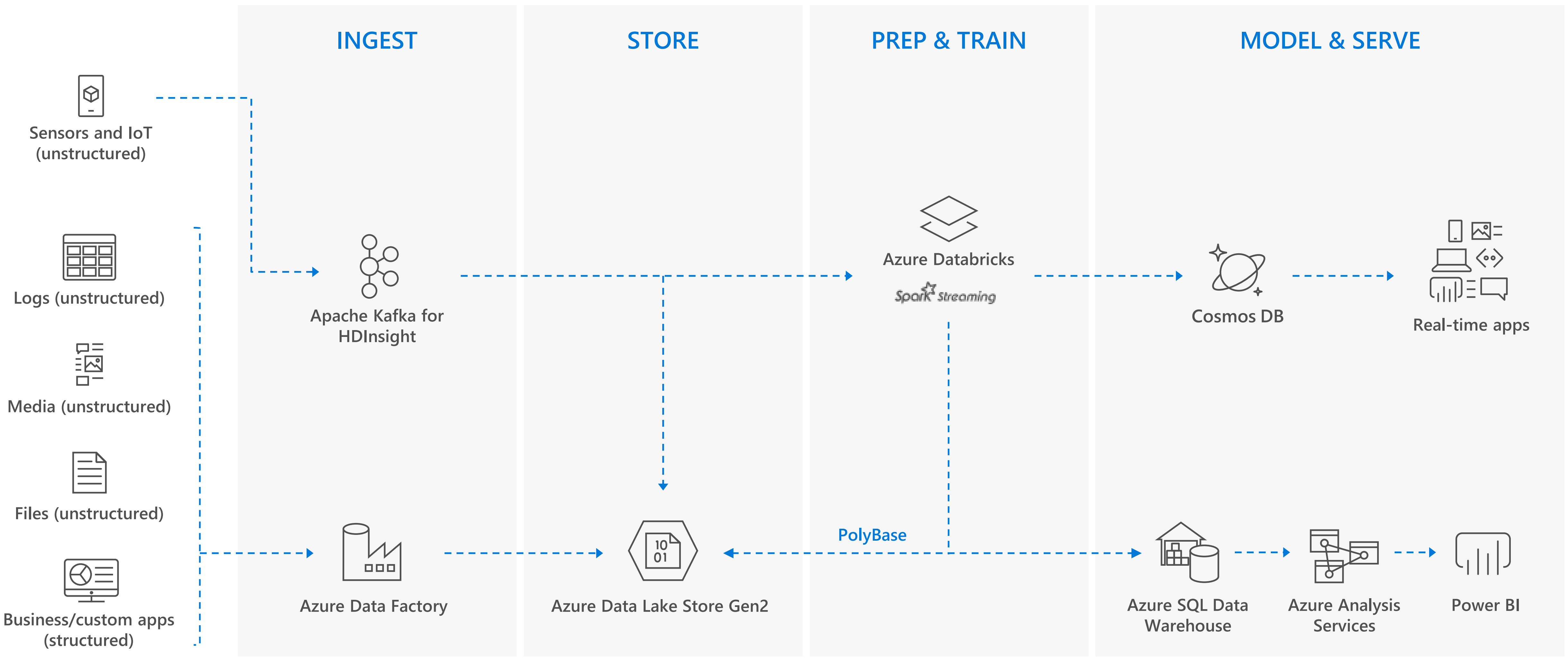
Microsoft Azure also supports other Big Data services like Azure HDInsight to allow customers to tailor the above architecture to meet their unique needs.

ADVANCED ANALYTICS ON BIG DATA



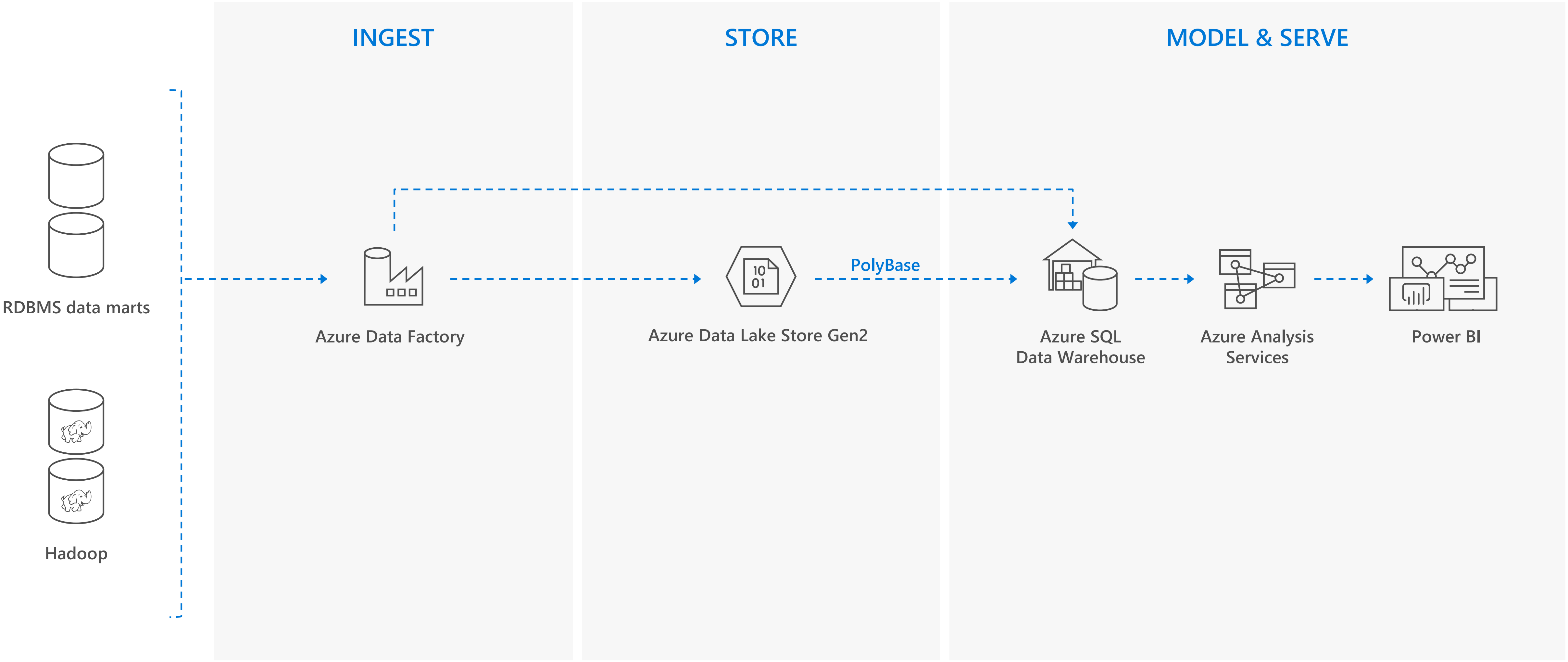
Microsoft Azure also supports other Big Data services like Azure HDInsight, Azure Machine Learning to allow customers to tailor the above architecture to meet their unique needs.

REAL TIME ANALYTICS



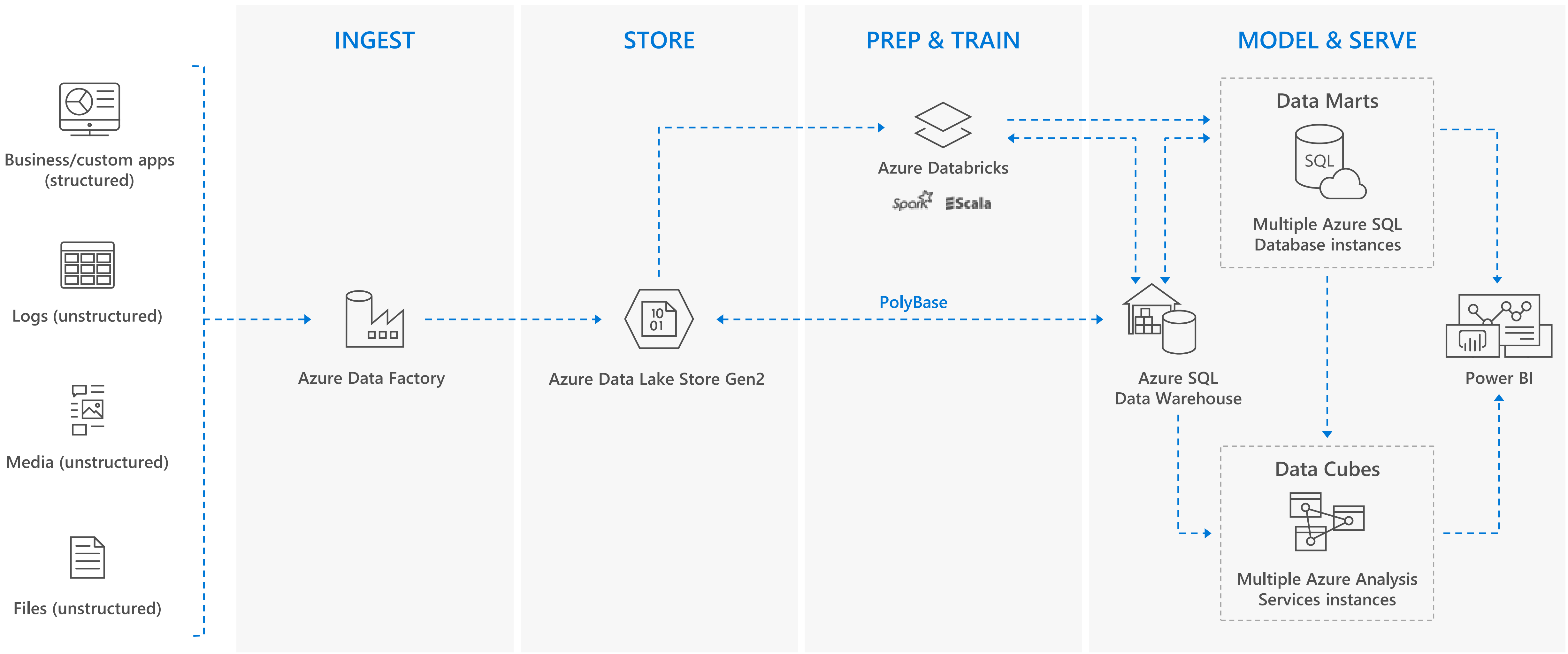
Microsoft Azure also supports other Big Data services like Azure IoT Hub, Azure Event Hubs, Azure Machine Learning to allow customers to tailor the above architecture to meet their unique needs.

DATA MART CONSOLIDATION



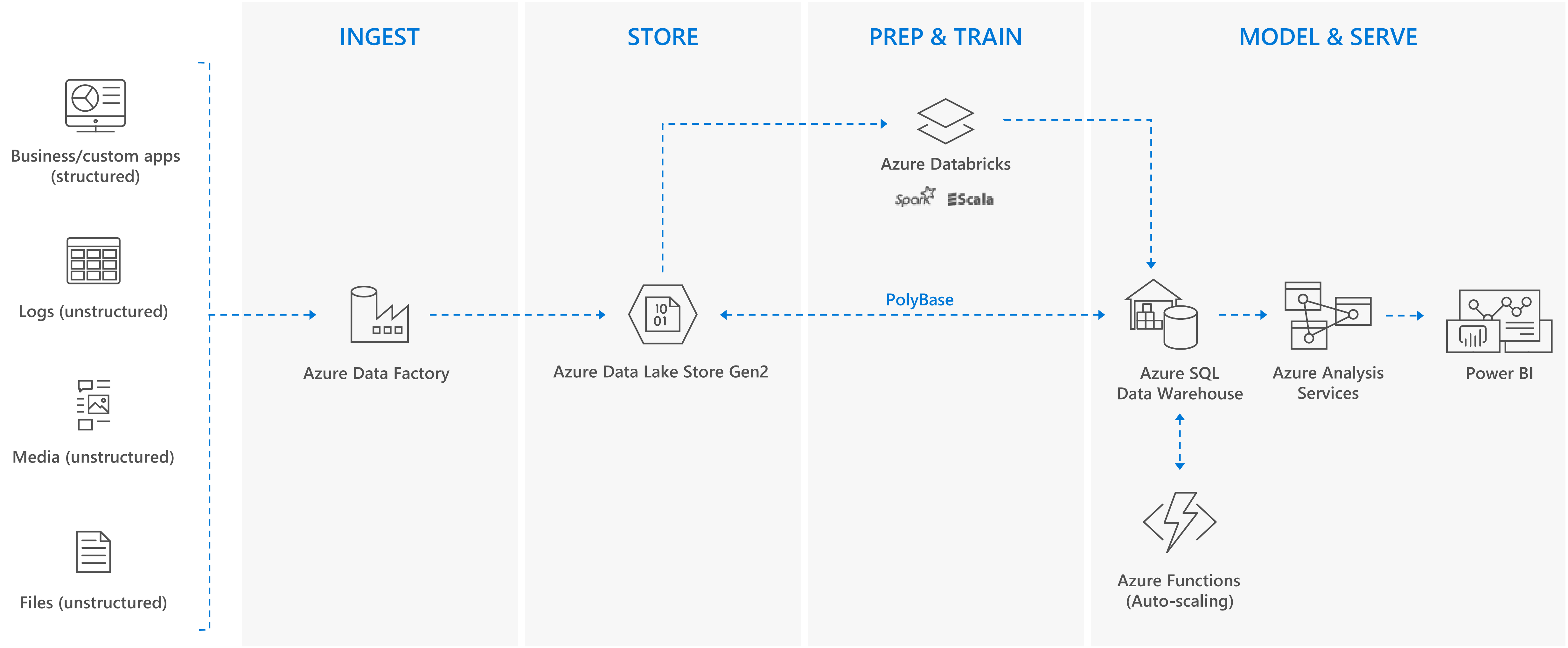
Microsoft Azure also supports other Big Data services like Azure HDInsight to allow customers to tailor the architecture to meet their unique needs.

HUB & SPOKE ARCHITECTURE FOR BI



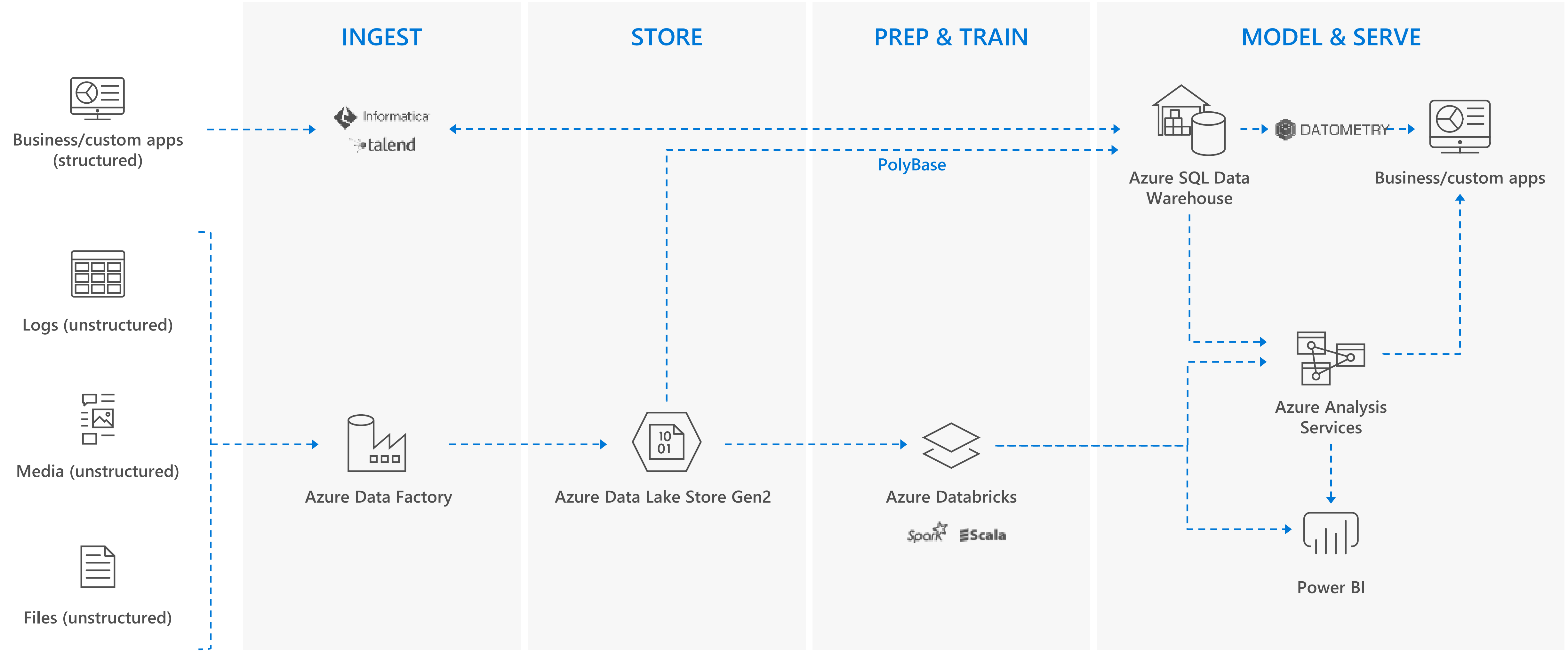
Microsoft Azure supports other services like Azure HDInsight to allow customers a truly customized solution.

AUTO SCALING DATA WAREHOUSE

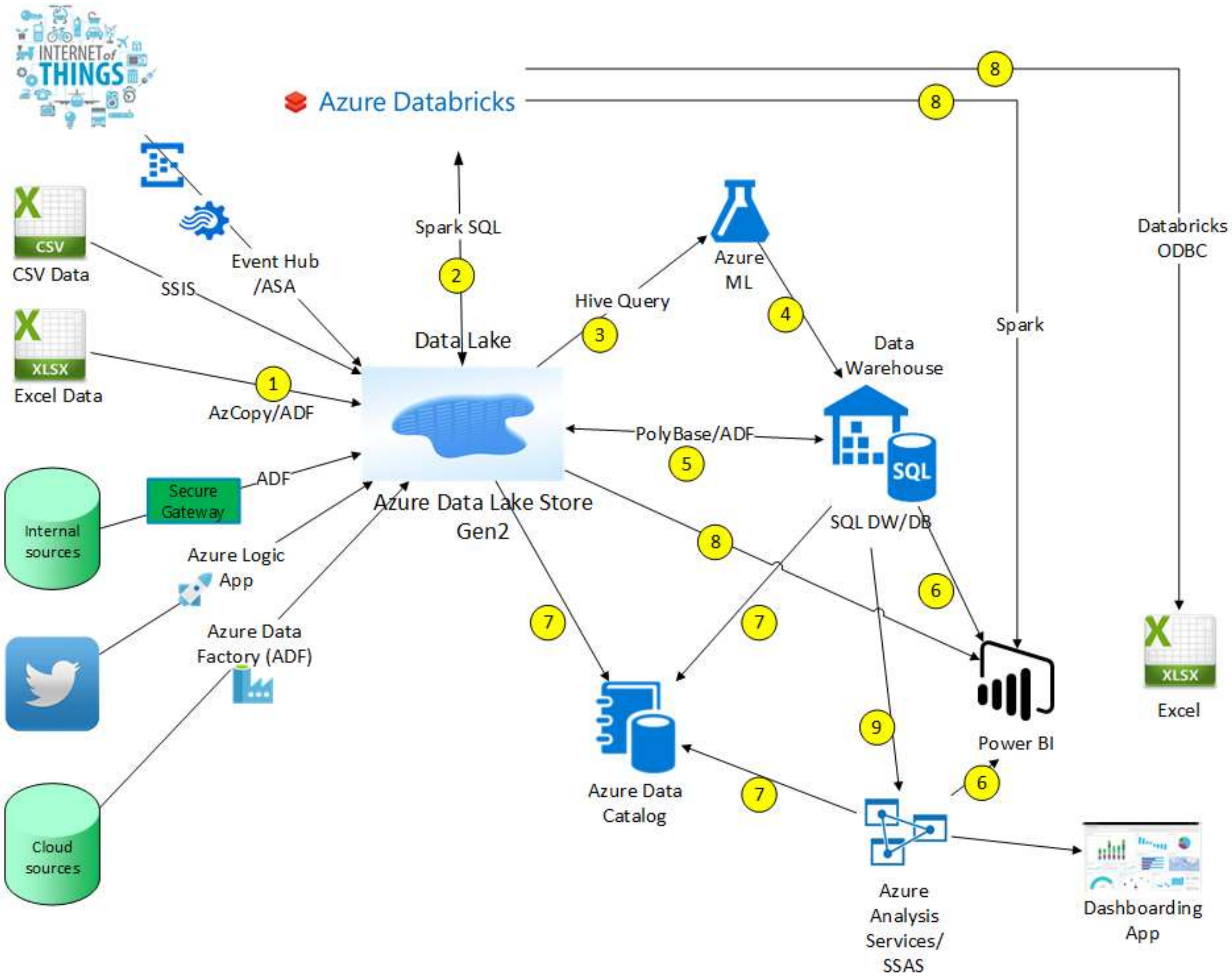


Microsoft Azure supports other services like Azure HDInsight to allow customers a truly customized solution.

DATA WAREHOUSE MIGRATION



Azure also supports other Big Data services like Azure HDInsight to allow customers to tailor the architecture to meet their unique needs.


INTERNET of THINGS
Azure Databricks
8
8
Databricks ODBC
Spark
CSV
CSV Data
SSIS
Event Hub /ASA
Spark SQL
2
Data Lake
Hive Query
3
Azure ML
4
Data Warehouse
XLSX
Excel Data
1
AzCopy/ADF
PolyBase/ADF
5
SQL
SQL DW/DB
ADF
Secure Gateway
Internal sources
Azure Data Lake Store Gen2
8
Azure Logic App
7
7
6
Azure Data Factory (ADF)
XLSX
Excel
9
Power BI
Azure Data Catalog
7
6
Azure Analysis Services/ SSAS
Dashboarding App
Cloud sources