Microsoft

Microsoft Partner Project Ready

Implement with Impact

Modern Data Platform with Azure Databricks

<Speaker name or subtitle>

<Date>

Day 1 of 3

Technofocus

Course Plan and Learning Objectives

Day 1

Module 1 - Introduction to Azure Databricks

- Azure Databricks: A Data Intelligent Platform
- Why Azure Databricks
- Decision guide: Azure Databricks vs. Microsoft Fabric

Module 2 - Migration to Azure Databricks

- Microsoft Cloud Adoption Framework for Azure
- Migration strategies
- Data landing zones
- Migration scenarios

Interactive Simulated Lab Experience

- End-to-End Streaming Pipeline with Lakeflow Declarative Pipelines in Azure Databricks

Day 2

Module 3 - Integration with Azure

- Seamless integration with Microsoft Azure services
- Connect to Azure Data Lake Storage (ADLS) Gen2 and Blob Storage
- Leverage Azure Databricks for Azure Cosmos DB Operations
- Secret management with Azure Key Vault
- Connect Azure Databricks to Azure Event Hubs

Module 4 - Integration with Microsoft Fabric and Power BI

- Data Intelligence with Azure Databricks and Microsoft Fabric
- Connect Power BI to Azure Databricks
- Integration with Azure Data Factory
- Mirroring Azure Databricks Unity Catalog

Interactive Simulated Lab Experience

- Setup and use Unity Catalog for Data Management in Azure Databricks
- Real-Time Streaming with Azure Databricks and Azure Event Hubs

Day 3

Module 5 - Integration with Azure AI Foundry

- Azure Databricks connector in Azure AI Foundry
- Mosaic AI and machine learning on Azure Databricks
- Query Generative AI model serving endpoints
- Databricks Assistant, AI/BI Genie and AI Functions on Azure Databricks
- Chat with LLMs and prototype GenAI apps using AI Playground
- Build and optimize agents on your data with Agent Bricks

Module 6 - Security and Governance

- Integrate Azure Databricks with Microsoft Purview
- Integration of Azure Databricks Unity Catalog with Microsoft Purview

Module 7 - Well-architected for Azure Databricks

- Lakehouse implementation: Principles and best practices
- Azure Databricks well-architected framework

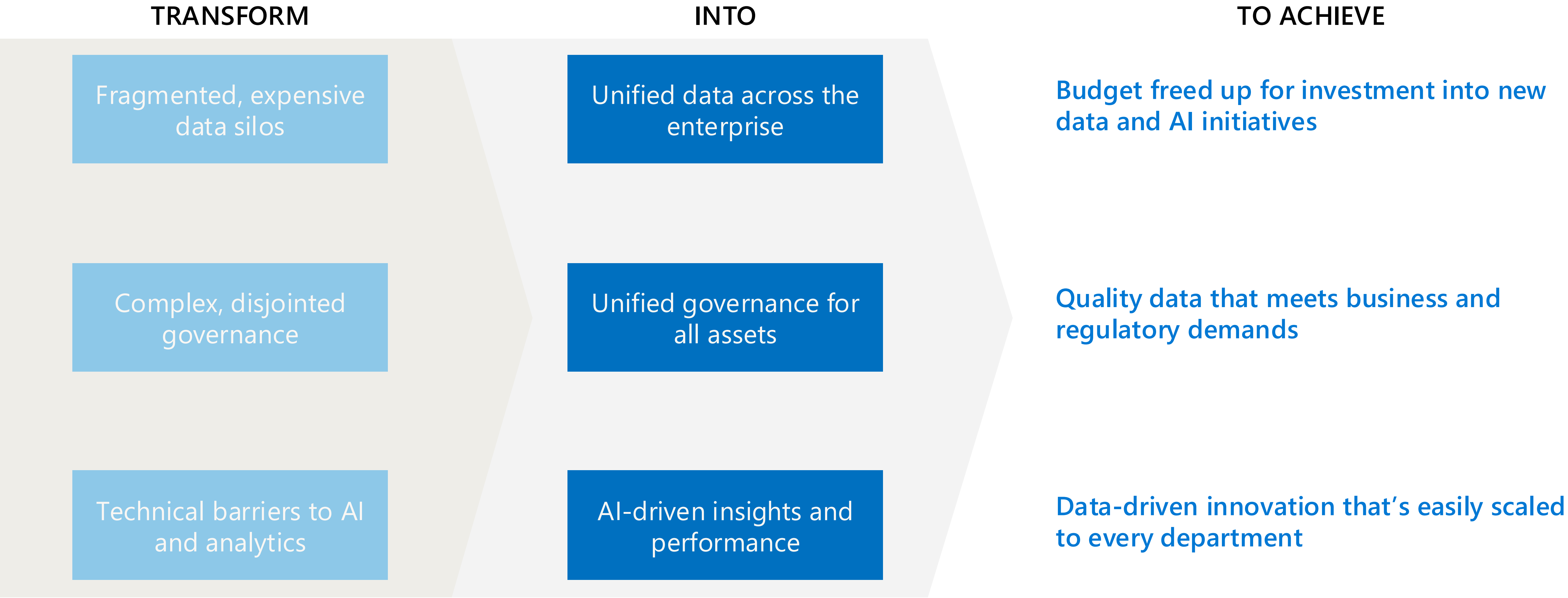
Interactive Simulated Lab Experience

- Responsible AI with Large Language Models using Azure Databricks and Azure OpenAI
- Connect to and manage Azure Databricks in Microsoft Purview

01
Introduction to Azure Databricks

Azure Databricks: A Data Intelligence Platform

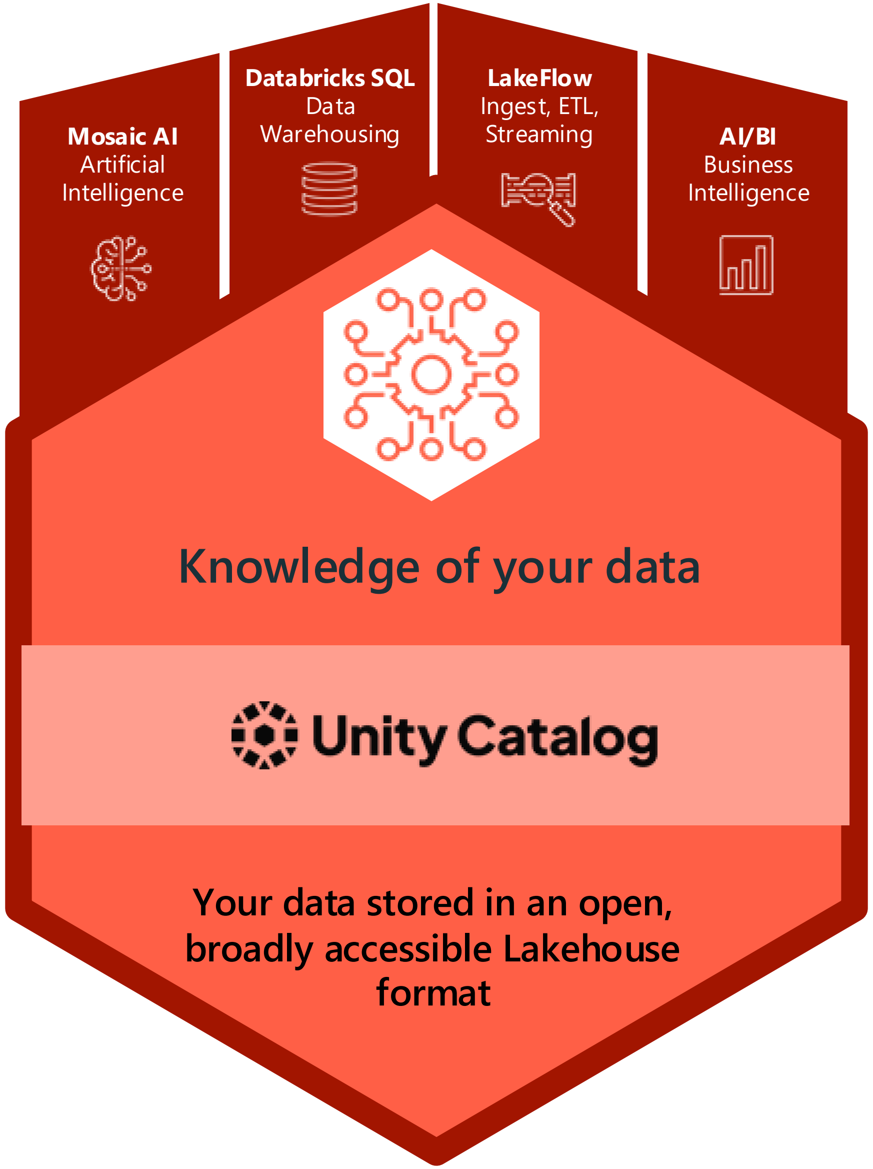
Data Intelligence accelerates your data and AI success



TRANSFORM	INTO	TO ACHIEVE
Fragmented, expensive data silos	Unified data across the enterprise	Budget freed up for investment into new data and AI initiatives
Complex, disjointed governance	Unified governance for all assets	Quality data that meets business and regulatory demands
Technical barriers to AI and analytics	AI-driven insights and performance	Data-driven innovation that’s easily scaled to every department

Azure Databricks

Data Intelligence Platform



Mosaic AI Artificial Intelligence

Databricks SQL Data Warehousing

LakeFlow Ingest, ETL, Streaming

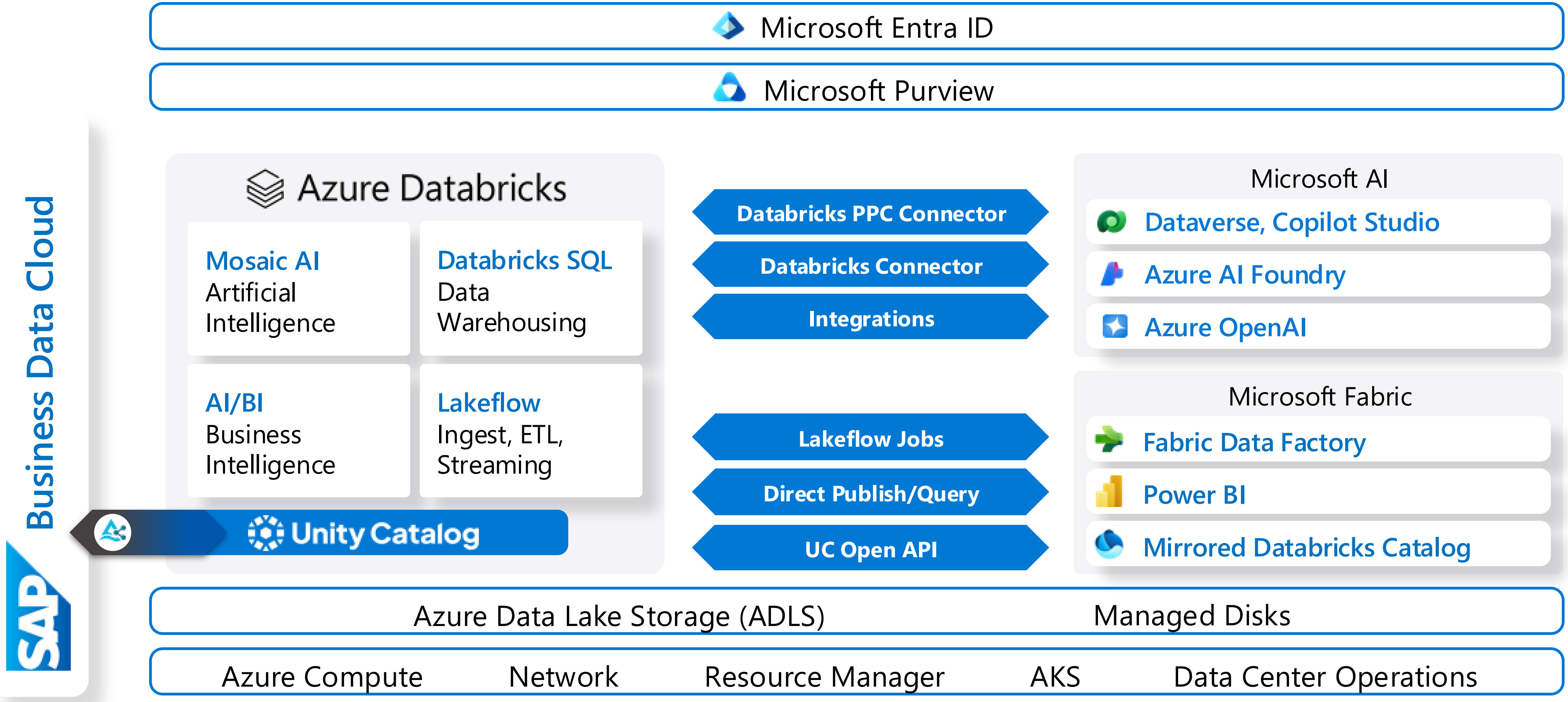
AI/BI Business Intelligence

Knowledge of your data

Unity Catalog

Your data stored in an open, broadly accessible Lakehouse format

Integrated with Azure Data & AI



Microsoft Entra ID

Microsoft Purview

Business Data Cloud

SAP

Azure Databricks

Mosaic AI
Artificial Intelligence

Databricks SQL
Data Warehousing

AI/BI
Business Intelligence

Lakeflow
Ingest, ETL, Streaming

Unity Catalog

Databricks PPC Connector

Databricks Connector

Integrations

Lakeflow Jobs

Direct Publish/Query

UC Open API

Microsoft AI

- Dataverse, Copilot Studio
- Azure AI Foundry
- Azure OpenAI

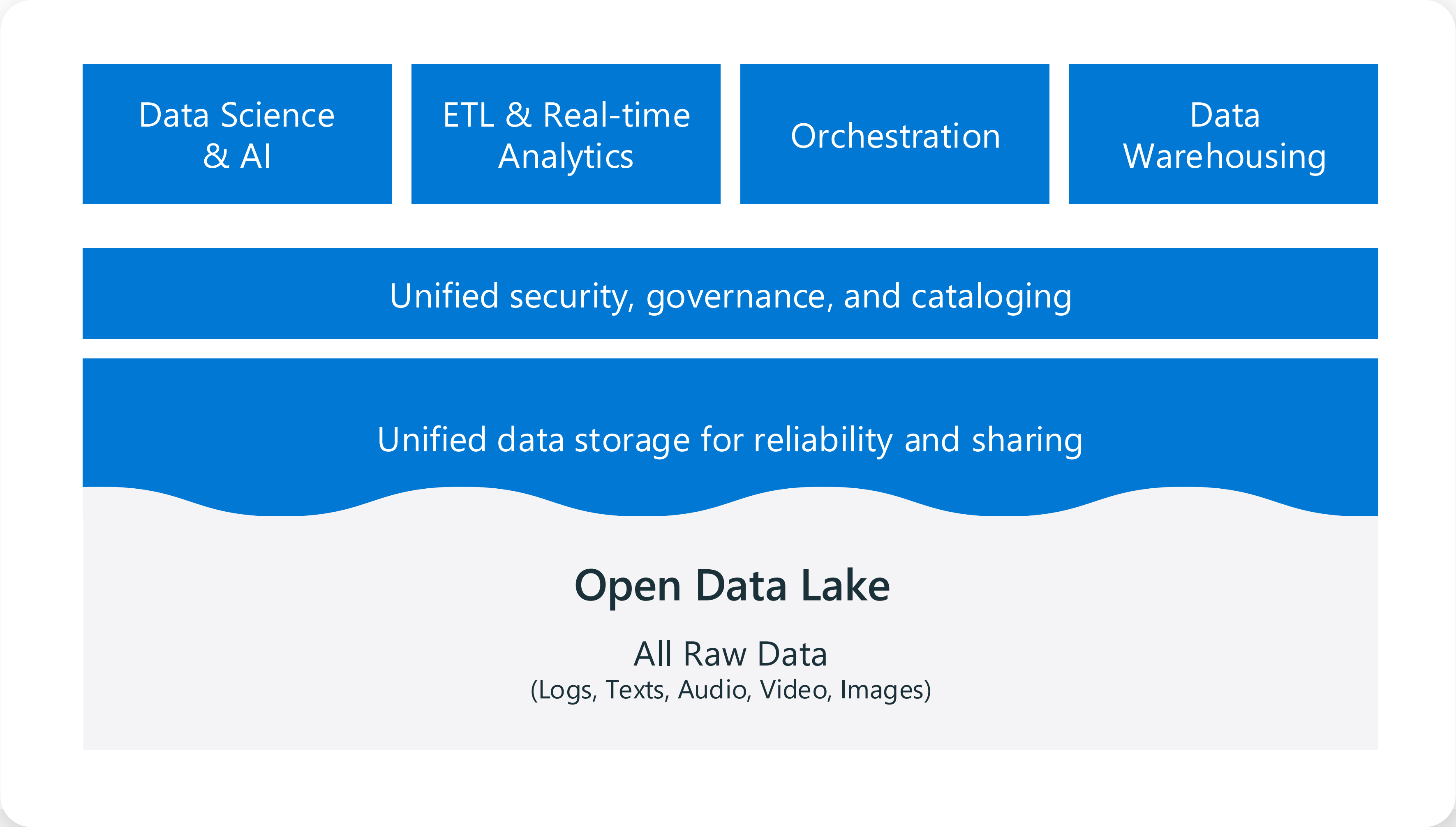
Microsoft Fabric

- Fabric Data Factory
- Power BI
- Mirrored Databricks Catalog

Azure Data Lake Storage (ADLS) Managed Disks

Azure Compute Network Resource Manager AKS Data Center Operations

The Data Lakehouse simplified and unified the architecture



Data Science & AI

ETL & Real-time Analytics

Orchestration

Data Warehousing

Unified security, governance, and cataloging

Unified data storage for reliability and sharing

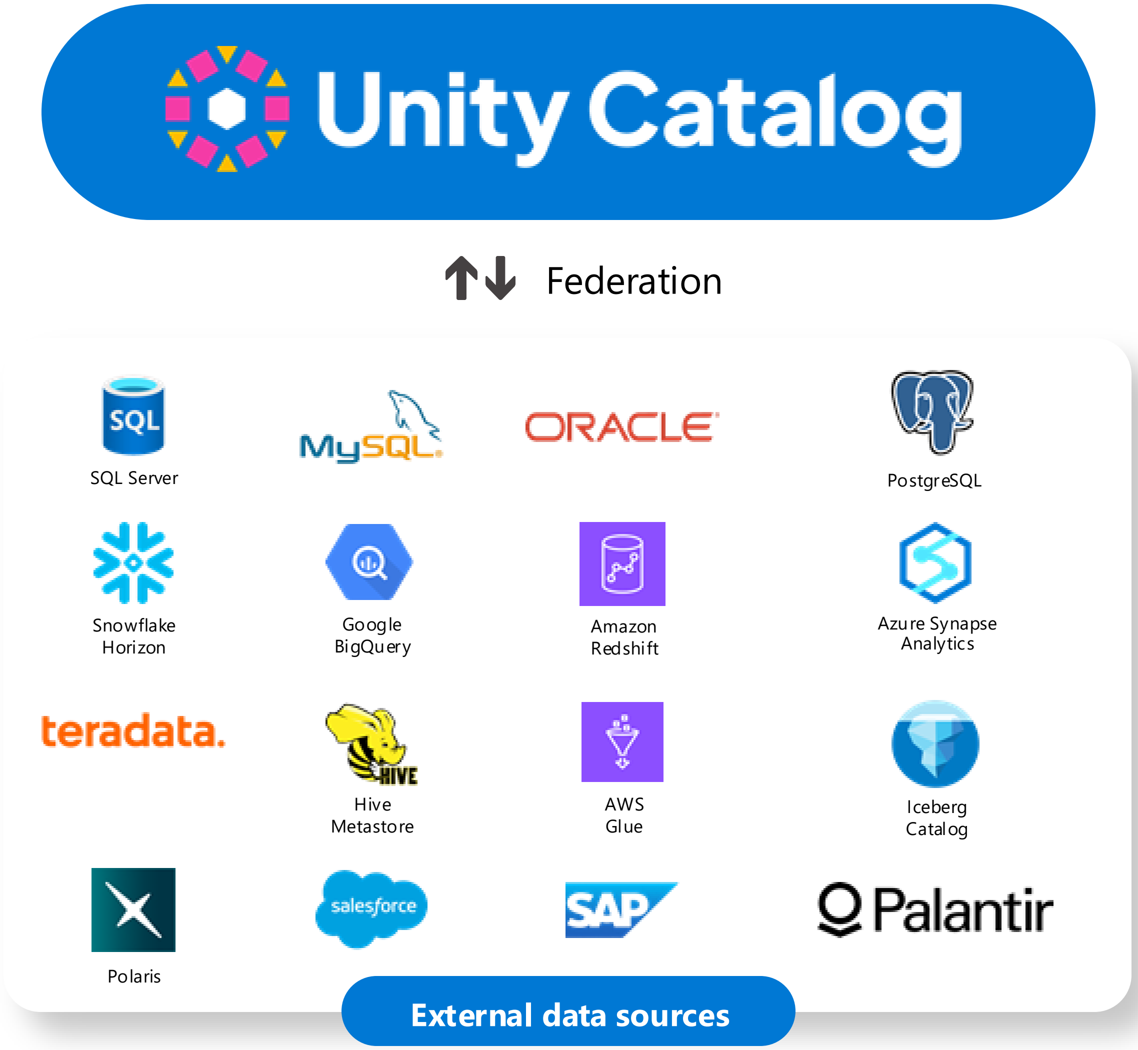
Open Data Lake

All Raw Data

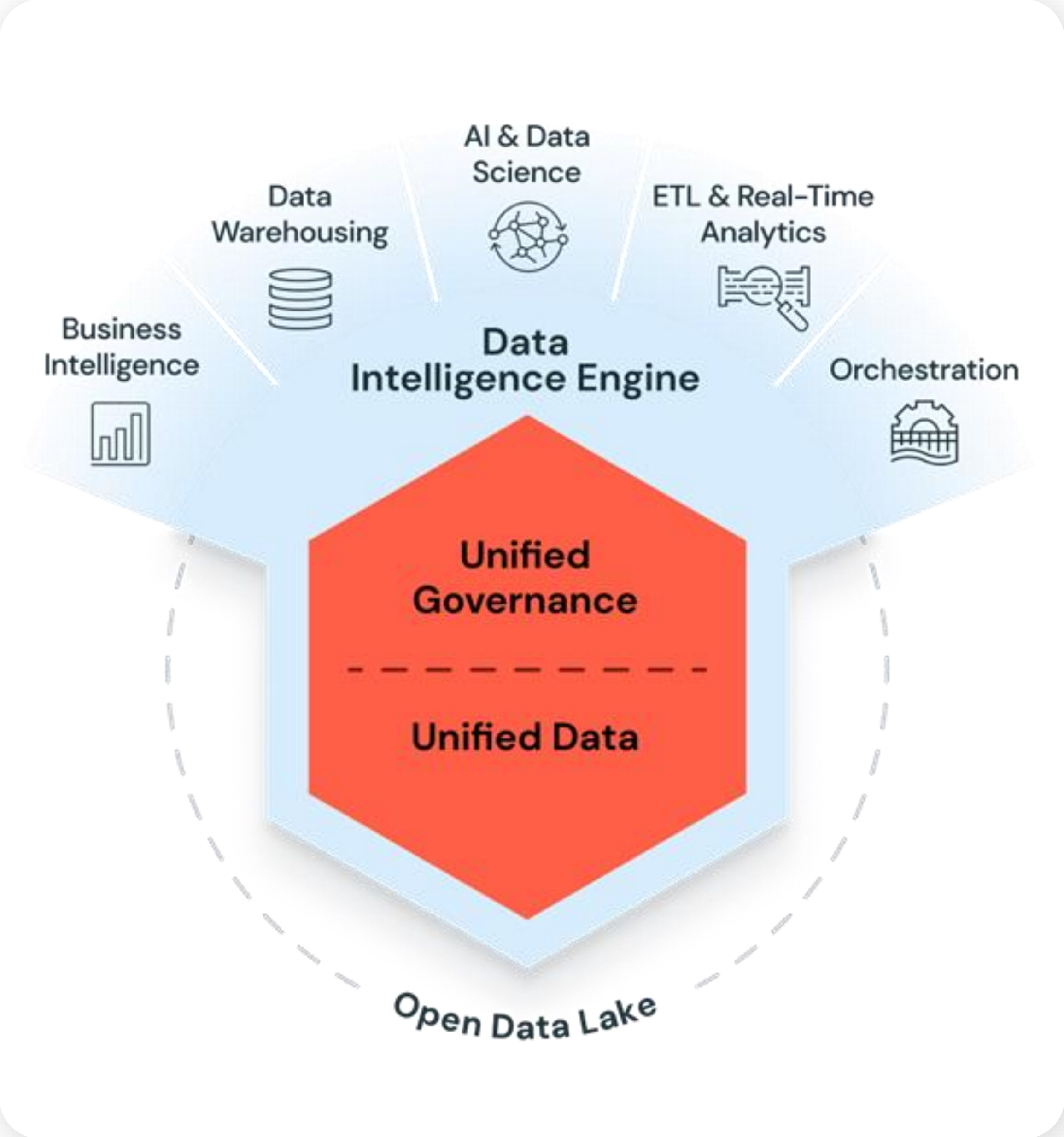
(Logs, Texts, Audio, Video, Images)

Make all of your data available to the business

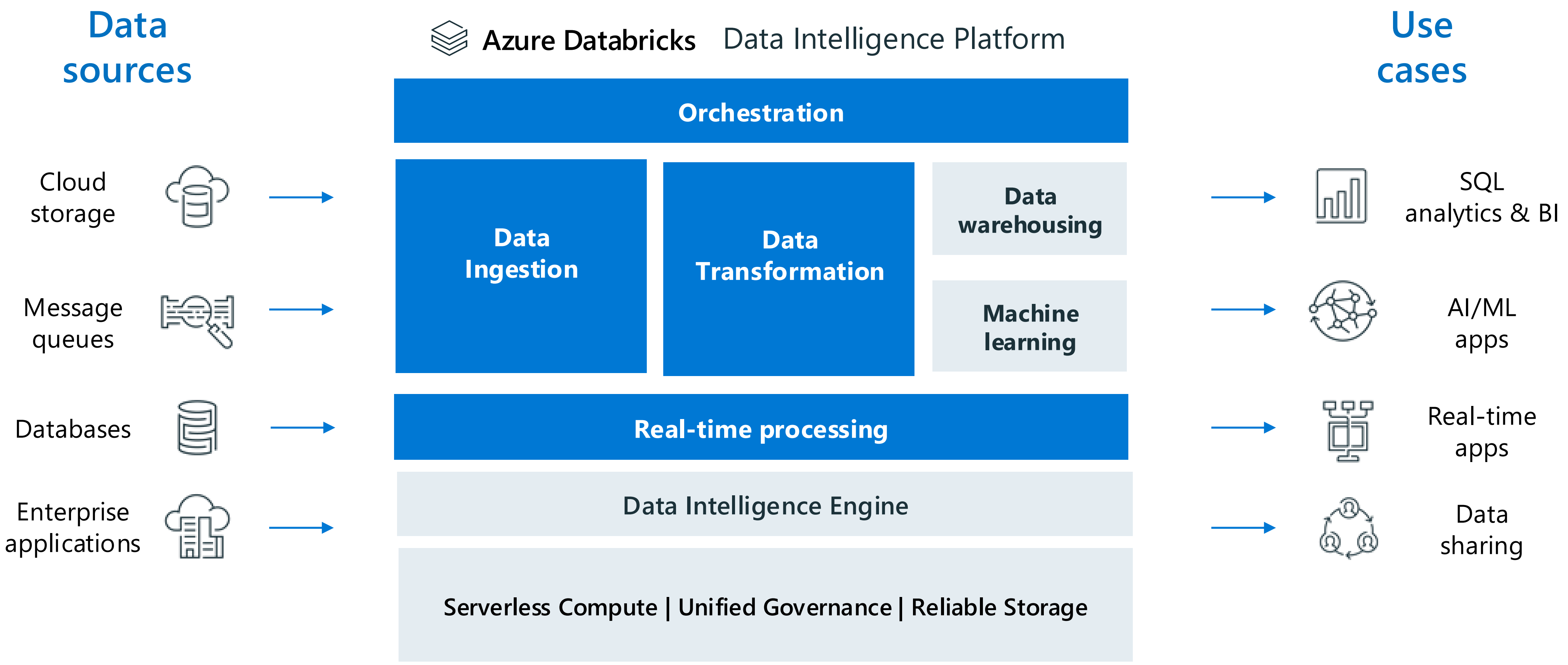
- Unify the data from every business system to answer bigger questions
- Support teams regardless of data format: Delta, Iceberg, Parquet, etc.
- Make your data easily discoverable



The Azure Databricks Data Intelligence Platform provides the foundation for Data Engineering in the age of AI

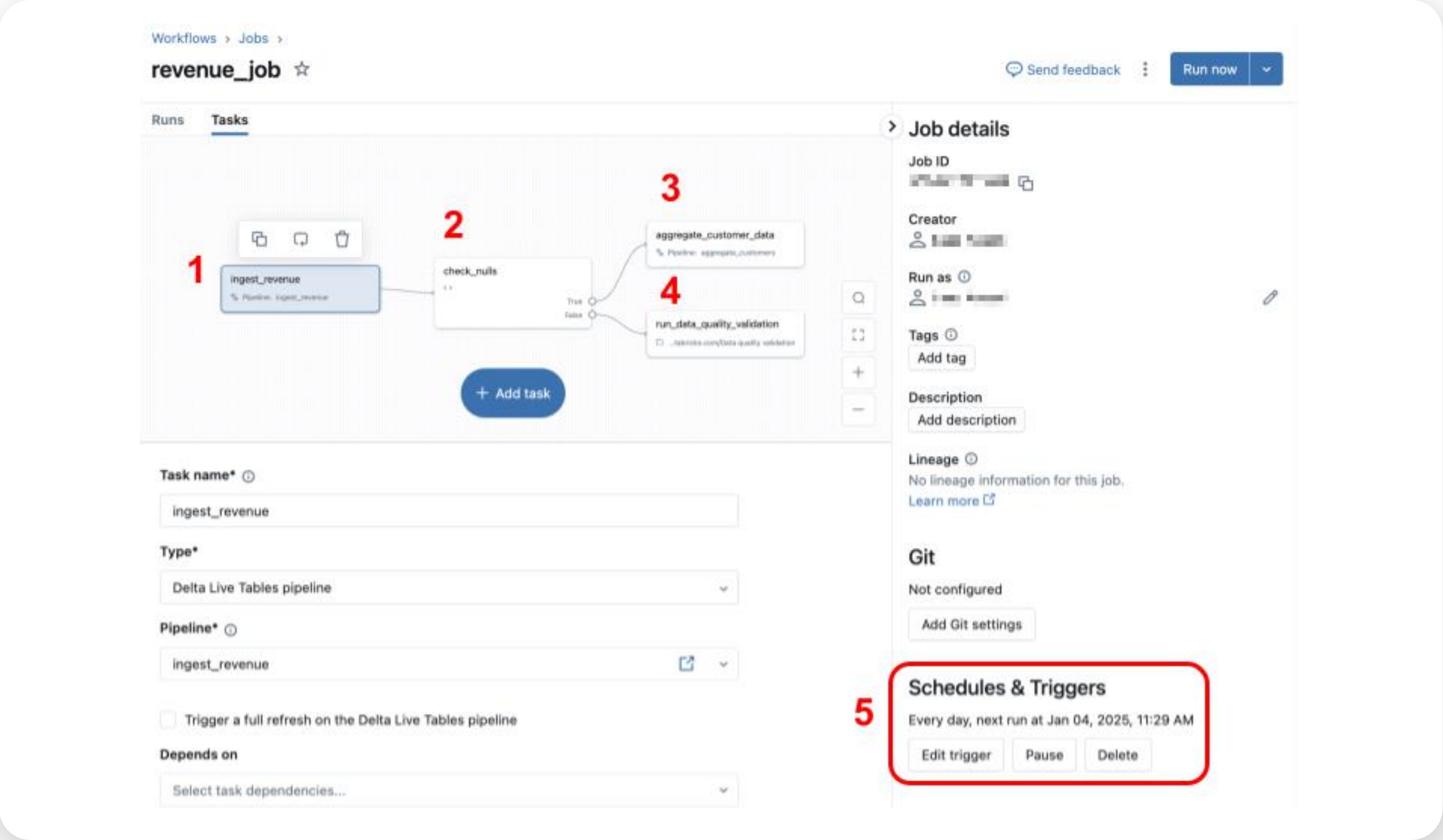


Data Engineering on Azure Databricks



Lakeflow Jobs

- Coordinate and run multiple tasks as part of a larger workflow
- You can optimize and schedule the execution of frequent, repeatable tasks and manage complex workflows



Task name* ingest_revenue

Type* Delta Live Tables pipeline

Pipeline* ingest_revenue

□ Trigger a full refresh on the Delta Live Tables pipeline

Depends on: Select task dependencies...

Job details

Job ID [redacted]

Creator [redacted]

Run as [redacted]

Tags: Add tag

Description: Add description

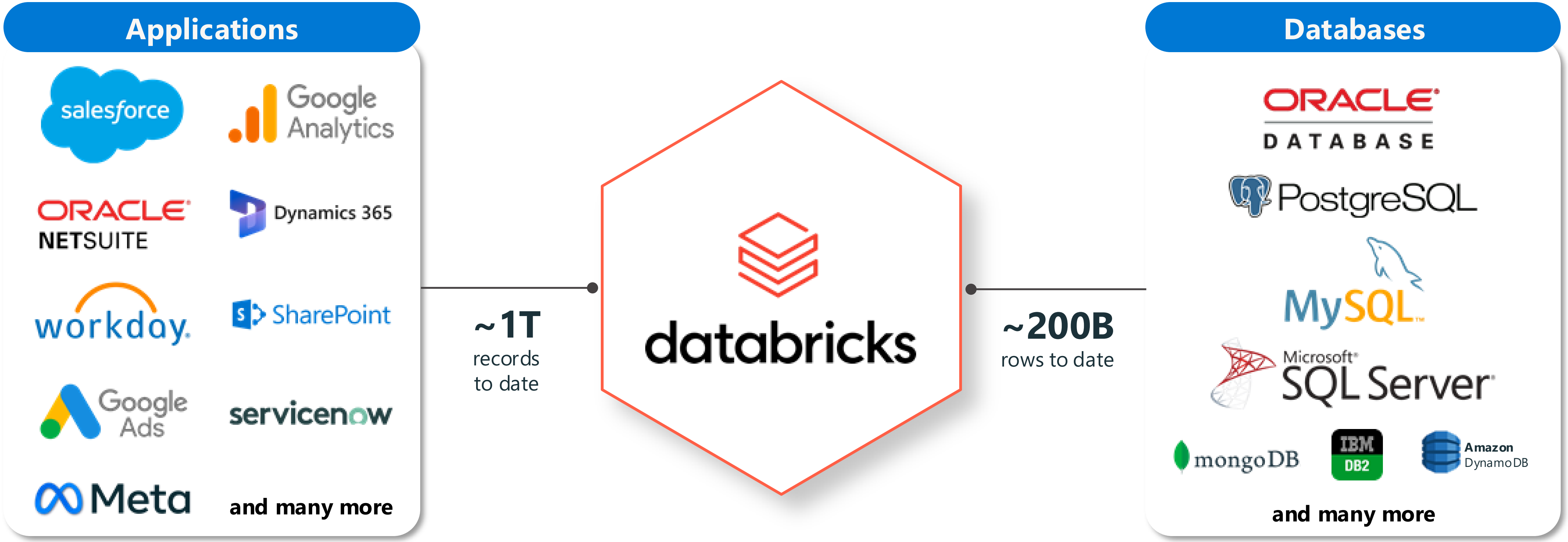
Lineage: No lineage information for this job. Learn more

Git: Not configured. Add Git settings

Schedules & Triggers: Every day, next run at Jan 04, 2025, 11:29 AM

Lakeflow Connect

Efficient native ingestion connectors



Applications

and many more

~1T records to date

~200B rows to date

Databases

and many more

No code, low maintenance design

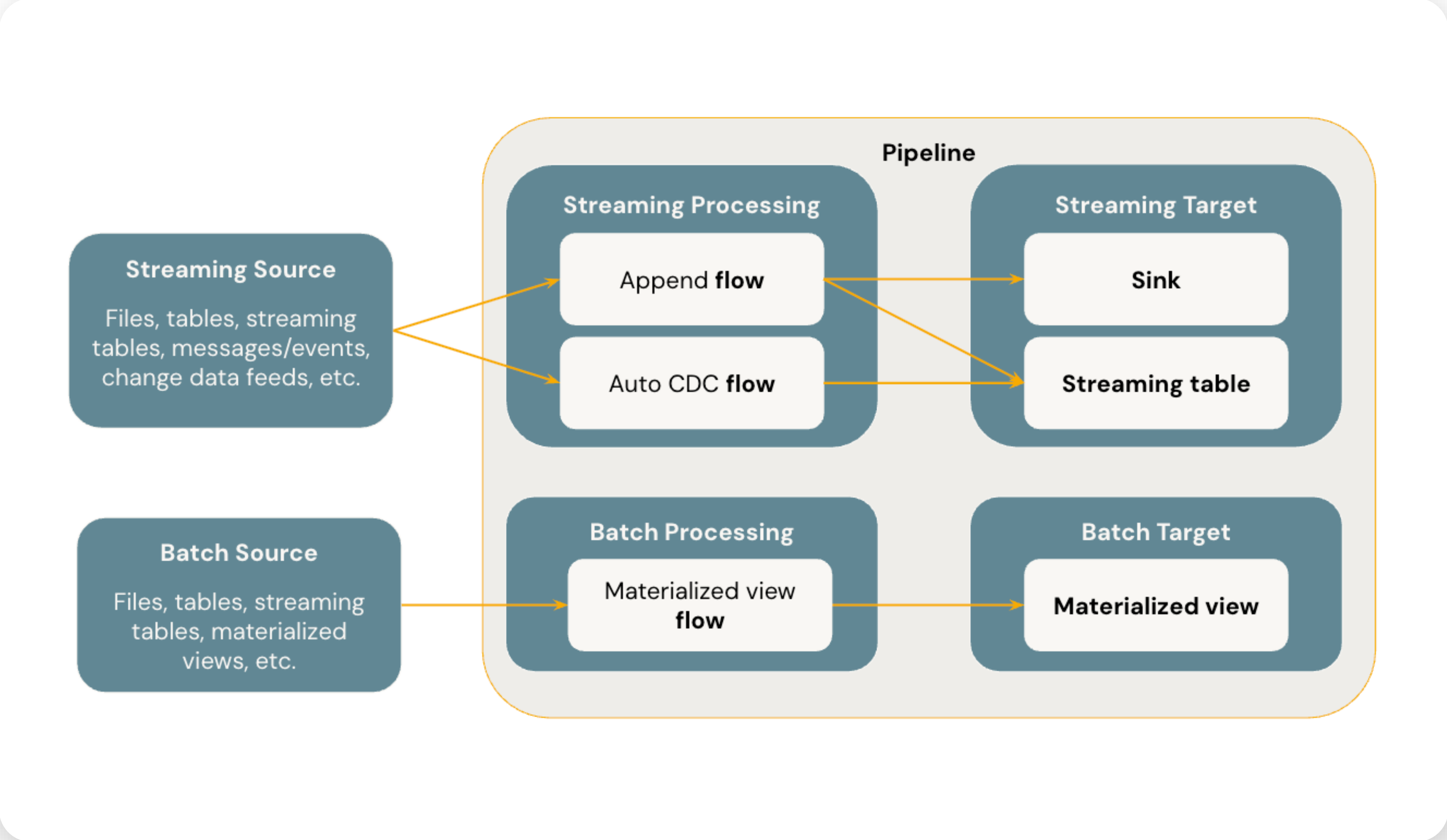
Safe and healthy pipelines

High scale, high performance

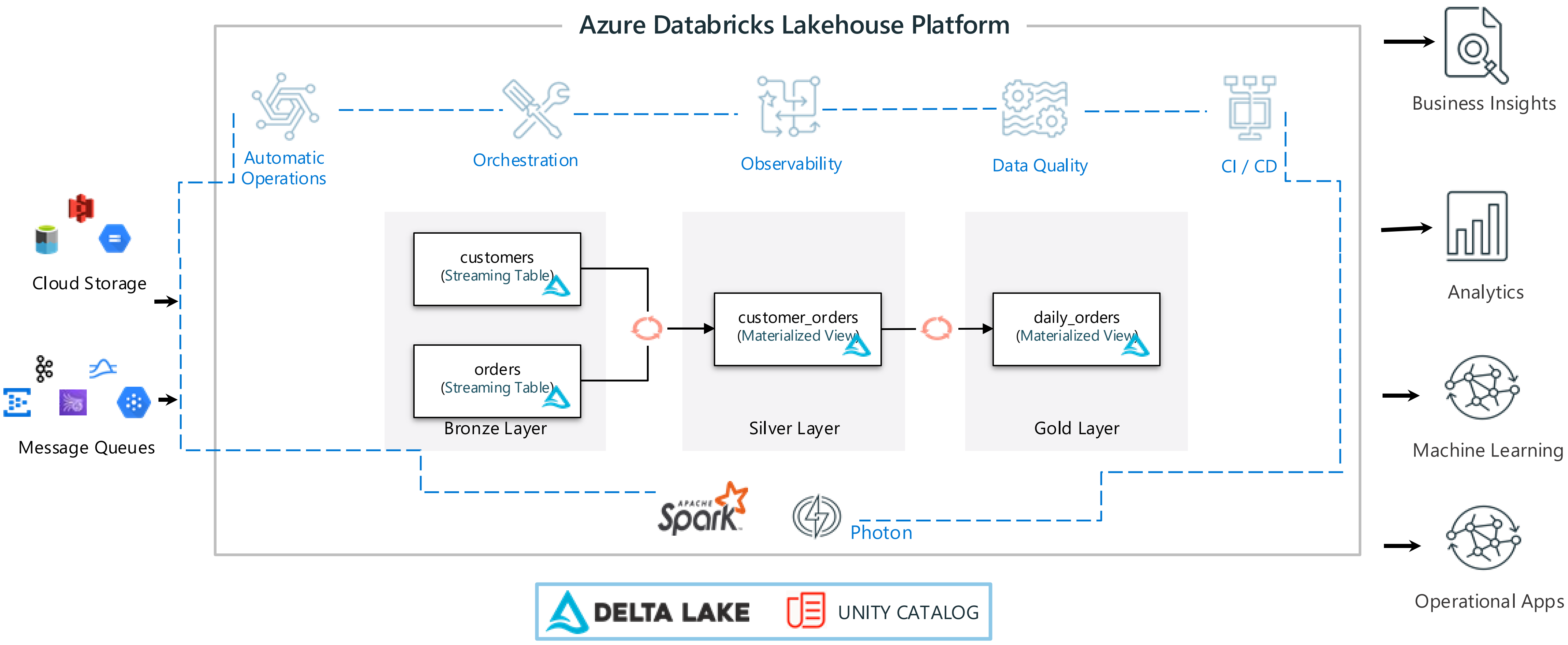
Lakeflow Declarative Pipelines

- A declarative framework for developing and running batch and streaming data pipelines in SQL and Python

- Runs on the performance-optimized Databricks Runtime (DBR)

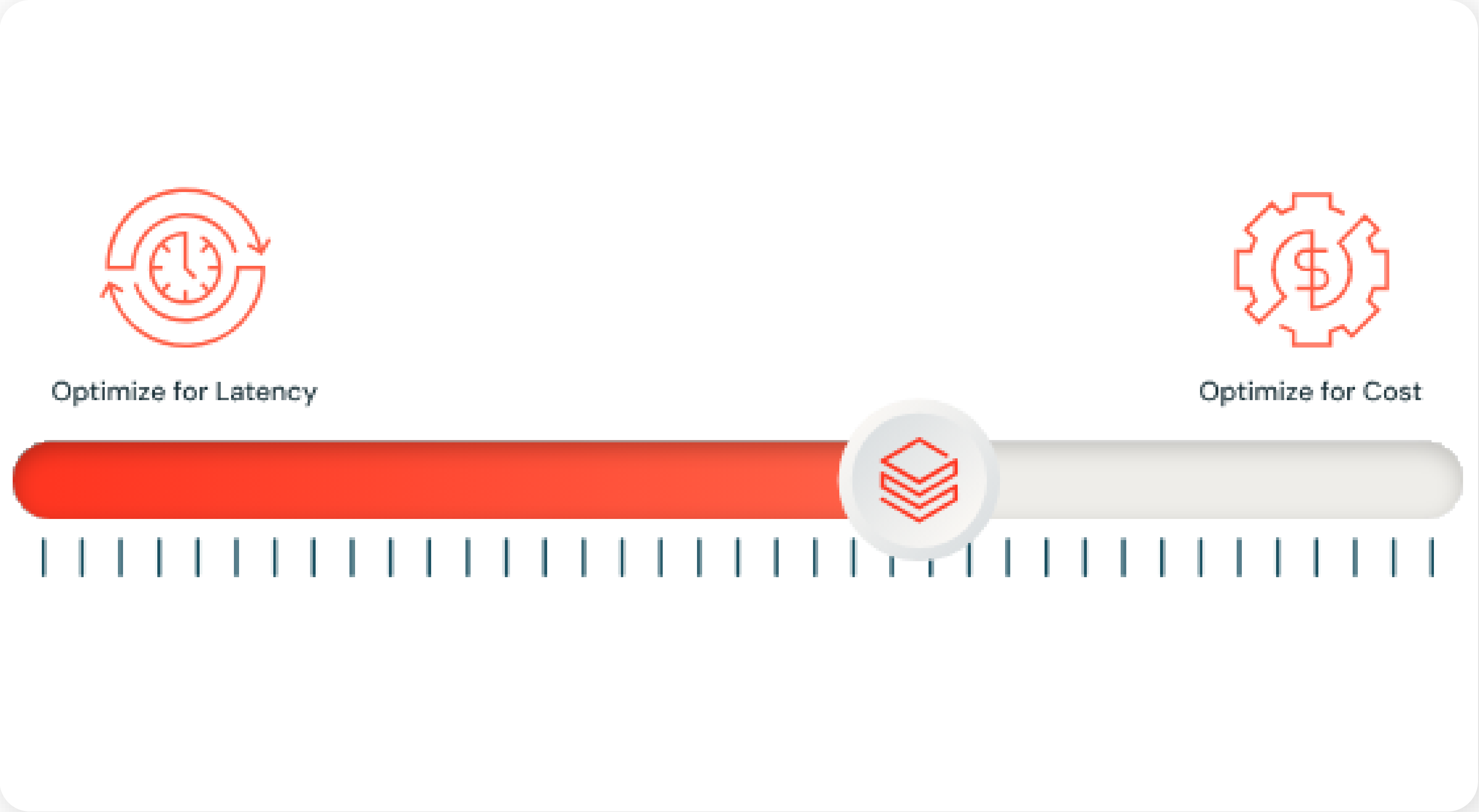


Build ETL Pipelines with Lakeflow Declarative Pipelines



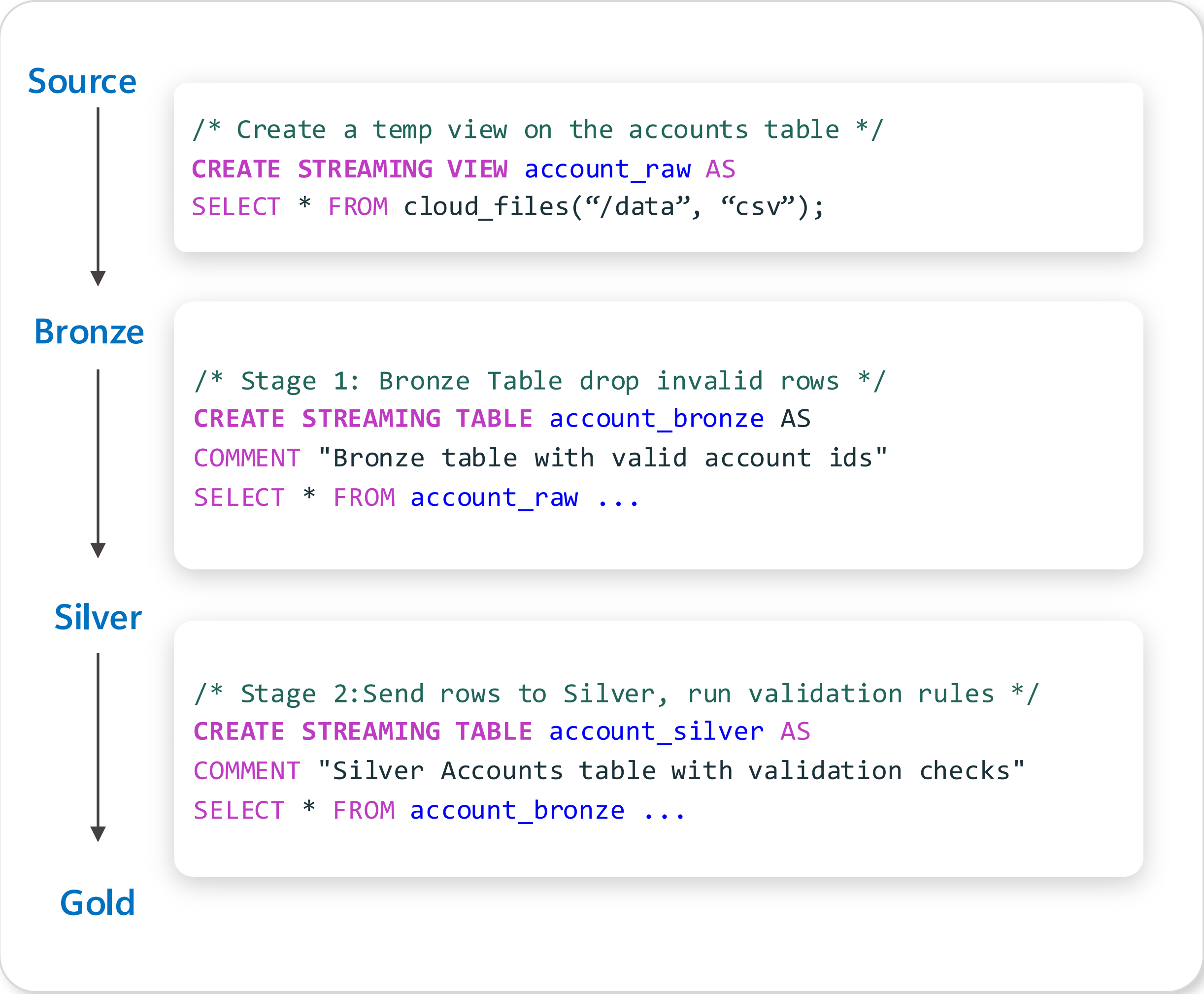
Apache Spark Structured Streaming

- Apache Spark™ Structured Streaming powers streaming data pipelines on Azure Databricks
- It provides a single, unified API for batch and stream processing



See https://www.apache.org/foundation/marks/guide

Declarative SQL & Python APIs



```
/* Create a temp view on the accounts table */
CREATE STREAMING VIEW account_raw AS
SELECT * FROM cloud_files(“/data”, “csv”);
```

```
/* Stage 1: Bronze Table drop invalid rows */
CREATE STREAMING TABLE account_bronze AS
COMMENT "Bronze table with valid account ids"
SELECT * FROM account_raw ...
```

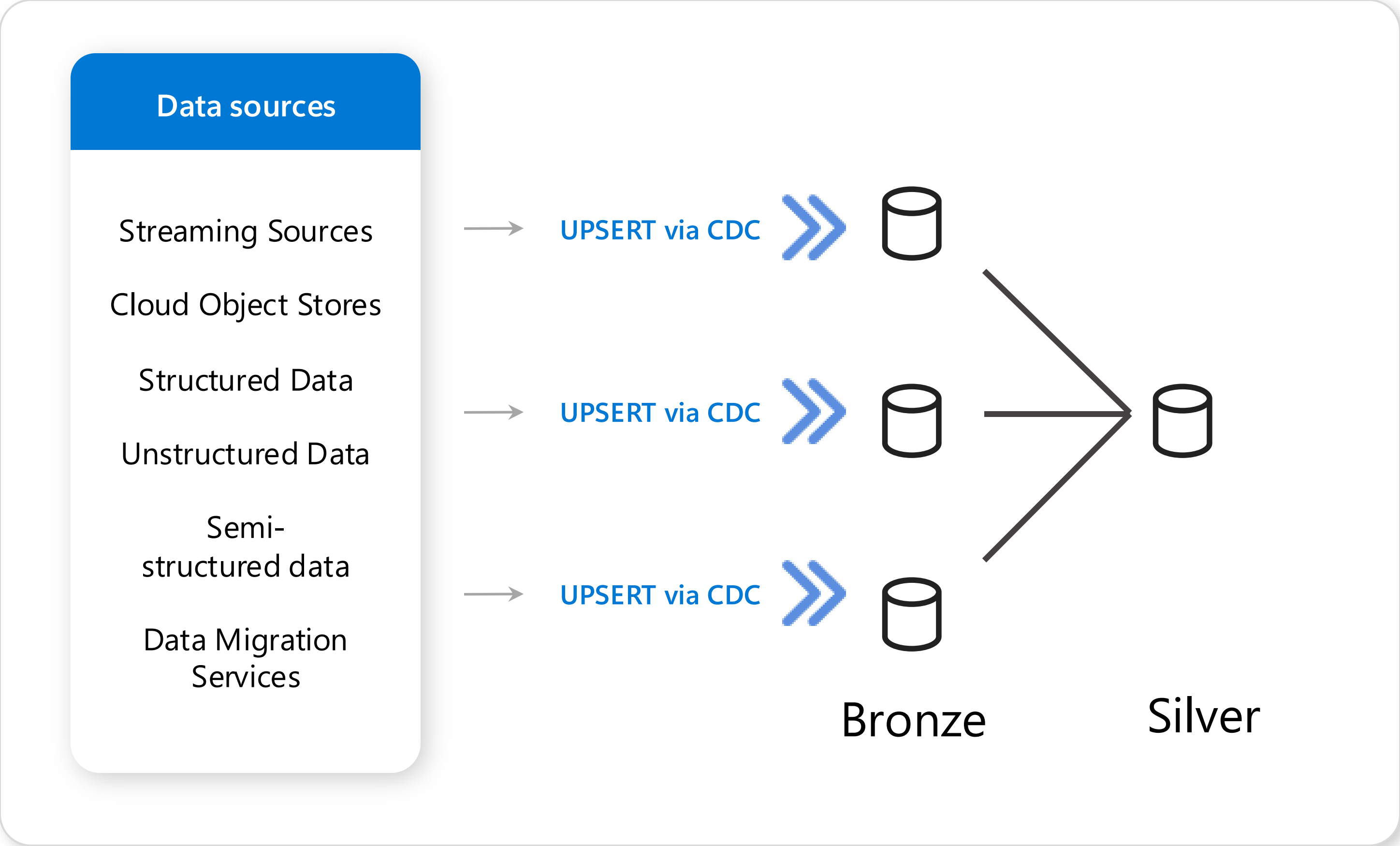
```
/* Stage 2:Send rows to Silver, run validation rules */
CREATE STREAMING TABLE account_silver AS
COMMENT "Silver Accounts table with validation checks"
SELECT * FROM account_bronze ...
```

Use intent-driven declarative development to abstract away the **“how”** and define **“what”** to solve

Automatically generate **lineage** based on table dependencies across the data pipeline

Automatically checks for errors, missing dependencies and syntax errors

Change data capture (CDC)



Stream change records (inserts, updates, deletes) from any data source supported by DBR, cloud storage, or DBFS

Simple, declarative "APPLY CHANGES INTO" API for SQL or Python

Handles out-of-order events

Schema evolution

SCD2 support

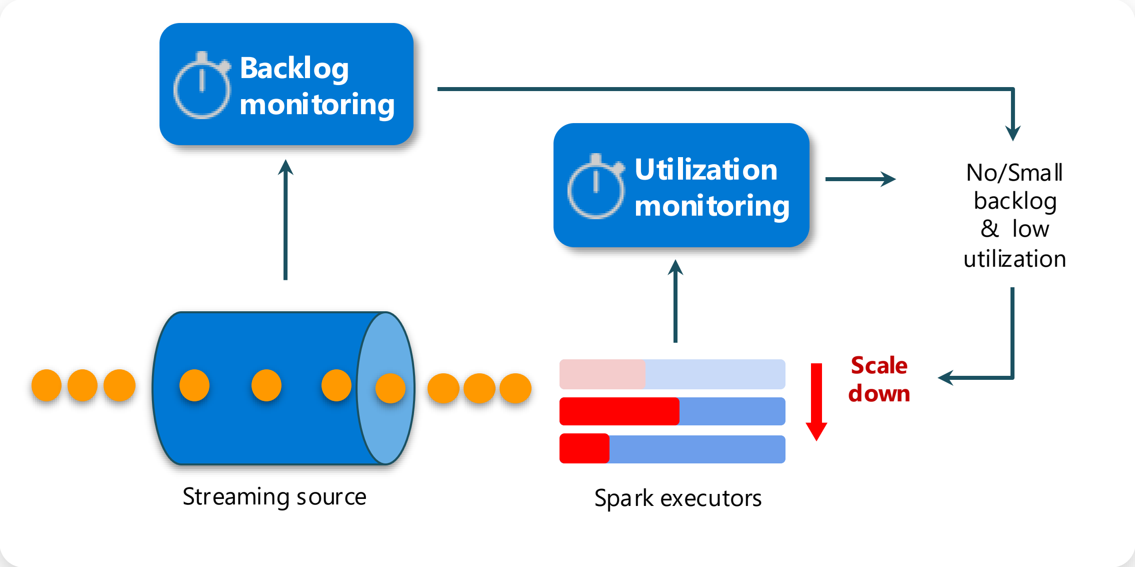
Enhanced Autoscaling

Save infrastructure costs while maintaining end-to-end latency SLAs for streaming workloads

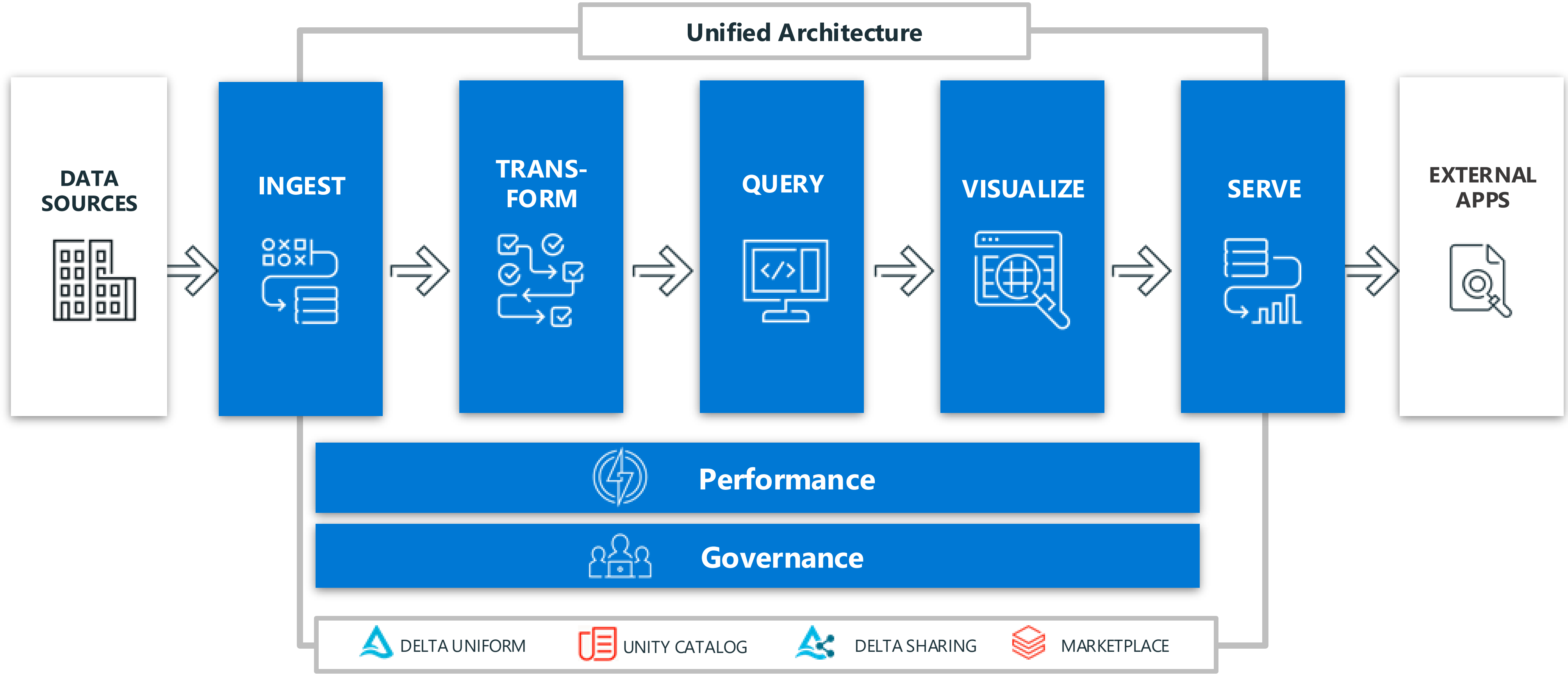
- Built to handle streaming workloads which are spiky and unpredictable
- Shuts down nodes when utilization is low while guaranteeing task execution
- Only scales up to needed # of nodes

Problem

Optimize infrastructure spend when making scaling decisions for streaming workloads



Data Warehousing with Databricks SQL



Unified Architecture

DATA SOURCES

INGEST

TRANS-FORM

QUERY

VISUALIZE

SERVE

EXTERNAL APPS

Performance

Governance

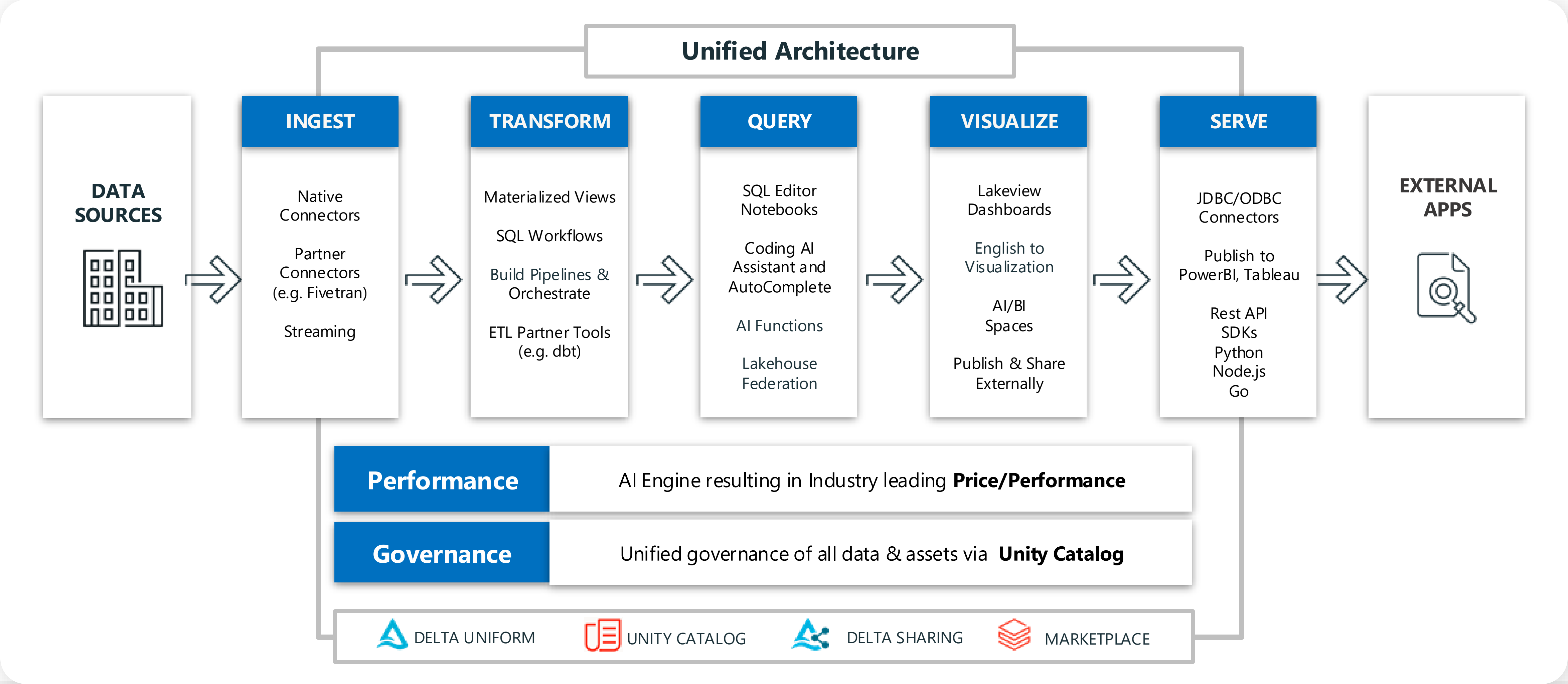
DELTA UNIFORM

UNITY CATALOG

DELTA SHARING

MARKETPLACE

Data Warehousing with Databricks SQL



Unified Architecture

DATA SOURCES

INGEST

Native Connectors

Partner Connectors (e.g. Fivetran)

Streaming

TRANSFORM

Materialized Views

SQL Workflows

Build Pipelines & Orchestrate

ETL Partner Tools (e.g. dbt)

QUERY

SQL Editor Notebooks

Coding AI Assistant and AutoComplete

AI Functions

Lakehouse Federation

VISUALIZE

Lakeview Dashboards

English to Visualization

AI/BI Spaces

Publish & Share Externally

SERVE

JDBC/ODBC Connectors

Publish to PowerBI, Tableau

Rest API
SDKs
Python
Node.js
Go

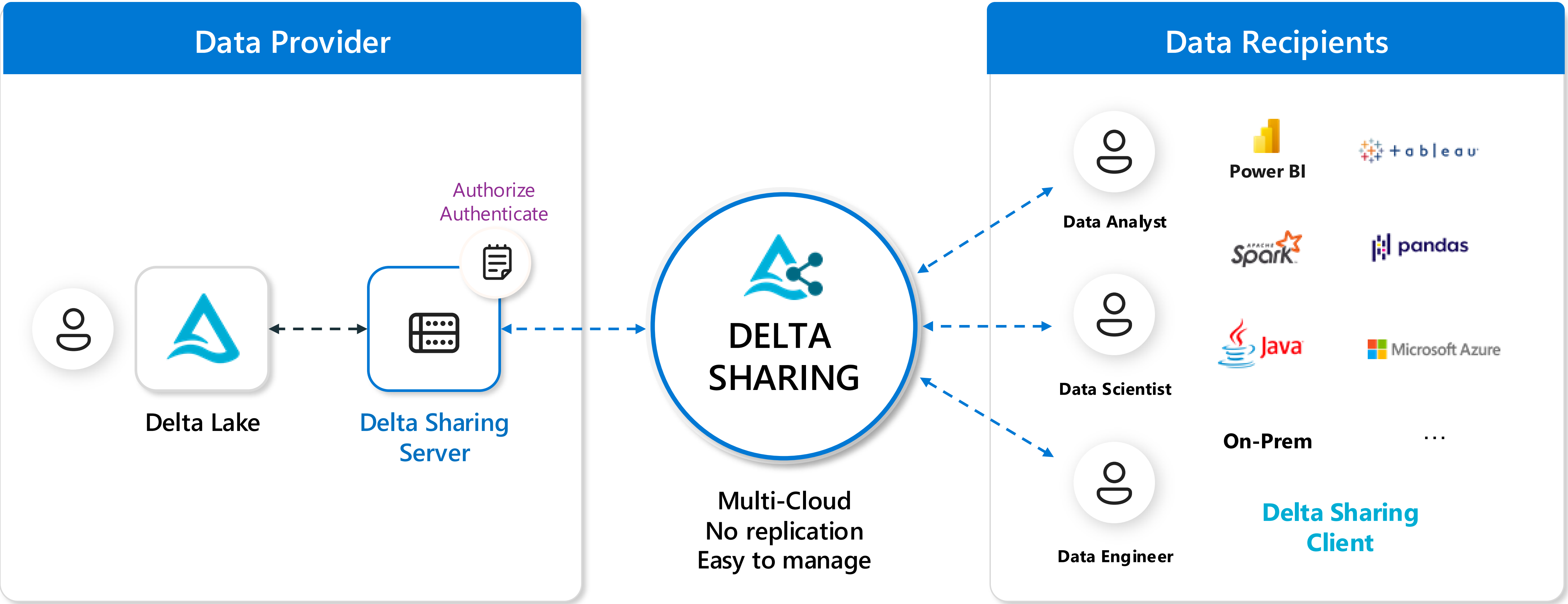
EXTERNAL APPS

Performance	AI Engine resulting in Industry leading **Price/Performance**
Governance	Unified governance of all data & assets via **Unity Catalog**

DELTA UNIFORM UNITY CATALOG DELTA SHARING MARKETPLACE

A simple, open and easy approach to data sharing

Reduce data sharing and collaboration from days to real-time



Data Provider

Authorize
Authenticate

Delta Lake

Delta Sharing
Server

DELTA
SHARING

Multi-Cloud
No replication
Easy to manage

Data Recipients

Data Analyst

Data Scientist

Data Engineer

Power BI

tableau

Spark

pandas

Java

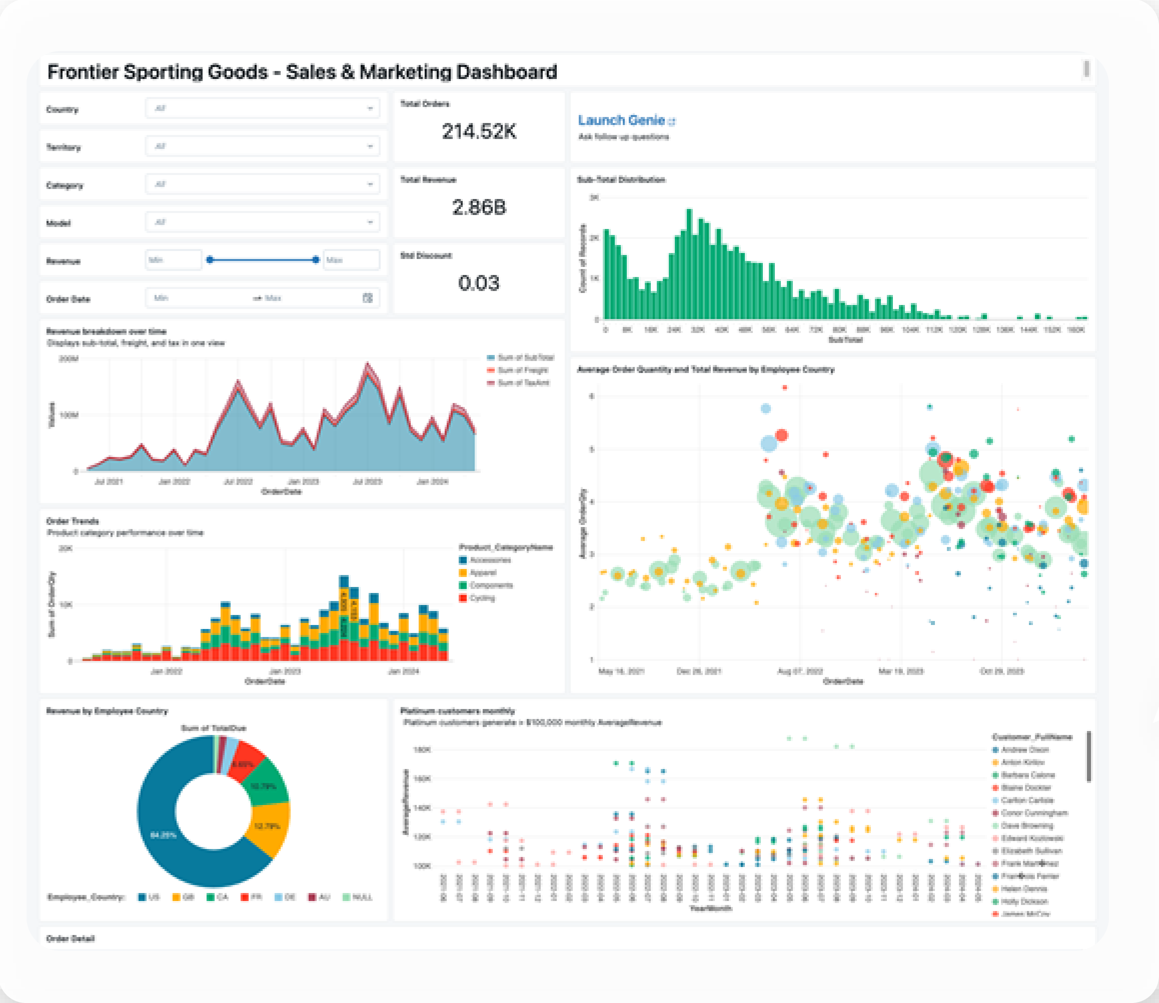
Microsoft Azure

On-Prem

...

Delta Sharing
Client

Data Intelligence for analytics and BI



Frontier Sporting Goods - Sales & Marketing Dashboard

Country: All

Territory: All

Category: All

Model: All

Revenue: Min – Max

Order Date: Min → Max

Total Orders: 214.52K

Total Revenue: 2.86B

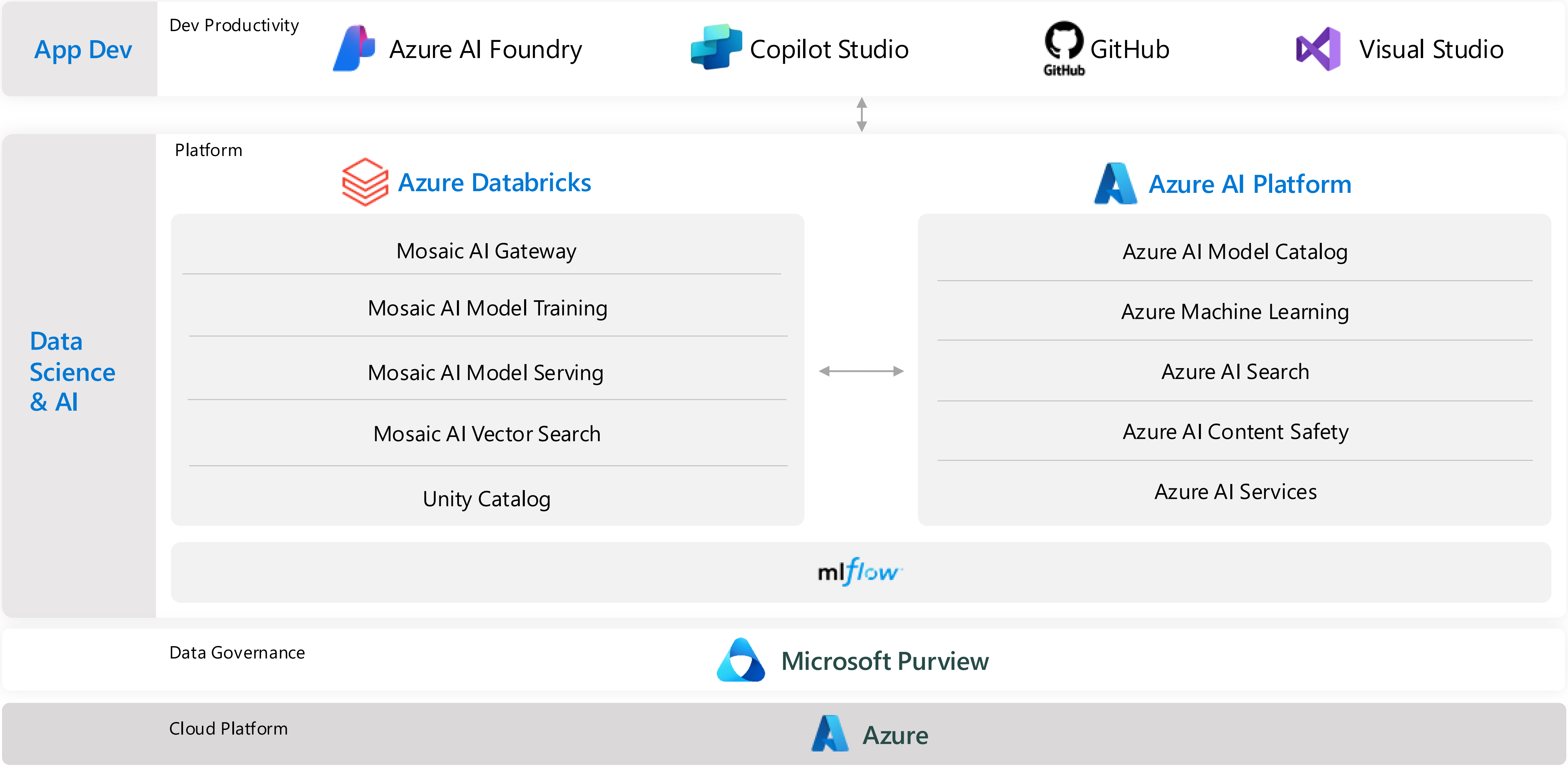
Std Discount: 0.03

Launch Genie

Ask follow up questions

Which customer segments outperformed the baseline in my top 10 most successful campaigns last quarter?

Azure Databricks and Azure AI – Best of both worlds



App Dev

Dev Productivity

Azure AI Foundry

Copilot Studio

GitHub

Visual Studio

Data Science & AI

Platform

Azure Databricks

- Mosaic AI Gateway
- Mosaic AI Model Training
- Mosaic AI Model Serving
- Mosaic AI Vector Search
- Unity Catalog

Azure AI Platform

- Azure AI Model Catalog
- Azure Machine Learning
- Azure AI Search
- Azure AI Content Safety
- Azure AI Services

mlflow

Data Governance

Microsoft Purview

Cloud Platform

Azure

Why Azure Databricks

Why Azure Databricks

Values	Azure Databricks	Databricks
Integration with Azure Services	Azure Databricks is deeply integrated with the Azure ecosystem, allowing seamless integration with other Azure services	Databricks is a standalone platform that can be used with other cloud providers
Managed Service	Offered as a managed service by Microsoft Azure	Can be deployed as a managed service on various cloud providers or as an on-premises solution
Security and Compliance	Leverages Azure's security features, including platform encryption, network isolation, and integration with Microsoft Entra ID	Provides robust security features, but the specific capabilities may vary depending on the cloud provider or deployment model
Pricing	Uses a consumption-based pricing model	Pricing varies based on the cloud provider and deployment model
Collaboration and Integration	Provides seamless collaboration through integration with Azure DevOps, Git repositories, and Azure Machine Learning	Integration capabilities depend on the chosen cloud provider and infrastructure

Decision guide: Azure Databricks vs. Microsoft Fabric

Decision guide: Azure Databricks vs. Microsoft Fabric

	Azure Databricks	Microsoft Fabric
Primary focus	A unified, open analytics platform for enterprise-grade data, analytics, and AI solutions at scale	A cloud-based SaaS platform providing low-code or no-code tools for end-to-end analytics
Technology	Delta Lake, Apache Iceberg, and unstructured data for storage, Spark optimized for data processing and AI/BI and Power BI for visualizations	OneLake is used for data storage, Power BI for visualizations, and Synapse for data engineering
Users	It requires more coding and expertise and is commonly used by tech professionals	Mostly used by business users and analysts to understand data insights
Data Engineering	Emphasis on advanced capabilities for complex data processing tasks	Emphasis on ease of use and integration
Machine Learning	Preferred choice for enterprises who focus on machine learning and AI workloads	While it supports machine learning, its primary strength lies in business intelligence and batch processing
Real-Time Analytics	Excels in real-time analytics, providing high-performance data pipelines	More focused on batch processing and creating business intelligence dashboards
Pricing	Uses a consumption-based pricing model	Uses a subscription-based pricing model tied to the Microsoft 365 ecosystem
Deployment	Supports CI/CD pipelines, Git, DABs	Seamlessly integrates with Git and Azure DevOps for deployments
Integration with Cloud Providers	Primarily integrates with external cloud providers leveraging their infrastructure for data storage and processing, integrated within the Microsoft ecosystem and Power BI	Designed to be deeply integrated within the Microsoft ecosystem, particularly with Power BI, for enhanced data visualization and reporting
Security Concerns	Provides encryption features to help protect your data in HIPAA and FedRAMP	Built-in security and reliability to secure your data at rest and transit

Coming up next...

Day 1

Module 1 - Introduction to Azure Databricks

- Azure Databricks: A Data Intelligent Platform
- Why Azure Databricks
- Decision guide: Azure Databricks vs. Microsoft Fabric

Module 2 - Migration to Azure Databricks

- Microsoft Cloud Adoption Framework for Azure
- Migration strategies
- Data landing zones
- Migration scenarios

Interactive Simulated Lab Experience

- End-to-End Streaming Pipeline with Lakeflow Declarative Pipelines in Azure Databricks

Day 2

Module 3 - Integration with Azure

- Seamless integration with Microsoft Azure services
- Connect to Azure Data Lake Storage (ADLS) Gen2 and Blob Storage
- Leverage Azure Databricks for Azure Cosmos DB Operations
- Secret management with Azure Key Vault
- Connect Azure Databricks to Azure Event Hubs

Module 4 - Integration with Microsoft Fabric and Power BI

- Data Intelligence with Azure Databricks and Microsoft Fabric
- Connect Power BI to Azure Databricks
- Integration with Azure Data Factory
- Mirroring Azure Databricks Unity Catalog

Interactive Simulated Lab Experience

- Setup and use Unity Catalog for Data Management in Azure Databricks
- Real-Time Streaming with Azure Databricks and Azure Event Hubs

Day 3

Module 5 - Integration with Azure AI Foundry

- Azure Databricks connector in Azure AI Foundry
- Mosaic AI and machine learning on Azure Databricks
- Query Generative AI model serving endpoints
- Databricks Assistant, AI/BI Genie and AI Functions on Azure Databricks
- Chat with LLMs and prototype GenAI apps using AI Playground
- Build and optimize agents on your data with Agent Bricks

Module 6 - Security and Governance

- Integrate Azure Databricks with Microsoft Purview
- Integration of Azure Databricks Unity Catalog with Microsoft Purview

Module 7 - Well-architected for Azure Databricks

- Lakehouse implementation: Principles and best practices
- Azure Databricks well-architected framework

Interactive Simulated Lab Experience

- Responsible AI with Large Language Models using Azure Databricks and Azure OpenAI
- Connect to and manage Azure Databricks in Microsoft Purview

Microsoft

Technofocus

Thank You!