Microsoft

Azure Data Services

Azure Data Services

Data

- Data modernization on-premises
- Data modernization to Azure
- Globally distributed data
- Cloud scale analytics

+

AI

- AI apps & agents
- Knowledge mining
- Machine learning

+

BI

- Self-service BI
- On-premises BI
- Apps with intelligence

Azure Data platform: intelligent performance, security and hyperscale

Data Modernization to Azure

- Power BI
- Azure App Services
- Azure Data Factory
- Azure Analysis Services
- Azure Machine Learning
- Azure Cognitive Services
- Azure Bot Service

Azure SQL Database | **Azure SQL Data Warehouse** | **Azure Database for PostgreSQL** | **Azure Database for MySQL** | **Azure Database for MariaDB**

Database Services Platform

Intelligent
Advisors, tuning, monitoring

Flexible
Independent scaling of compute & storage, resource governance

Trusted
HA/DR, backup/restore, security, audit, isolation

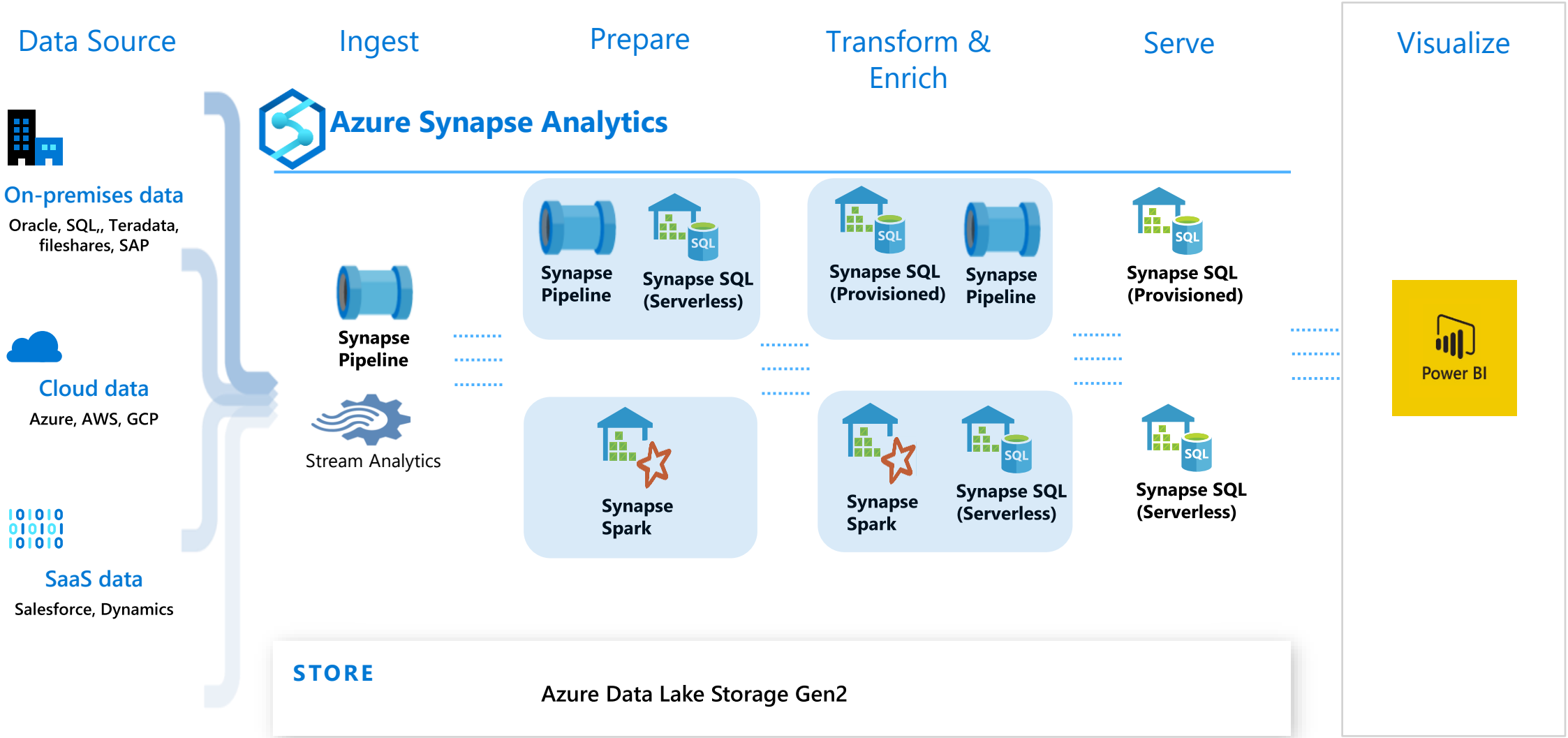
GEO Replication

Azure compute

Azure storage

Globally Available in 54 Regions

Modern Data Platform



MICROSOFT'S APPROACH TO OPEN SOURCE

Enable

An open and flexible platform that meets you where you are and adds value to your existing investments

Integrate

Embracing leading ecosystems, increase agility and offer consistent open source offerings

Release

Support a strong ecosystem to achieve more through Microsoft's own portfolio investments

R Server
.NET Core
Roslyn
TypeScript
F#

PowerShell
VS Code
PowerBI Visuals
Office UI Fabric
MSR projects

Contribute

Extend the community and reach to more people, and partner for first-class experiences

Open source partners and ecosystem

Azure Database for MySQL, PostgreSQL, and MariaDB

Azure databases for MySQL, PostgreSQL, and MariaDB offer enterprise-ready and fully managed community versions of the popular open source database engines.

Released

- Intelligent Performance for PostgreSQL—GA
- Intelligent Performance for MySQL and MariaDB—Preview
- MariaDB v10.3 support—GA
- Read Replica across regions for MySQL and PostgreSQL—Preview
- 16TB/20K IOPS for MySQL and PostgreSQL—Preview

Coming Soon

- Read Replica across regions for MariaDB
- Azure Active Directory integration for MySQL, Maria DB and PostgreSQL

Learn more about MySQL

Learn more about PostgreSQL

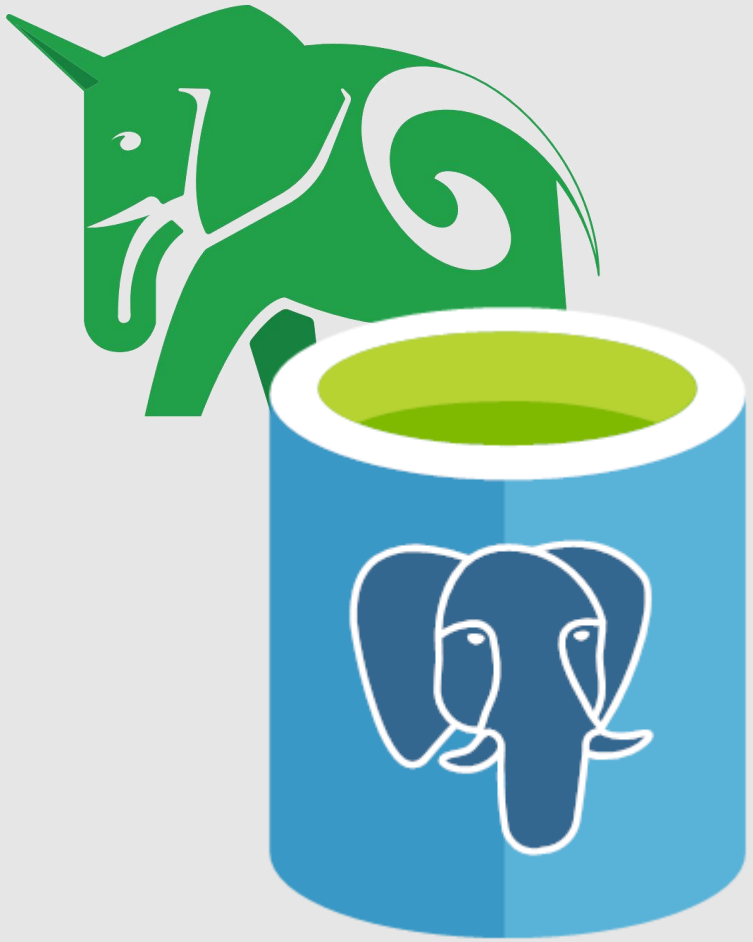
Learn more about MariaDB

Azure Database for PostgreSQL

Azure Database for PostgreSQL offers enterprise-ready and fully managed services for the community version of PostgreSQL.

Hyperscale (Citus) is a deployment option for Azure Database for PostgreSQL that scales out Postgres horizontally.

Hyperscale (Citus) is ideal for multi-tenant and SaaS applications that are growing fast, real-time operational analytics, and building high throughput transactional applications that need sub-second response times across billions of rows. With Hyperscale (Citus) developed as an extension to PostgreSQL, customers don't have to worry about scaling their databases again, and will stay current with the latest innovations, versions, tools of PostgreSQL.



Learn more about PostgreSQL

Azure Data Factory—Visual Data Transformation

Visually author, debug & monitor data flows at scale

Transform data at scale

Platform-as-a-Service data transformation

Zero-code environment

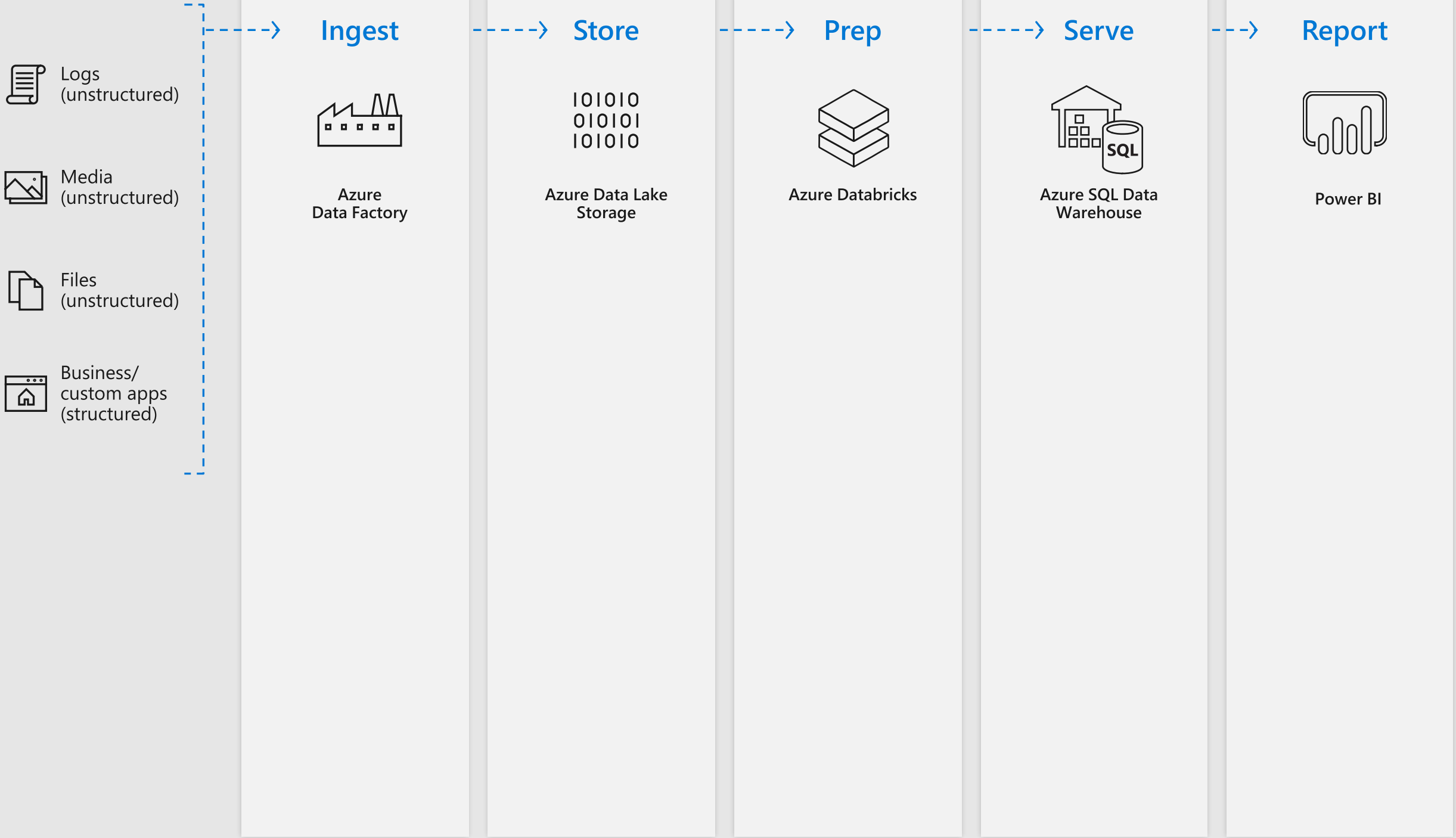
No need to lean Spark, Scala, Java, Python … Focus on business logic

Built-in Resiliency

Schema drift options provide support for flexible/shifting schemas

Elastic Scale

Zero-infrastructure Spark transformations that scale with your project needs


Logs (unstructured)
Media (unstructured)
Files (unstructured)
Business/ custom apps (structured)
Ingest
Azure Data Factory
Store
Azure Data Lake Storage
Prep
Azure Databricks
Serve
SQL
Azure SQL Data Warehouse
Report
Power BI

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

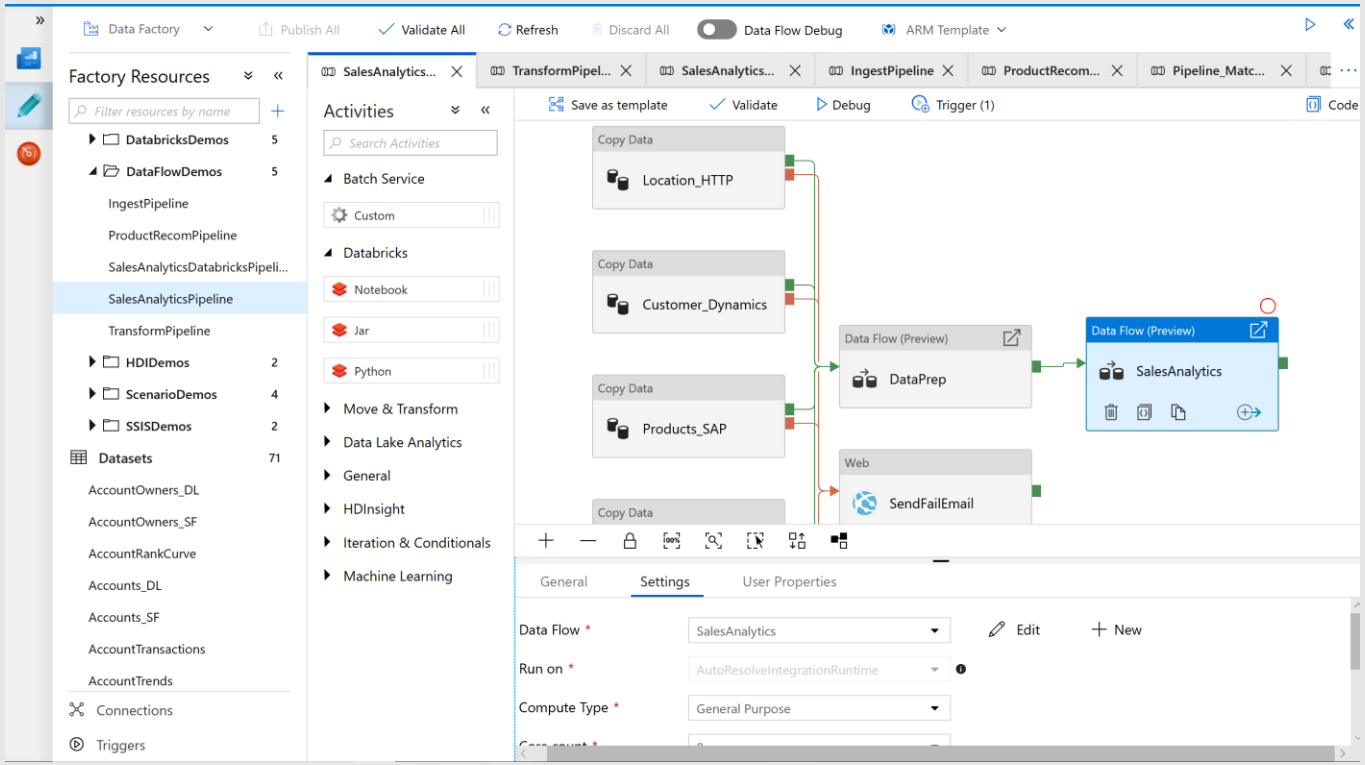
- All-inclusive connectivity, >80 + connectors
- Trusted, global cloud presence
- Multiple language support for coders

[Documentation]

[Product Page]

[ADF Videos]

[Azure Data Factory Blog]



Azure Cosmos DB

Azure Cosmos DB is a fully-managed NoSQL database service with global distribution and support for multiple data models.

//build announcements—May 2019

- Operational analytics across globally-distributed data
  - Built-in support for Apache Spark (Preview)
  - Jupyter notebook experience for all APIs (Preview)
- API for etcd (self-managed Kubernetes)—Preview

Upcoming announcements—NDA

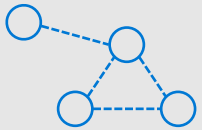
- Azure Directory integration
- Azure Advisor portal integration
- Backup and Point-in-time Restore (PITR)
- Cassandra driver extensions for Java
- Availability Zone support
- Improvements to Gremlin API functionality

Deliver cloud-native, multi-model apps with Azure Cosmos DB

Global distribution

1st and only database with turnkey global distribution capability

Multi-model + multi API



Use key-value, graph, and document with a schema-agnostic service

Elastic scale-out

Independently and elastically scale storage and throughput

Choice of consistency

Choose from five defined consistency levels

Guaranteed single-digit latency

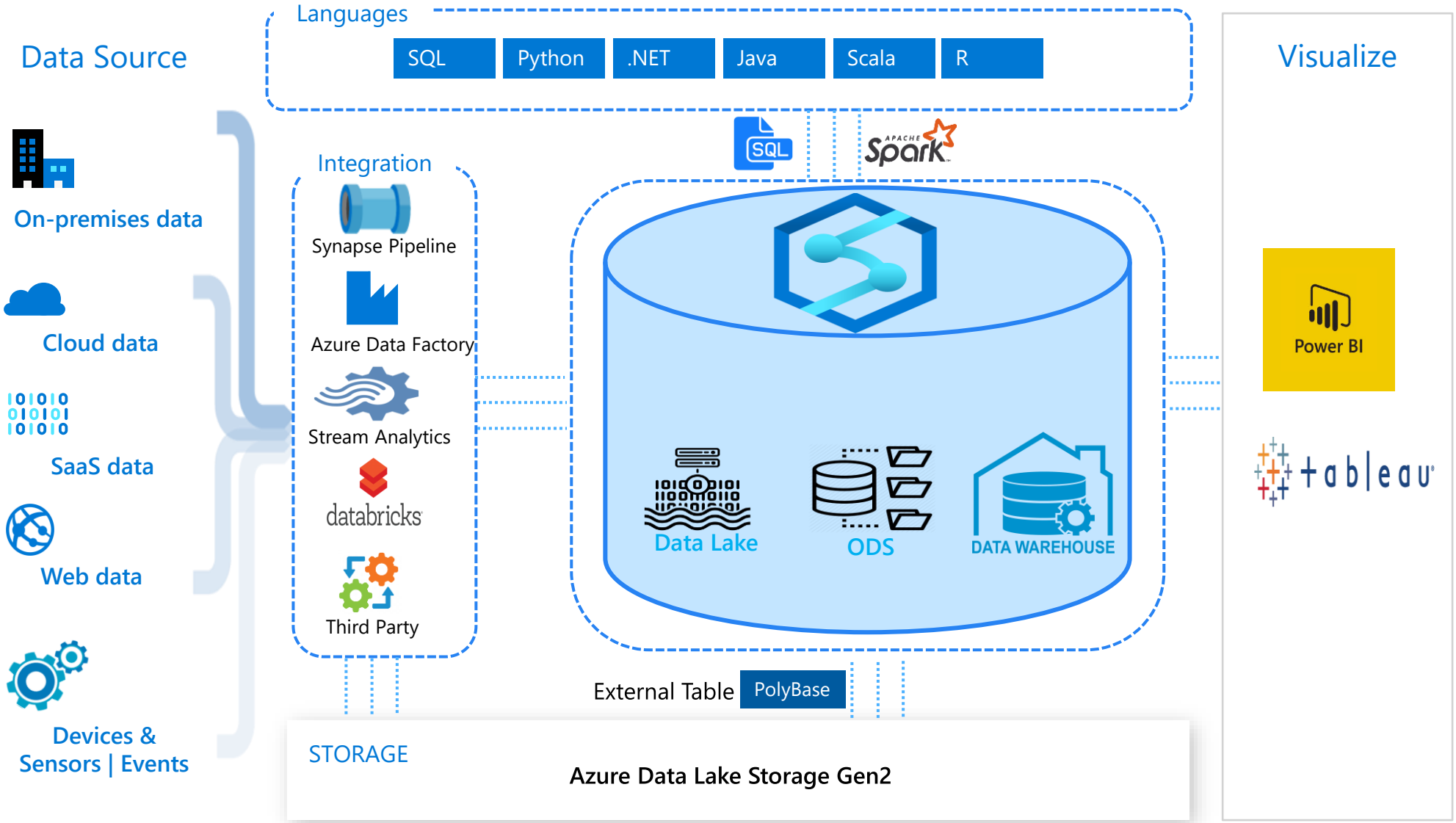
<10ms read and write requests at 99th percentile with money-back guarantee

Enterprise-level SLAs

Industry-leading 99.999 SLAs

Save up to big with Azure Cosmos DB reserved capacity pricing

Azure Synapse Analytics



Why Spark?

- Open-source data processing engine built around **speed, ease of use, and sophisticated analytics**
- In memory engine that is up to **100 times faster than Hadoop**
- **Largest open-source data project** with 1000+ contributors
- **Highly extensible** with support for Scala, Java and Python alongside Spark SQL, GraphX, Streaming and Machine Learning Library (Mllib)

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 Databricks key audiences & benefits

Data scientist

Integrated workspace

Easy data exploration

Collaborative experience

Interactive dashboards

Faster insights

- Best Spark & serverless
- Databricks managed Spark

Data engineer

Improved ETL performance

- Zero management clusters, serverless

Easy to schedule jobs

Automated workflows

Enhanced monitoring & troubleshooting

- Automated alerts & easy access to logs

Zero Management Spark

Cluster democratization (serverless)

CDO, VP of analytics

Fast, collaborative analytics platform accelerating time to market

No dev-ops required

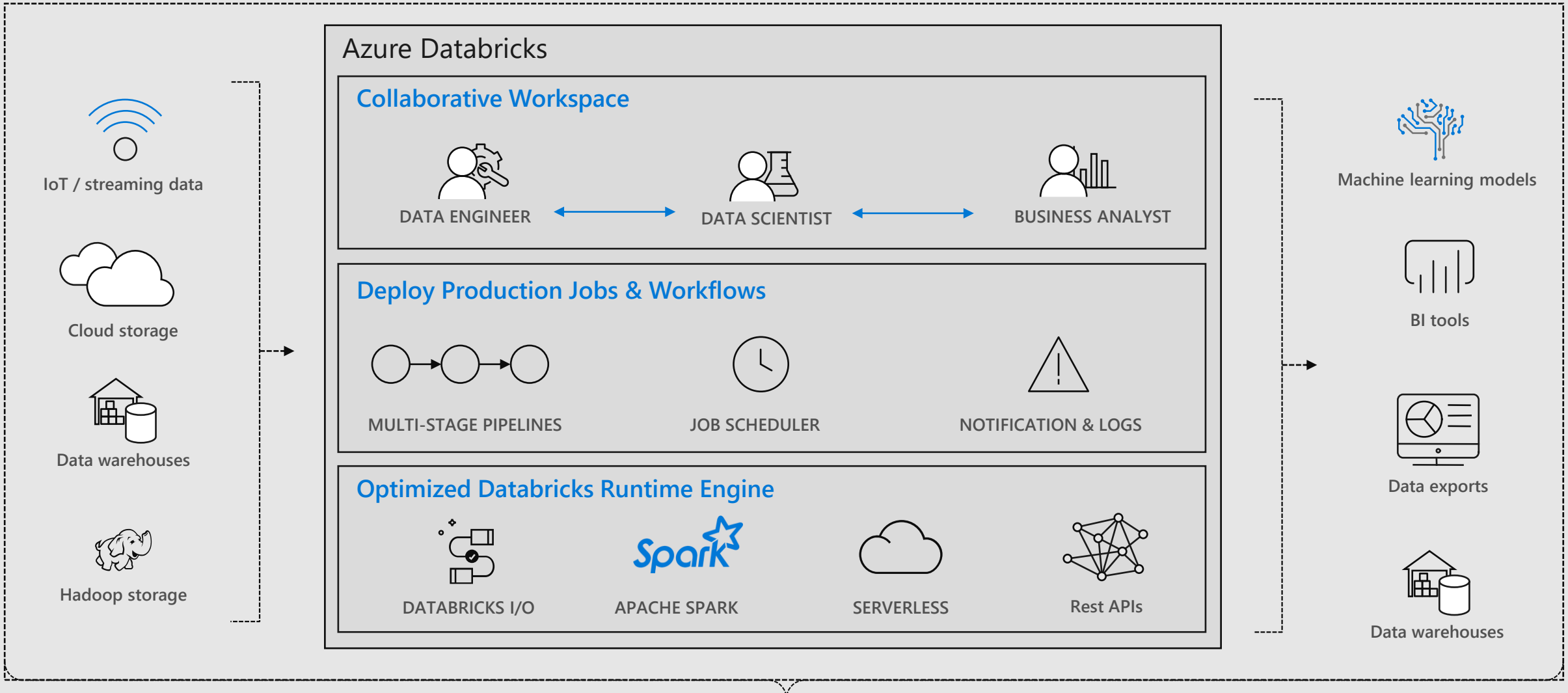
Enterprise grade security

- Encryption
- End-to-end auditing
- Role-based control
- Compliance

Unified analytics platform

Provided by Microsoft and Databricks under NDA

Azure Databricks



Enhance Productivity

Build on secure & trusted cloud

Scale without limits

Optimized Databricks Runtime Engine

OPTIMIZED I/O PERFORMANCE

The Databricks I/O module (DBIO) takes processing speeds to the next level — significantly improving the performance of Spark in the cloud

FULLY-MANAGED PLATFORM ON AZURE

Reap the benefits of a fully managed service and remove the complexity of big data and machine learning

SERVERLESS INFRASTRUCTURE

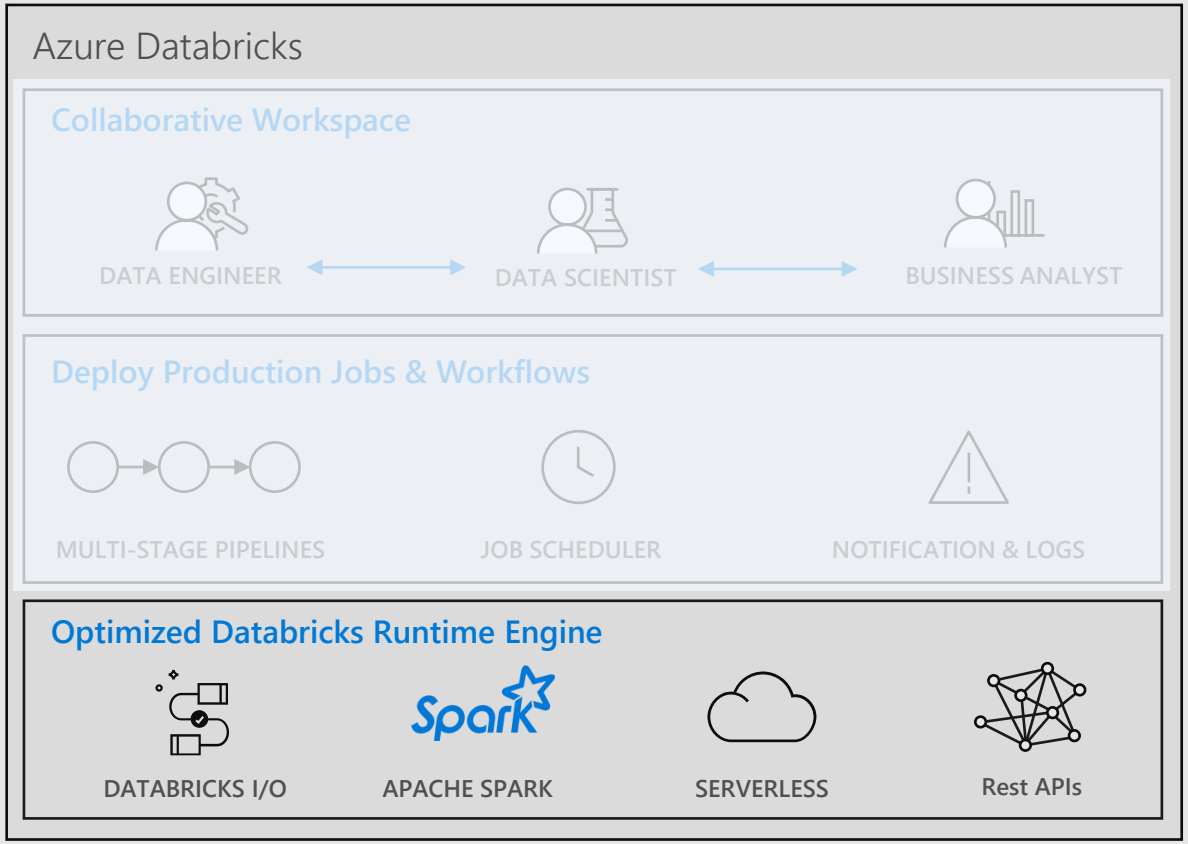
Databricks' serverless and highly elastic cloud service is designed to remove operational complexity while ensuring reliability and cost efficiency at scale

OPERATE AT MASSIVE SCALE

Without limits globally

SUPPORT FOR GPU ENABLED VMS

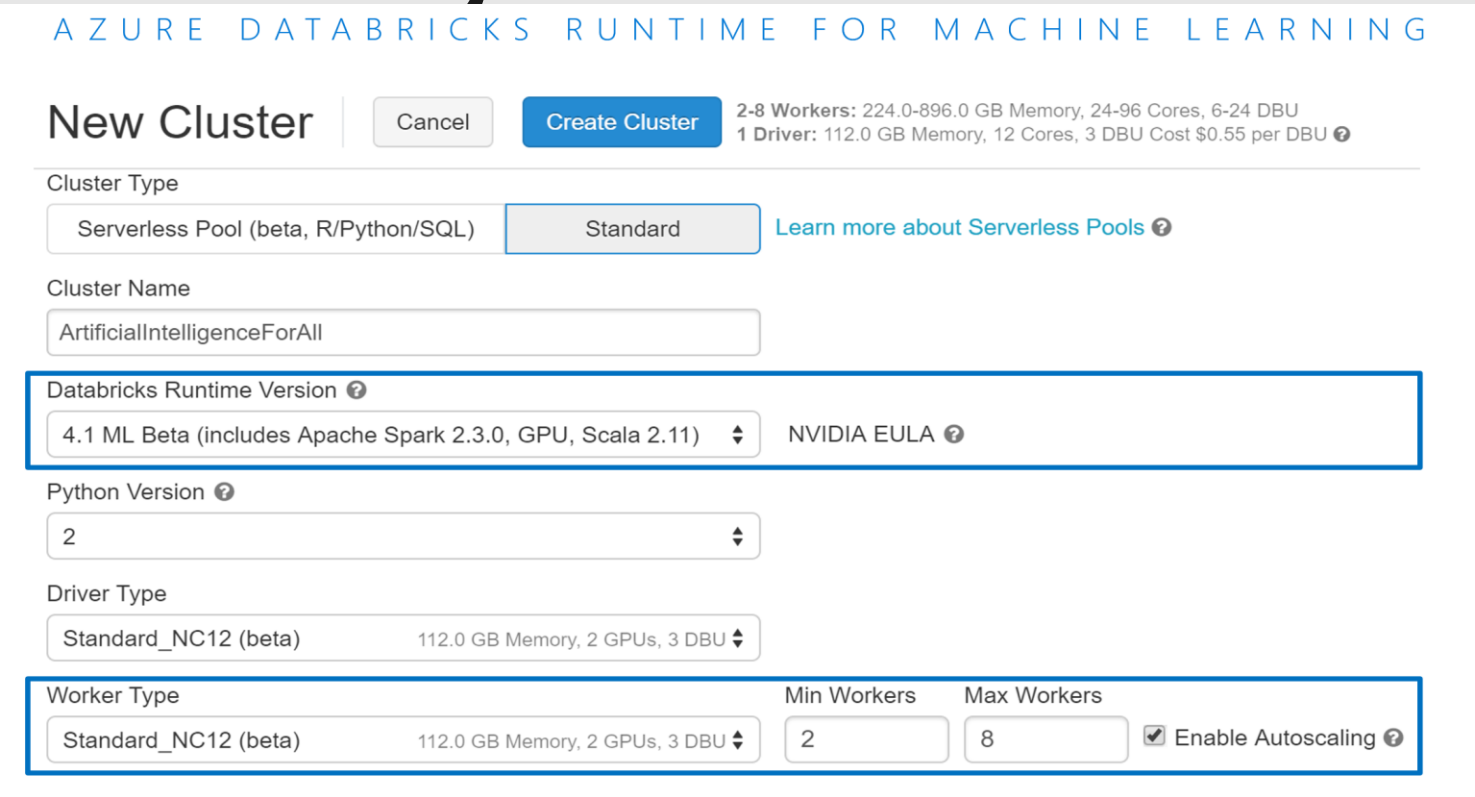
Specialized compute for your deep learning needs



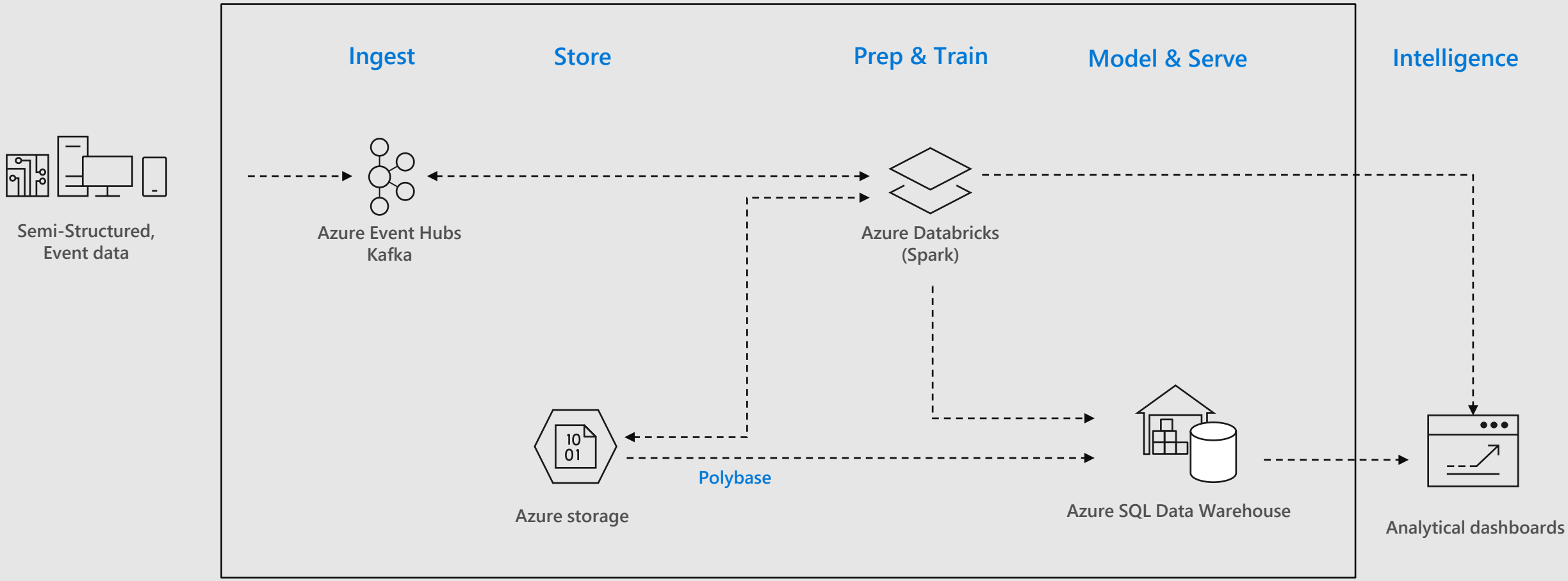
Azure Databricks Runtime for Machine Learning

AZURE Databricks Runtime for Machine Learning

- Pre-installed packages for machine learning like Tensorflow, Keras, Horovod and XGBoost
- Pre-configured HorovodEstimator for seamless integration of Horovod with the Spark DataFrames
- Support for GPU enabled VMs for specialized compute for your deep learning needs
- Multi-GPU trainings of deep neural networks using Horovod
- Unlock complex machine learning and deep learning scenarios with a few lines of code

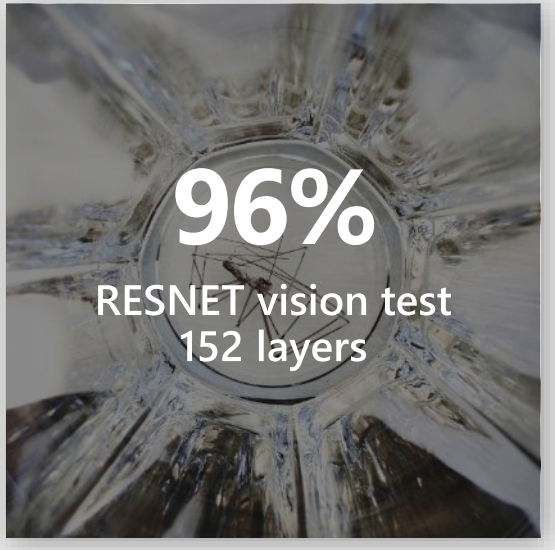


Real-time analytics on Big Data



Microsoft AI breakthroughs

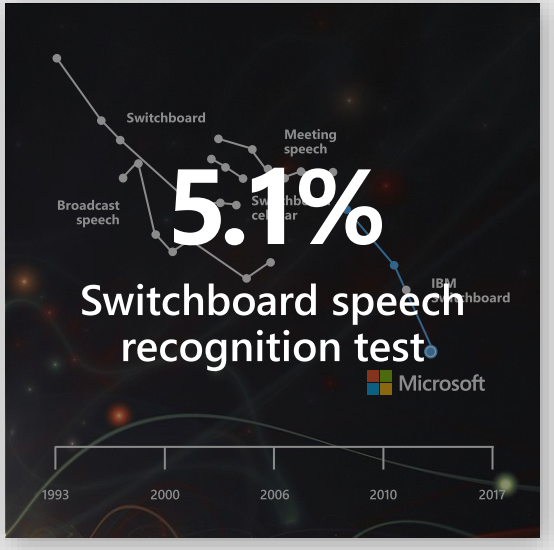
Vision



2016

Object recognition
Human parity

Speech



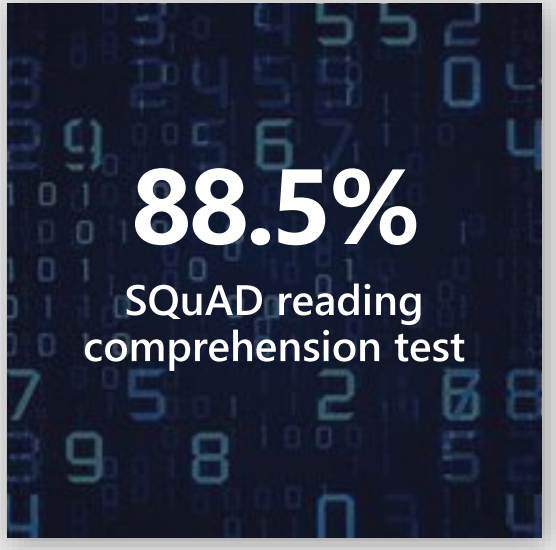
2017

Speech recognition
Human parity

March 2018

Machine translation
Human parity

Language



January 2018

Machine reading comprehension
Human parity

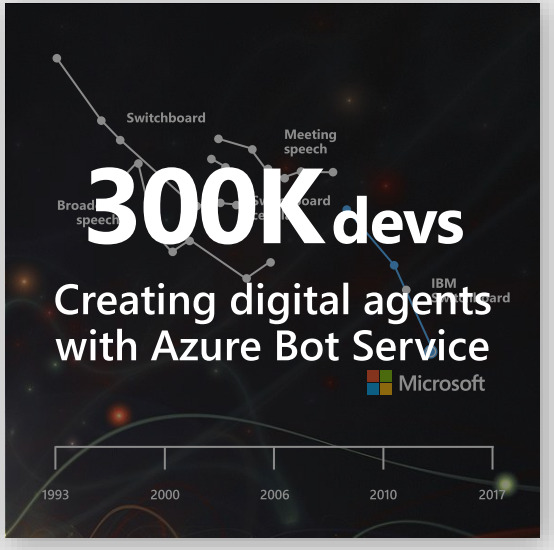
AI momentum

Vision

2016

Object recognition
Human parity

Speech



2017

Speech recognition
Human parity

March 2018

Machine translation
Human parity

Language

January 2018

Machine reading comprehension
Human parity

Azure is the best place for machine learning

Sophisticated pretrained models
Most comprehensive set of pretrained services

Vision | Speech | Language | Search

Popular frameworks
Open & interoperable

Pytorch | TensorFlow | Keras | Onnx

Productive services
Machine learning at scale

Azure Databricks | Azure Machine Learning | Machine Learning VMs

Powerful infrastructure
Most comprehensive
Lowest cost inferencing using FPGAs

CPU | GPU | FPGA

Flexible deployment
From cloud to edge

On-premises | Cloud | Edge

Powerful infrastructure

Accelerate deep learning

CPUs

General purpose machine learning

D, F, L, M, H Series

GPUs

Deep learning

N Series

FPGAs

Specialized hardware accelerated deep learning

Project Brainwave

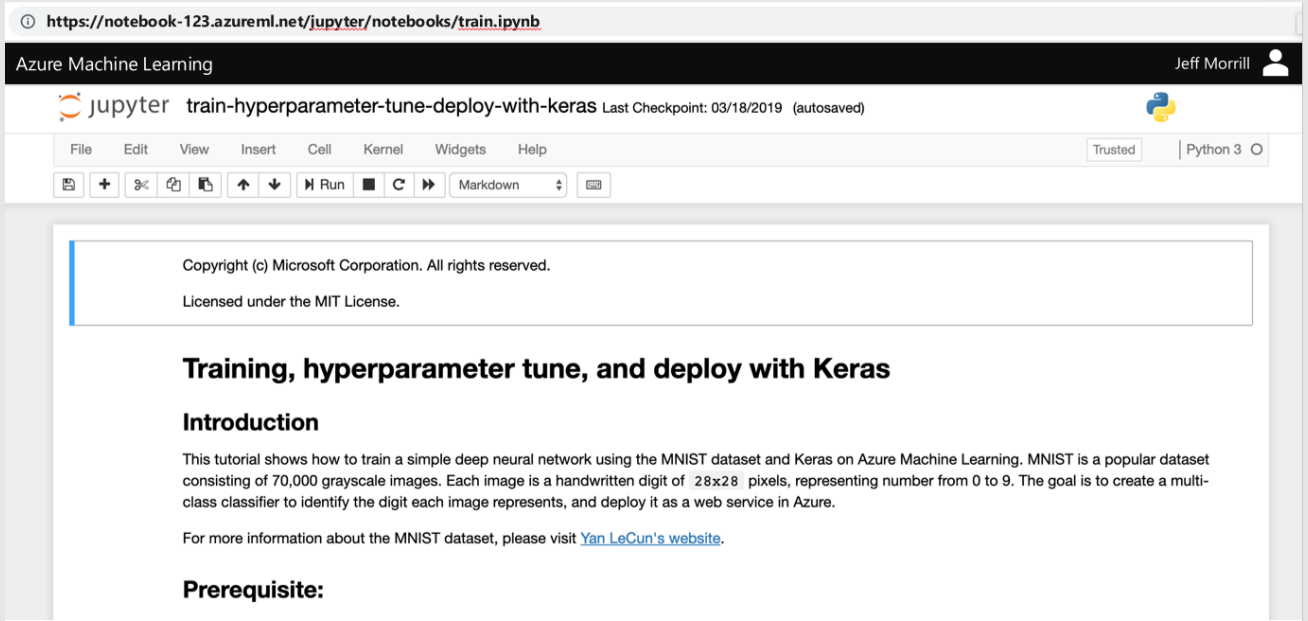
Optimized for flexibility ← → Optimized for performance

Notebook VM

Hosted notebooks provide a code-first experience where users can perform every operation supported by the Azure Machine Learning Python SDK using a familiar Jupyter notebook.

Hosted notebooks simplify the getting-started process by providing a secure, enterprise-ready environment for ML practitioners. In the private preview customers will be able to: access a notebook integrated into the Azure ML workspace, use preconfigured Azure ML notebooks with no set up required, fully customize their notebook VMs including having the ability to add packages and drivers.

Documentation: [Quickstart]



Automated machine learning user interface in Azure portal

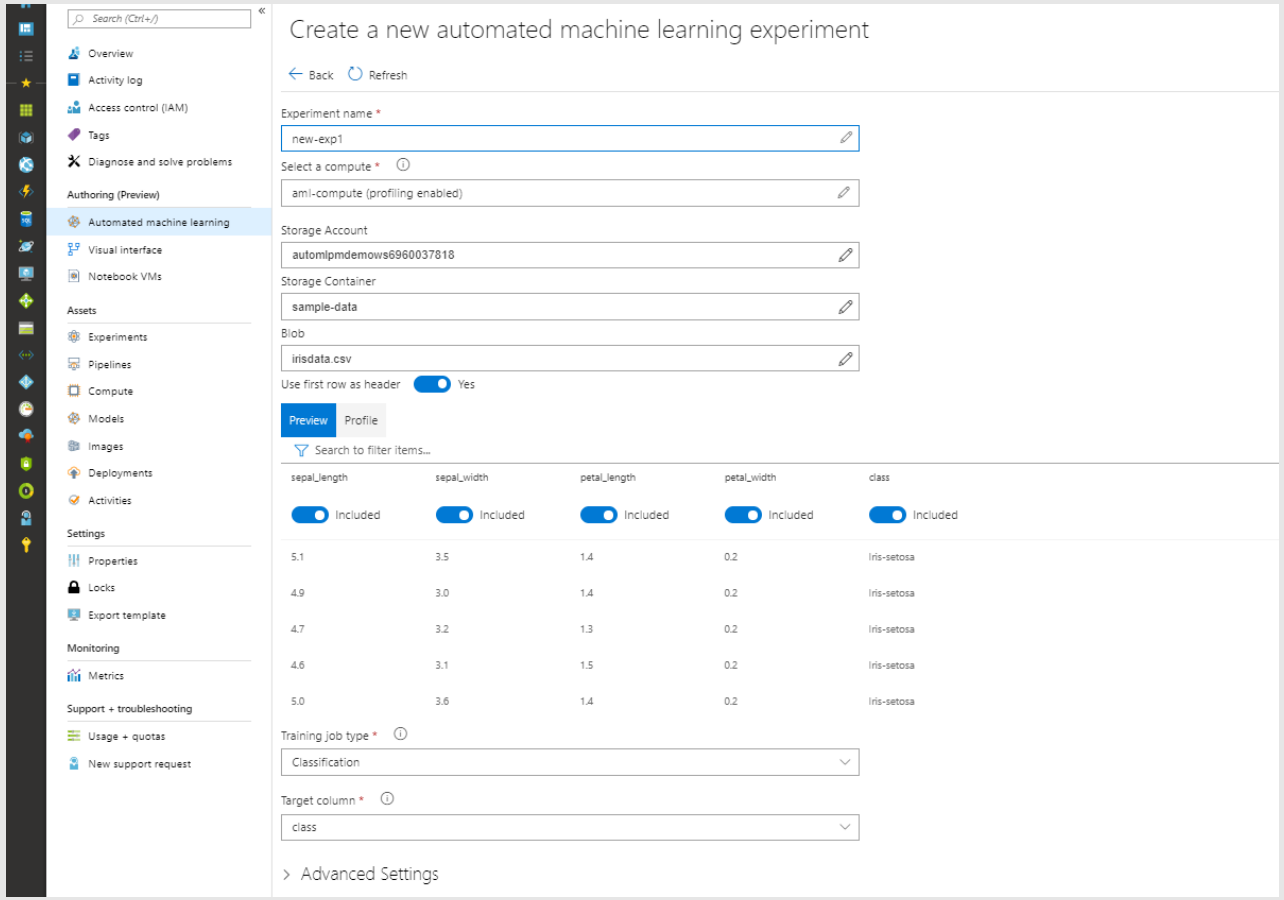
Run automated ML without writing code

- Get the best model for classification, regression, or forecasting problems with a few clicks of a button
- Analyze the generated models
- Deploy as web services to predict on new data

Citizen data scientists: Generate ML models without needing to write Python code (or any type of code).

Data scientists: Explore and generate hundreds of models quickly, then continue to optimize the best ones in Jupyter notebook.

Documentation: [Experiment in UI]

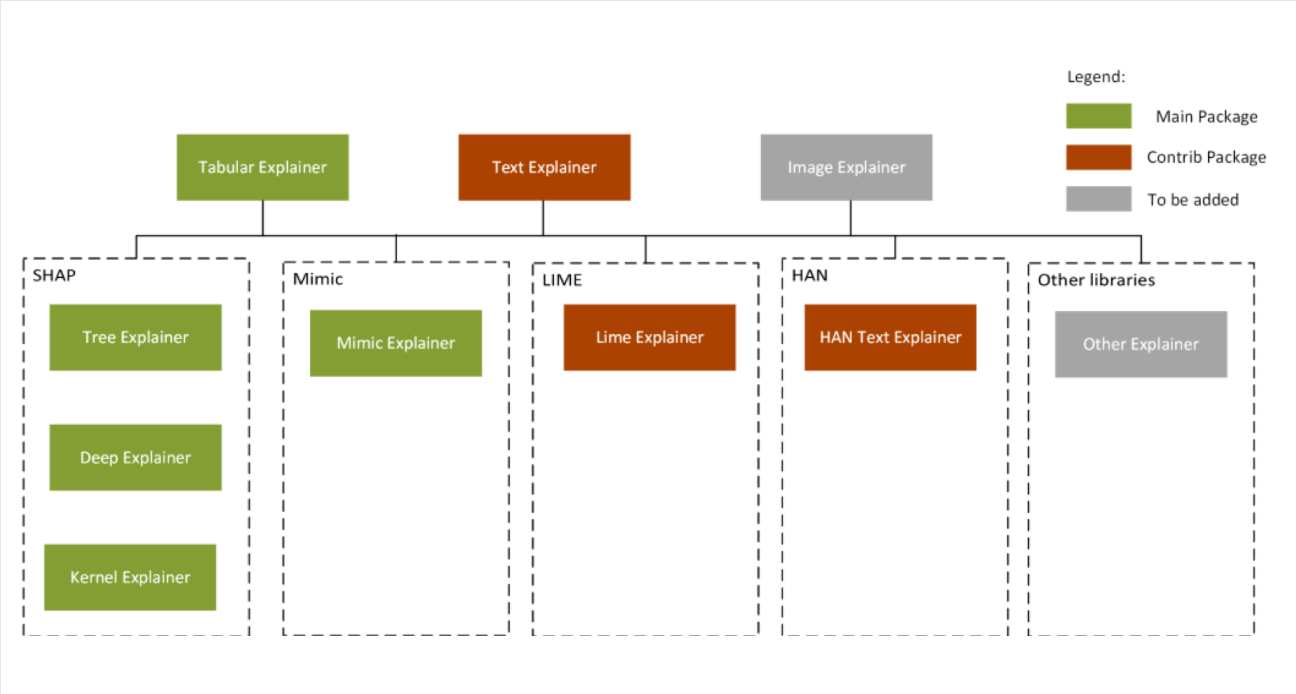


Model interpretability in the AzureML SDK

Machine learning interpretability allows data scientists to explain machine learning models globally on all data, or locally on a specific data point using the state-of-art technologies in an easy-to-use and scalable fashion.

Machine Learning interpretability incorporates technologies developed by Microsoft and proven third-party libraries (for example, SHAP and LIME). The SDK creates a common API across the integrated libraries and integrates Azure Machine Learning services. Using this SDK, you can explain machine learning models globally on all data, or locally on a specific data point using the state-of-art technologies in an easy-to-use and scalable fashion.

Documentation: [Overview & examples]

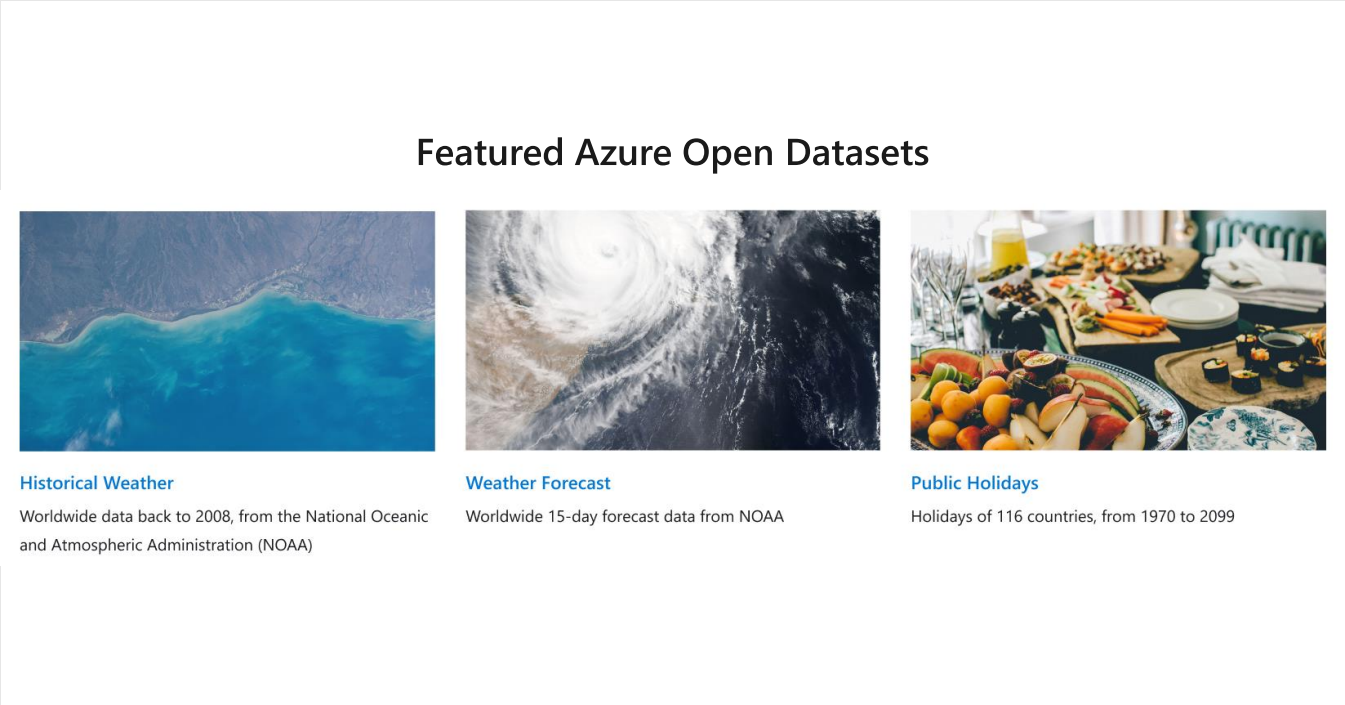


Open Datasets

Open Datasets are a collection of datasets from open data to accelerate the development of machine learning models built in Azure. Open Datasets integrates with Machine Leaning Studio or can be accessed from python notebooks in Azure Machine Learning Service.

Azure Open Datasets offer good quality data from the public domain, which is often hard to find and expensive to curate. Data scientists will be more productive focusing on model building rather than data preparation.

Learn more



Microsoft Azure