

Azure Big Data & Machine Learning Overview

Rui Carmo

Architecture & Industry Services
Office of the CTO, EMEA

 @rcarmo,  [linkedin.com/in/ruicarmo](https://www.linkedin.com/in/ruicarmo)

Agenda

- ➔ Drivers for Analytics & ML Adoption
- ➔ (Short) Azure Overview
- ➔ Azure Data Stores and ML Services
- ➔ Processes & Pipelines
- ➔ Modern Data Warehousing
- ➔ Estimating Cloud OPEX
- ➔ Use Case + Q & A

Drivers for Analytics and ML Adoption



Evolution towards “Data Maturity”

STAGE 1:

Traditional

Query historical, relational data from a variety of sources

STAGE 2:

Operational

Gain real-time insights without impacting performance

STAGE 3:

Logical

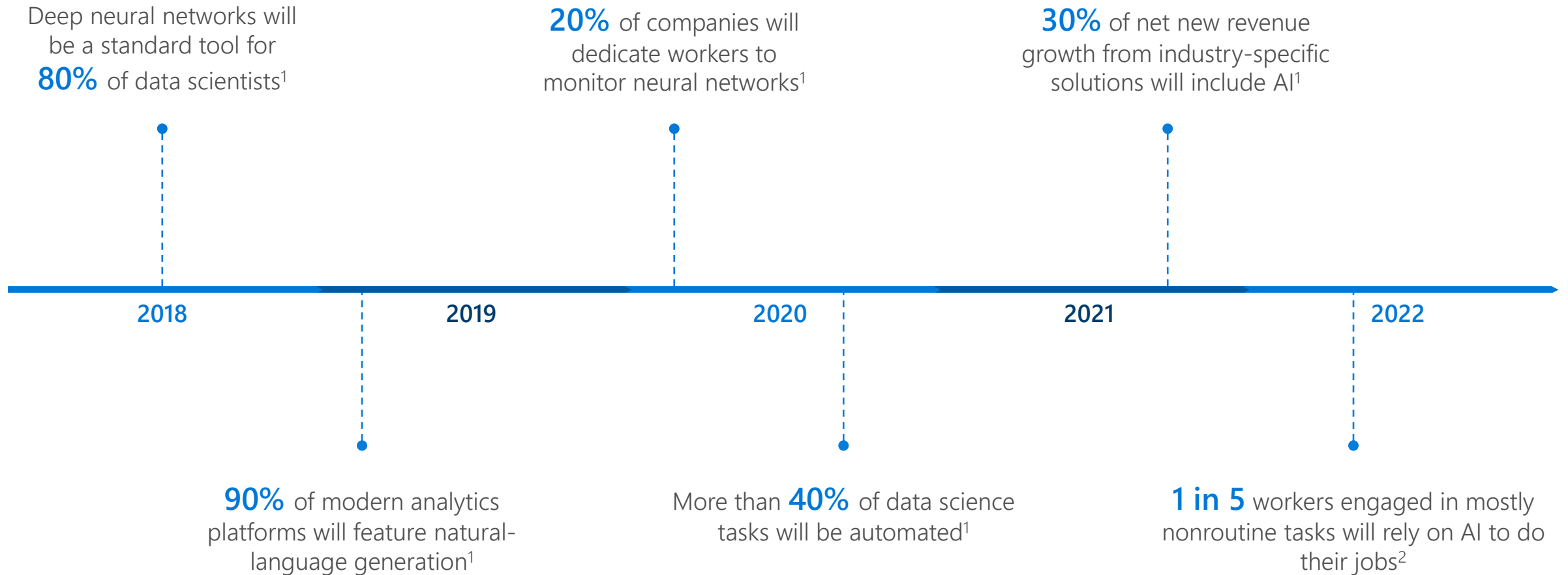
Ask questions of big data—all types, volumes and locations

STAGE 4:

Free-form

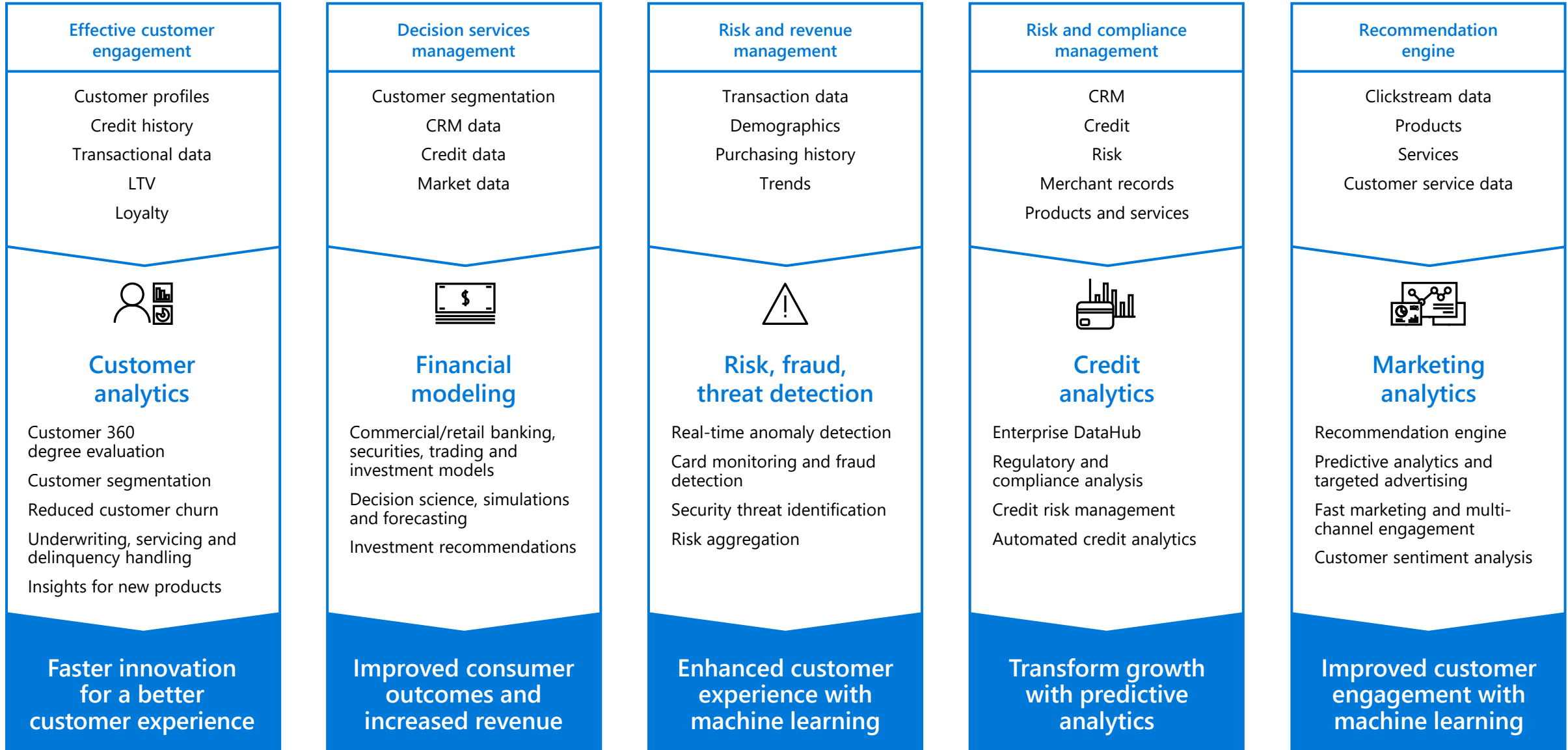
Establish enterprise-wide data lake and run advanced analytics and deep learning on unstructured data that arrives in real-time

What are companies looking to do next?



¹ "100 Data and Analytics Predictions Through 2021", Gartner, 2017. ² "Predicts 2018: AI and the Future of Work", Gartner, 2018.

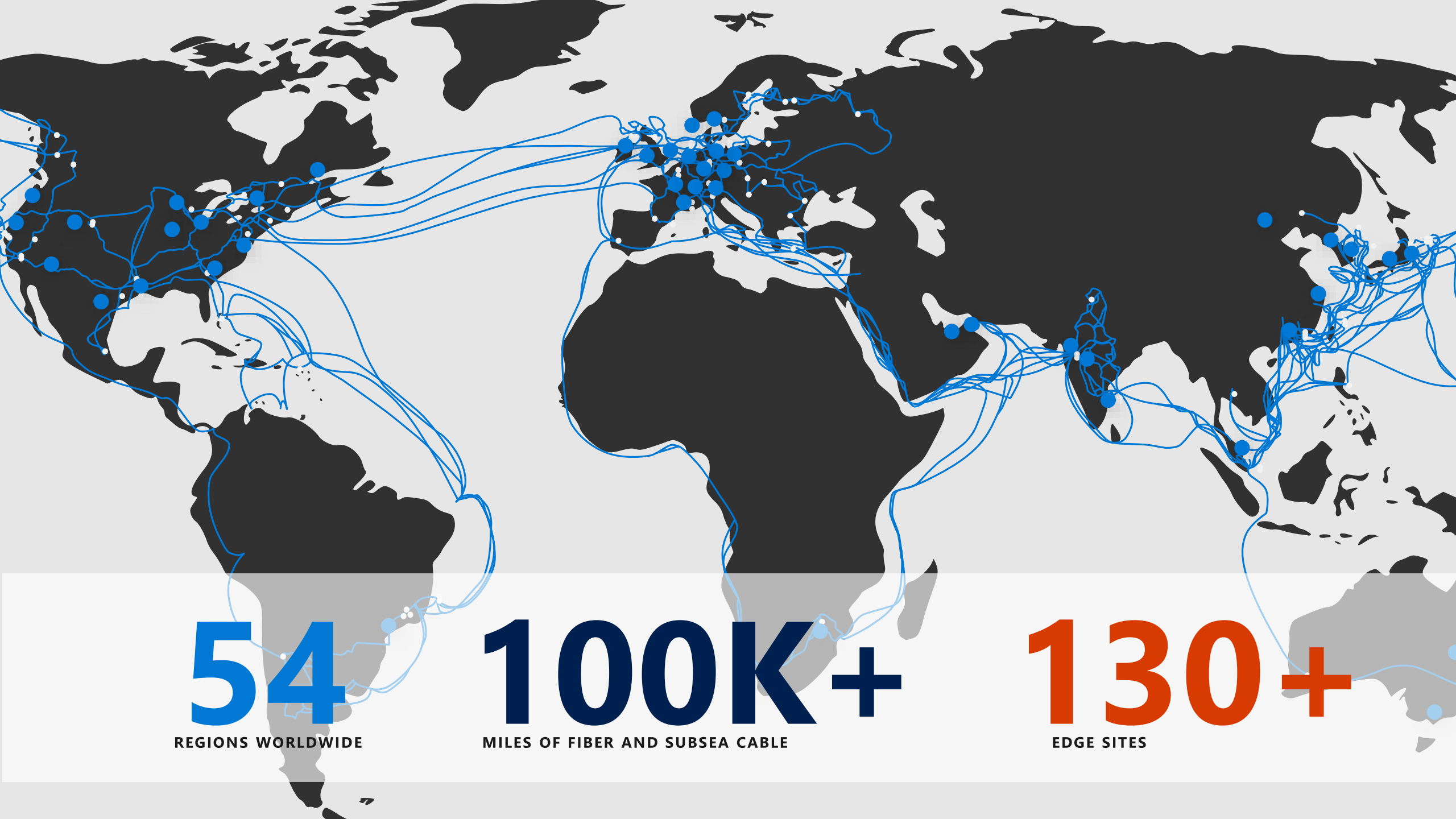
Financial services use cases (focus: retail banking)



Azure

Quincy, WA





54

REGIONS WORLDWIDE

100K+

MILES OF FIBER AND SUBSEA CABLE

130+

EDGE SITES

Microsoft Trusted Cloud

→ <https://aka.ms/AzureCompliance>

Azure has the deepest and most comprehensive compliance coverage in the industry

Global

- | | | | |
|--|--|--|--|
| <input checked="" type="checkbox"/> ISO 27001:2013 | <input checked="" type="checkbox"/> ISO 22301:2012 | <input checked="" type="checkbox"/> SOC 1 Type 2 | <input checked="" type="checkbox"/> CSA STAR Certification |
| <input checked="" type="checkbox"/> ISO 27017:2015 | <input checked="" type="checkbox"/> ISO 9001:2015 | <input checked="" type="checkbox"/> SOC 2 Type 2 | <input checked="" type="checkbox"/> CSA STAR Attestation |
| <input checked="" type="checkbox"/> ISO 27018:2014 | | <input checked="" type="checkbox"/> SOC 3 | <input checked="" type="checkbox"/> CSA STAR Self-Assessment |

US Gov

- | | | | |
|--|--|---|--|
| <input checked="" type="checkbox"/> FedRAMP High | <input checked="" type="checkbox"/> DoD DISA SRG Level 5 | <input checked="" type="checkbox"/> DoE 10 CFR Part 810 | <input checked="" type="checkbox"/> ITAR |
| <input checked="" type="checkbox"/> FedRAMP Moderate | <input checked="" type="checkbox"/> DoD DISA SRG Level 4 | <input checked="" type="checkbox"/> NIST SP 800-171 | <input checked="" type="checkbox"/> CJIS |
| | <input checked="" type="checkbox"/> DoD DISA SRG Level 2 | <input checked="" type="checkbox"/> FIPS 140-2 | <input checked="" type="checkbox"/> IRS 1075 |
| | <input checked="" type="checkbox"/> DFARS | <input checked="" type="checkbox"/> Section 508 VPATs | |

Industry

- | | | | |
|--|--|---|---|
| <input checked="" type="checkbox"/> PCI DSS Level 1 | <input checked="" type="checkbox"/> HIPAA BAA | <input checked="" type="checkbox"/> IG Toolkit (UK) | <input checked="" type="checkbox"/> CDSA |
| <input checked="" type="checkbox"/> GLBA | <input checked="" type="checkbox"/> HITRUST | <input checked="" type="checkbox"/> NEN 7510:2011 (Netherlands) | <input checked="" type="checkbox"/> MPAA |
| <input checked="" type="checkbox"/> FFIEC | <input checked="" type="checkbox"/> 21 CFR Part 11 (GxP) | <input checked="" type="checkbox"/> FERPA | <input checked="" type="checkbox"/> FACT (UK) |
| <input checked="" type="checkbox"/> Shared Assessments | <input checked="" type="checkbox"/> MARS-E | | |
| <input checked="" type="checkbox"/> FISC (Japan) | | | |

Regional

- | | | | |
|---|---|---|--|
| <input checked="" type="checkbox"/> Argentina PDPA | <input checked="" type="checkbox"/> China TRUCS / CCCPPF | <input checked="" type="checkbox"/> India MeitY | <input checked="" type="checkbox"/> Singapore MTCS Level 3 |
| <input checked="" type="checkbox"/> Australia CCSL / IRAP | <input checked="" type="checkbox"/> EU ENISA IAF | <input checked="" type="checkbox"/> Japan CS Mark Gold | <input checked="" type="checkbox"/> Spain ENS |
| <input checked="" type="checkbox"/> Canada Privacy Laws | <input checked="" type="checkbox"/> EU Model Clauses | <input checked="" type="checkbox"/> Japan My Number Act | <input checked="" type="checkbox"/> Spain DPA |
| <input checked="" type="checkbox"/> China GB 18030:2005 | <input checked="" type="checkbox"/> EU – US Privacy Shield | <input checked="" type="checkbox"/> Netherlands BIR 2012 | <input checked="" type="checkbox"/> UK G-Cloud |
| <input checked="" type="checkbox"/> China DJCP (MLPS) Level 3 | <input checked="" type="checkbox"/> Germany IT-Grundschutz workbook | <input checked="" type="checkbox"/> New Zealand Gov CIO Fwk | |

Azure Portfolio



Edge Devices

Azure Stack

Azure Data Box

Azure Sphere

Azure Kinect

HoloLens



Serverless

Web

Mobile

Mixed Reality

Containers

Events + Integration

Databases

Analytics

AI + Machine Learning

Internet of Things

Media



Tools

Visual Studio

GitHub

PowerApps

Power BI



Infrastructure

Compute

Networking

Storage

Security

Identity

Let's Go Deeper

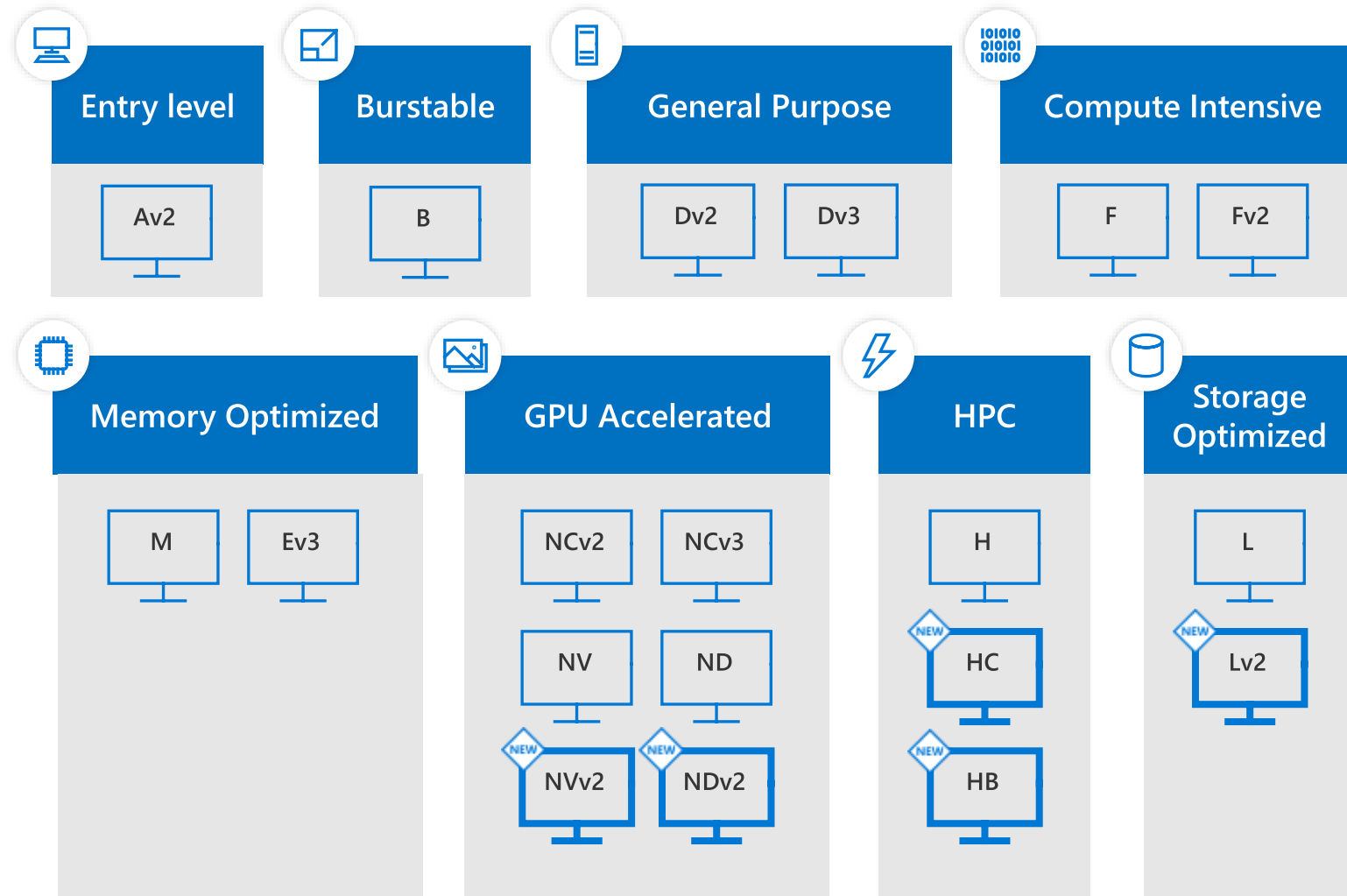


Azure Compute

(VM Families, pure Infrastructure as a Service)

Virtual machines

Purpose-built



Azure Data Stores

(Platform as a Service)

Tables of related data
with rows of identical
records (think of an
Excel sheet)

Sets of documents
with similar, but not
necessarily uniform
fields (think of a Word
document with an
outline)

Relational

Non-relational

Structured



SQL Server



SQL Managed
Instance
(compatibility)



Azure
SQL Database



Azure Database
for PostgreSQL



Azure Database
for MySQL



Azure Database
for MariaDB



Azure Synapse
Analytics
(formerly SQL DW)



Azure
Cosmos DB

Unstructured

Images, audio, video,
standalone files of any
type



Storage Accounts
(Blobs, Files,
Tables, Queues)

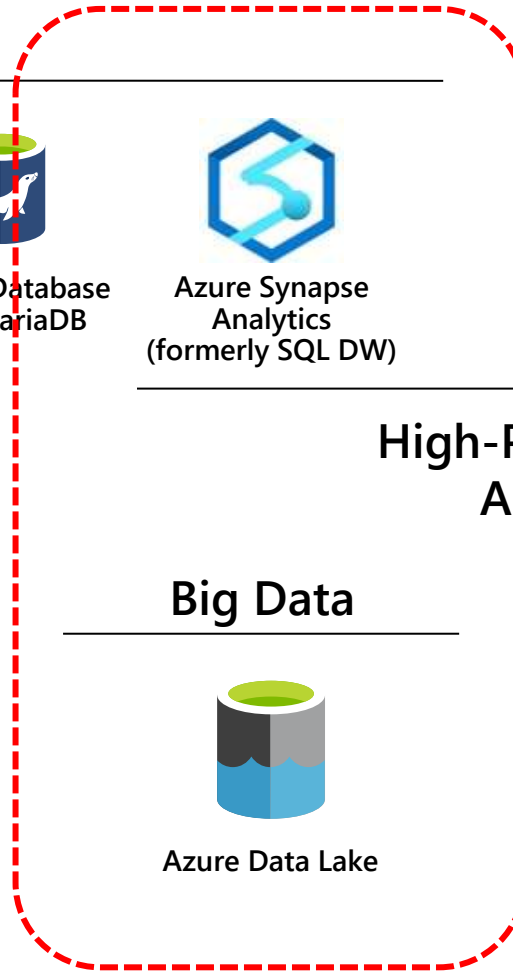
General Purpose

Big Data



Azure Data Lake

High-Performance Analytics



Machine Learning on Azure

Domain specific pretrained models

To simplify solution development



Vision



Speech



Language



Search

Familiar Data Science tools

To simplify model development



Visual Studio Code



Azure Notebooks



Jupyter



Command line

Popular frameworks

To build advanced deep learning solutions



PyTorch



TensorFlow



Scikit-Learn



ONNX

Productive services

To empower data science and development teams



Azure
Databricks



Azure Machine
Learning



Machine
Learning VMs

Powerful infrastructure

To accelerate deep learning



CPU



GPU



FPGA



From the Intelligent Cloud to the Intelligent Edge



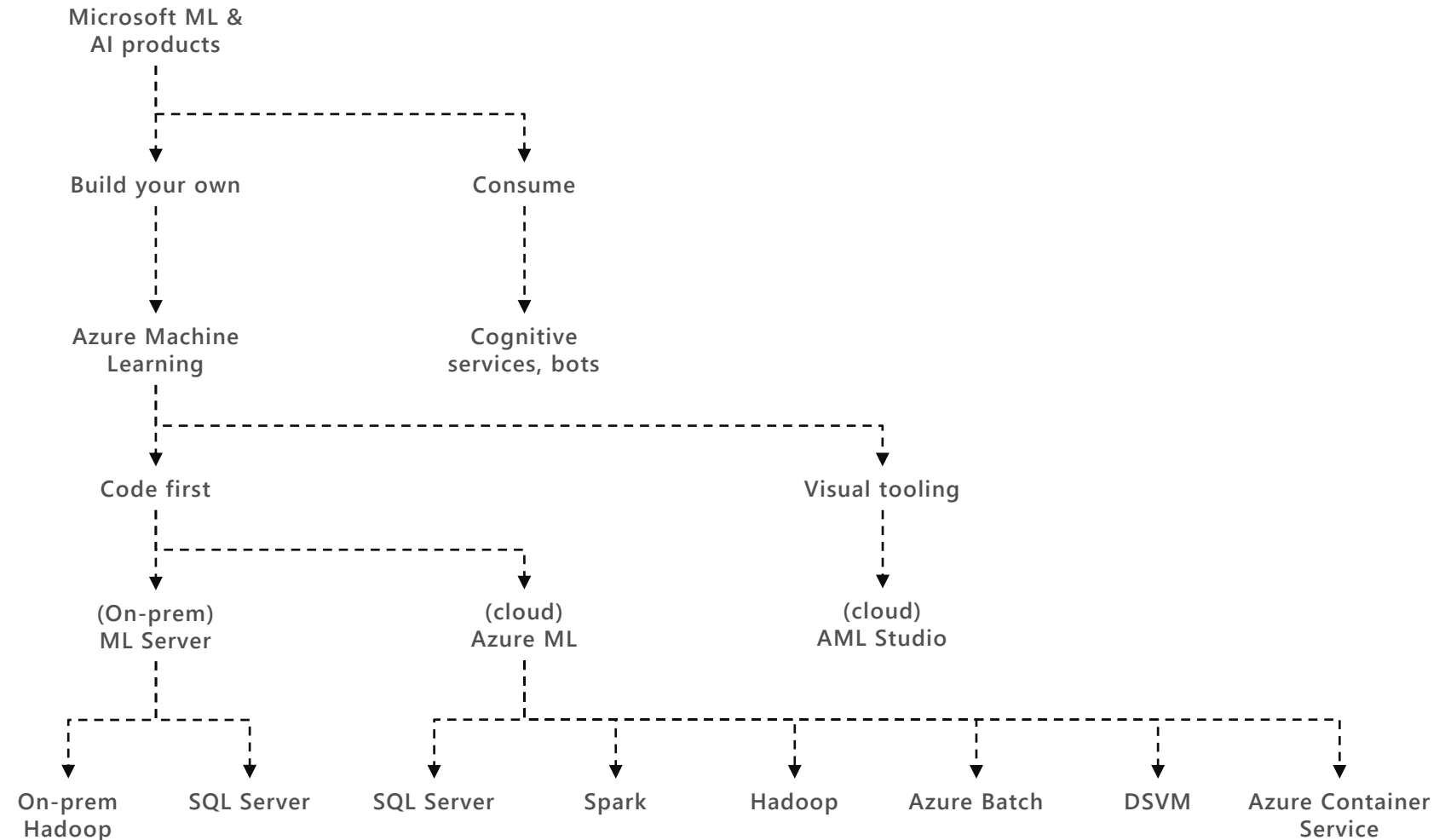
That is too much stuff! Which should I use?

Build your own or consume pre-trained models?

Which experience do you want?

Deployment target

What data engine(s) do you want to use?
(typical choices)



Familiar Data Science tools

Choose any python development environment



Visual Studio Code



Azure Notebooks



Jupyter



PyCharm

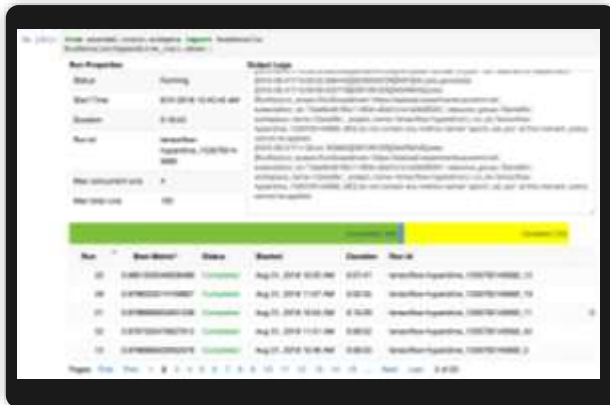


Zeppelin

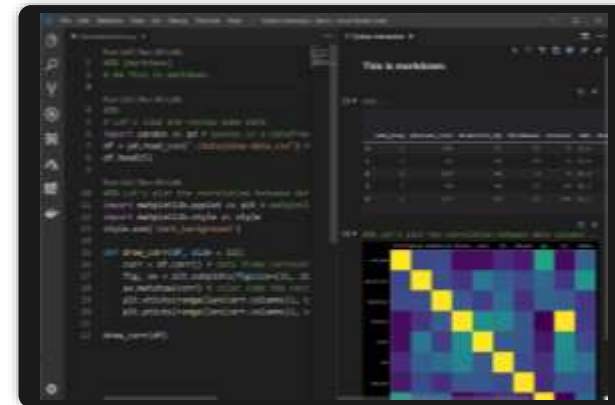


Command line

And improve data science productivity



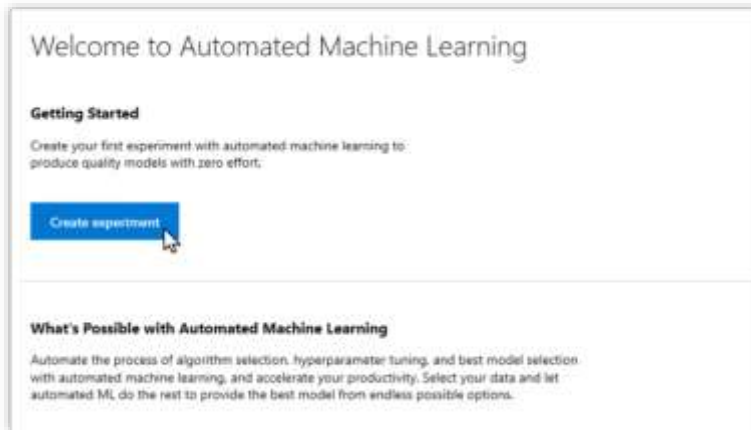
Interactive widgets for Jupyter Notebooks



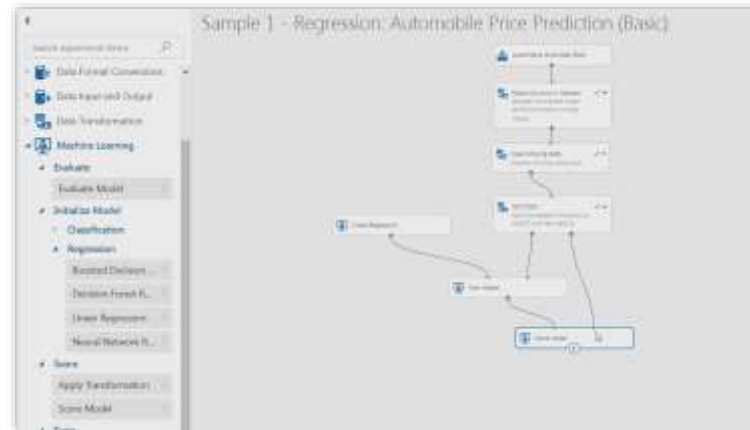
Azure Machine Learning for Visual Studio Code extension

➤ Get started with AML on Azure Notebooks: <http://aka.ms/aznotebooks-aml>

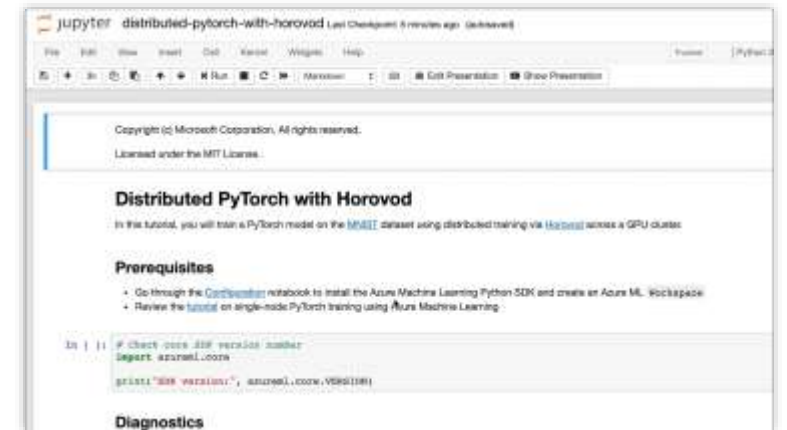
Machine learning for any skill level



Automated
machine learning UI

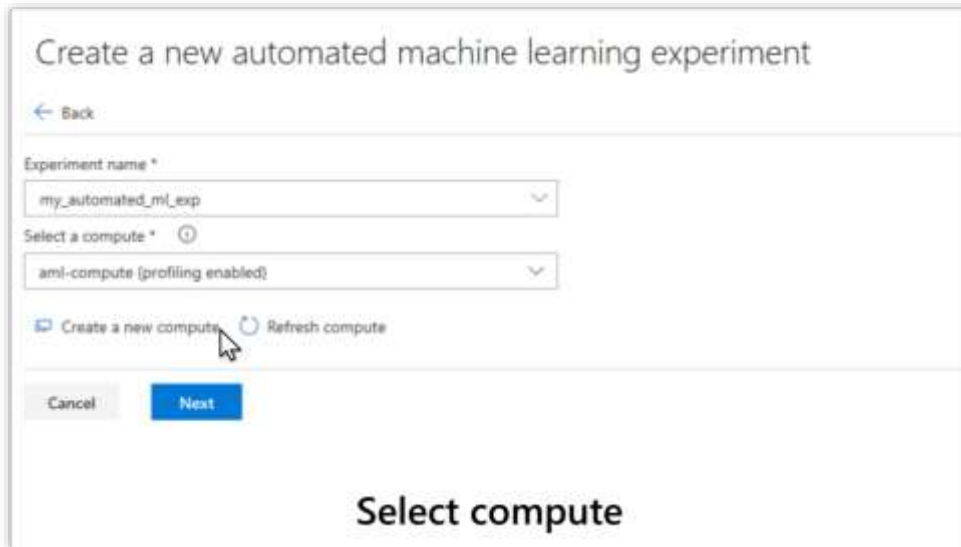


Visual interface



Machine learning notebooks

Machine learning for any skill level



Create a new automated machine learning experiment

← Back

Experiment name *

my_automated_ml_exp

Select a compute * ⓘ

aml-compute (profiling enabled)

Create a new compute Refresh compute

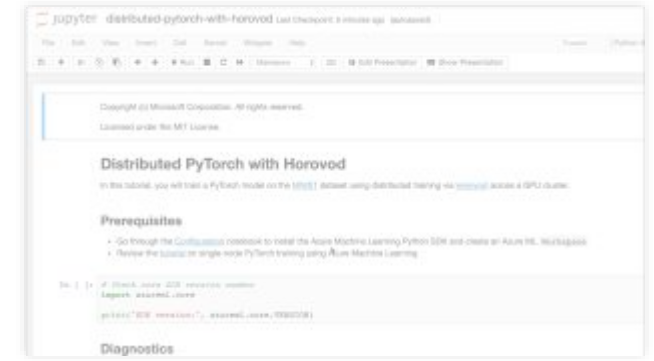
Cancel Next

Select compute

Automated
machine learning UI



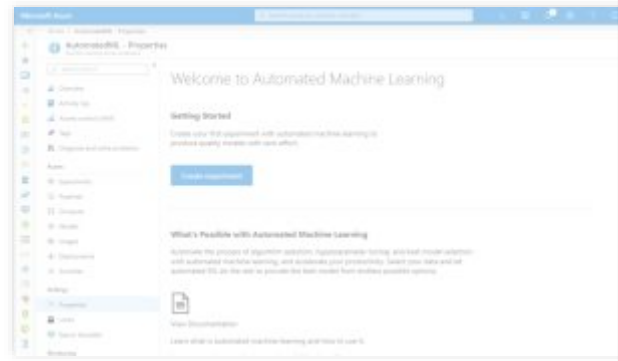
Visual interface



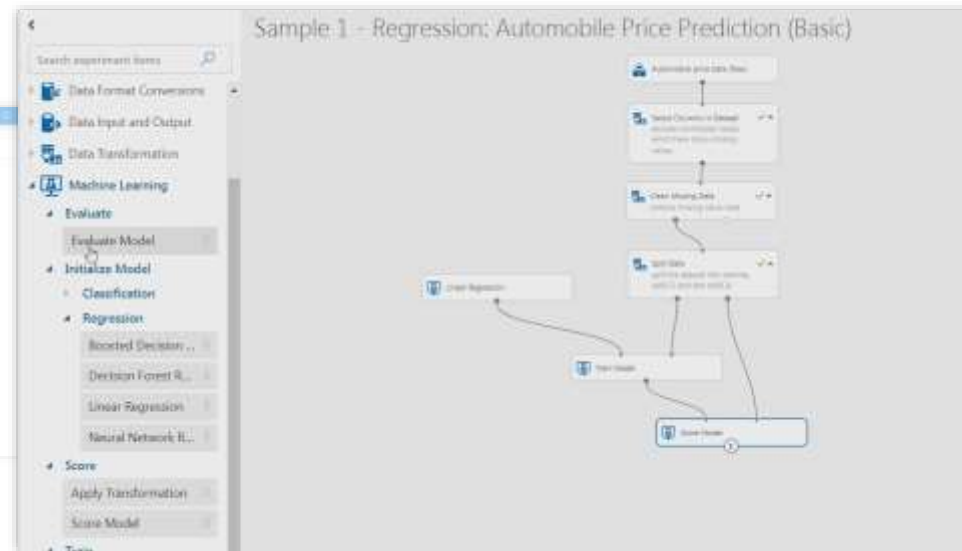
Machine learning notebooks

Machine learning for any skill level

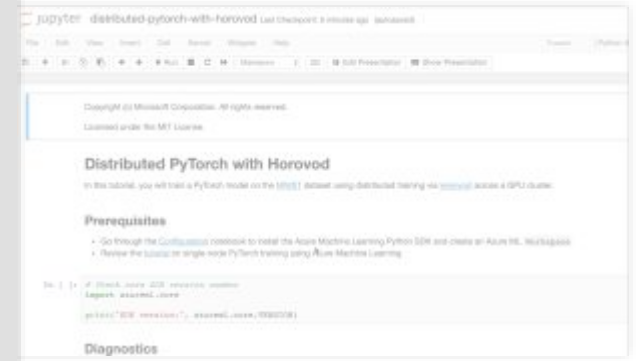
New capabilities in Azure Machine Learning service



Automated
machine learning UI



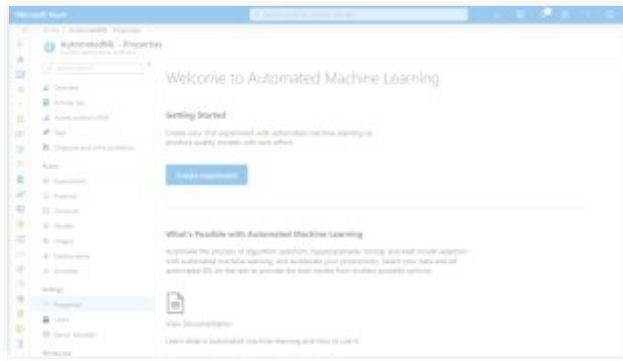
Visual interface



Machine learning notebooks

Machine learning for any skill level

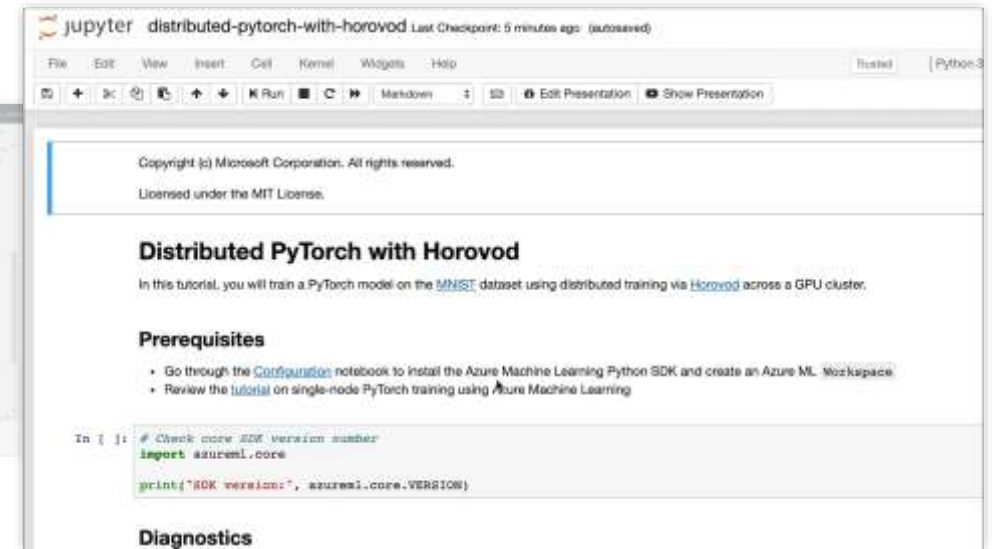
New capabilities in Azure Machine Learning service



Automated
machine learning UI



Visual interface

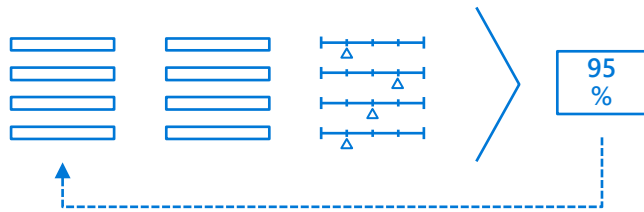


Machine learning notebooks

Differentiators

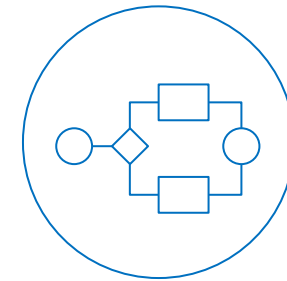
Machine Learning

Automated machine learning



Accelerated model building

Machine learning DevOps

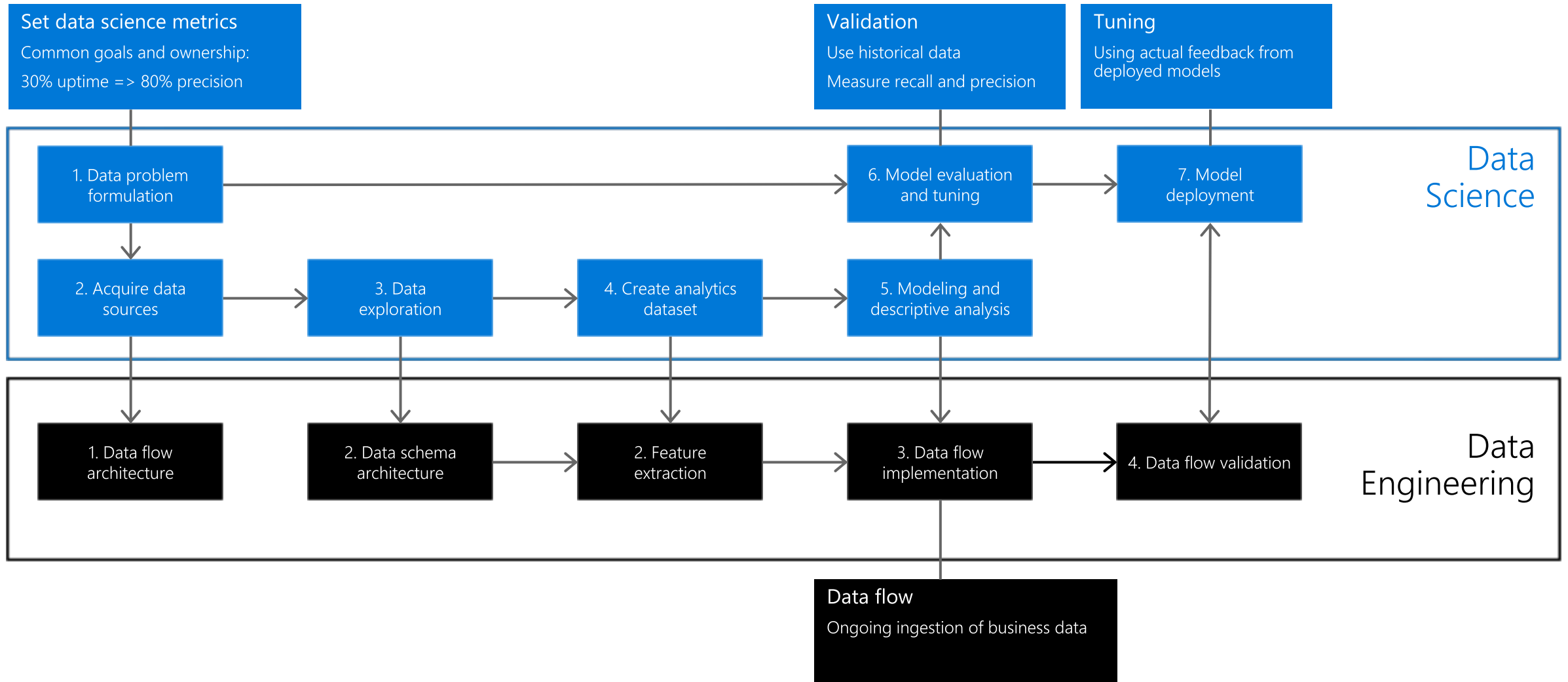


Azure DevOps integration for CI/CD

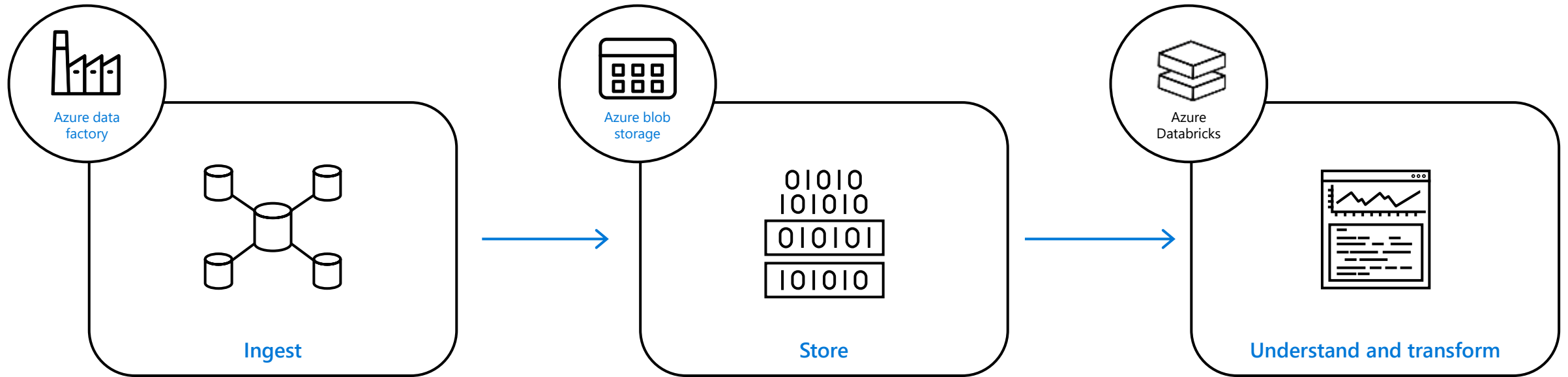
Processes & Pipelines



Data Science vs Data Engineering

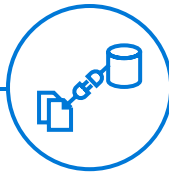


Typical Pipeline (now folded into Azure DataFlow)



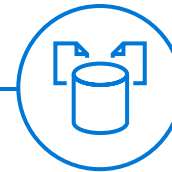
Connect to data from any source

Integrate with all of your data sources
Create hybrid pipelines
Orchestrate in a code-free environment



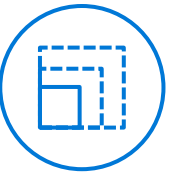
Leverage best-in-class analytics capabilities

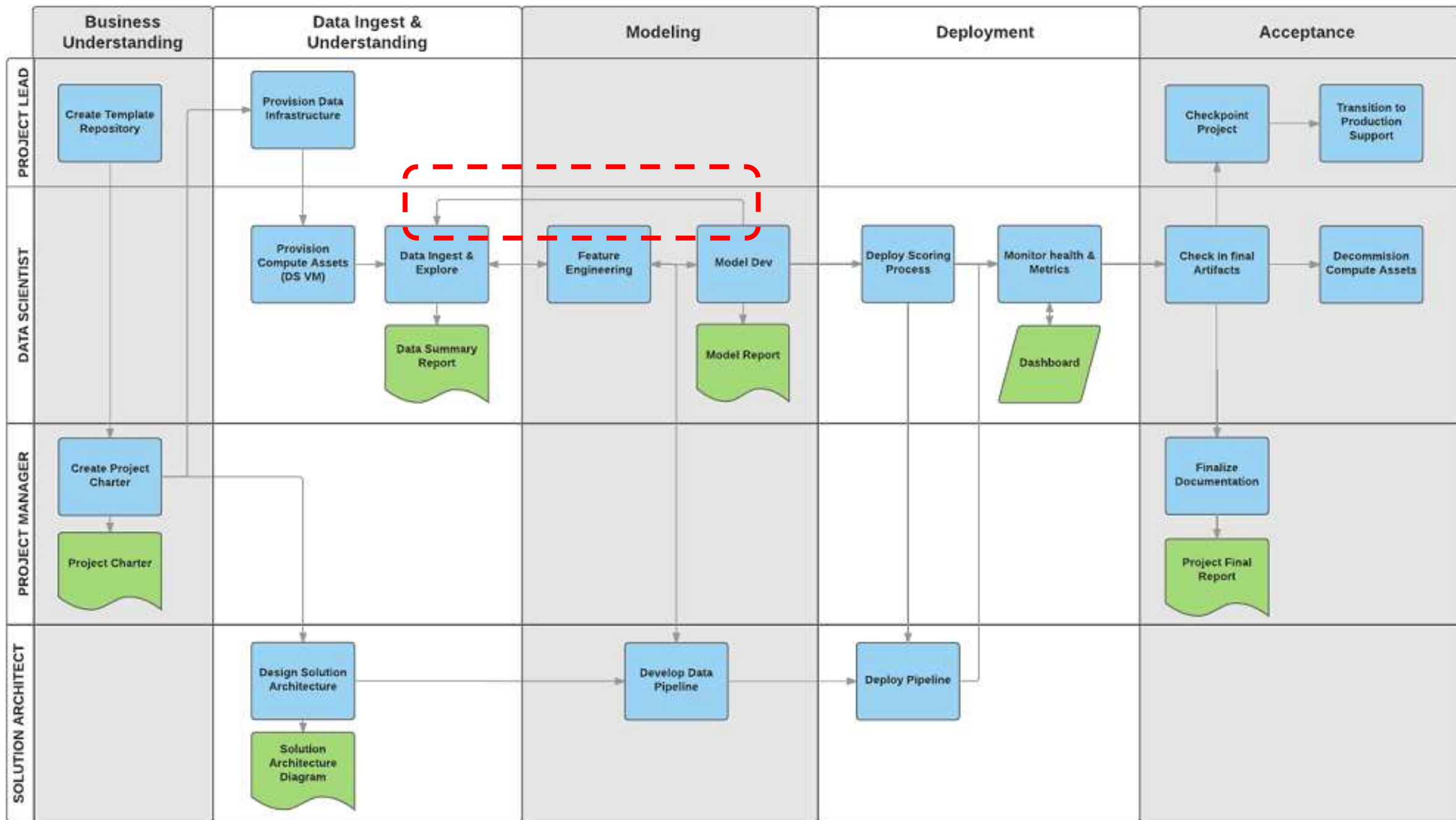
Leverage open source technologies
Collaborate within teams
Use ML (machine learning) on batch streams



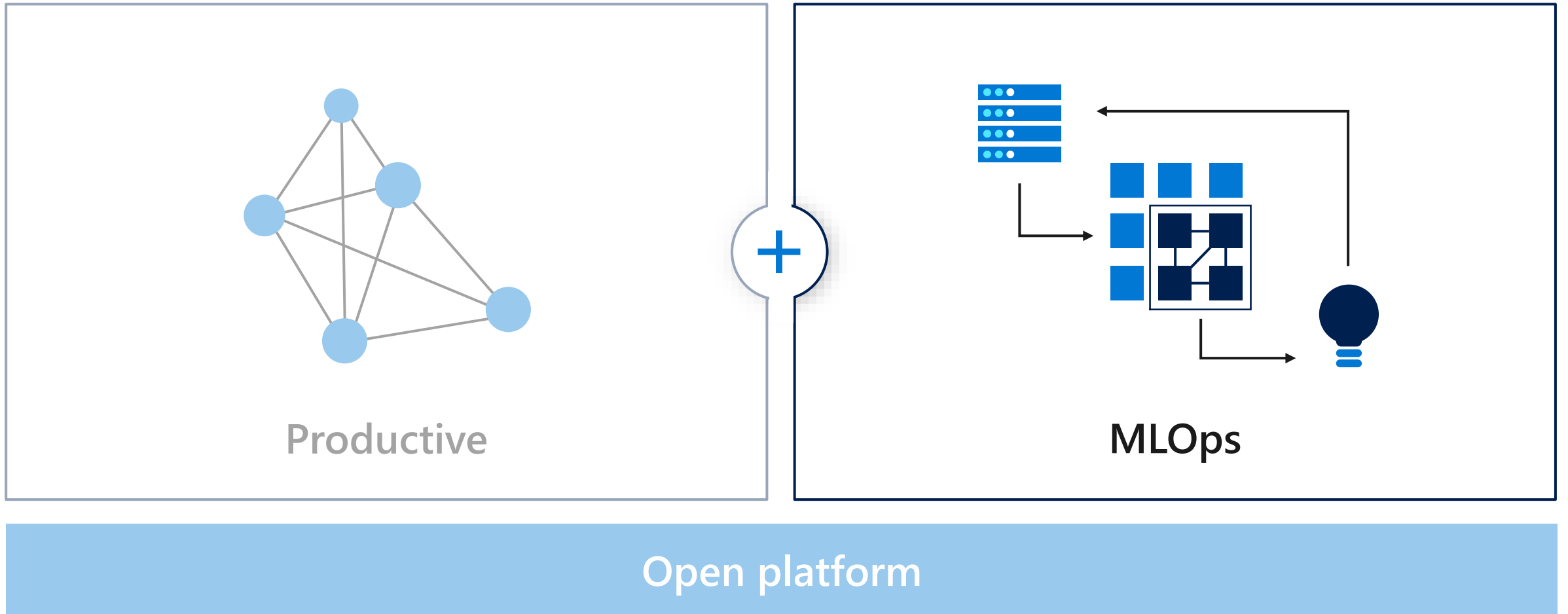
Scale without limits

Build in the language of your choice
Leverage scale out topology
Scale compute and storage separately





Azure Machine Learning service



Azure Machine Learning accelerates model development

with automated machine learning

Input

101010
010101
101010

Enter data

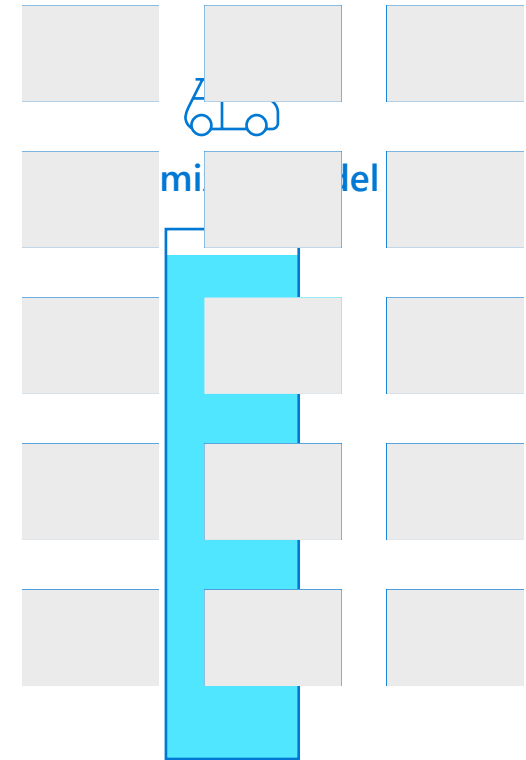
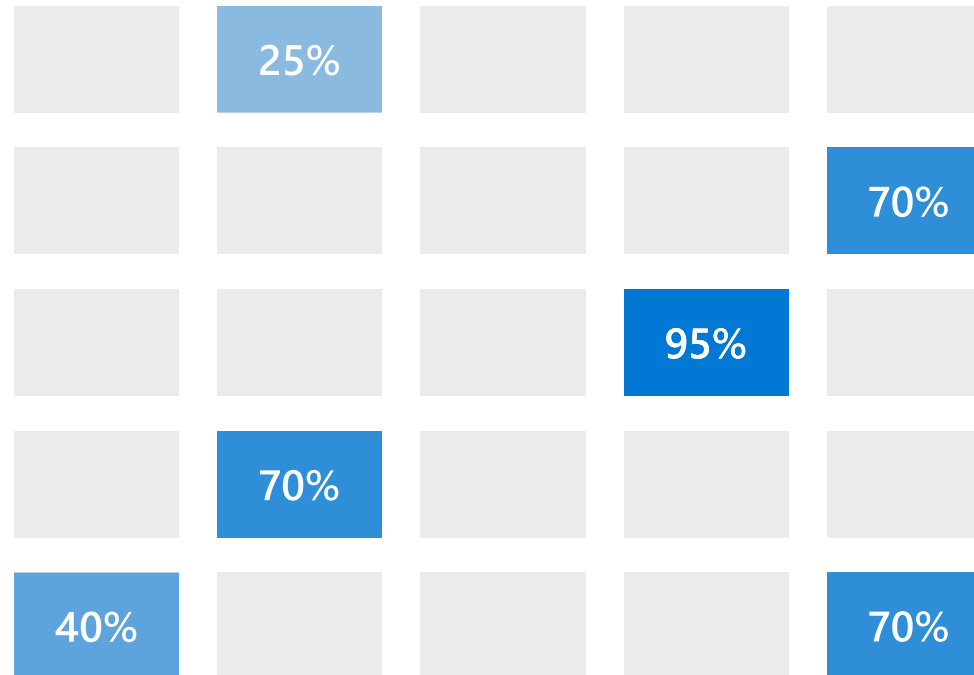


Define goals



Apply constraints

Intelligently test multiple models in parallel



DevOps



Code reproducibility



Code testing



App deployment

MLOps



Model reproducibility



Model validation

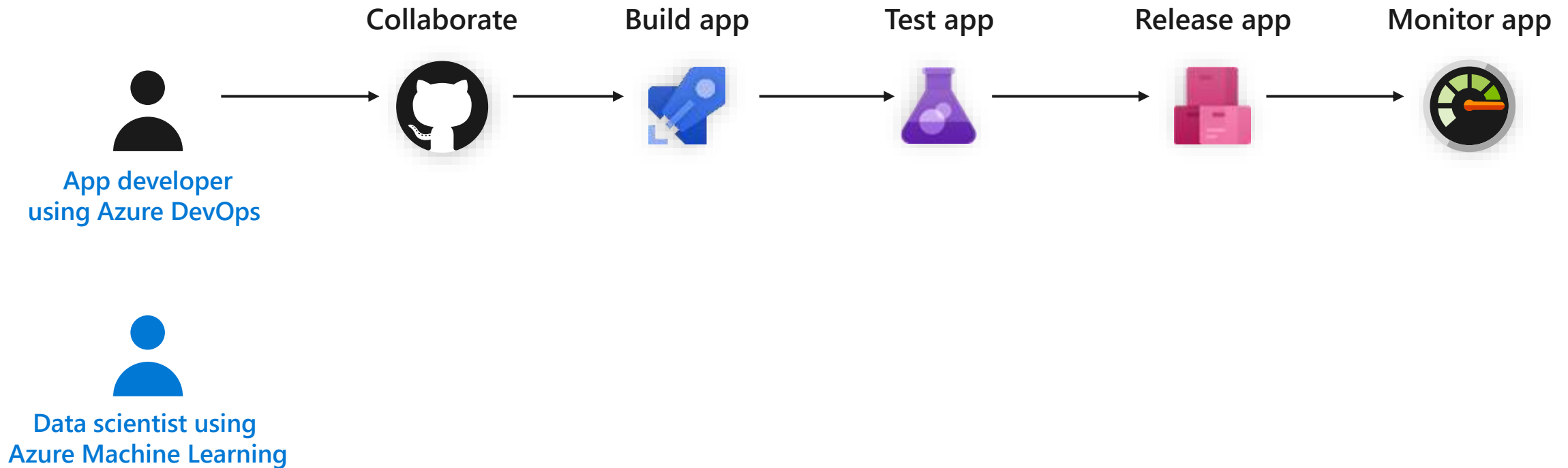


Model deployment



Model retraining

MLOps with Azure Machine Learning



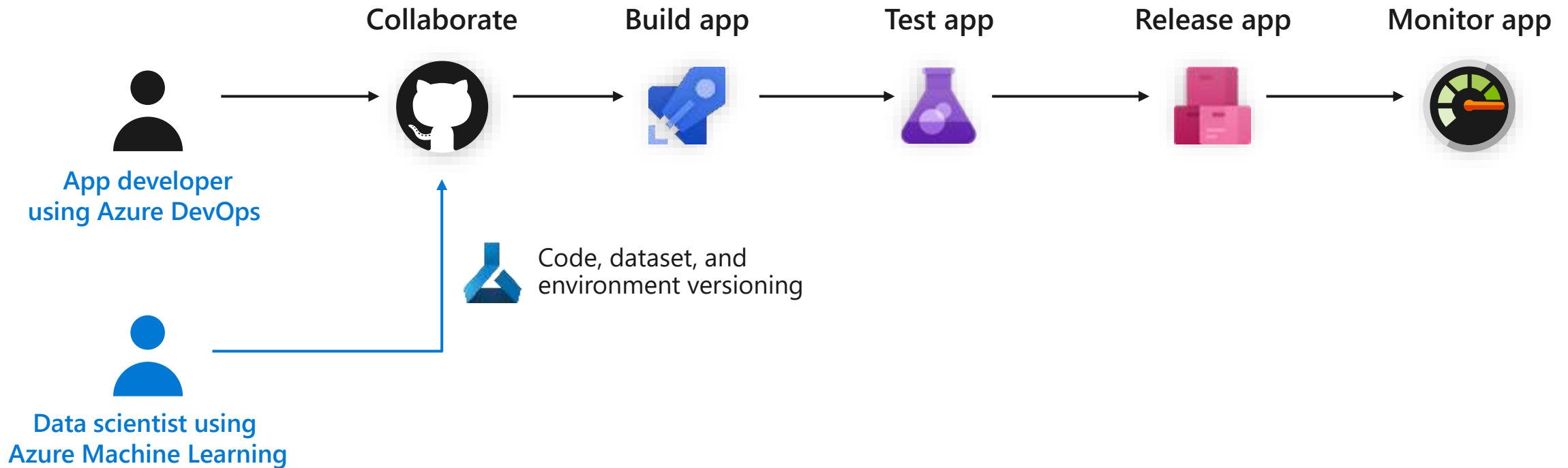
☐ Model reproducibility

☐ Model validation

☐ Model deployment

☐ Model retraining

MLOps with Azure Machine Learning



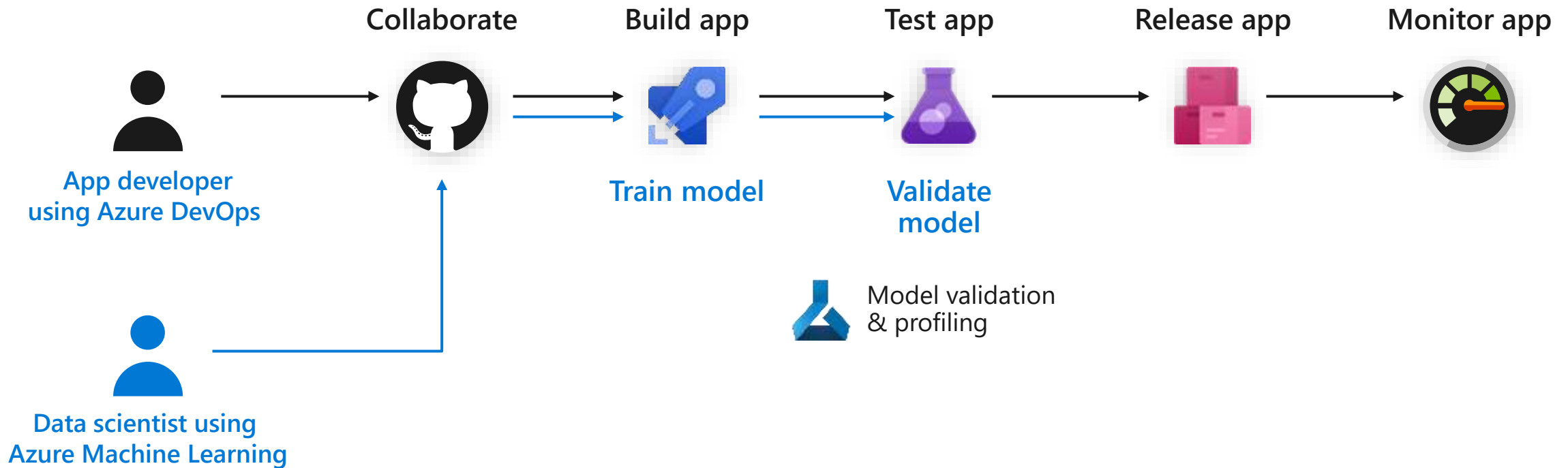
☒ Model reproducibility

☐ Model validation

☐ Model deployment

☐ Model retraining

MLOps with Azure Machine Learning



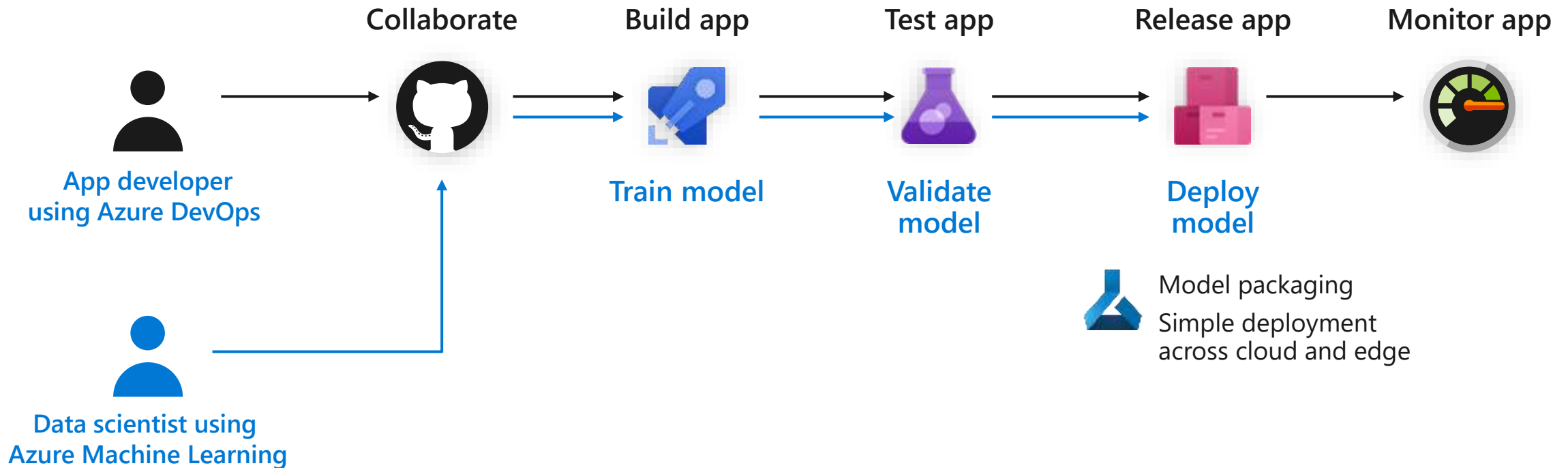
☒ Model reproducibility

☒ Model validation

☐ Model deployment

☐ Model retraining

MLOps with Azure Machine Learning



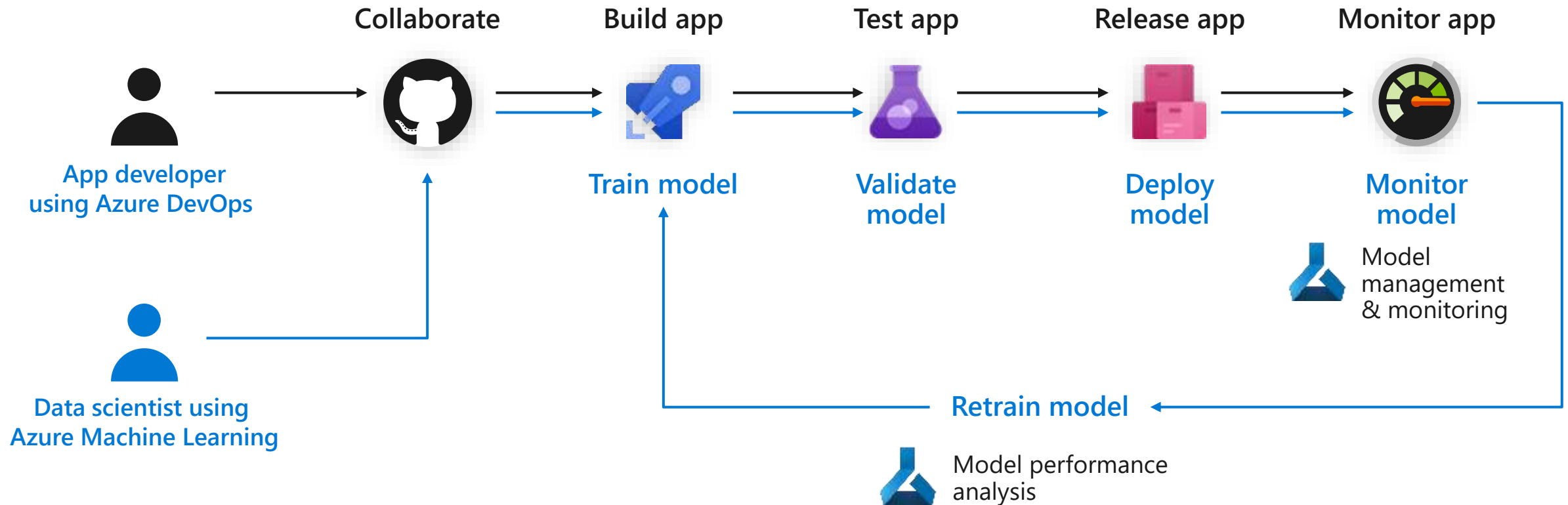
☒ Model reproducibility

☒ Model validation

☒ Model deployment

☐ Model retraining

MLOps with Azure Machine Learning



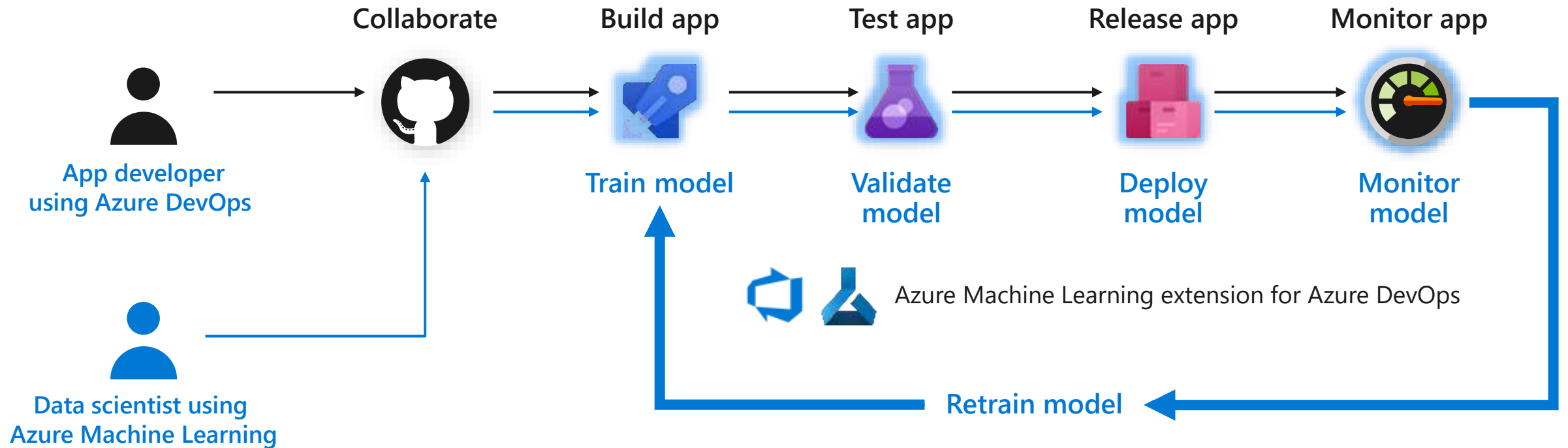
✓ Model reproducibility

✓ Model validation

✓ Model deployment

✓ Model retraining

MLOps with Azure Machine Learning



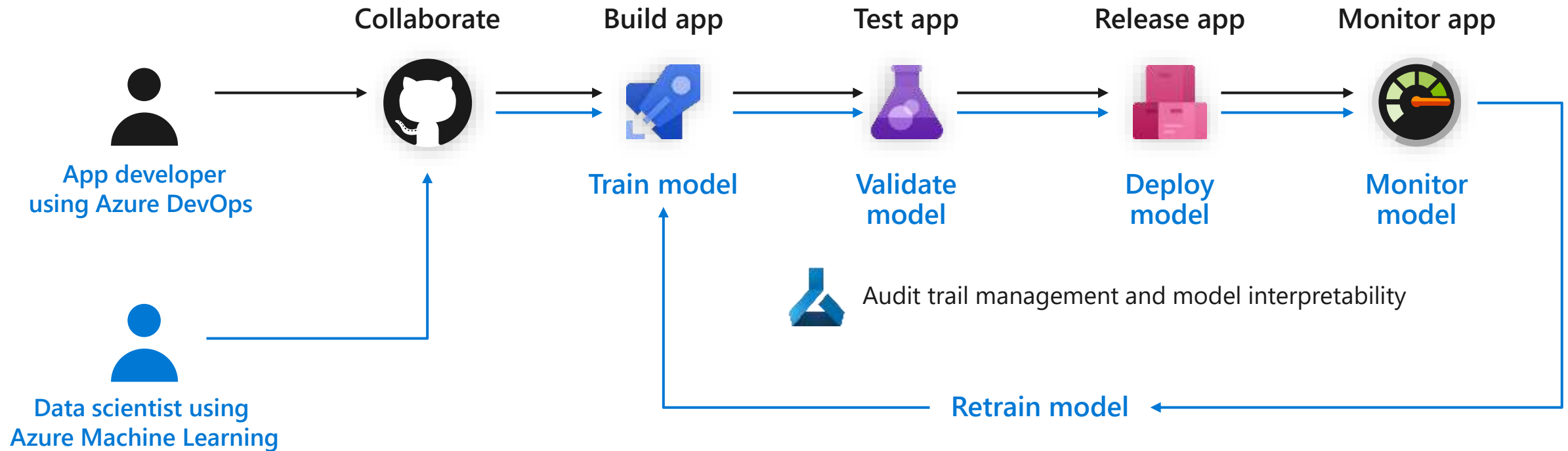
✓ Model reproducibility

✓ Model validation

✓ Model deployment

✓ Model retraining

MLOps with Azure Machine Learning



✓ Model reproducibility

✓ Model validation

✓ Model deployment

✓ Model retraining

Modern Data Warehousing



BIG DATA, BIG BLOCKS

Relational data



OLTP



ERP



CRM



LOB

Non-relational data



Web



Media



Social media



Devices

Data virtualization

Data warehousing

Big Data processing

Any BI tool

Dashboards | Reporting
Mobile BI | Cubes

Advanced Analytics

Machine Learning
Stream analytics | Cognitive | AI

Any language

.NET | Java | R | Python
Ruby | PHP | Scala



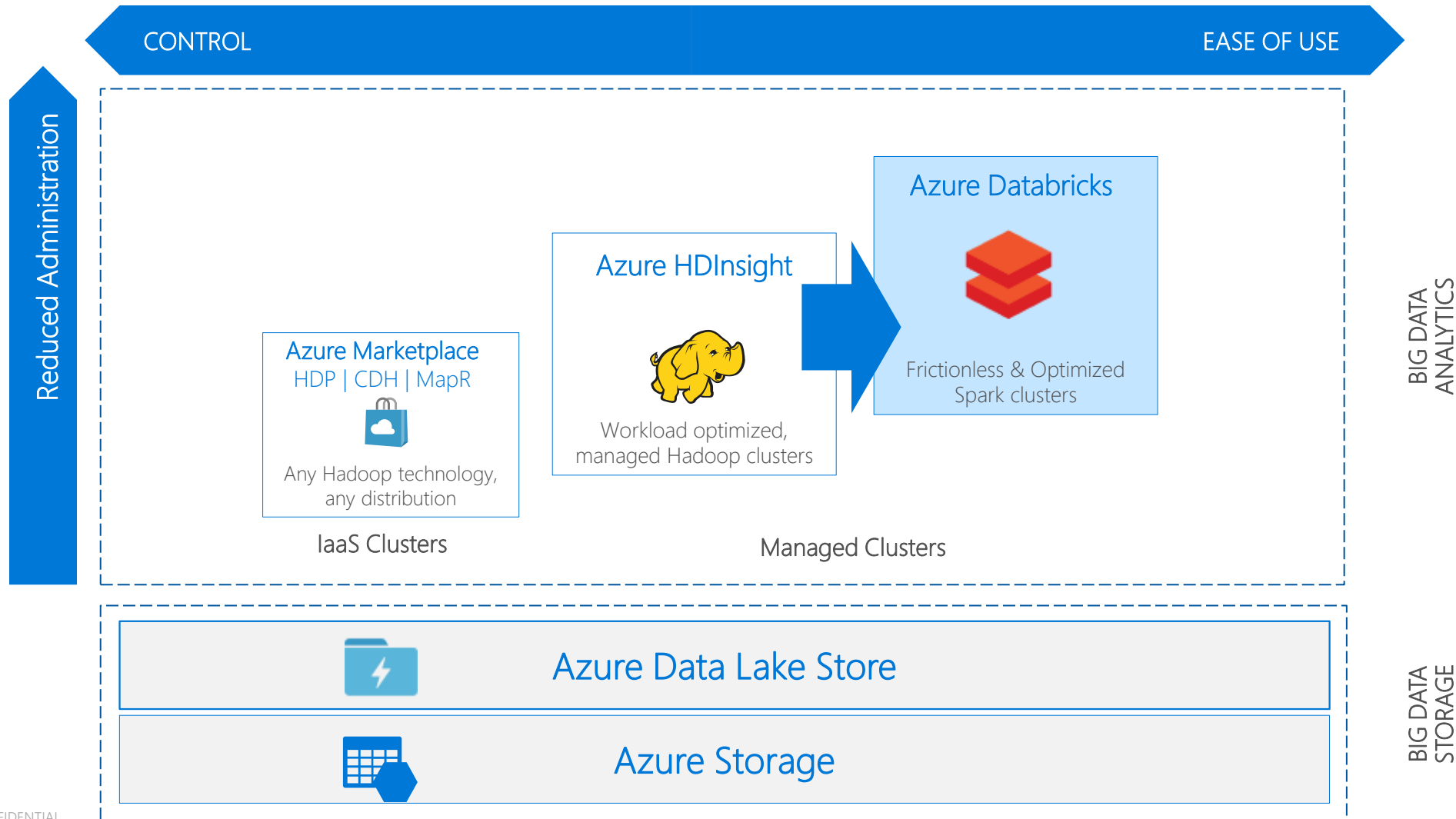
On-premises

Cloud



Building Blocks

How we see Big Data on Azure today



Data Warehouse Paradigms (Hadoop vs Spark)



Hadoop is a good example for old-school data warehousing designed for conventional datacenters

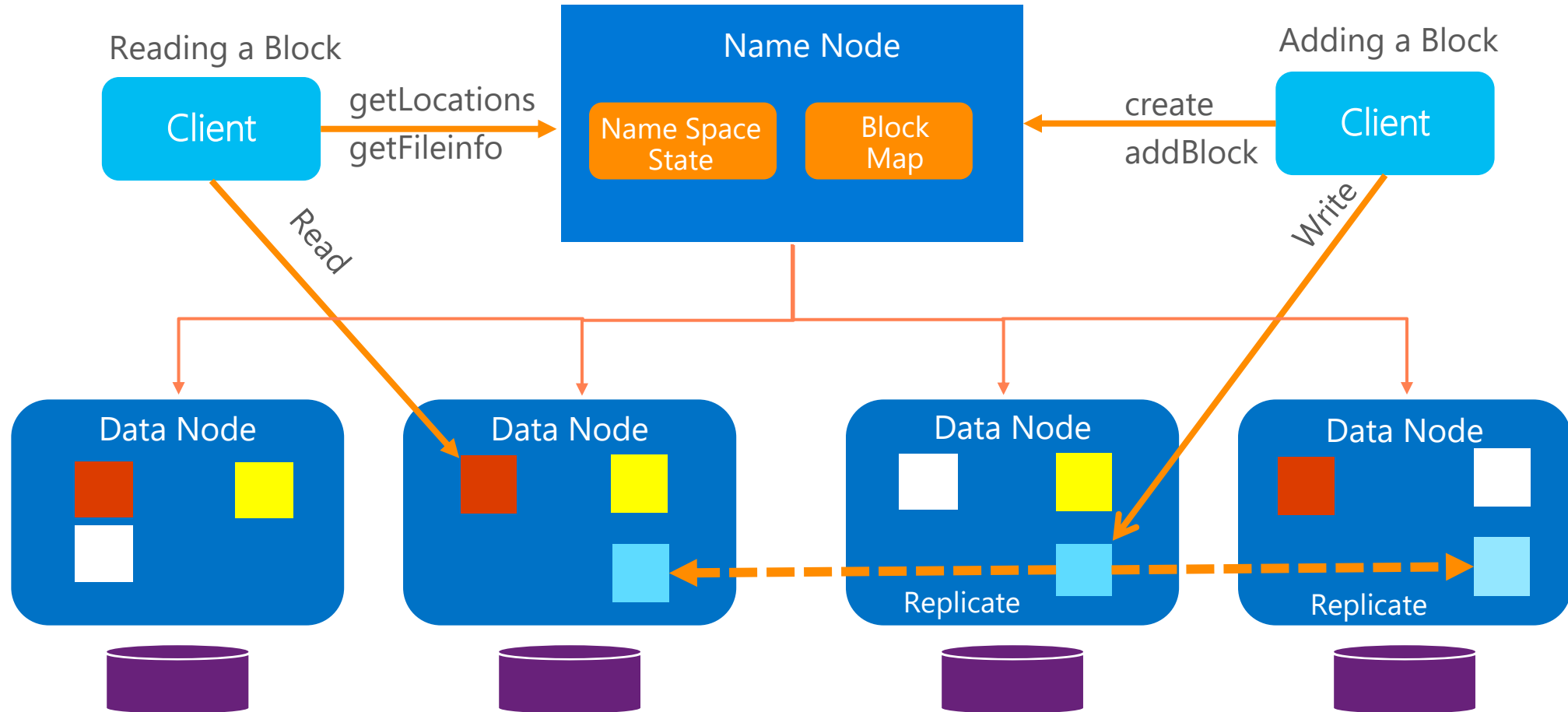
- Designed for redundancy and built-in storage
- Clusters typically run 24/7, and it is expensive to have more than one
- Buy/build all the CPU and storage you need in advance
- Storage is (roughly) triple the amount of live data



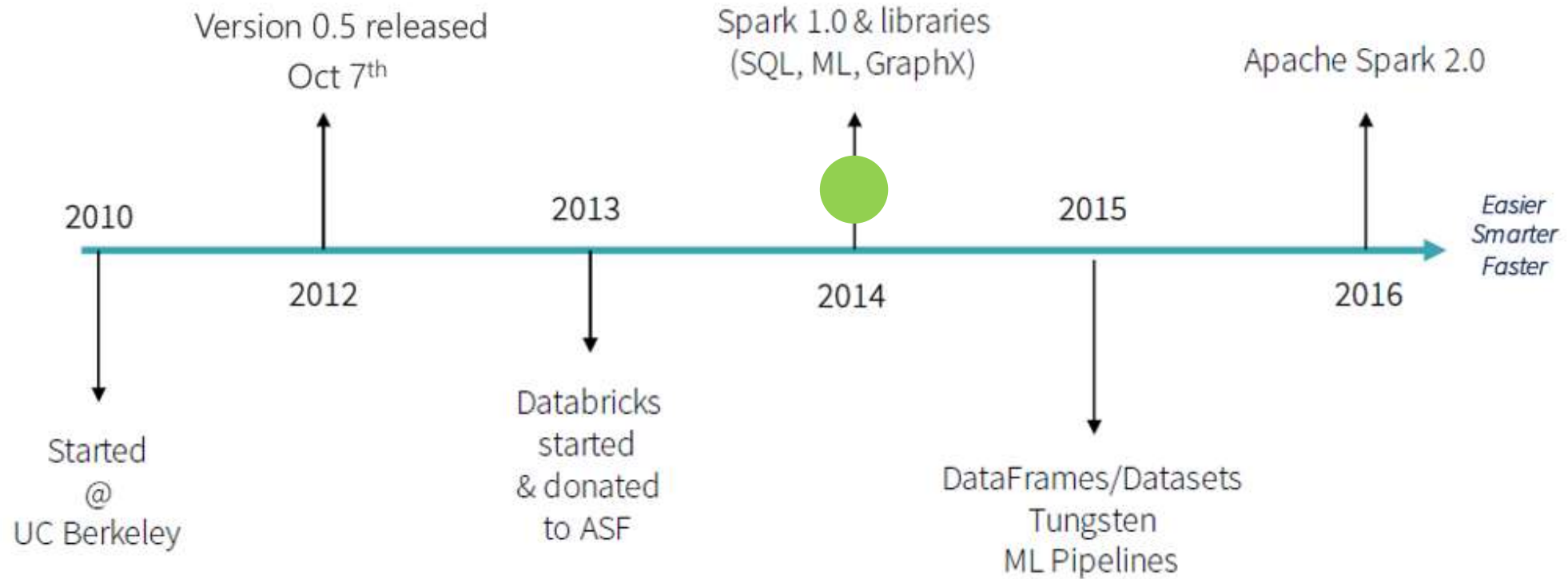
Spark/Databricks is a different paradigm that is more suited to the cloud

- Designed for performance and access to external storage
- Clusters typically run on-demand, and you can run many against the same storage
- One-click creation of “right-sized” ephemeral clusters
- Pay only for the storage you actually need for live data

Apache Hadoop Architecture

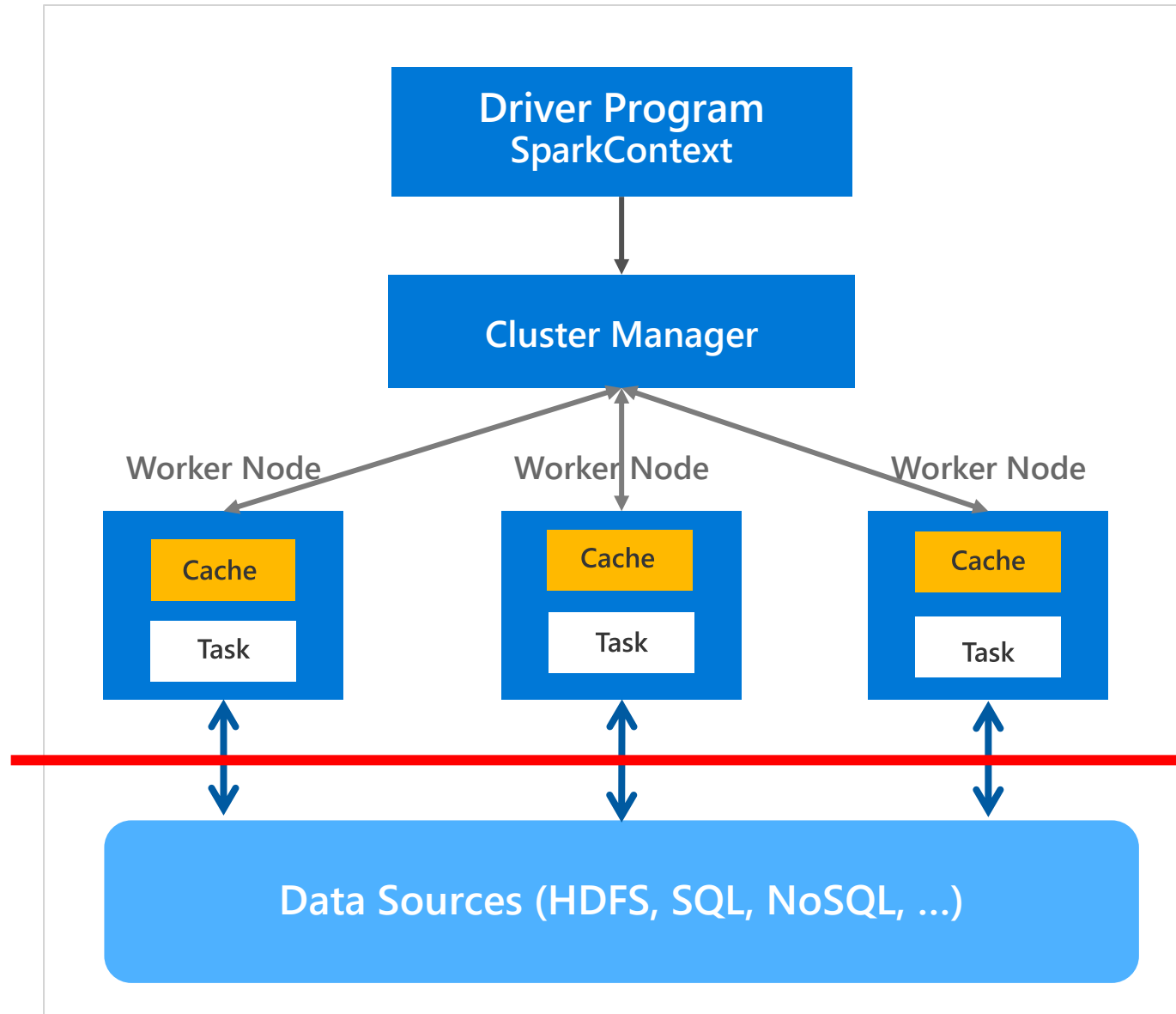


Spark: A Brief History

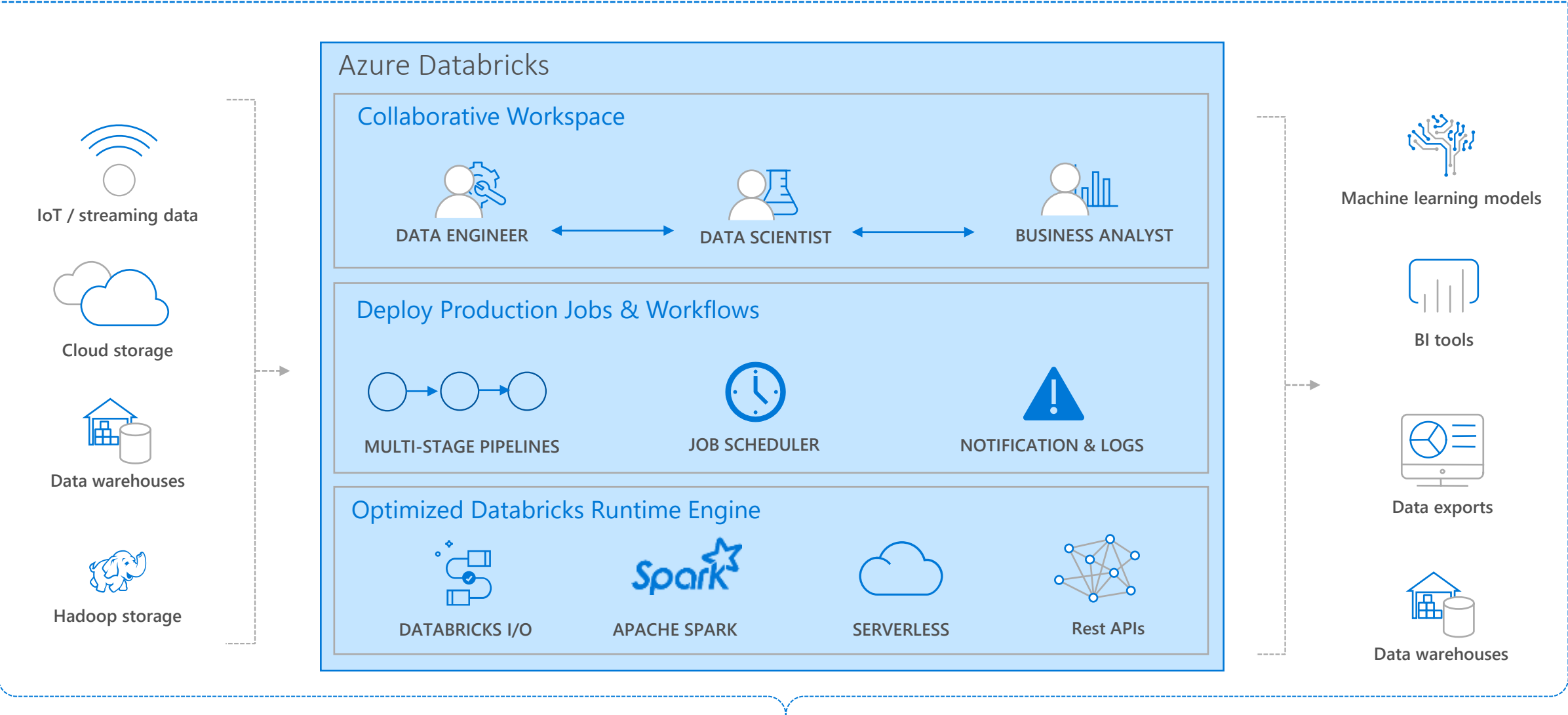


Apache Spark Architecture

- 'Driver' runs the user's 'main' function and executes the various parallel operations on the worker nodes.
- The results of the operations are collected by the driver
- The worker nodes read and write data from/to Data Sources including HDFS.
- Worker node also cache transformed data in memory as RDDs (Resilient Data Sets). Spark is especially suited for **distributed in-memory data processing**.
- Worker nodes and the Driver Node execute as throwaway VMs in the cloud
- Databricks provides a managed Spark service that allows you to build and tear down clusters automatically
- **Storage is completely separate** from the cluster when deployed on Azure



Spark/Databricks - Azure Managed Service



Enhance Productivity

Build on secure & trusted cloud

Scale without limits

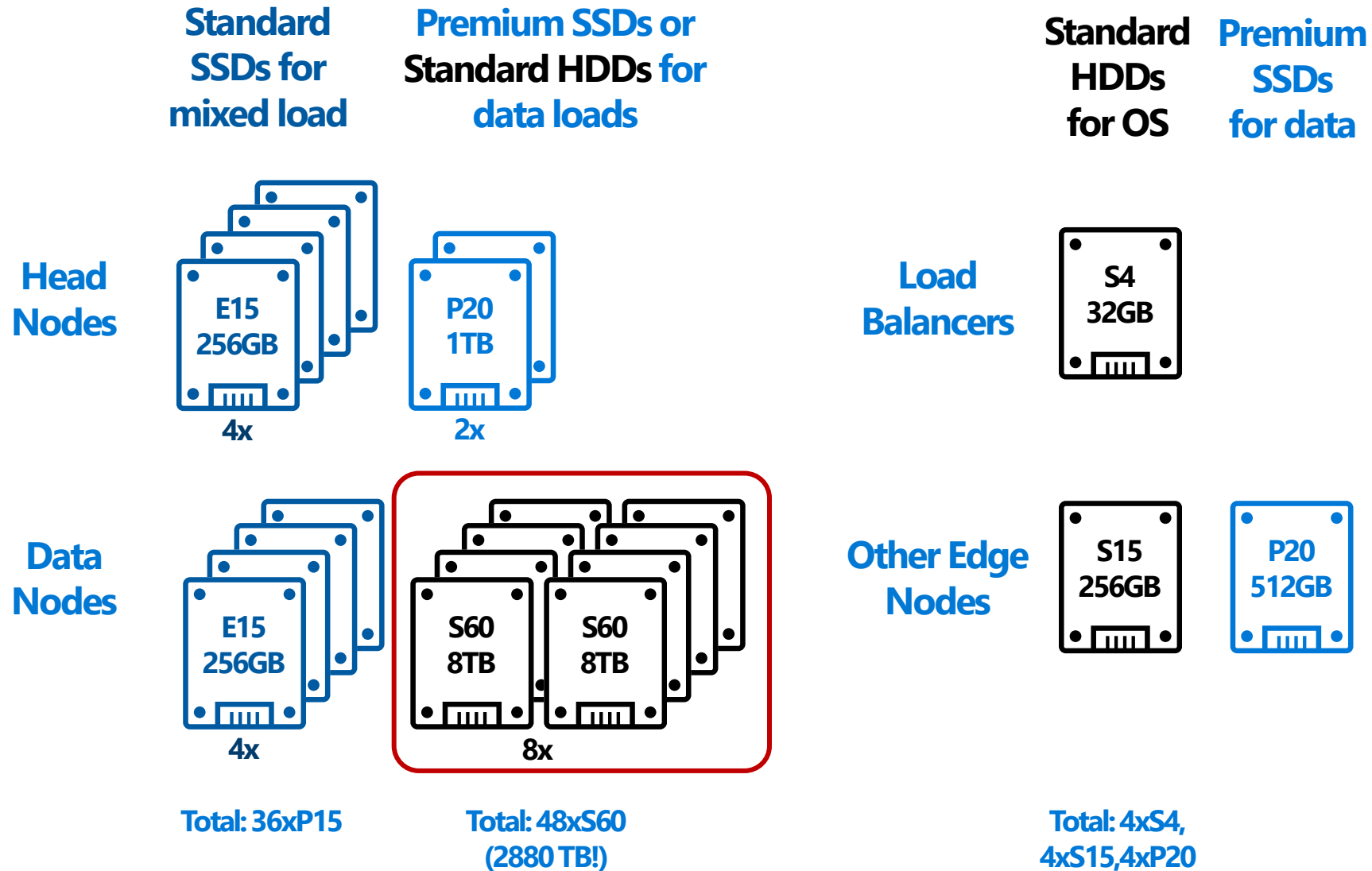
Spark vs Hadoop on VMs

Databricks	Third-Party Hadoop on VMs (not HDInsight)
PaaS (zero-touch setup, scale, manage, patch, etc.)	IaaS (you maintain and patch the VMs)
Managed by Microsoft	Managed by customer
Storage is completely separate (Blob or ADLS) and can scale independently	Storage in VM (local disk), but can also have storage in Azure blob or ADLS
Delete VM keeps data	Delete VM deletes data (unless external)
Automatic, seamless auto scaling (during job execution)	Manual scaling (need to stop jobs)
Up to 30 days behind latest stable Apache Spark release	Hadoop version supplied by vendor
Microsoft supports VM and Databricks	Microsoft: VM, Hadoop: Third-Party Support
Python, R, SQL, Scala, SparkML	Impala, Hive, PiG,....
No on-prem version (build your own Spark)	On-prem version
Business continuity through globally redundant storage in Data Lake	Roll your own disaster recovery (IaaS-based)

A Standard Hadoop Deployment

VM Group	CPUxCores	RAM	Storage	VM Series	T-Shirt S	T-Shirt M	T-Shirt L
3 Head Nodes	1x14 HT	128GB	2x300GB SSD+ 2x200GB SSD	Dv3 (HT) or E (in-mem HT)	D16sv3 (64GB)	L16sv2 (128GB)	L16sv2 (128GB)
6 Worker Nodes	2x14 HT	512GB	4x200GB SSD+ 10x6TB HDD	Dv3 (HT) or E (in-mem HT), L (storage opt.)	E32v3 (256GB)	E64v3 (432GB)	L64sv2 (512GB)
4 Load Balancers	1x2	2GB	30GB HDD	B (burstable) or F (compute optimized)	B1ms (2GB)	F1 (2GB)	F1 (2GB)
2 ETL	1x16	64GB	200GB HDD+ 500GB SSD	F or Dv3	F16sv2 (32GB)	D16sv3 (64GB)	D16sv3 (64GB)
2 CM+CN	1x8	32GB	200GB HDD+ 500GB SSD	B or Dv3	B8ms (32GB)	D8sv3 (32GB)	D8sv3 (32GB)

Hadoop Storage Scenario



The PaaS Alternative: Spark (Databricks) Clusters

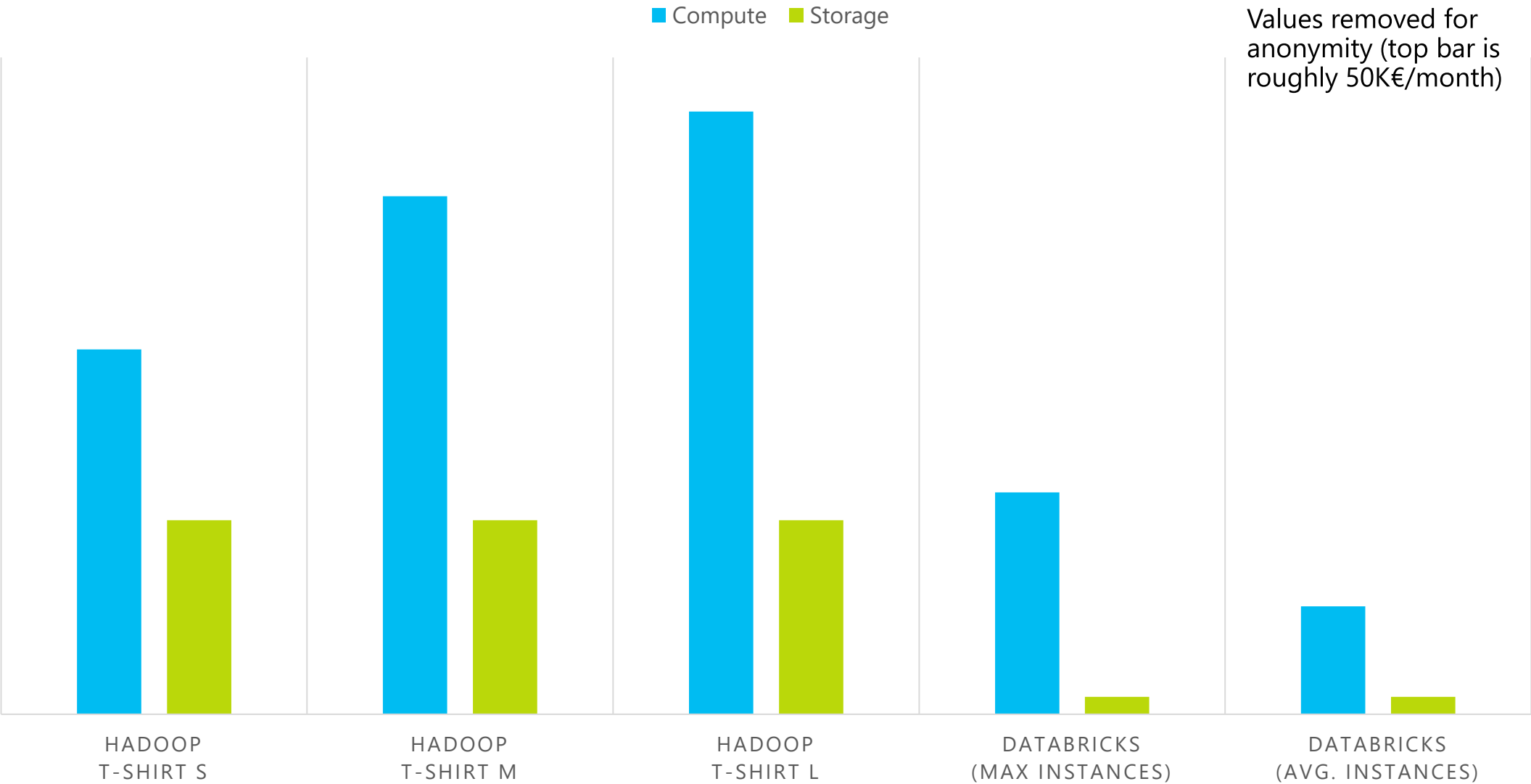
Typical Profiles

Profile	Instance Size	Average Instances	Max Instances	Hours
High-Performance batches	E32v3 (256GB)	4	9	480h (~30x16h)
Data Science	L8 (64GB)	2	6	160h (20x8h)
Data Science Light	L4 (32GB)	1	3	160h (20x8h)
Streaming (CDC, etc.)	F4 (8GB)	2	2	730h (~30x24h)

Average vs Max:

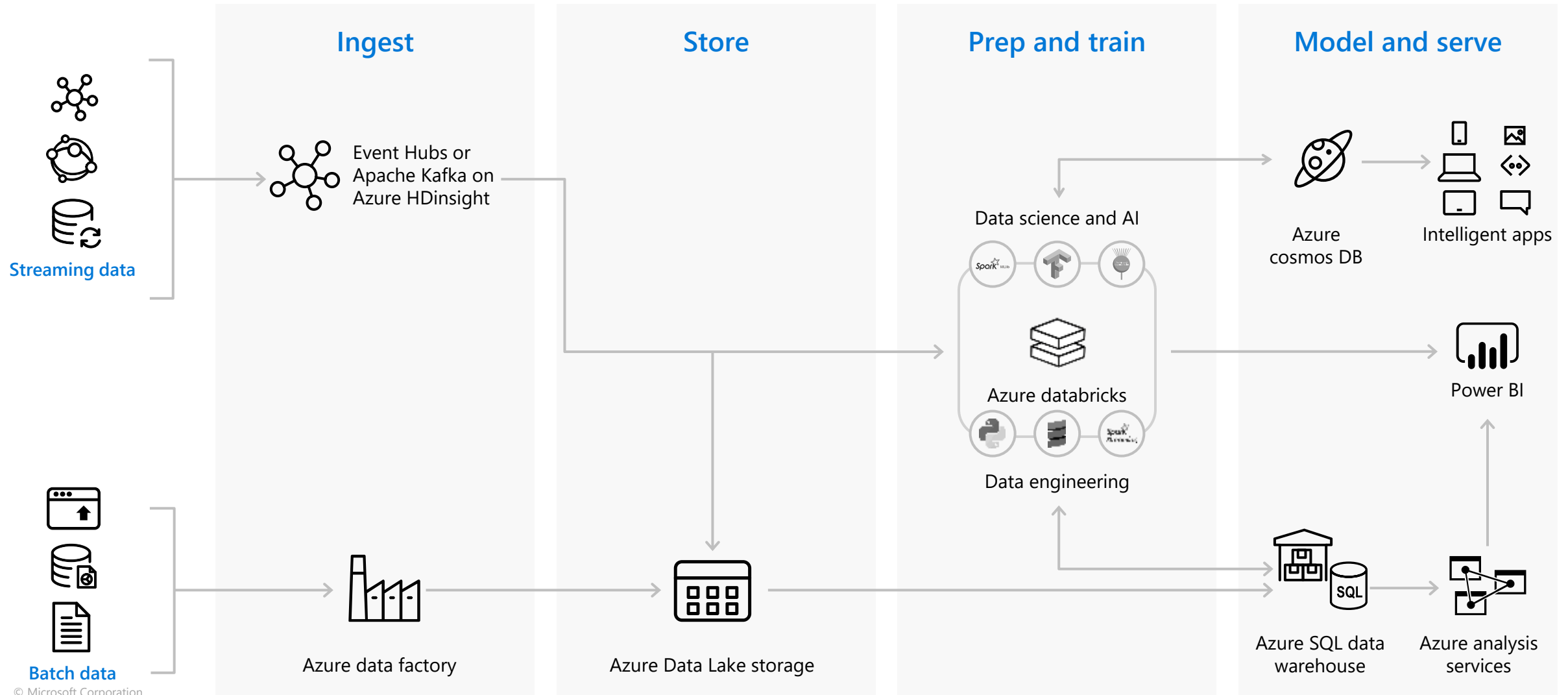
You can run more machines when required (peak usage), but you'd typically run less

Cost Comparison (Compute & Storage)

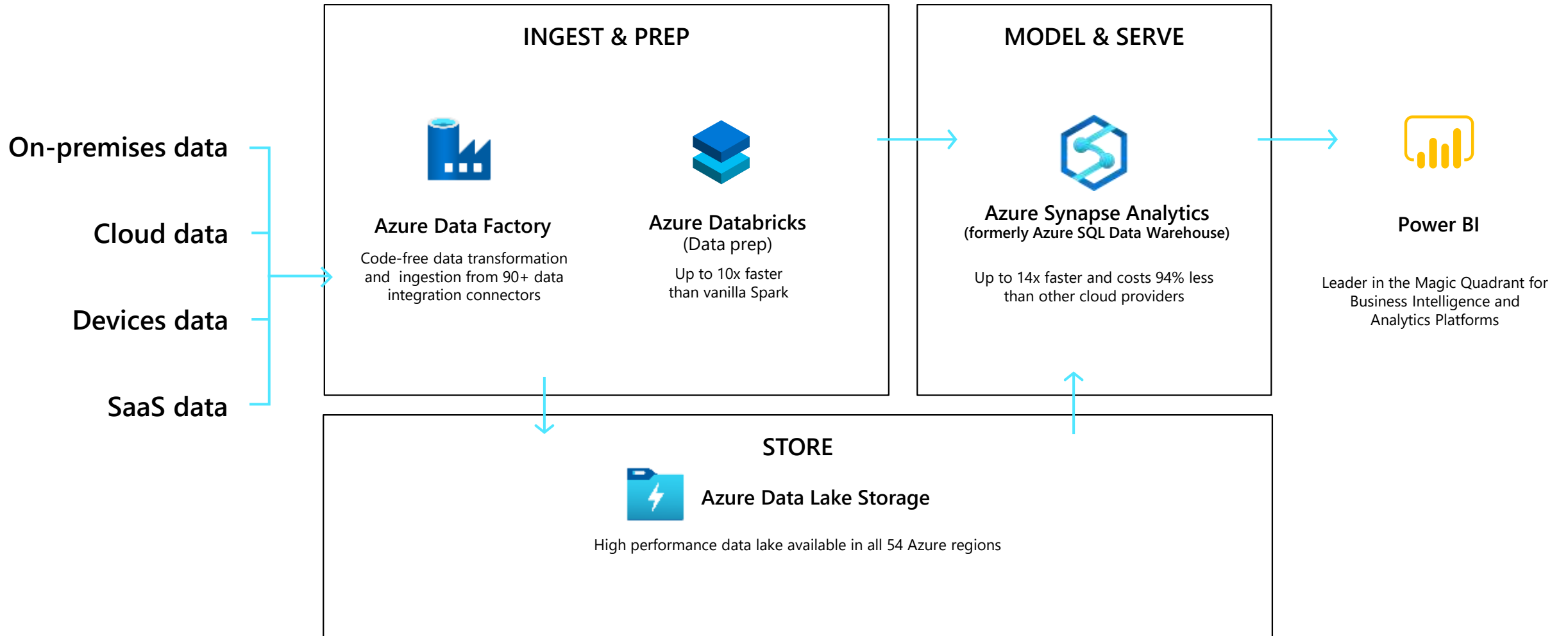


Microsoft Reference Architecture

Based on the hot/cold lambda architecture pattern



Azure Analytics



Estimating Cloud OPEX



Rule of Thumb:

1 Core × 1 GB RAM × 1 month ≈ US\$14*

* West Europe Price, A-series VMs

1 Core × 1 GB RAM × ~~1-month~~ ≈ US\$14
730 h

How accurate is this?

1 Core × 1 GB RAM × 730 h ≈ US\$14

How accurate is this?

1 Core × 1 GB RAM × 730 h ≈ US\$14

Not at all accurate!
(depends on location, machine type, etc.)

just a way to:

- Do quick mental calculations/ballpark figures
- Understand **three dimensions** of pricing:

Performance × Capacity × Usage ≈ Cost

What about persistent storage?

Rule of Thumb #2:

1 TB of Standard Storage \times 1 month \approx US\$19*

*** West Europe Price, Locally Redundant Storage**

What dimension has changed?

1 TB of Standard Storage \times 1 month \approx US\$19*

1 TB of Premium Storage \times 1 month \approx US\$195*

*** West Europe Price, Locally Redundant Storage**

What dimension has changed?

1 TB of HDD Storage \times 1 month \approx US\$19*

1 TB of SSD Storage \times 1 month \approx US\$195*

Performance

* West Europe Price, Locally Redundant Storage

The Fourth Dimension:

1 TB of **Locally Redundant** Storage \times 1 month \approx US\$19*

1 TB of **Globally Redundant** Storage \times 1 month \approx US\$39*

Redundancy

* West Europe Price

Deliver results on time (either in batches or real time)

Performance

- How Fast

Capacity

- How Much

Remember you only pay for the storage you actually use

Design
Parameters

Day-to-Day
OPEX

Redundancy

- How Tolerant

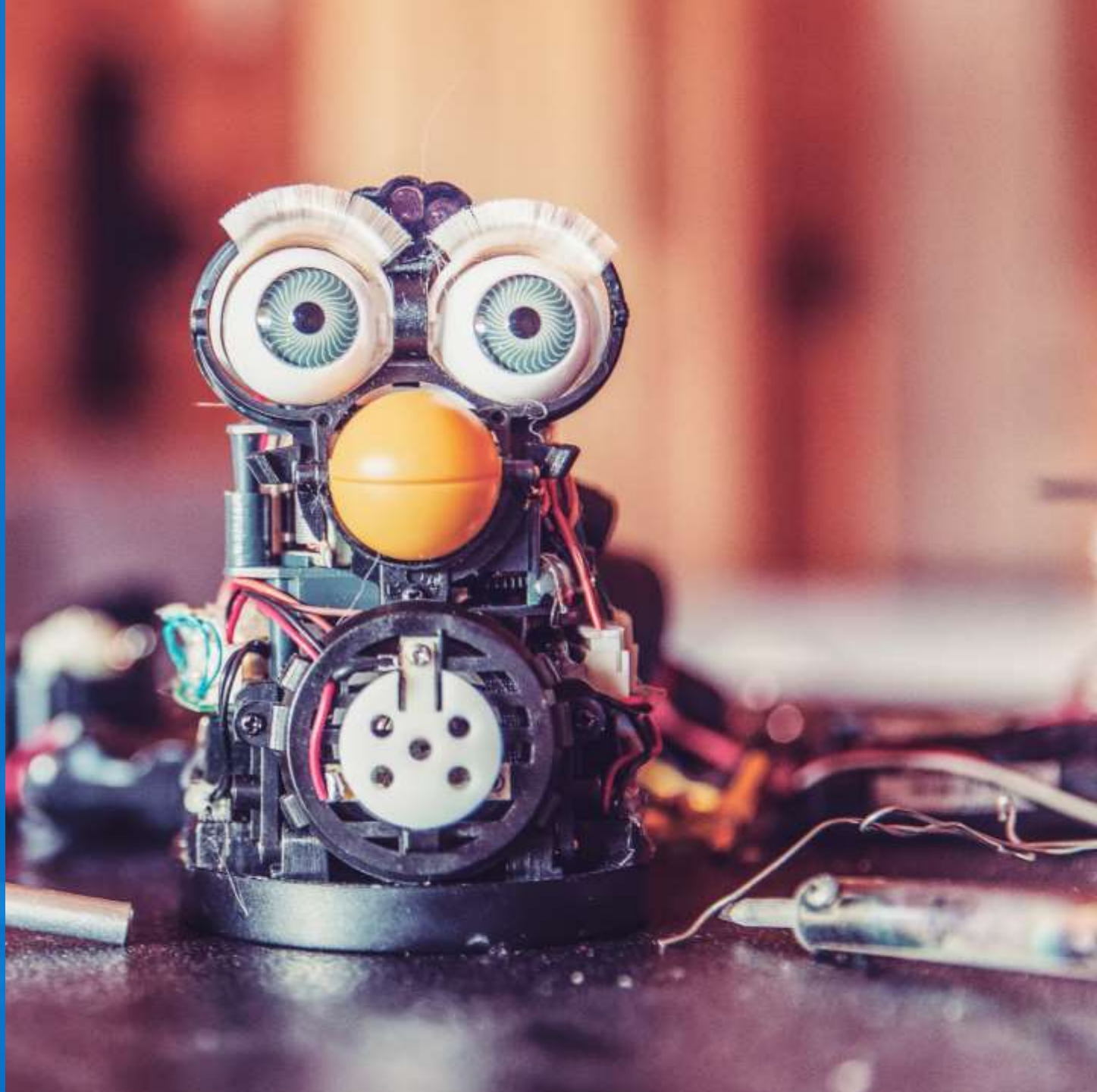
Ensure data and processing capacity is available in case of failure

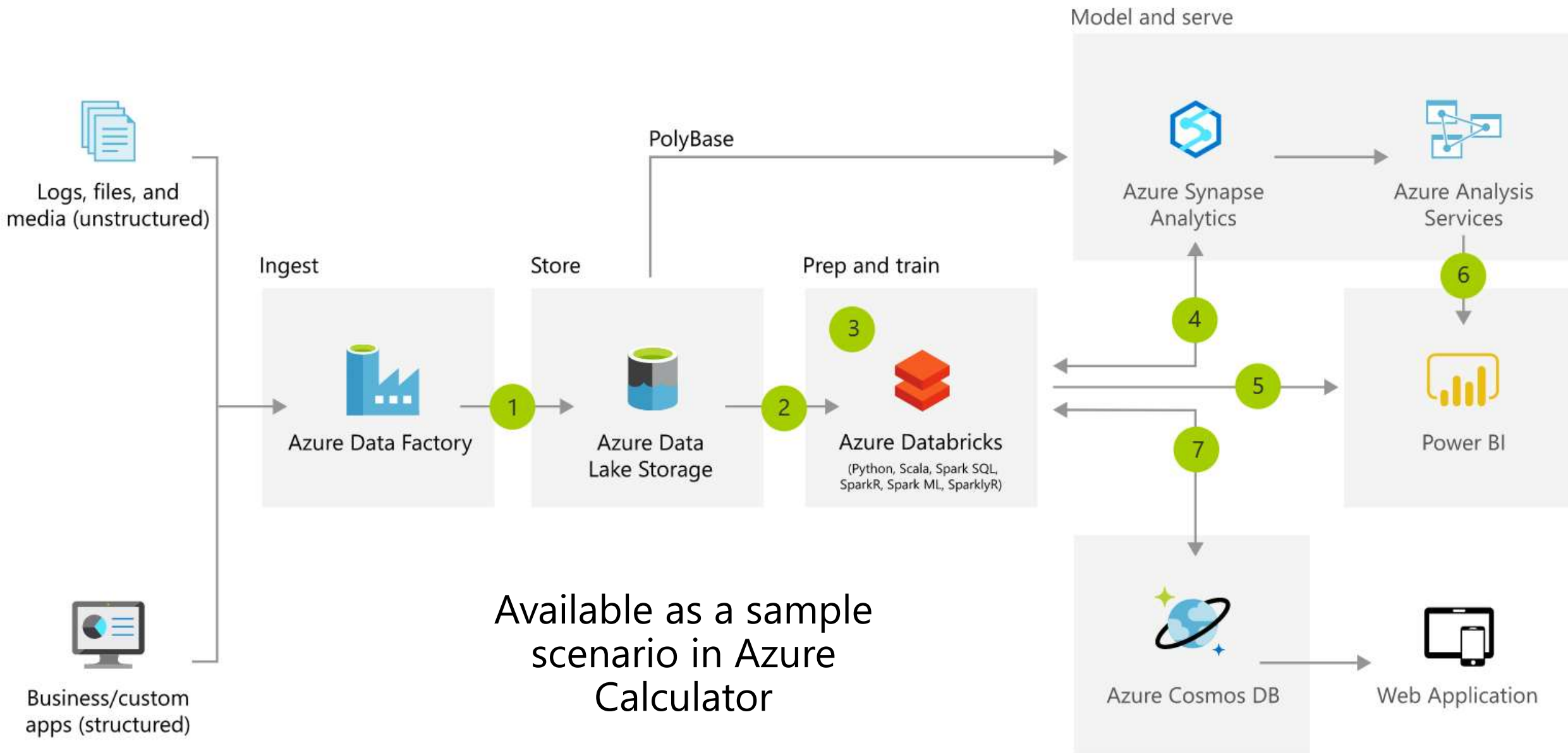
Usage

- How Long

You can run business processes and turn off services when not needed

Use Case





Select an example scenario to include in your estimate. You may add or remove products in your example scenario.

Advanced analytics on big data

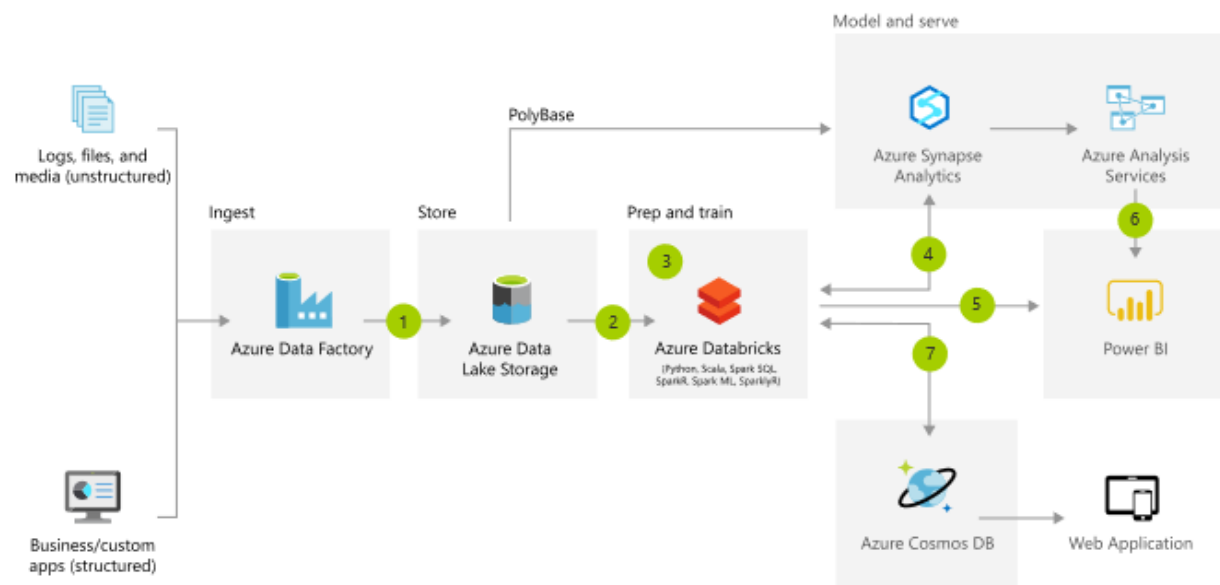
CI/CD for Azure Web Apps

CI/CD for Containers







Modern data warehouse

Real-time analytics

Transform your data into actionable insights using the best-in-class machine learning tools. This architecture allows you to combine any data at any scale, and to build and deploy custom machine learning models at scale.



Products

-  Azure Analysis Services
-  Azure Cosmos DB
-  Data Factory
-  Azure Databricks
-  Power BI Embedded
-  Storage Accounts

[Learn more >](#)

Add to estimate

Thank you

Appendix

Azure Data Factory

Hybrid data integration service that simplifies ETL & ELT at scale

Work efficiently

- Fully accessible visual environment
- Simplified ETL/ELT with automation, visual workflows & data prep, templates
- Continuous integration & delivery (CI/CD)

Cost-effective integration

- Serverless & fully managed, scales on demand
- Scale-out transformations via Spark
- Reduced overhead, SSIS in the cloud

Connect with confidence

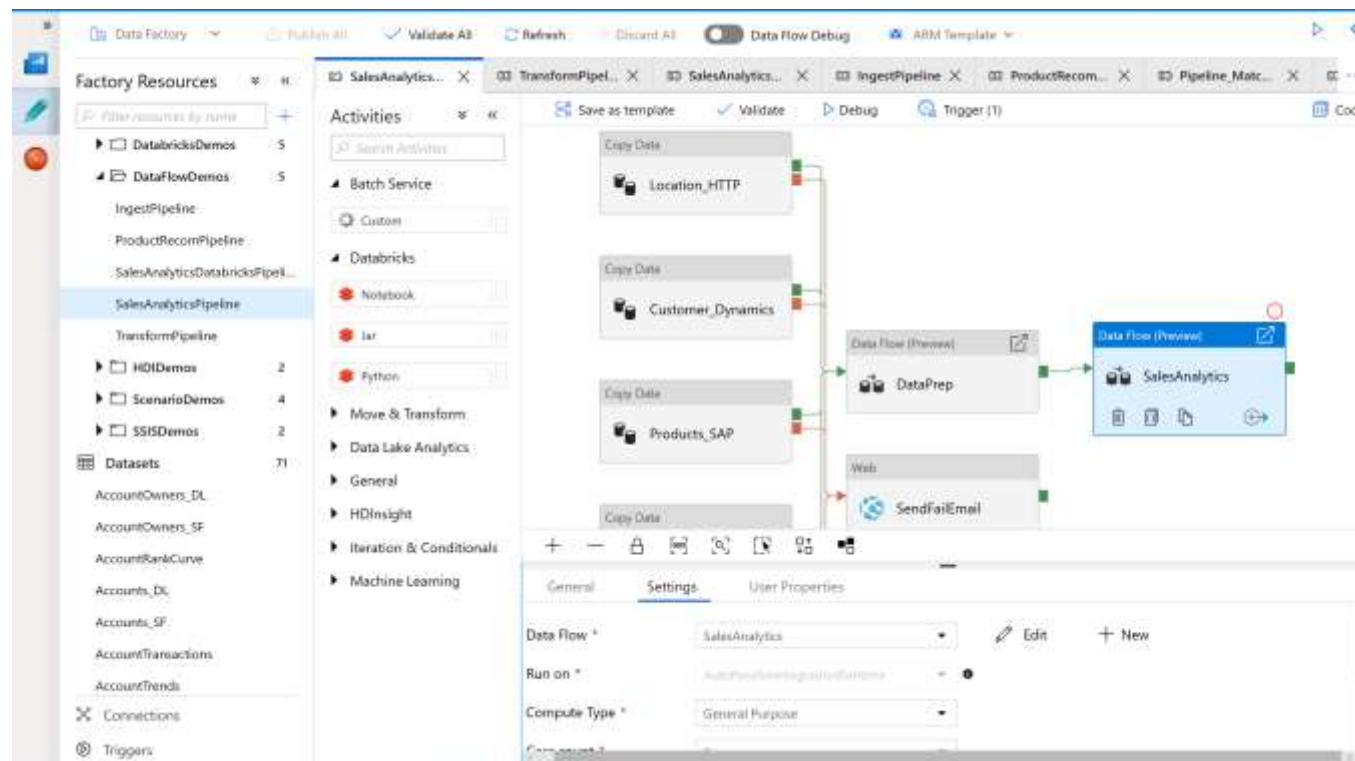
- All-inclusive connectivity, >80 + connectors
- Trusted, global cloud presence
- Multiple language support for coders

[Documentation](#)

[Product Page](#)

[ADF Videos](#)

[Azure Data Factory Blog](#)



Archive Storage

Cold Storage for rarely accessed data needing long term retention

Data is expected to be stored for several months

From milliseconds (hot) to hours (archive) to retrieve

Lowest storage cost, higher access costs

Object-level tiering between hot, cool, and archive

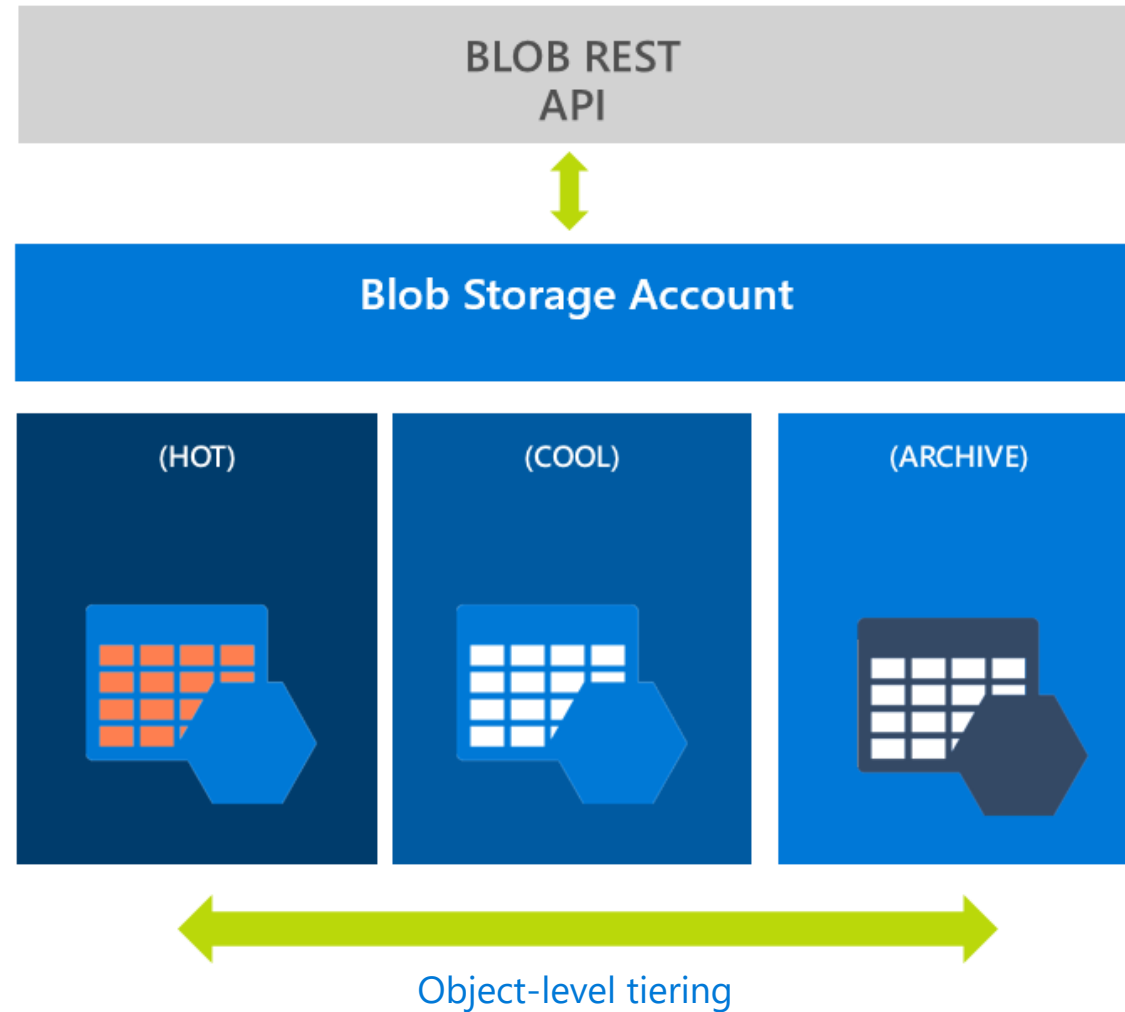
Consistent API Among Tiered Storage

Access through Blob REST API

All tiers of blobs co-exist in a storage account

Blob can move between any tiers within the storage account

[Learn more.](#)



Azure Data Lake Storage Gen2

Brings together the best of Azure Data Lake Store and Blob Storage

- Hadoop compatible file system interface for Azure Blob Storage
- Fine grained file and folder permissions (ACLs)
- Atomic file system operations
- Full support for all Blob features (AAD Integration, Zone Redundant and RA-Geo Redundant Storage)
- Pricing at Blob Storage levels
- Available in all 54 Azure regions

[Learn more.](#)

Azure Data Lake Storage



Scalable, secure storage
that speeds time to insight

Scale and
Availability

Speed to
Insight

Cost
Effectiveness

Rich Security



Upgrade path for
existing ADLS
Customers



Strong Partner
Support



Optimized for
performance with
Spark and Hadoop
analytics engines

Azure SQL Database reserved capacity

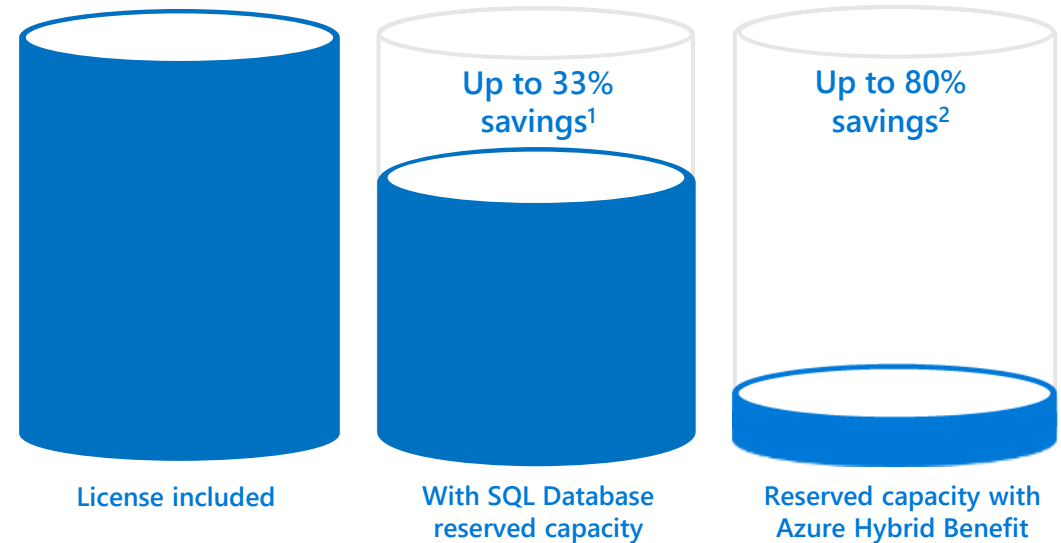
Reserve Azure SQL Database resources in advance and save up to 33%¹

Budget and forecast better with upfront payment for one-year or three-year terms

Exchange or cancel reservations as your needs evolve

Scale up or down within a performance tier and region with auto-fit

Move SaaS apps between elastic pools and single databases and keep your reserved instance benefit



¹ Savings based on eight vCore SQL Database managed instance general purpose in West2 US region, running 730 hours per month. Savings are calculated from on demand full price (license included) against 3-year reserved capacity license Included. Actual savings may vary based on region, instance size, and performance tier. Prices as of May 2018, subject to change.

² Savings based on eight vCore SQL Database managed instance business critical in West2 US region, running 730 hours per month. Savings are calculated from on demand full price (license included) against base rate with Azure Hybrid Benefit plus 3-year reserved capacity. Savings excludes Software Assurance cost for SQL Server Enterprise edition, which may vary based on EA agreement. Actual savings may vary based on region, instance size, and performance tier. Prices as of May 2018, subject to change.

[Learn more.](#)

Azure SQL Data Warehouse

- Compute Optimized Gen 2 is now Generally Available
- **5 times** the performance of our Gen1 offer
- **4 times** the concurrency up to 128 concurrent queries – the highest of any cloud data warehousing service
- **5 times** the compute headroom (over 4000 compute cores)
- **Infinite** storage of columnar data

The fast, flexible, and secure hub for all your data



Fast

Unlimited scale & performance



Flexible

Fits your needs



Secure

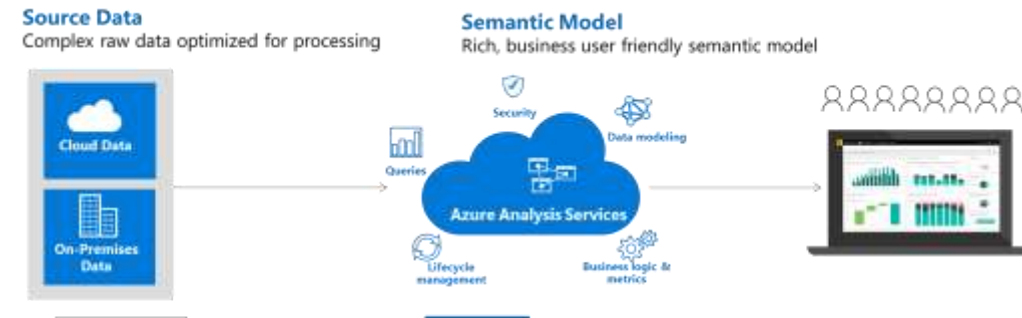
Trusted. Compliant.
Reliable.

Seamlessly compatible across Microsoft and other leading BI & Data Integration services

New update: Flexible restore points – GA Sept 2018

[Learn more.](#)

Azure Analysis Services



Azure Analysis Services is a fully managed platform as a service (PaaS) that provides enterprise-grade data models in the cloud. Use advanced mashup and modeling features to combine data from multiple data sources, define metrics, and secure your data in a single, trusted tabular semantic data model. The data model provides an easier and faster way for users to browse massive amounts of data for ad hoc data analysis.

Get up and running quickly

- In Azure portal, you can [create a server](#) within minutes. And with Azure Resource Manager [templates](#) and PowerShell, you can create servers using a declarative template. With a single template, you can deploy server resources along with other Azure components such as storage accounts and Azure Functions.
- Azure Analysis Services integrates with many Azure services enabling you to build sophisticated analytics solutions. Integration with [Azure Active Directory](#) provides secure, role-based access to your critical data. Integrate with [Azure Data Factory](#) pipelines by including an activity that loads data into the model. [Azure Automation](#) and [Azure Functions](#) can be used for lightweight orchestration of models using custom code.

The right tier when you need it

- Azure Analysis Services is available in **Developer**, **Basic**, and **Standard** tiers. Within each tier, plan costs vary according to processing power, QPUs, and memory size. When you create a server, you select a plan within a tier. You can change plans up or down within the same tier, or upgrade to a higher tier, but you can't downgrade from a higher tier to a lower tier.

Scale to your needs

- Scale up\down, pause, and resume:** Go up, down, or pause your server. Use the Azure portal or have total control on-the-fly by using PowerShell. You only pay for what you use.
- Scale out resources for fast query responses:** With scale out, client queries are distributed among multiple *query replicas* in a query pool. Query replicas have synchronized copies of your tabular models. By spreading the query workload, response times during high query workloads can be reduced. Model processing operations can be separated from the query pool, ensuring client queries are not adversely affected by processing operations.

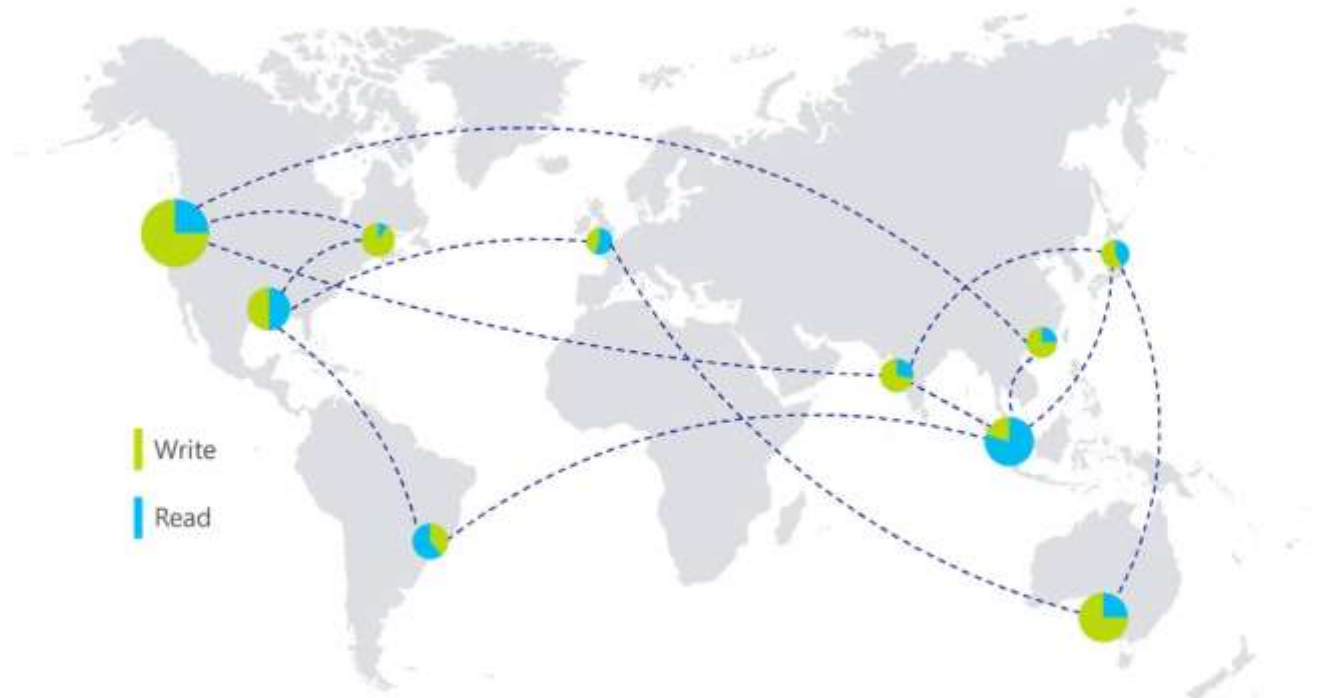
Azure Cosmos DB



Multi-Master Support

Multi-master enables developers to write data in any region, and enjoy <10ms reads and writes around the world.

- <10 ms low latency writes anywhere in the world
- High write availability >99.999%
- Comprehensive conflict resolution support
- Compatible with all existing consistency models





Azure Cosmos DB Multi-Master

[Learn more.](#)


Azure Databricks


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


 Designed in collaboration with the founders of
Apache Spark

 Autoscaling and auto termination of Spark
clusters

 Interactive workspace for data scientists,
engineers, and analysts

 Native integration with Azure services like
Power BI, SQL DW, Cosmos DB

 Enterprise grade Azure security (Active
Directory integration, compliance, enterprise-
grade SLAs)

 Faster and reliable Spark query performance
and simplified batch and data pipeline with
[Azure Databricks Delta](#)

[Learn more.](#)



Data serving

A side-by-side comparison of general capabilities and features

	SQL Database	SQL Data Warehouse	Azure Analysis Services
Is a managed service	Yes (Azure SQL Database)	Yes	Yes
Primary database model	Relational (columnar format when using columnstore indexes)	Relational tables with columnar storage	Tabular and MOLAP semantic models
SQL language support	Yes	Yes	No
Optimized for speed serving layer	Yes, using memory-optimized tables and hash or nonclustered indexes	No	No

Data serving

A side-by-side comparison of scalability capabilities

	SQL Database	SQL Data Warehouse	Azure Analysis Services
Redundant regional servers for high availability	Yes (Azure SQL Database)	Yes	No
Supports query scale out	No	Yes	Yes
Dynamic scalability (scale up)	Yes (Azure SQL Database)	Yes	Yes
Supports in-memory caching of data	Yes	Yes	Yes

Business Continuity

A side-by-side comparison of availability alternatives

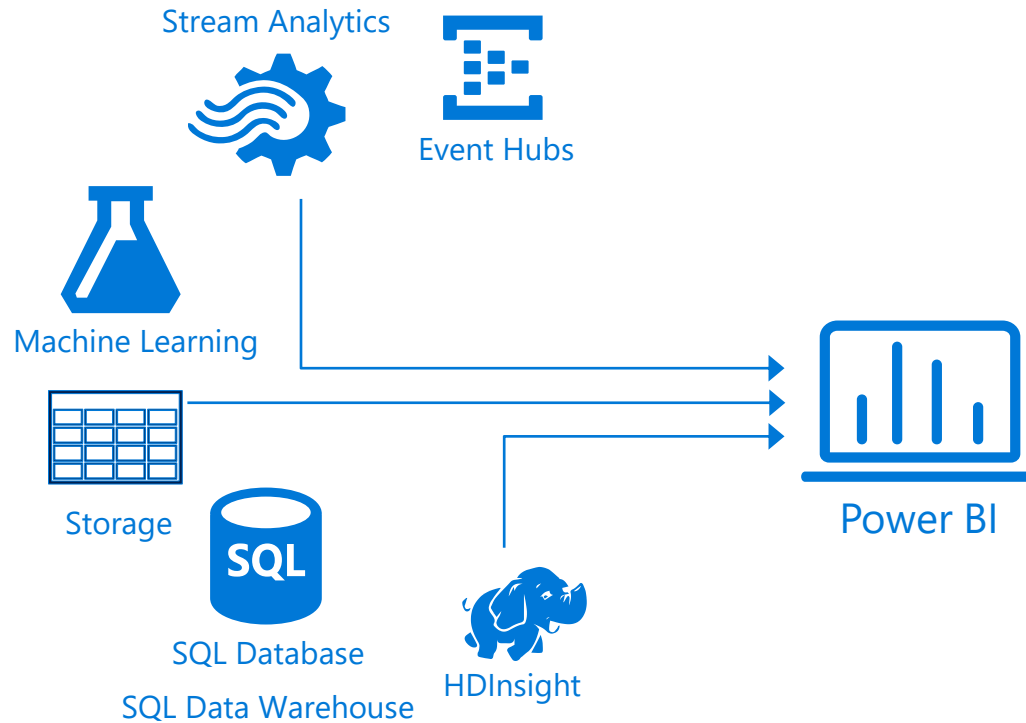
	SQL Data Warehouse	Azure Data Lake	Data Factory + SSIS	Azure SQL DB	Azure Analysis Services	Azure Databricks
High Availability	Built-in to PaaS	Built-in to PaaS	Built-in to PaaS	Built-in to PaaS	Built-in to PaaS	Built-in to PaaS
Backups & Data Protection	7-day restore points Daily geo-backup to paired data center (on by default)	Storage Geo-replication to paired data center	N/A	Geo-backups	Need to backup data independently	N/A (storage is in Data Lake)
Geo-Redundancy	See above	See above	N/A	Active geo-replication	No	N/A
Recovery process	New DW in paired region	Manual fail-over to paired region	Rebuild DF+SSIS in paired region from Git and SSISDB	Auto-failover or restore DB	Set up new instance and restore data	Rebuild workspace from Git and deploy new cluster

Power BI

Dashboards & Visualizations



Power BI



- Analytics for everyone, even non-data experts
- Your whole business on one dashboard
- Create stunning, interactive reports

- Drive consistent analysis across your organization
- Embed visuals in your applications
- Get real-time alerts when things change