

THE MODERN DATA ESTATE



LOB



CRM



Graph



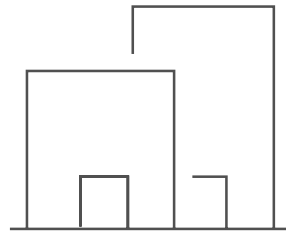
Image



Social



IoT



Operational databases

Data warehouses

Data Lakes

← Hybrid →



Operational databases

Data warehouses

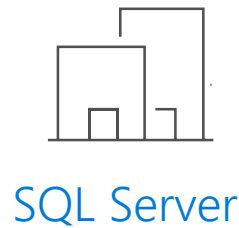
Data Lakes

Reason over any data, anywhere

Flexibility of choice

Security and performance

THE MICROSOFT OFFERING



← Hybrid →



Easiest lift and shift
with no code changes

Industry leader 2 years in a row Operational databases
#1 TPC-H performance Data warehouses
T-SQL query over any data Data lakes

Operational databases **70% faster than Aurora**
Data warehouses **2x global reach than Redshift**
Data lakes **No Limits Analytics with 99.9% SLA**

AI built-in | Most secure | Lowest TCO



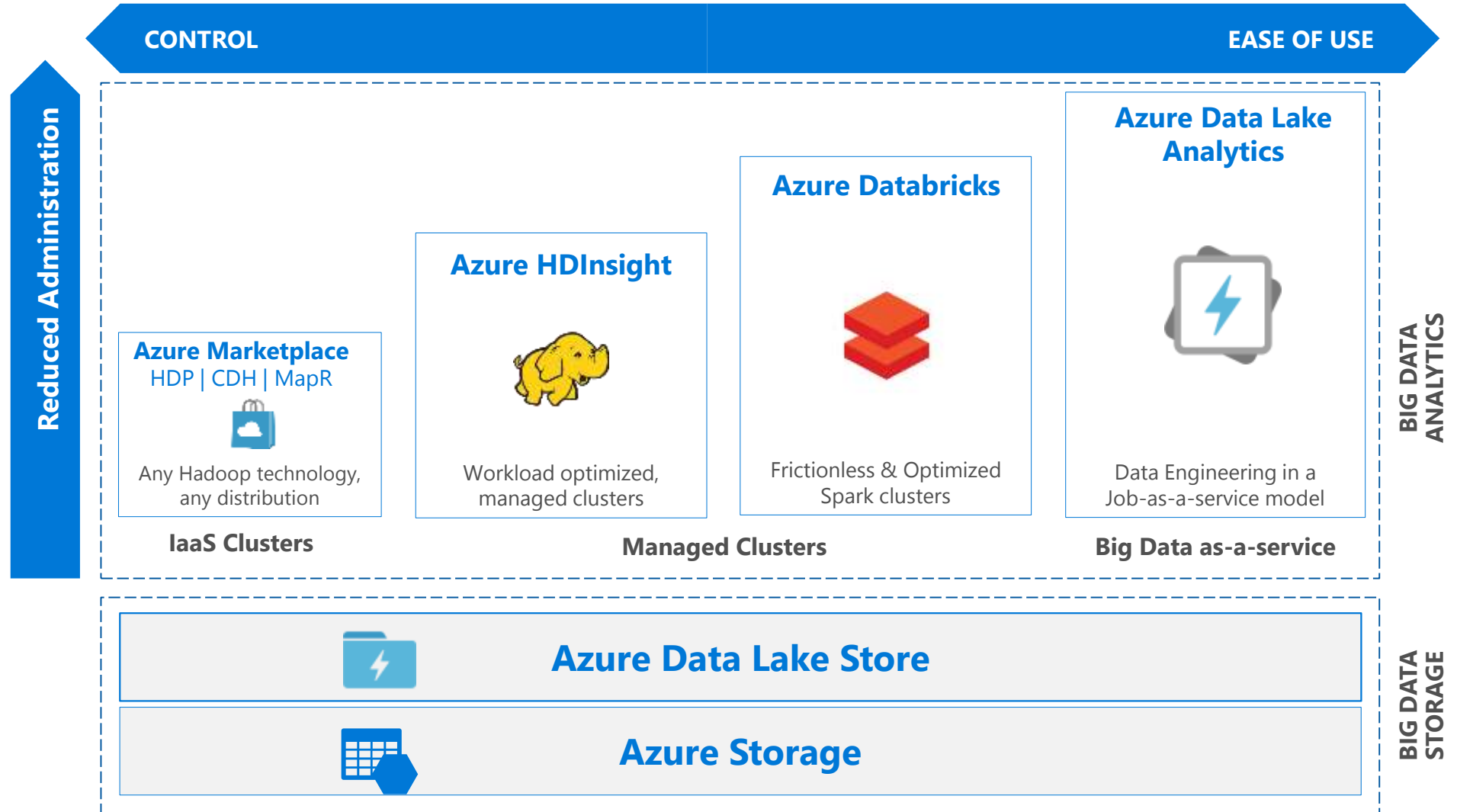
Reason over any data, anywhere

Flexibility of choice

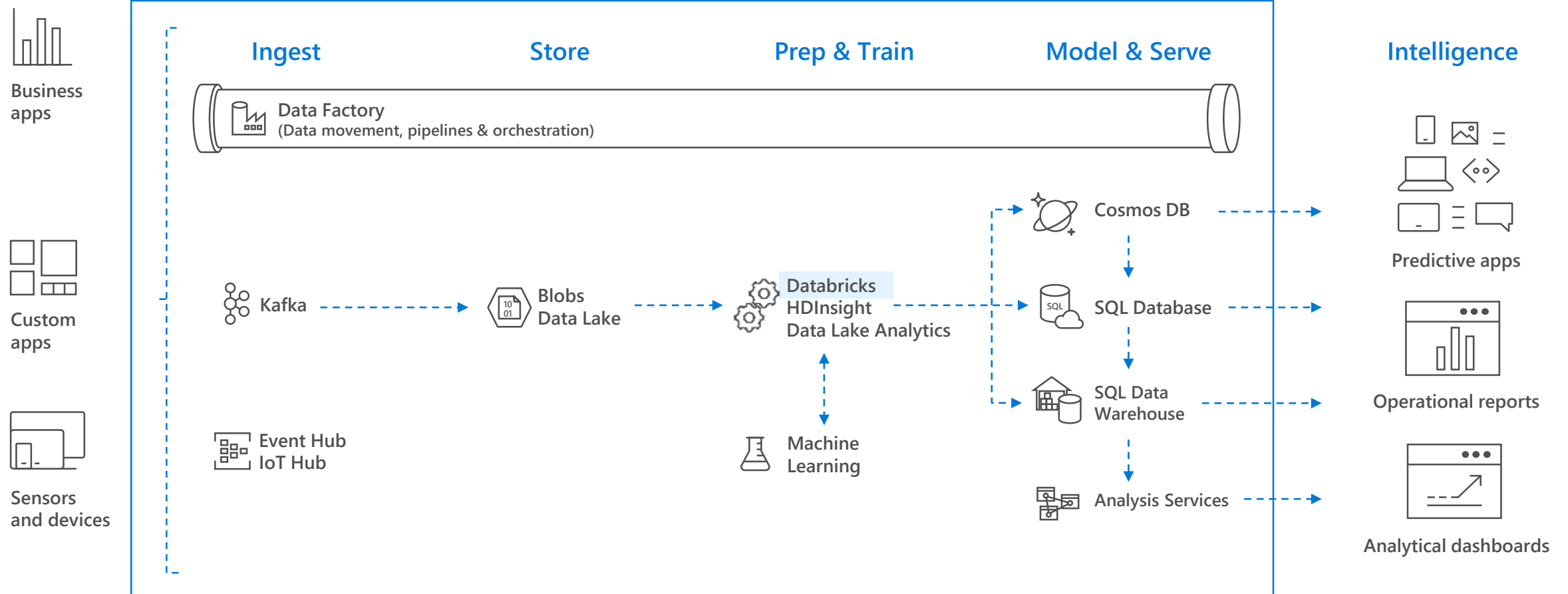
Security and performance

Big Data & Advanced Analytics in Azure

KNOWING THE VARIOUS BIG DATA SOLUTIONS



BIG DATA & ADVANCED ANALYTICS AT A GLANCE



Azure Databricks

Powered by Apache Spark

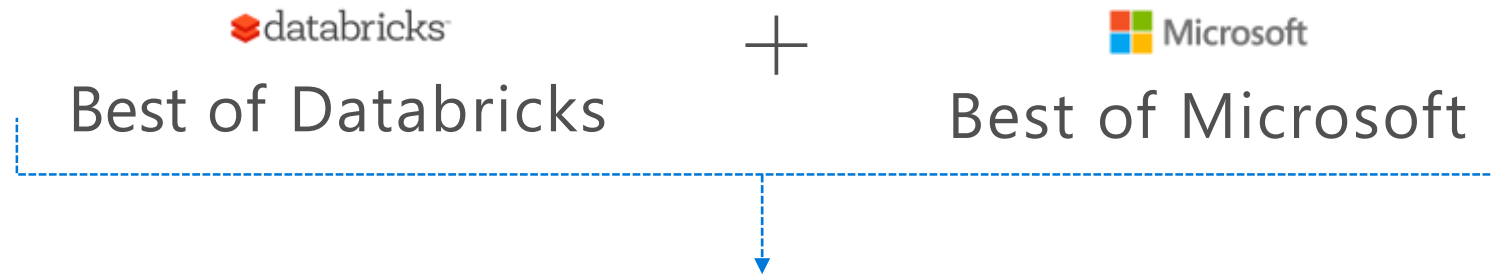
Why Spark?



- 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



 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)



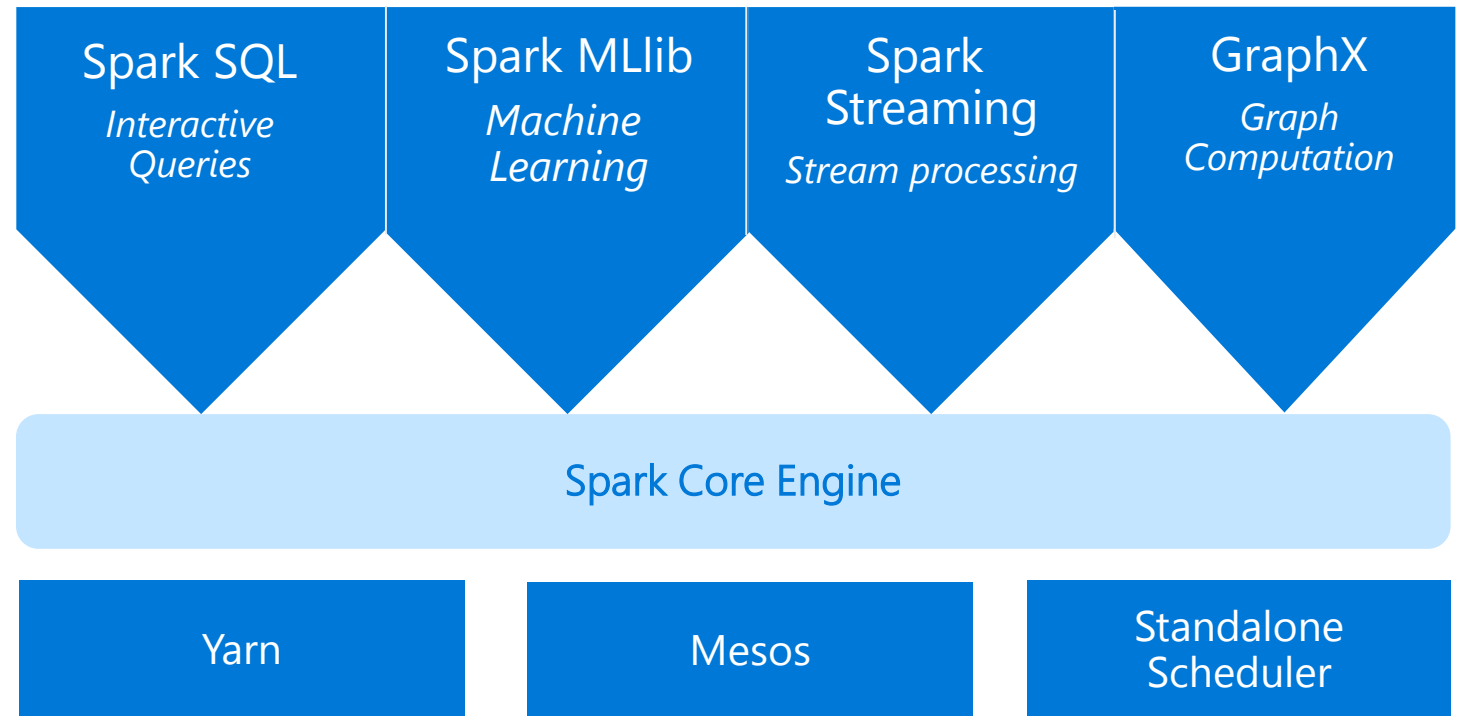
Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs)

A P A C H E S P A R K

An unified, open source, parallel, data processing framework for Big Data Analytics

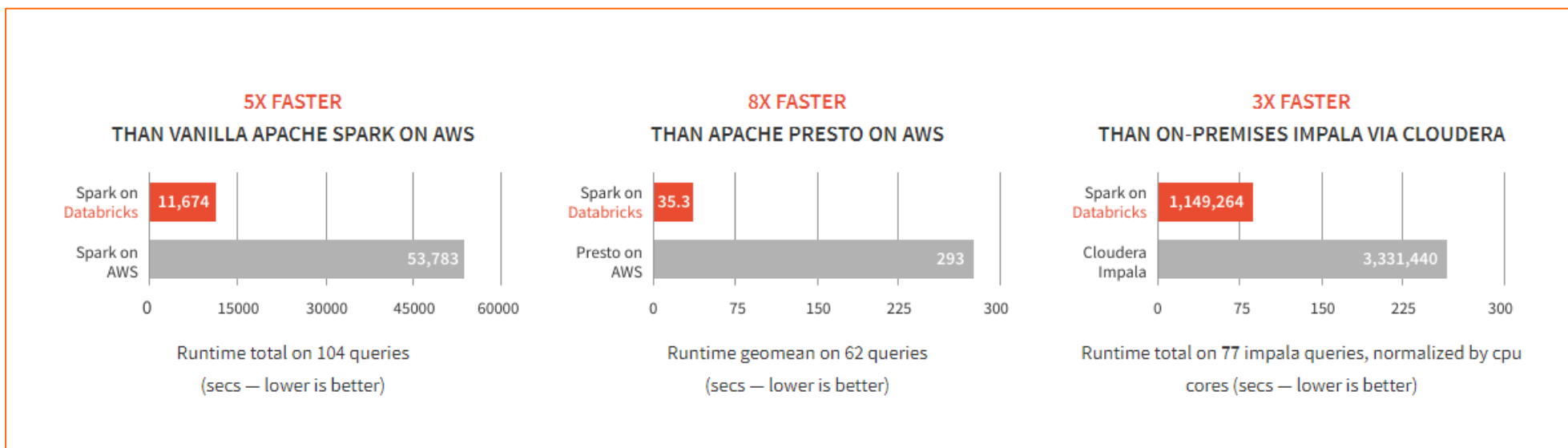
Spark Unifies:

- Batch Processing
- Interactive SQL
- Real-time processing
- Machine Learning
- Deep Learning
- Graph Processing



DATABRICKS SPARK IS FAST

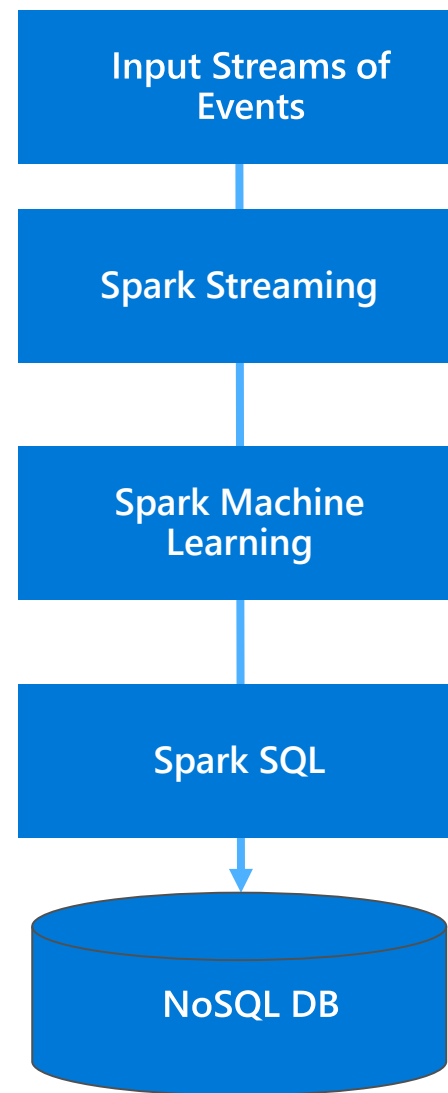
Benchmarks have shown Databricks to often have better performance than alternatives



SOURCE: [Benchmarking Big Data SQL Platforms in the Cloud](#)

ADVANTAGES OF A UNIFIED PLATFORM

- Improves developer productivity—a single consistent set of APIs
- All different systems in Spark share the same abstraction – RDDs (Resilient Distributed Datasets)
- Developers can mix and match different kind of processing in the same application. This is a common requirement for many big data pipelines.
- Performance improves because unnecessary movement of data across engines is eliminated. In many pipelines, data exchange between engines is the dominant cost



Differentiated experience on Azure

ENHANCE PRODUCTIVITY

Get started quickly by launching your new Spark environment with one click.

Share your insights in powerful ways through rich integration with Power BI.

Improve collaboration amongst your analytics team through a unified workspace.

Innovate faster with native integration with rest of Azure platform

BUILD ON THE MOST COMPLIANT CLOUD

Simplify security and identity control with built-in integration with Active Directory.

Regulate access with fine-grained user permissions to Azure Databricks' notebooks, clusters, jobs and data.

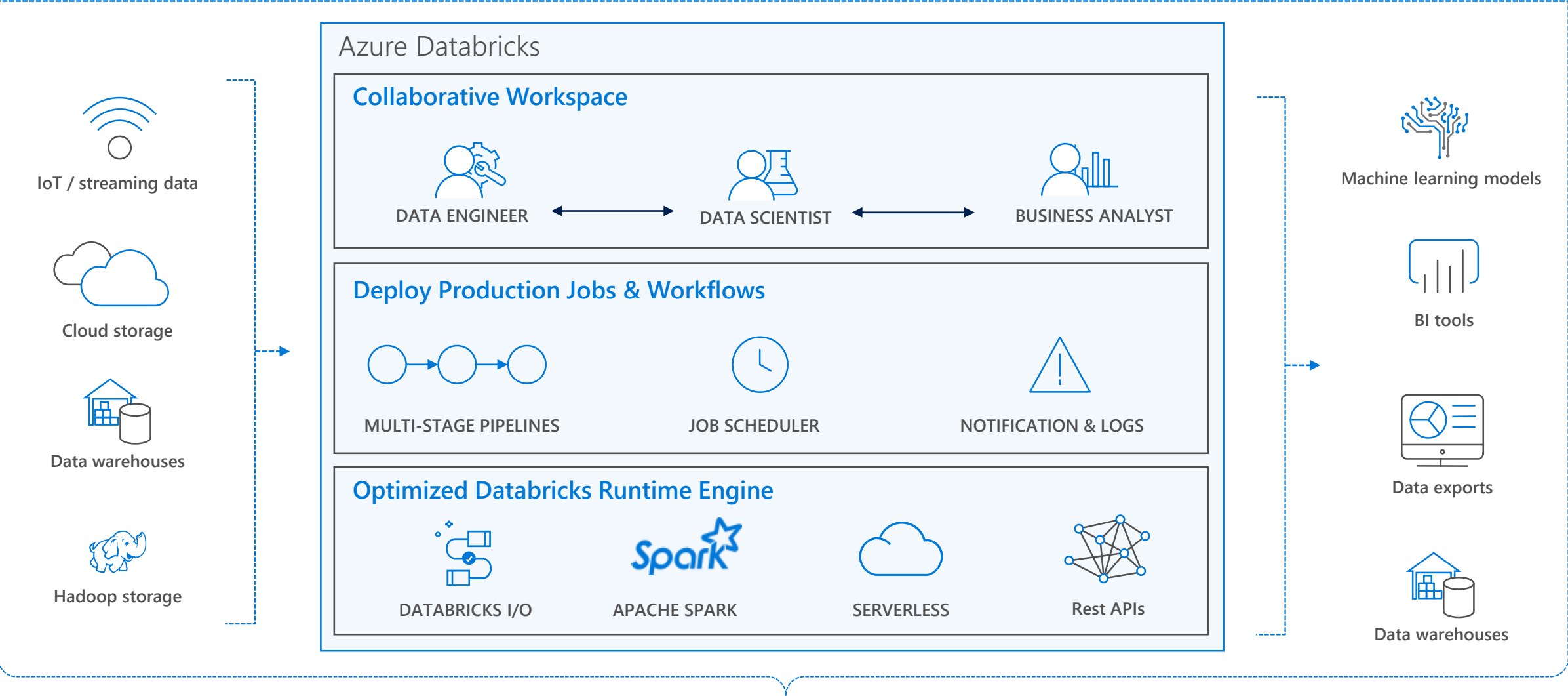
Build with confidence on the trusted cloud backed by unmatched support, compliance and SLAs.

SCALE WITHOUT LIMITS

Operate at massive scale without limits globally.

Accelerate data processing with the fastest Spark engine.

Azure Databricks



Enhance Productivity

Build on secure & trusted cloud

Scale without limits

Collaborative Workspace

GET STARTED IN SECONDS

Single click to launch your new Spark environment

INTERACTIVE EXPLORATION

Explore data using interactive notebooks with support for multiple programming languages including R, Python, Scala, and SQL

COLLABORATION

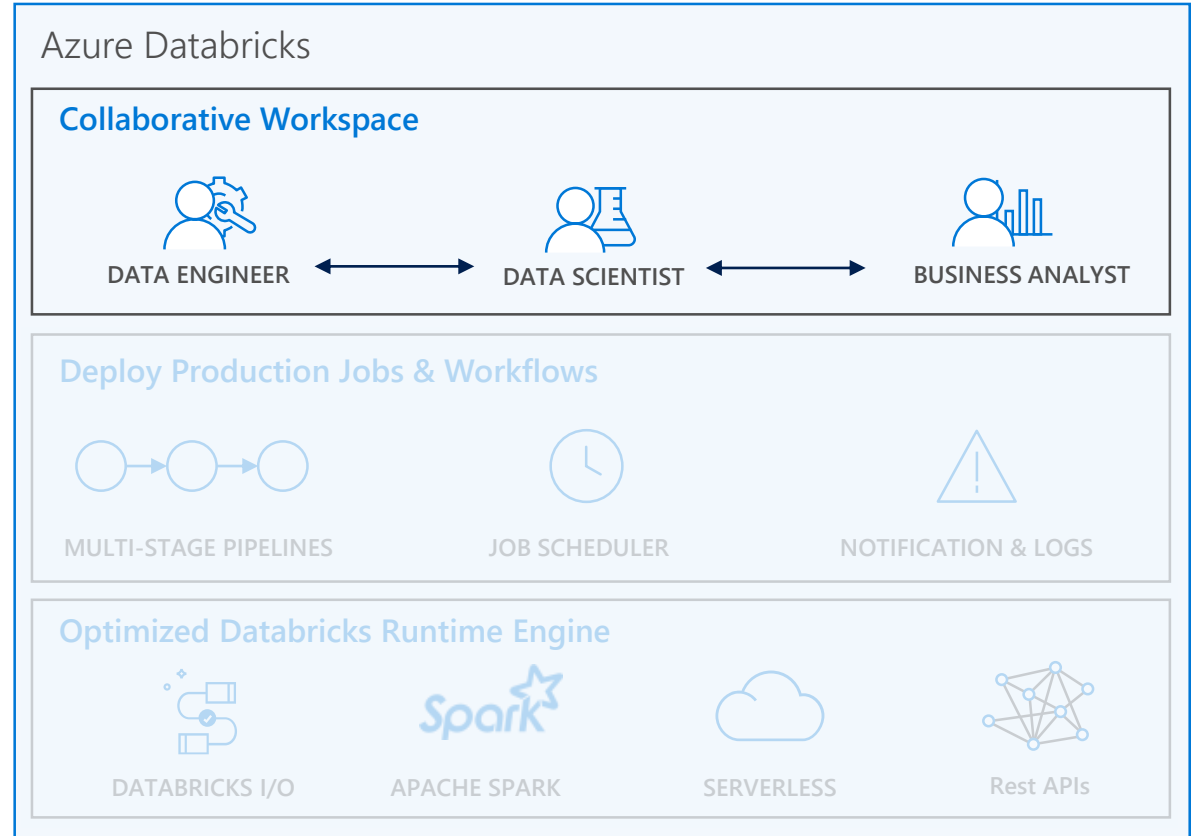
Work on the same notebook in real-time while tracking changes with detailed revision history, GitHub, or Bitbucket

VISUALIZATIONS

Visualize insights through a wide assortment of point-and-click visualizations. Or use powerful scriptable options like matplotlib, ggplot, and D3

DASHBOARDS

Rich integration with PowerBI to discover and share your insights in powerful new ways



Deploy Production Jobs & Workflows

JOBS SCHEDULER

Execute jobs for production pipelines on a specific schedule

NOTEBOOK WORKFLOWS

Create multi-stage pipelines with the control structures of the source programming language

RUN NOTEBOOKS AS JOBS

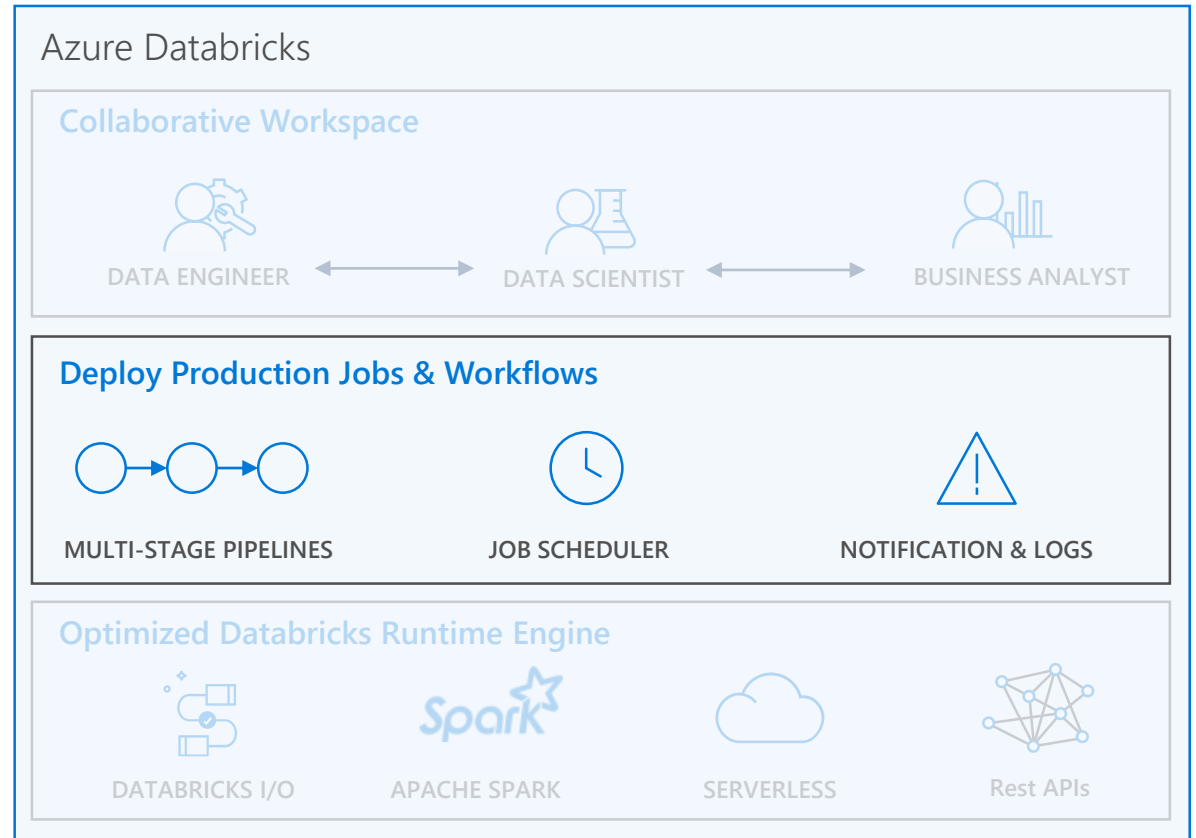
Turn notebooks or JARs into resilient Spark jobs with a click or an API call

NOTIFICATIONS AND LOGS

Set up alerts and quickly access audit logs for easy monitoring and troubleshooting

INTEGRATE NATIVELY WITH AZURE SERVICES

Deep integration with Azure SQL Data Warehouse, Cosmos DB, Azure Data Lake Store, Azure Blob Storage, and Azure Event Hub



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

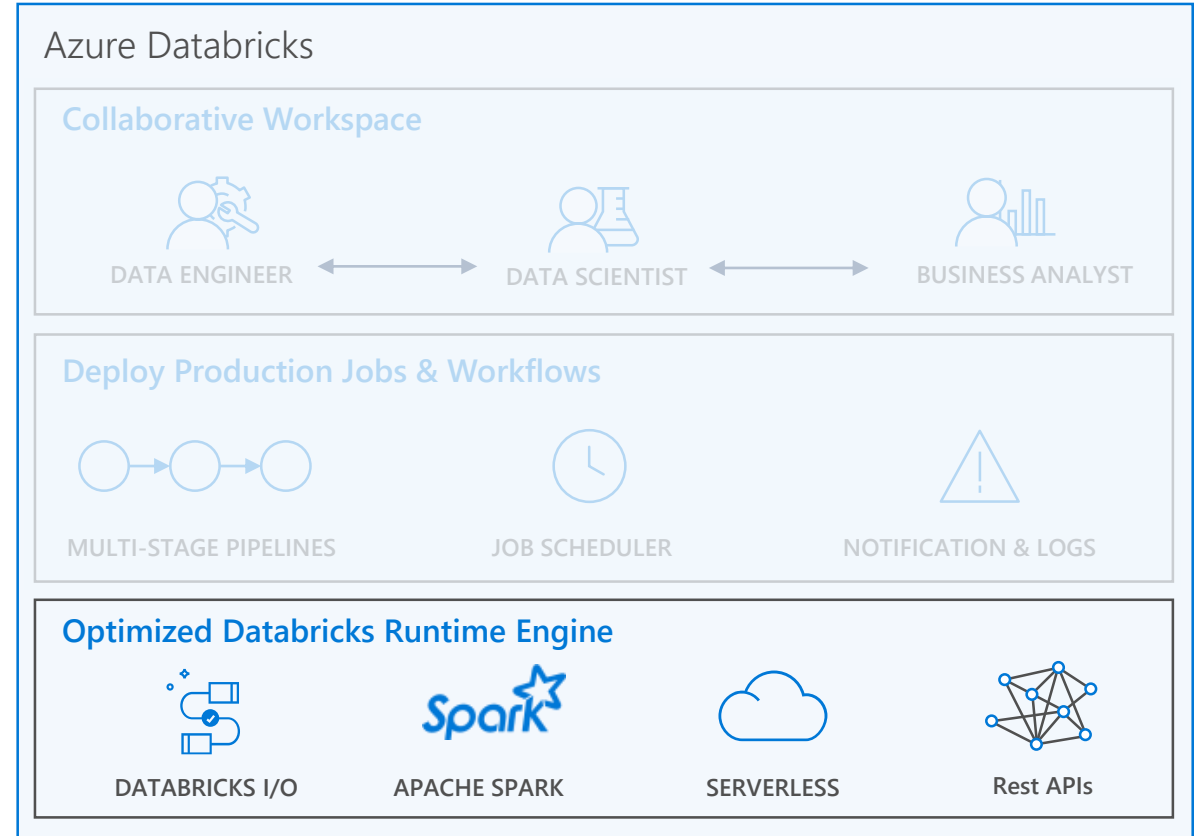
Reap the benefits of a fully managed service and remove the complexity of big data and machine learning

SERVERLESS INFRASTRUCTURE

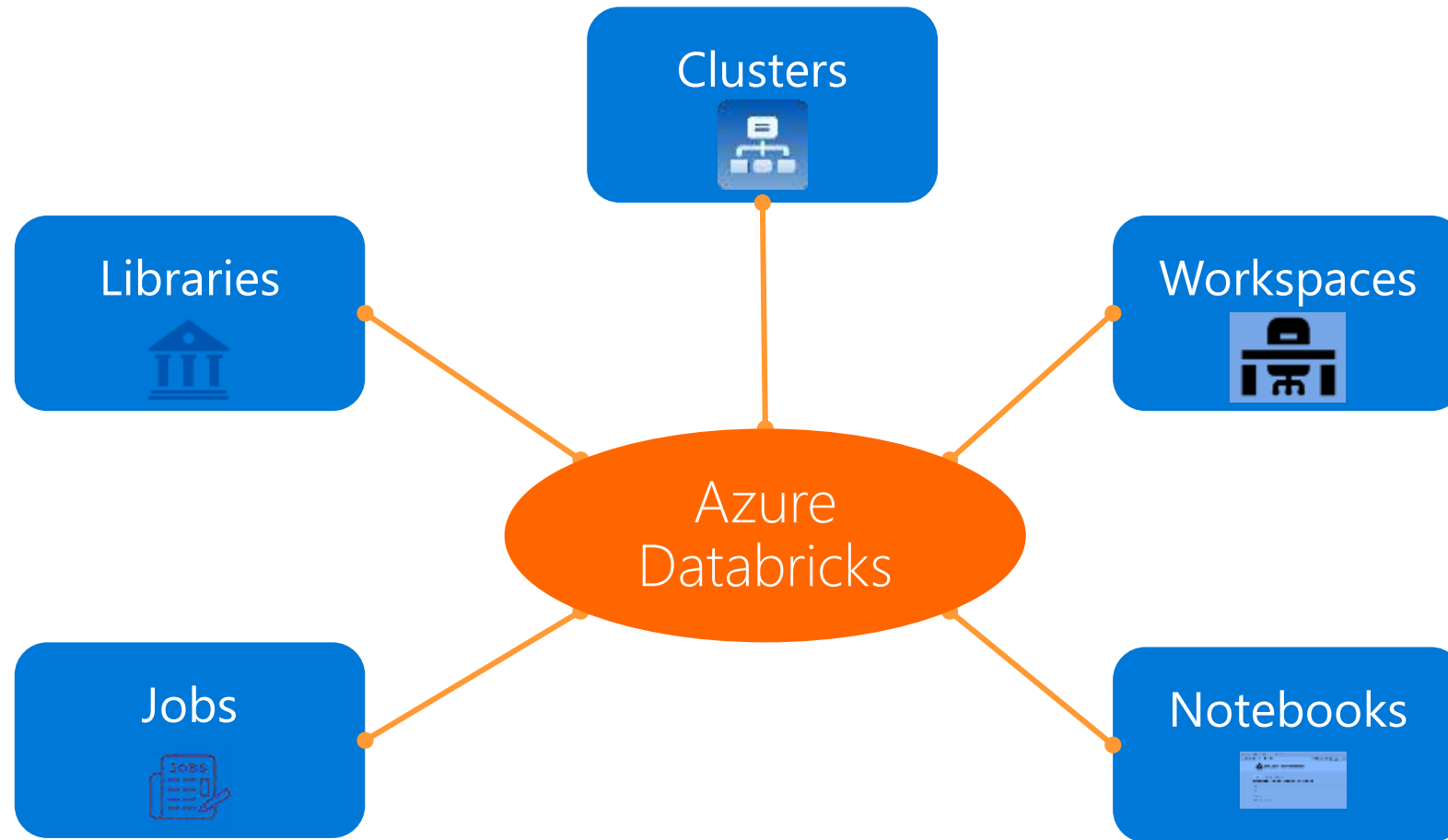
Databricks' serverless and highly elastic cloud service is designed to remove operational complexity while ensuring reliability and cost efficiency at scale

OPERATE AT MASSIVE SCALE

Without limits globally

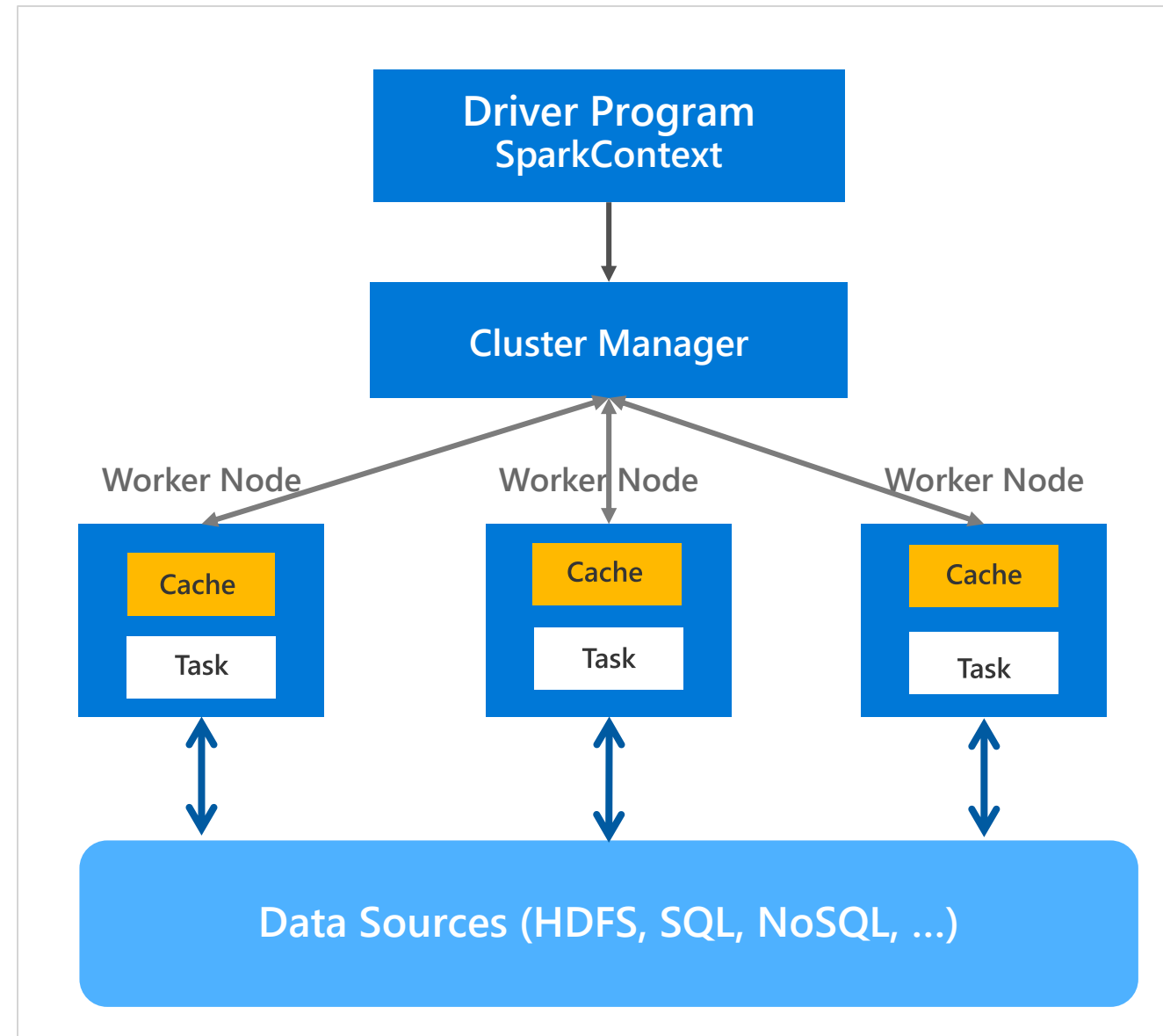


AZURE DATABRICKS CORE ARTIFACTS



GENERAL SPARK CLUSTER ARCHITECTURE

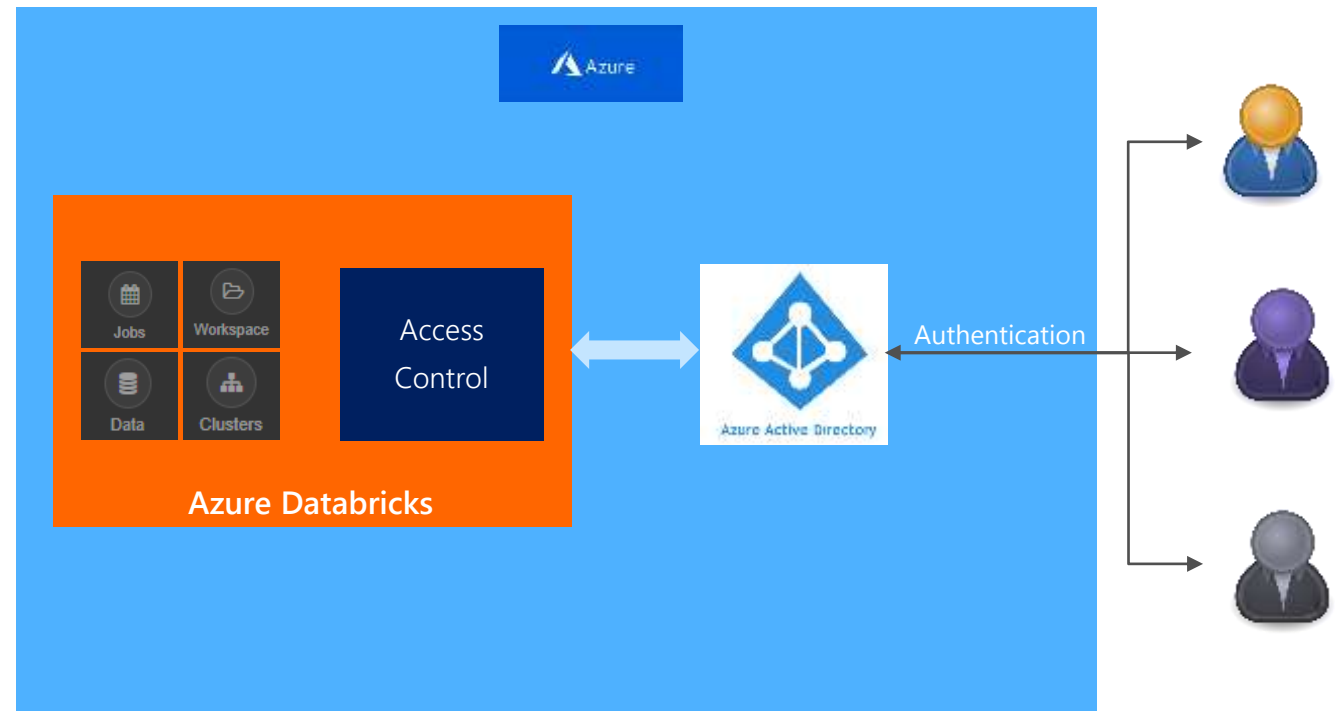
- 'Driver' runs the user's 'main' function and executes the various parallel operations on the worker nodes.
- The results of the operations are collected by the driver
- The worker nodes read and write data from/to Data Sources including HDFS.
- Worker node also cache transformed data in memory as RDDs (Resilient Data Sets).
- Worker nodes and the Driver Node execute as VMs in public clouds (AWS, Google and Azure).



A Z U R E D A T A B R I C K S I N T E G R A T I O N W I T H A A D

Azure Databricks is integrated with AAD—so Azure Databricks users are just regular AAD users

- There is no need to define users—and their access control—separately in Databricks.
- AAD users can be used directly in Azure Databricks for all user-based access control (Clusters, Jobs, Notebooks etc.).
- Databricks has delegated user authentication to AAD enabling single-sign on (SSO) and unified authentication.
- *Notebooks, and their outputs, are stored in the Databricks account.* However, AAD-based access-control ensures that only authorized users can access them.



CLUSTERS: AUTO SCALING AND AUTO TERMINATION

Simplifies cluster management and reduces costs by eliminating wastage

When creating Azure Databricks clusters you can choose Autoscaling and Auto Termination options.

Autoscaling: Just specify the min and max number of clusters. Azure Databricks automatically scales up or down based on load.

Auto Termination: After the specified minutes of inactivity the cluster is automatically terminated.

Benefits:

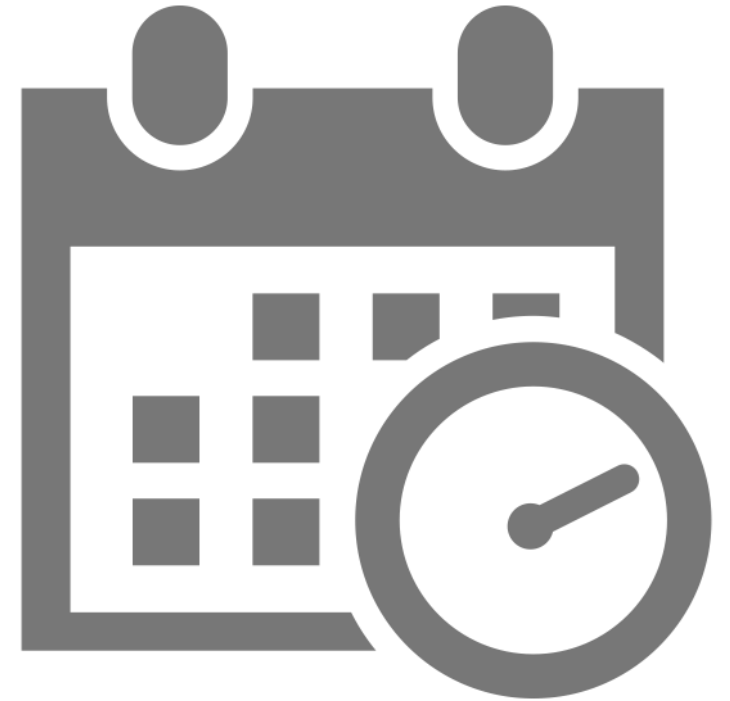
- You do not have to guess, or determine by trial and error, the correct number of nodes for the cluster
- As the workload changes you do not have to manually tweak the number of nodes
- You do not have to worry about wasting resources when the cluster is idle. You only pay for resource when they are actually being used
- You do not have to wait and watch for jobs to complete just so you can shutdown the clusters

The screenshot displays the 'Create Cluster' interface in the Microsoft Azure Databricks portal. The left sidebar contains navigation links for Azure, Databricks, Home, Workspace, Recent, Data, Clusters, Jobs, and Search. The main content area is titled 'Create Cluster' and 'New Cluster'. It includes a 'Cancel' button and a 'Create Cluster' button. A summary bar at the top right indicates '2-8 Workers: 28.0-112.0 GB Memory, 8-32 Cores' and '1 Driver: 14.0 GB Memory, 4 Cores'. The configuration section includes: 'Cluster Type' with 'Serverless Pool (beta, Python/SQL)' and 'Standard' options; 'Cluster Name' with the text 'MyDemoCluster'; 'Databricks Runtime Version' set to '3.3 (includes Apache Spark 2.2.0, Scala 2.11)'; 'Driver Type' set to 'Same as worker' with '14.0 GB Memory, 4 Cores'; 'Worker Type' set to 'Standard_DS3_v2 (beta)' with '14.0 GB Memory, 4 Cores'; 'Min Workers' set to '2' and 'Max Workers' set to '8', with a checked 'Enable Autoscaling' option; and 'Auto Termination' set to '120 minutes of inactivity' with a checked 'Terminate after' option. Red boxes highlight the 'Min Workers', 'Max Workers', 'Enable Autoscaling', and 'Auto Termination' settings.

J O B S

Jobs are the mechanism to submit Spark application code for execution on the Databricks clusters

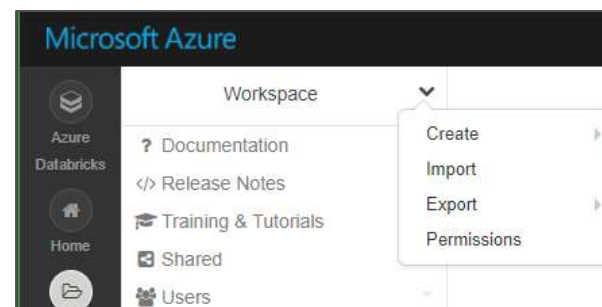
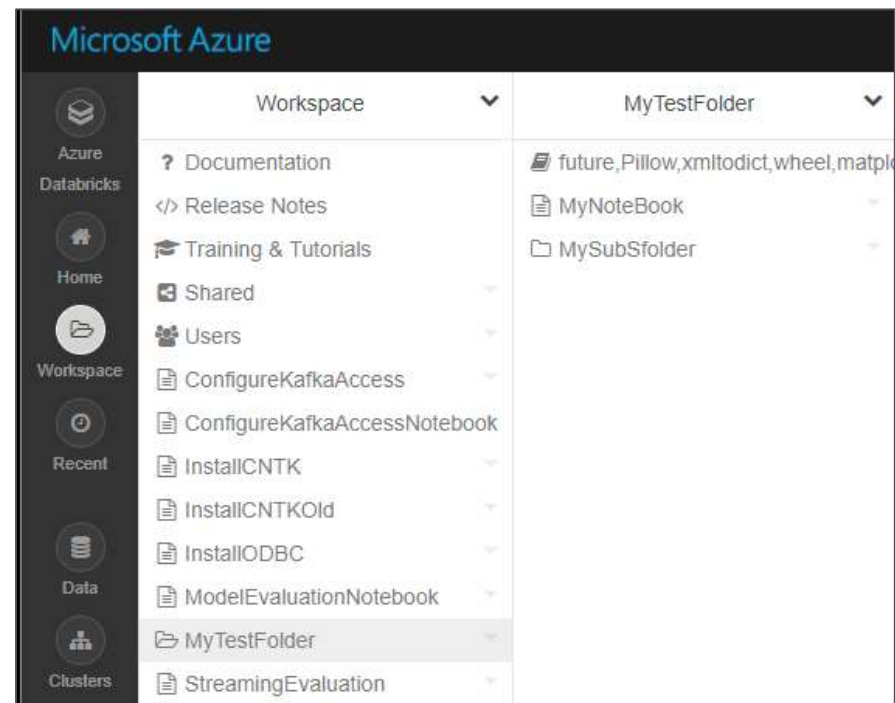
- Spark application code is submitted as a 'Job' for execution on Azure Databricks clusters
- Jobs execute either 'Notebooks' or 'Jars'
- Azure Databricks provide a comprehensive set of graphical tools to create, manage and monitor Jobs.



WORKSPACES


Workspaces enables users to organize—and share—their Notebooks, Libraries and Dashboards

- Workspaces—sort of like Directories—are a convenient way to organize an user's Notebook, Libraries and Dashboards.
- Everything in a workspace is organized into hierarchical folders. Folders can hold Libraries, Notebooks, Dashboard or more (sub) folders.
 - Icons indicate the type of the object contained in a folder
- Every user has one directory that is private and unshared.
 - By default, the workspace and all its contents are available to users.
- Fine grained access control can be defined on workspaces (next slide) to enable *secure collaboration with colleagues*.



AZURE DATABRICKS NOTEBOOKS OVERVIEW

Notebooks are a popular way to develop, and run, Spark Applications

- Notebooks are not only for authoring Spark applications but can be *run/executed directly* on clusters
 - **Shift+Enter**
 - click the  at the top right of the cell in a notebook
 - Submit via Job
- Notebooks support fine grained permissions—so they can be *securely shared* with colleagues for collaboration (see following slide for details on permissions and abilities)
- Notebooks are well-suited for prototyping, rapid development, exploration, discovery and iterative development



Notebooks typically consist of code, data, visualization, comments and notes

LIBRARIES OVERVIEW

Enables external code to be imported and stored into a Workspace

- Libraries are containers to hold all your *Python, R, Java/Scala* libraries.
- Libraries resides within workspaces or folders.
- Libraries are created by importing the source code
- After importing libraries are immutable—can be deleted or overwritten only.
- You can customize installation of libraries via [Init Scripts](#) by writing custom UNIX scripts
- Libraries can also be managed via the [Library API](#)

This screenshot shows the 'Create Library' page in the Microsoft Azure Portal, specifically for creating a new Python library. The 'Language' dropdown is set to 'Upload Python Egg or PyPI'. Under the 'Install PyPI Package' section, there is a 'PyPI Name' field with the example 'PyPI Package (e.g. simplejson or simplejson==3.8.0)' and an 'Install Library' button. The 'Upload Egg' section has a 'Library Name' field and a large area for an 'Egg File' with the instruction 'Drop library egg here to upload' and a 'Create Library' button at the bottom.

This screenshot shows the 'Create Library' page in the Microsoft Azure Portal for creating a new R library. The 'Source' dropdown is set to 'R Library'. The 'Install from' dropdown is set to 'CRAN-like Repository'. The 'Repository' field contains the URL 'https://cloud.r-project.org'. There is a 'Package' field and a 'Create Library' button at the bottom.

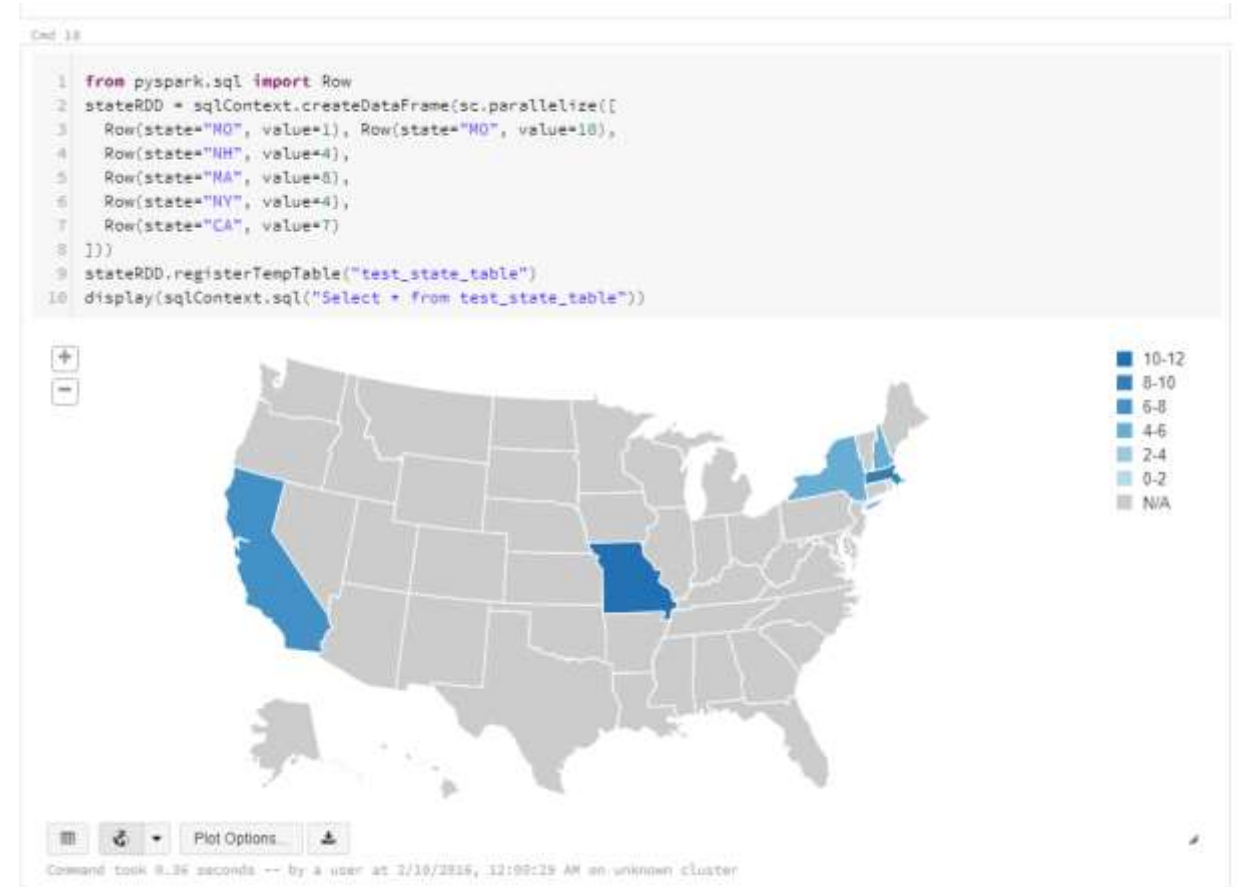
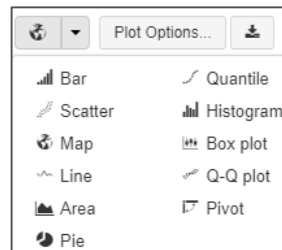
This screenshot shows the 'Create Library' page in the Microsoft Azure Portal for creating a new Java/Scala library. The 'Source' dropdown is set to 'Upload Java/Scala JAR'. The 'Library Name' field contains 'My Library'. There is a large area for a 'JAR File' with the instruction 'Drop library JAR here to upload' and a 'Create Library' button at the bottom.

This screenshot shows the 'Create Library' page in the Microsoft Azure Portal for creating a new library from Maven artifacts. The 'Source' dropdown is set to 'Maven Coordinate'. The 'Install Maven Artifacts' section has a 'Coordinate' field with the example 'Maven Coordinate (e.g. com.databricks:spark-csv_2.10:1.0.0)'. Below this is a button that says 'Search Spark Packages and Maven Central'. There is also an 'Advanced Options' link and a 'Create Library' button at the bottom.

VISUALIZATION

Azure Databricks supports a number of visualization plots out of the box

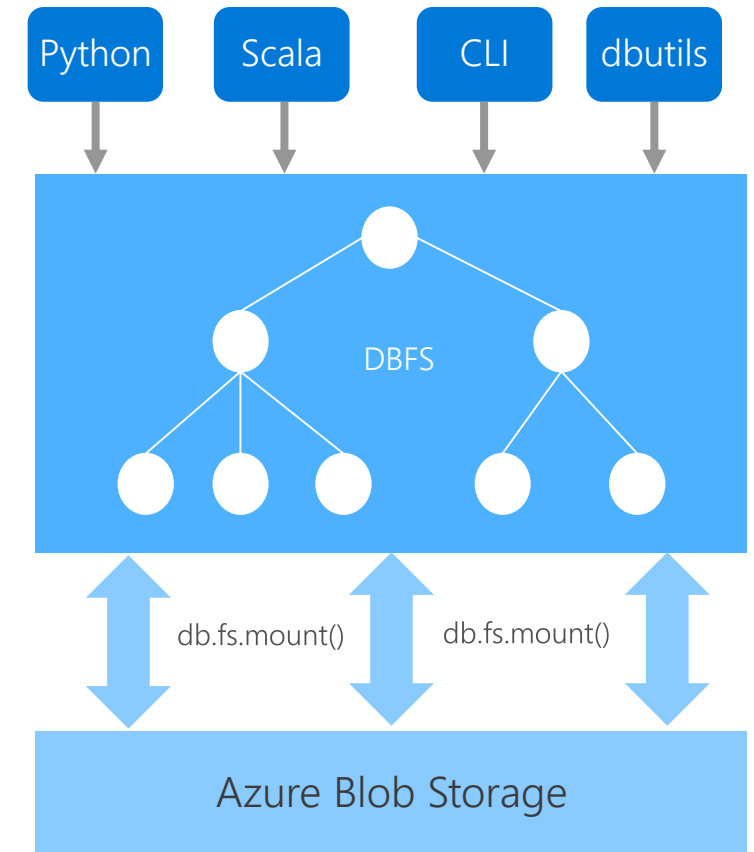
- All notebooks, *regardless of their language*, support Databricks visualizations.
- When you run the notebook the visualizations are rendered inside the notebook in-place
- The visualizations are written in HTML.
 - You can save the HTML of the entire notebook by exporting to HTML.
 - If you use Matplotlib, the plots are rendered as images so you can just right click and download the image
- You can change the plot type just by picking from the selection



DATABRICKS FILE SYSTEM (DBFS)

Is a distributed File System (DBFS) that is a layer over Azure Blob Storage

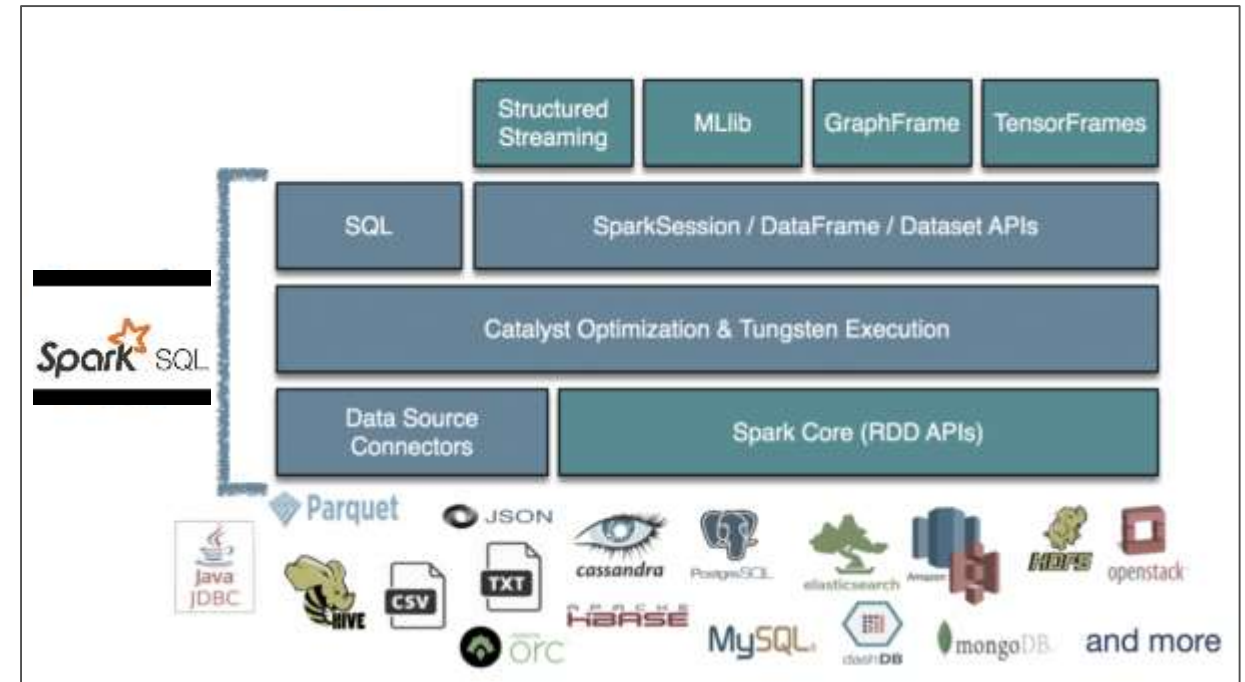
- Azure Storage buckets can be mounted in DBFS so that users can directly access them without specifying the storage keys
- DBFS mounts are created using `dbutils.fs.mount()`
- Azure Storage data can be cached locally on the SSD of the worker nodes
- Available in both Python and Scala and accessible via a DBFS CLI
- Data persist in Azure Blob Storage – is not lost even after cluster termination
- Comes pre-installed on Spark clusters in Databricks



SPARK SQL OVERVIEW

Spark SQL is a distributed SQL query engine for processing structured data

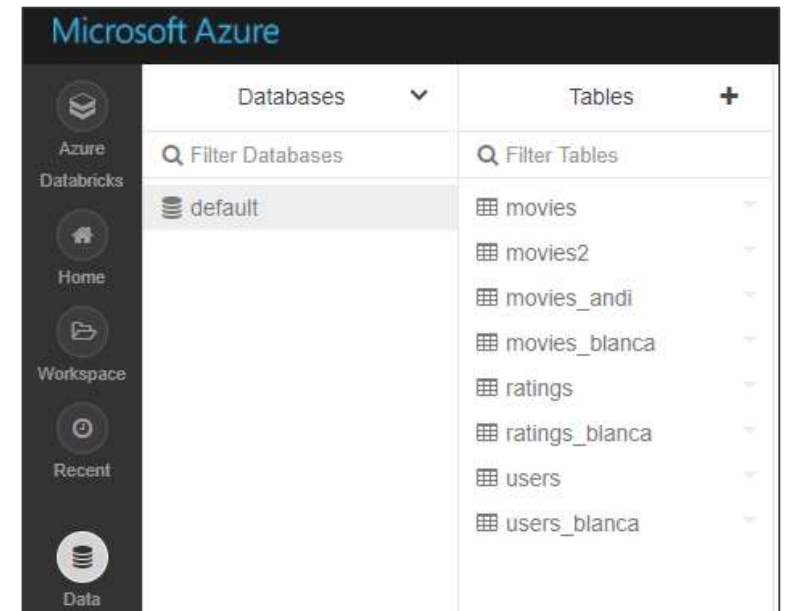
- Can query data stored in wide variety of data sources—external databases, structured data files, Hive tables and more.
- Data can be queried using either SQL or HiveQL
- Has bindings in Python, Scala and Java
- Has built-in support for structured streaming.
- Built using the [Catalyst optimizer](#) and [Tungsten execution](#)



DATABASES AND TABLES OVERVIEW

Tables enable data to be structured and queried using Spark SQL or any of the Spark's language APIs

- Databases are a collection of related tables
- Tables are defined using the GUI in the console or programmatically using APIs or Notebooks
- Databricks uses the Hive metastore to manage tables, and supports all file formats and Hive data sources.
- There are multiple ways to create tables (see next slide).
- Like Apache Spark DataFrames, any Spark operation can be applied to Tables (including caching, filtering).
- Partitioned Tables and Partition Pruning: Spark SQL is able to dynamically generate partitions at the file storage level to provide partition columns for tables. When the table is scanned, Spark pushes down the filter predicates involving the [partitionBy](#) keys for partition pruning.



SPARK MACHINE LEARNING (ML) OVERVIEW

Enables Parallel, Distributed ML for large datasets on Spark Clusters

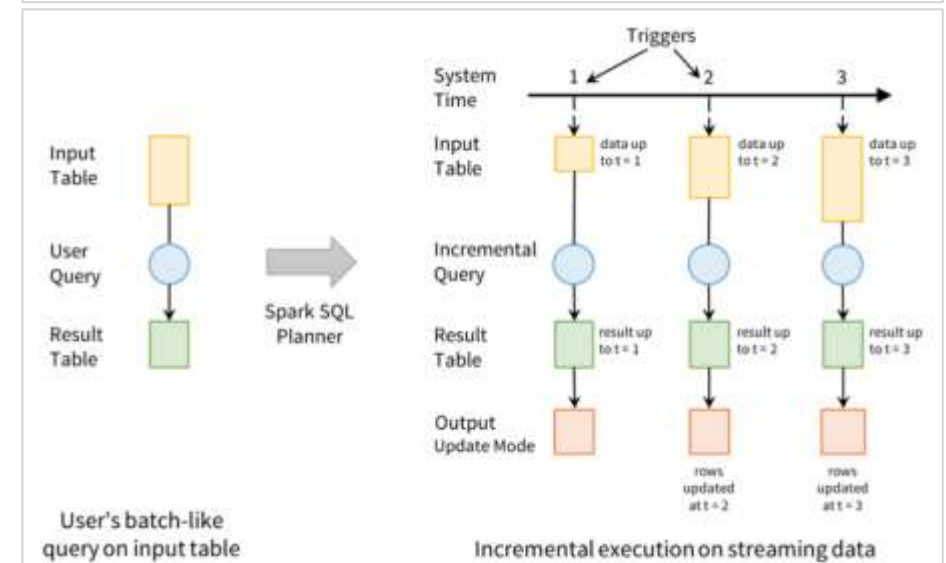
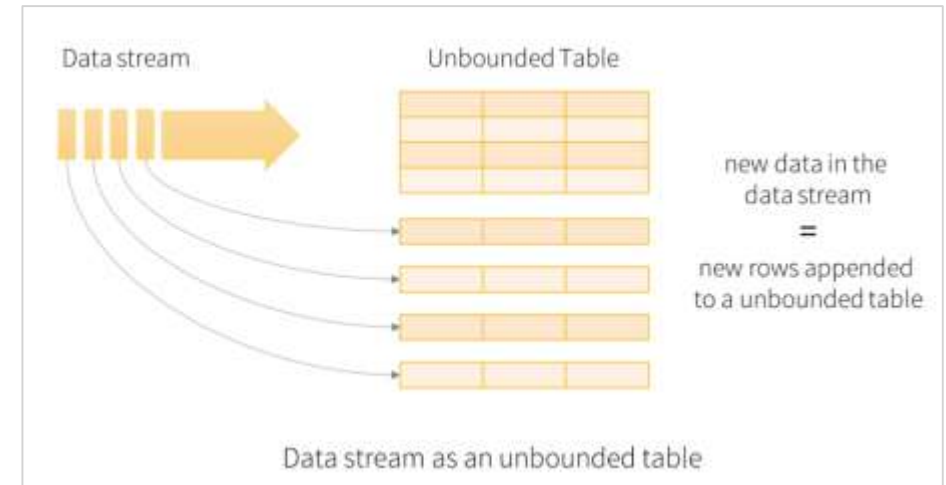
- Offers a set of parallelized machine learning algorithms (MMLSpark, Spark ML, Deep Learning, SparkR)
- Supports [Model Selection](#) (hyperparameter tuning) using [Cross Validation](#) and [Train-Validation Split](#).
- Supports Java, Scala or Python apps using [DataFrame](#)-based API (as of Spark 2.0). Benefits include:
 - An uniform API across ML algorithms and across multiple languages
 - Facilitates [ML pipelines](#) (enables combining multiple algorithms into a single pipeline).
 - Optimizations through Tungsten and Catalyst
- Spark MLlib comes pre-installed on Azure Databricks
- 3rd Party libraries supported include: [H2O Sparkling Water](#), [SciKit-learn](#) and [XGBoost](#)



SPARK STRUCTURED STREAMING OVERVIEW

A unified system for end-to-end fault-tolerant, exactly-once stateful stream processing

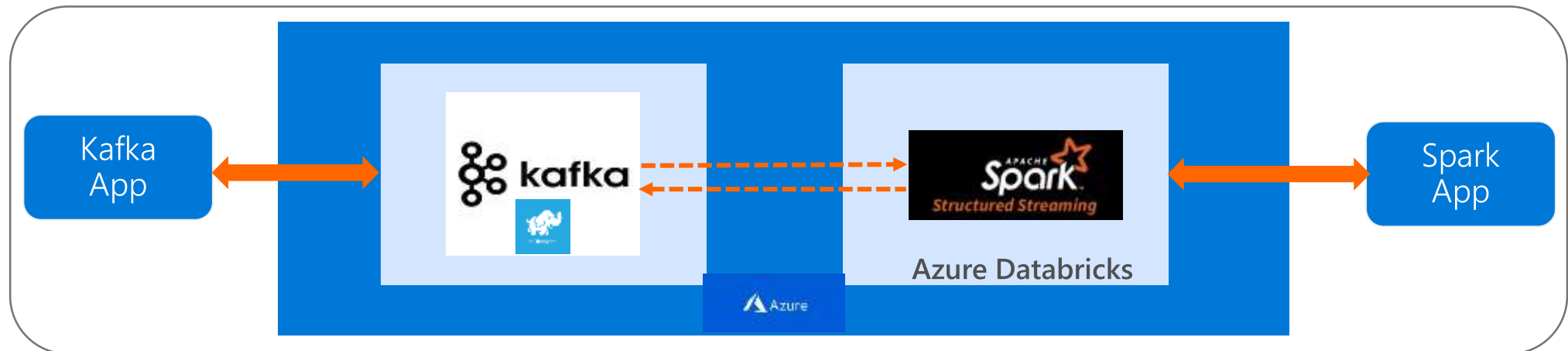
- Unifies streaming, interactive and batch queries—a single API for both static bounded data and streaming unbounded data.
- Runs on Spark SQL. Uses the Spark SQL [Dataset/DataFrame](#) API used for batch processing of static data.
- Runs incrementally and continuously and updates the results as data streams in.
- Supports app development in Scala, Java, Python and R.
- Supports streaming aggregations, event-time windows, windowed grouped aggregation, stream-to-batch joins.
- Features streaming deduplication, multiple output modes and APIs for managing/monitoring streaming queries.
- Built-in sources: Kafka, File source (json, csv, text, parquet)



APACHE KAFKA FOR HDINSIGHT INTEGRATION

Azure Databricks Structured Streaming integrates with Apache Kafka for HDInsight

- Apache Kafka for Azure HDInsight is an enterprise grade streaming ingestion service running in Azure.
- Azure Databricks Structured Streaming applications can use Apache Kafka for HDInsight as a data source or sink.
- No additional software (gateways or connectors) are required.
- Setup: Apache Kafka on HDInsight does not provide access to the Kafka brokers over the public internet. So the Kafka clusters and the Azure Databricks cluster must be located in the same Azure Virtual Network.



Note: Azure Databricks Structured Streaming integration with **Azure Event Hubs** is forthcoming

SPARK GRAPHX OVERVIEW

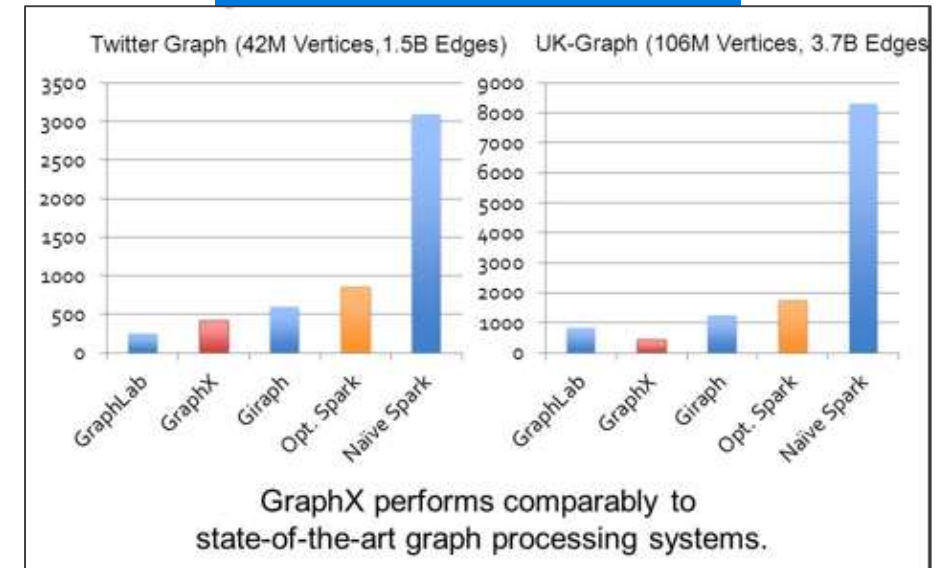
A set of APIs for graph and graph-parallel computation.

- Unifies ETL, exploratory analysis, and iterative graph computation within a single system.
- Developers can:
 - [view](#) the same data as both graphs and collections,
 - [transform](#) and [join](#) graphs with RDDs, and
 - write custom iterative graph algorithms using the [Pregel API](#).
- Currently only supports using the Scala and RDD APIs.

Algorithms

- PageRank
- Connected components
- Label propagation
- SVD++
- Strongly connected components
- Triangle count

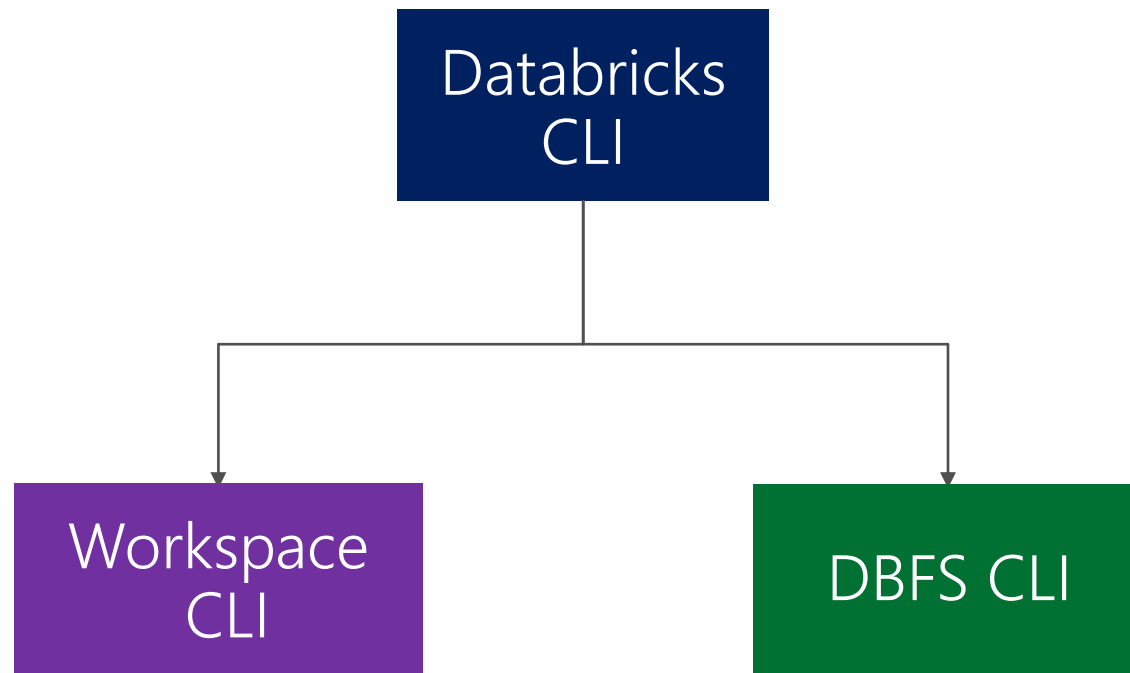
PageRank Benchmark



Source: [AMPLab](#)

DATABRICKS CLI

An easy to use interface built on top of the Databricks [REST API](#)



Currently, the CLI fully implements the [DBFS API](#) and the [Workspace API](#)

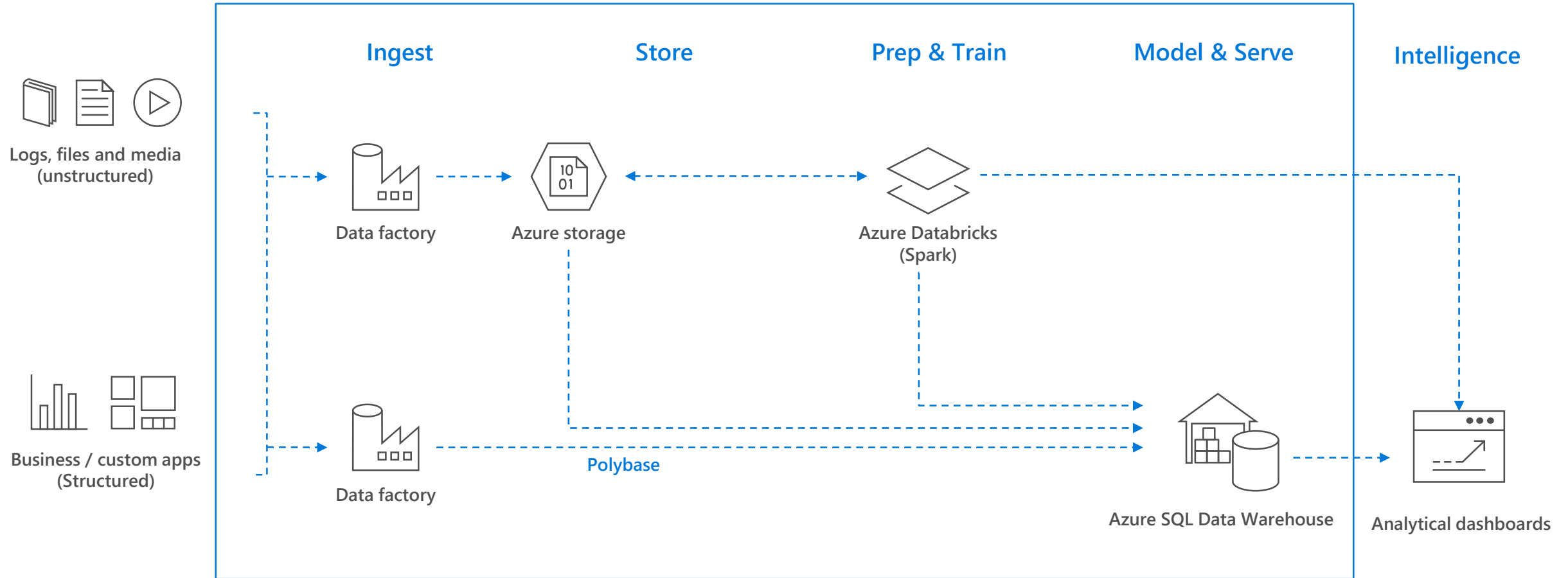
DATABRICKS REST API

Databricks REST API

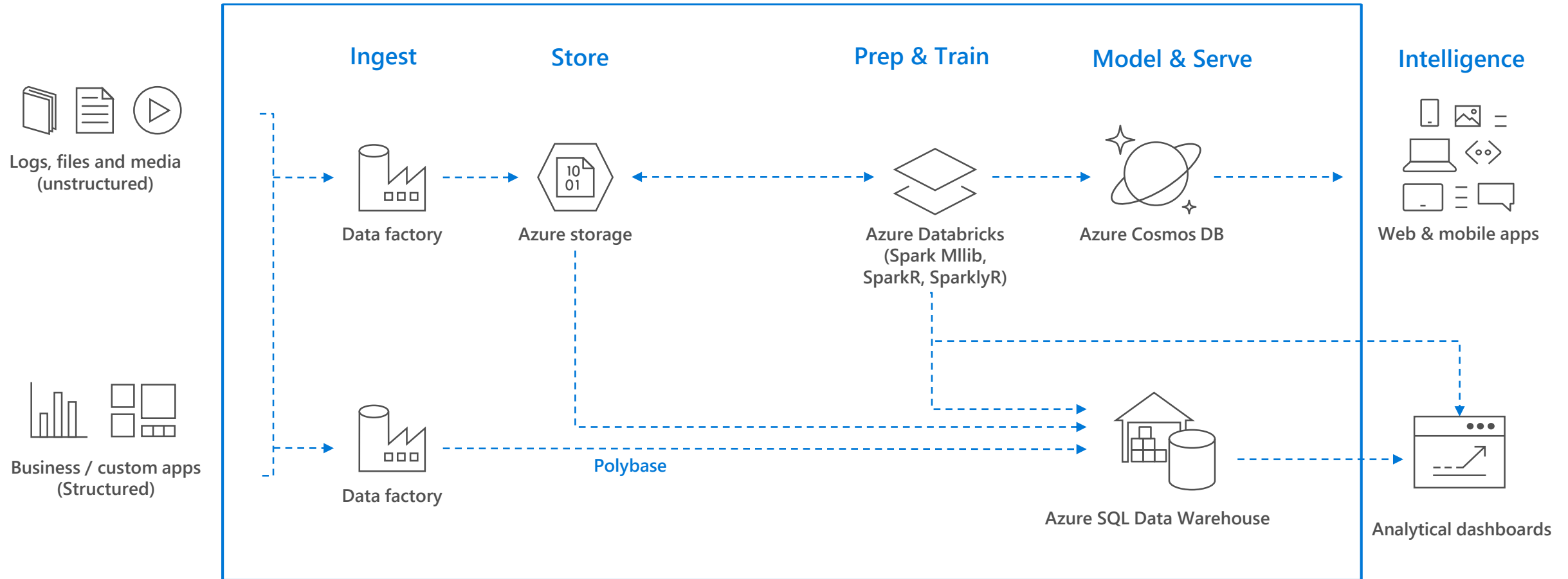
Cluster API	Create/edit/delete clusters
DBFS API	Interact with the Databricks File System
Groups API	Manage groups of users
Instance Profile API	Allows admins to add, list, and remove instances profiles that users can launch clusters with
Job API	Create/edit/delete jobs
Library API	Create/edit/delete libraries
Workspace API	List/import/export/delete notebooks/folders

Use Cases

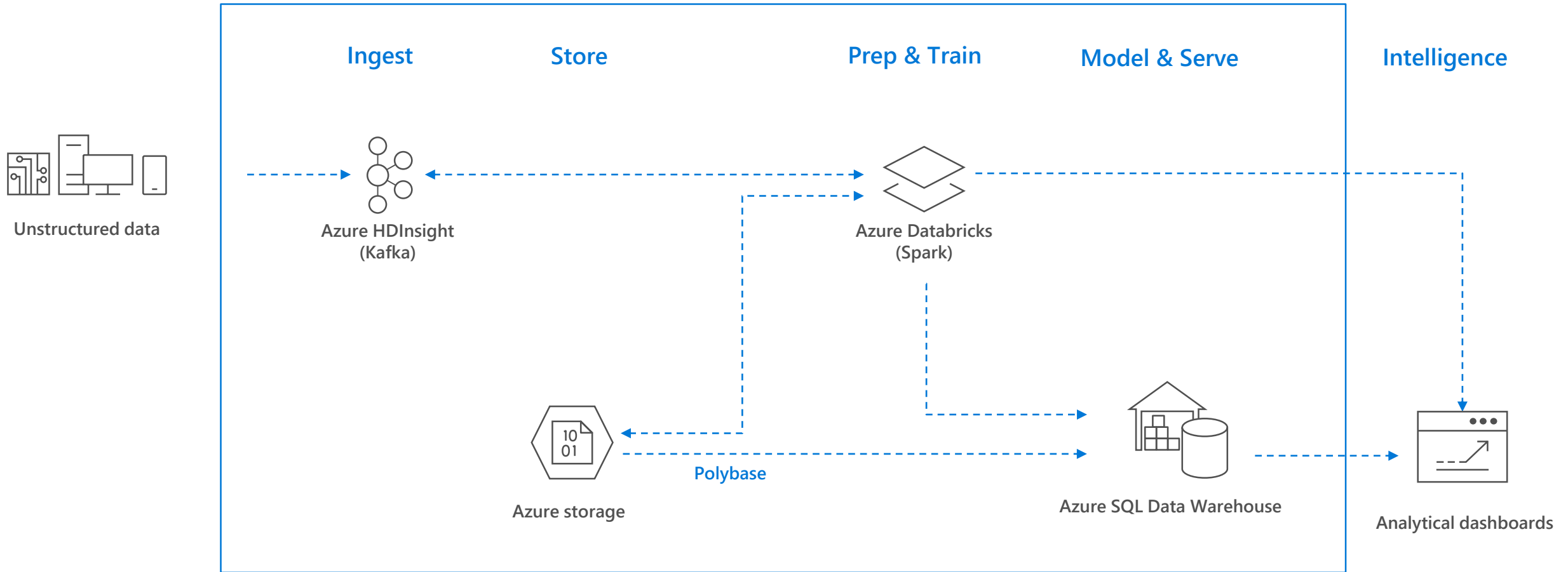
Modern Big Data Warehouse



Advanced Analytics on Big Data



Real-time analytics on Big Data



Pricing & Product Guidance

Big Data OSS - Comparison

Azure HDInsight (1st party + Support)

What it is

- **Hadoop** (Hortonworks' Distribution) as a managed service supporting a variety of open-source analytics engines such as Apache Spark, Hive LLAP, Storm, Kafka, HBase.
- Security via Ranger (Kerberos based)

Pricing

- Priced to compete with AWS EMR. Standard offering.

Use When

- Customer prefers a **PaaS** like experience to address big data use cases by working with different OSS analytics engines to address big data use cases. Cost sensitive.

Azure Databricks (1st party + Support)

What it is

- Databricks **Spark**, the most popular open-source analytics engine, as a managed service providing an easy and fast way to unlock big data use cases. Offers best-in-class notebooks experience for productivity and collaboration as well integration with Azure Data Warehouse, Power BI, etc
- Security via native Azure AD integration

Pricing

- Priced to match Databricks on AWS. Premium offering.

Use When

- Customer prefers **SaaS** like experience to address big data use cases and values Databricks' ease of use, productivity & collaboration features.

3rd Party Offerings

What it is

Hadoop distributions from Cloudera, MapR & Hortonworks available on Azure Marketplace as IaaS VMs.

Pricing

- N/A. Vendor prices their products.

Use When

- Customer wants to move their on premises Hadoop distribution to Azure IaaS using their existing licenses.

LOOKING ACROSS THE OFFERINGS

Azure HDInsight

What It Is

- Hortonworks distribution as a first party service on Azure
- Big Data engines support – Hadoop Projects, Hive on Tez, Hive LLAP, Spark, HBase, Storm, Kafka, R Server
- Best-in-class developer tooling and Monitoring capabilities
- **Enterprise Features**
 - VNET support (join existing VNETs)
 - Ranger support (Kerberos based Security)
 - Log Analytics via OMS
 - Orchestration via Azure Data Factory
 - Available in most Azure Regions (27) including Gov Cloud and Federal Clouds

Guidance

- Customer needs Hadoop technologies other than, or in addition to Spark
- Customer prefers Hortonworks Spark distribution to stay closer to OSS codebase and/or 'Lift and Shift' from on-premises deployments
- Customer has specific project requirements that are only available on HDInsight

Azure Databricks

What It Is

- Databricks' Spark service as a first party service on Azure
- Single engine for Batch, Streaming, ML and Graph
- Best-in-class notebooks experience for optimal productivity and collaboration
- **Enterprise Features**
 - Native Integration with Azure for Security via AAD (OAuth)
 - Optimized engine for better performance and scalability
 - RBAC for Notebooks and APIs
 - Auto-scaling and cluster termination capabilities
 - Native integration with SQL DW and other Azure services
 - Serverless pools for easier management of resources

Guidance

- Customer needs the best option for Spark on Azure
- Customer teams are comfortable with notebooks and Spark
- Customers need Auto-scaling and
- Customer needs to build integrated and performant data pipelines
- Customer is comfortable with limited regional availability (3 in preview, 8 by GA)

Azure ML

What It Is

- Azure first party service for Machine Learning
- Leverage existing ML libraries or extend with Python and R
- Targets emerging data scientists with drag & drop offering
- Targets professional data scientists with
 - Experimentation service
 - Model management service
 - Works with customers IDE of choice

Guidance

- Azure Machine Learning Studio is a GUI based ML tool for emerging Data Scientists to experiment and operationalize with least friction
- Azure Machine Learning Workbench is not a compute engine & uses external engines for Compute, including SQL Server and Spark
- AML deploys models to HDI Spark currently
- AML should be able to deploy Azure Databricks in the near future

Demo

Azure Databricks – service home page

The screenshot displays the Azure Databricks (preview) service page within the Microsoft Azure portal. The interface includes a top navigation bar with the Microsoft Azure logo, a breadcrumb trail (New > Marketplace > Data + Analytics > Azure Databricks (preview)), a search bar, and user information (nishanth@microsoft.c... MICROSOFT (MICROSOFT.ONL...)).

On the left, a sidebar lists various Azure services: New, Dashboard, All resources, Resource groups, App Services, SQL databases, SQL data warehouses, Azure Cosmos DB, Virtual machines, Load balancers, Storage accounts, Virtual networks, Azure Active Directory, Monitor, Security Center, Cost Management + Bil..., Help + support, Subscriptions, and Data Lake Store. A 'More services >' link is at the bottom.

The main content area for 'Azure Databricks (preview)' features the Microsoft logo and the following text:

- DATABRICKS IS A TRULY UNIFIED APPROACH TO DATA ANALYTICS AT SCALE**
- Founded by the team who created Apache Spark, Databricks provides a Unified Analytics Platform that accelerates innovation by unifying data science, engineering, and business.
- UNIFIED EXPERIENCE ACROSS TEAMS**
- A collaborative workspace for data science teams to work with data engineering and lines of business.
- UNIFIED ANALYTICS WORKFLOWS**
- One environment from data preparation to exploration and model building to production.
- UNIFIED INFRASTRUCTURE**
- Fully managed, serverless cloud infrastructure for isolation, automation, and cost control.

Below the text are social media icons for Twitter, Facebook, LinkedIn, YouTube, and Google+. A smaller inset image shows the Databricks workspace interface with a 'Create' button at the bottom.

Azure Databricks – creating a workspace

The screenshot shows the Azure Databricks Service creation page. The left sidebar contains a list of services, with 'Azure Databricks' at the bottom. The main area displays the 'Azure Databricks Service' configuration form. The form includes fields for 'Workspace name' (ntedemodbr12252017), 'Subscription' (Azure conversion - External), 'Resource group' (ntedemorg), and 'Location' (West US). A 'Create' button is visible at the bottom left of the form area. The top navigation bar shows the path: Microsoft Azure > New > Marketplace > Everything > Azure Databricks (preview) > Azure Databricks Service. The user's profile 'nishanth@microsoft.com' is visible in the top right corner.

Microsoft Azure New > Marketplace > Everything > Azure Databricks (preview) > Azure Databricks Service

Search resources, services and docs

nishanth@microsoft.com MICROSOFT (MICROSOFT.ON...)

Azure Databricks Service

New

- App Services
- SQL databases
- SQL data warehouses
- Azure Cosmos DB
- Virtual machines
- Load balancers
- Storage accounts
- Virtual networks
- Azure Active Directory
- Monitor
- Security Center
- Cost Management + Bill...
- Help + support
- Subscriptions
- Data Lake Store
- Data Lake Analytics
- Advisor
- Azure Databricks

More services >

* Workspace name
ntedemodbr12252017 ✓

* Subscription
Azure conversion - External ▼

* Resource group ⓘ
☒ Create new ☐ Use existing
ntedemorg ✓

* Location
West US ▼

☒ Pin to dashboard

Create Automation options

Azure Databricks – workspace deployment

The screenshot displays the Microsoft Azure portal dashboard. The left sidebar contains a navigation menu with options like 'New', 'Dashboard', 'All resources', 'Resource groups', 'App Services', 'SQL databases', 'SQL data warehouses', 'Azure Cosmos DB', 'Virtual machines', 'Load balancers', 'Storage accounts', 'Virtual networks', 'Azure Active Directory', 'Monitor', 'Security Center', 'Cost Management + Billing', 'Help + support', 'Subscriptions', and 'Data Lake Store'. The main content area is titled 'Dashboard' and includes a search bar at the top right. Below the search bar, there are several sections: 'All resources' (listing various Azure services like Azure Databricks, App Service, Azure Cosmos DB, App Service plan, Search service, Storage account, SQL server, and SQL database), 'Quickstart tutorials' (providing links to Windows/Linux Virtual Machines, App Service, Functions, and SQL Database), and 'Service Health' (offering personalized guidance and support). A 'Deploying Azure Databricks (preview)' button is also visible on the right side of the dashboard.

Microsoft Azure

Search resources, services and docs

nishanth@microsoft.c...
MICROSOFT (MICROSOFT ON...

Dashboard

New dashboard Edit dashboard Share Fullscreen Clone Delete

All resources
AZURE CONVERSION - EXTERNAL

Refresh

ntedemodbr12252017	Azure Databricks Serv...
02082017tj	App Service
02082017tj-dcdcb	Azure Cosmos DB acc...
02082017tj-hosting-plan	App Service plan
02082017tjs	Search service
02082017tjso	Storage account
02082017tj-sqlserver	SQL server
AdventureWorks	SQL database

See more...

Service Health

Personalized guidance and support when issues in Azure services affect you. [Learn more](#)

Quickstart tutorials

- Windows Virtual Machines
Provision Windows Server, SQL Server, SharePoint VMs
- Linux Virtual Machines
Provision Ubuntu, Red Hat, CentOS, SUSE, CoreOS VMs
- App Service
Create Web Apps using .NET, Java, Node.js, Python, PHP
- Functions
Process events with a serverless code architecture
- SQL Database
Managed relational SQL Database as a Service

Marketplace

Deploying Azure Databricks (preview)

Azure Databricks – launching the workspace

The screenshot displays the Microsoft Azure portal interface for the resource group 'ntedemodbr12252017'. The left sidebar contains a navigation menu with various services, including App Services, SQL databases, SQL data warehouses, Azure Cosmos DB, Virtual machines, Load balancers, Storage accounts, Virtual networks, Azure Active Directory, Monitor, Security Center, Cost Management + Billing, Help + support, Subscriptions, Data Lake Store, Data Lake Analytics, Advisor, and Azure Databricks. The main content area shows the 'Overview' page for the 'ntedemodbr12252017' resource group. It includes a search bar, a list of services (Overview, Activity log, Access control (IAM), Tags), and a 'Launch Workspace' button. Below the button, there are links to 'Documentations', 'Getting Started', 'Import Data from File', 'Import Data from Azure Storage', 'Notebook', and 'Admin Guide'. The top of the page shows the user's profile 'nishanth@microsoft.c...' and the search bar.

Microsoft Azure ntedemodbr12252017

Search resources, services and docs

nishanth@microsoft.c... MICROSOFT (MICROSOFT.ONL...)

ntedemodbr12252017
Azure Databricks Service - PREVIEW

Search (Ctrl+V)

Overview

Activity log

Access control (IAM)

Tags

SETTINGS

Locks

Automation script

SUPPORT + TROUBLESHOOTING

New support request

Delete

Resource group (change)
ntedemo

Subscription (change)
Azure conversion - External

Subscription ID
15c5cb6e-191a-40ea-9f69-08207a17fe97

Managed Resource Group
databricks-rg-ntedemodbr12252017-6kt7r3v4ehflu

URL
<https://westeurope.azuredatabricks.net>

Guides
Documentations

Launch Workspace

Documentations

Getting Started

Import Data from File

Import Data from Azure Storage

Notebook

Admin Guide

Azure Databricks – workspace home page

The screenshot displays the Azure Databricks workspace home page. At the top, a dark header bar contains the 'Microsoft Azure' logo on the left, and 'PORTAL' and the user email 'nishanth@microsoft.com' on the right. A vertical sidebar on the left side of the page lists navigation options: Azure Databricks, Home, Workspace, Recent, Data, Clusters, Jobs, and Search. The main content area features the 'Azure Databricks' logo at the top center. Below the logo, the 'Featured Notebooks' section displays three notebook thumbnails: 'Introduction to Apache Spark on Databricks', 'Databricks for Data Scientists', and 'Introduction to Structured Streaming'. The bottom section is divided into three columns: 'New' (containing links for Notebook, Job, Cluster, Table, and Library), 'Documentation' (containing links for Databricks Guide, Python, R, Scala, SQL, and Importing Data), and 'Open Recent' (with a message about recent files and a link to the welcome guide).

Microsoft Azure

PORTAL nishanth@microsoft.com

Azure Databricks

Home Workspace Recent Data Clusters Jobs Search

Featured Notebooks

- Introduction to Apache Spark on Databricks
- Databricks for Data Scientists
- Introduction to Structured Streaming

New

- Notebook
- Job
- Cluster
- Table
- Library

Documentation

- Databricks Guide
- Python, R, Scala, SQL
- Importing Data

Open Recent

Recent files appear here as you work. Get started with the [welcome guide](#).