



# Unlocking Data Synergy:

Integrating Databricks with SAP Datasphere  
for Scalable Analytics



# About the presenter



## Kamalendu Biswas (Kamal/Biswas)

Specialist Solutions Architect @ Databricks, Amsterdam

- 2+ years at Databricks – Benelux and Nordics
- Data Engineering and Migrations
- Previously worked with Accenture, Oracle, JustEat

X Not an SAP expert



# Preface

Nederlands is moeilijk

“SAPanese” is moeilijker



**10,000+**  
global customers

**\$1.5B+**  
in revenue

**\$4B**  
in investment

Inventor of the **lakehouse**  
&  
Pioneer of **generative AI**



**databricks**

The data and AI company

**Gartner-recognized Leader**  
Database Management Systems  
Data Science and Machine Learning Platforms

Creator of



**mlflow**<sup>TM</sup>

APACHE  
**Spark**<sup>TM</sup>

 **Unity Catalog**

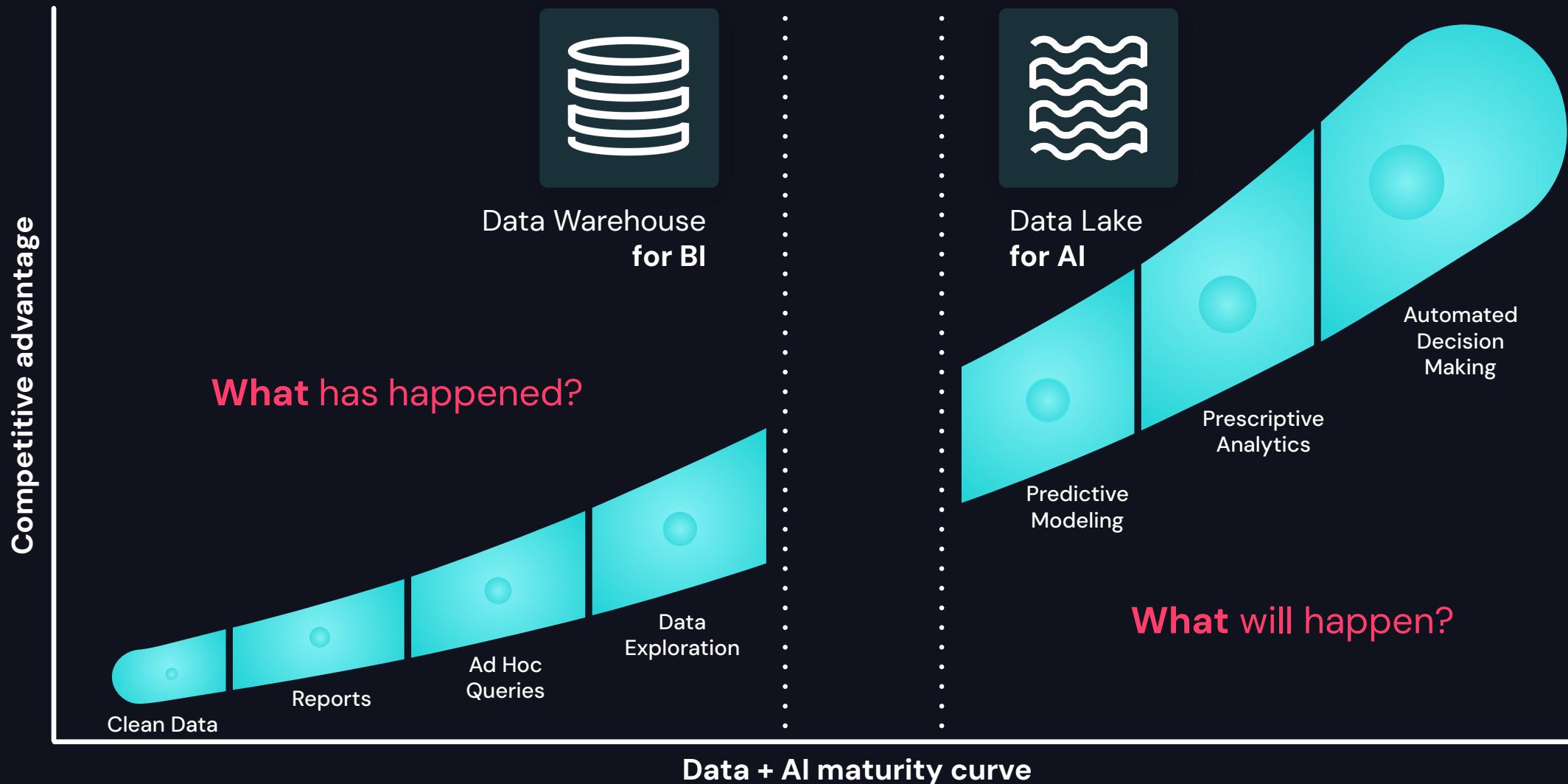
**D B R X**



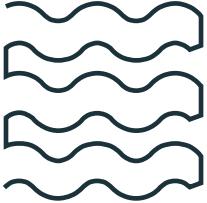
The winners in every industry will be  
data + AI companies



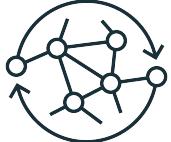
# Two in-compatible architectures



Data Lake



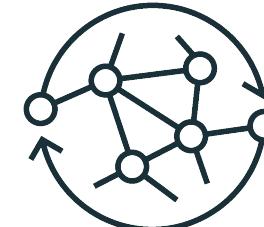
Machine Learning



Streaming



Generative AI



Data Science



Governance



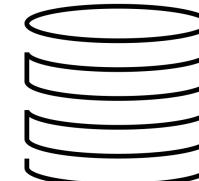
Orchestration & ETL



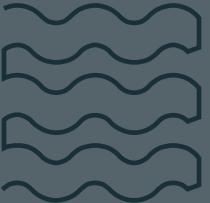
BI



Data Warehouse



Data Lake



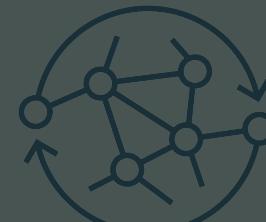
Machine Learning



Streaming



Generative AI



Data S...

Data and AI  
are siloed

Data privacy & control  
are challenged

Dependent on highly  
technical staff

Governance



Orchestration  
& ETL



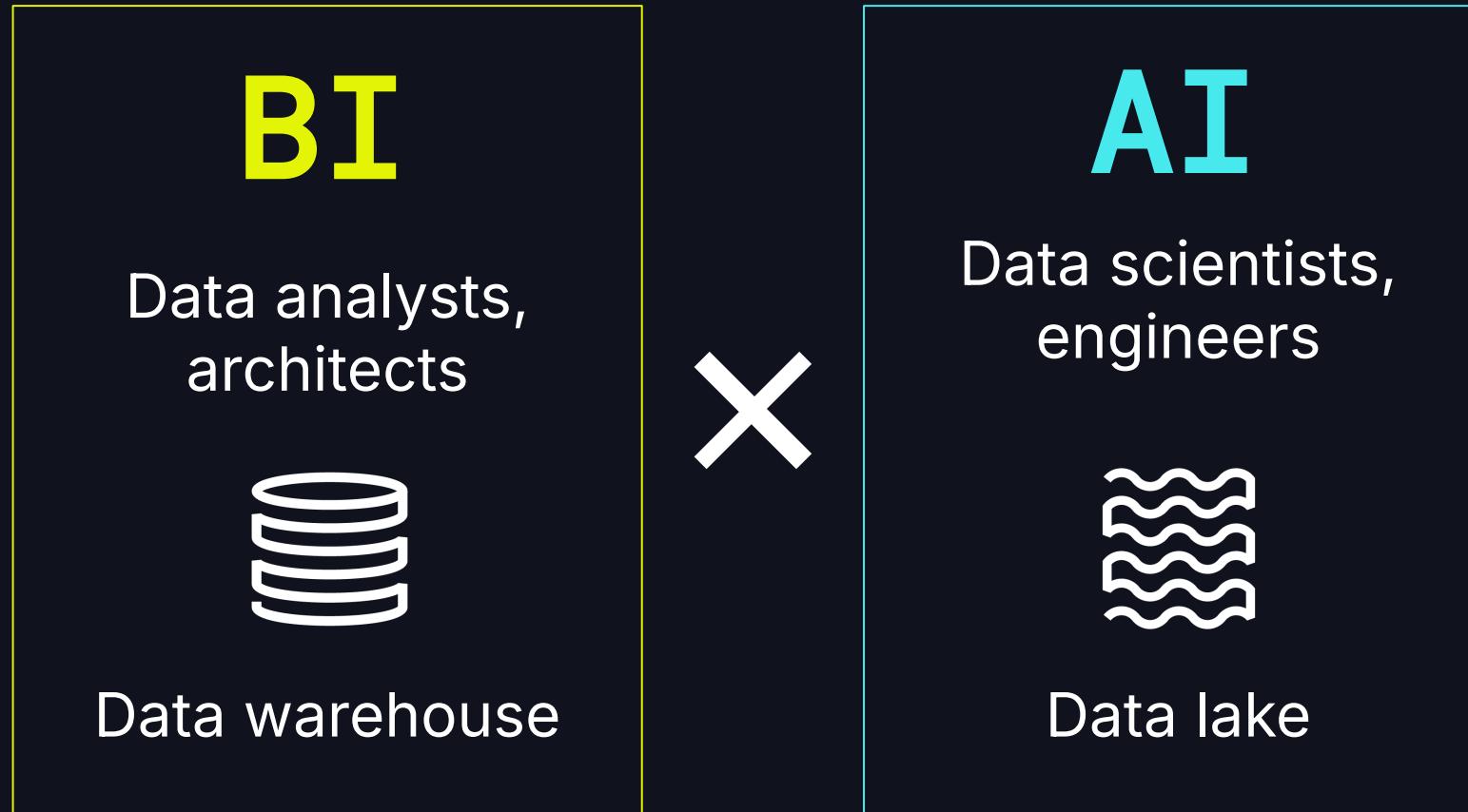
BI



Warehouse



# Legacy “split” data stacks



# We pioneered the lakehouse

Combines the best of data warehousing and data lakes

## What Is a Lakehouse?



by Ben Lorica, Michael Armbrust, Reynold Xin, Matei Zaharia and Ali Ghodsi

January 30, 2020 in Engineering Blog

Share this post



Over the past few years at Databricks, we've seen a new data management architecture that emerged independently across many customers and use cases: [the lakehouse](#). In this post we describe this new architecture and its advantages over previous approaches.

Data warehouses have a [long history](#) in decision support and business intelligence applications. Since its inception in the late 1980s, data warehouse technology continued to evolve and MPP architectures led to systems that were able to handle larger data sizes. But while warehouses were great for structured data, a lot of modern enterprises have to deal with unstructured data, semi-structured data, and data with high variety, velocity, and volume. Data warehouses are not suited for many of these use cases, and they are

# The industry followed

## What is a lakehouse in Microsoft Fabric?

Article • 08/30/2023 • 5 contributors

In this article

- Lakehouse SQL endpoint
- Automatic table discovery and registration
- Interacting with the Lakehouse item

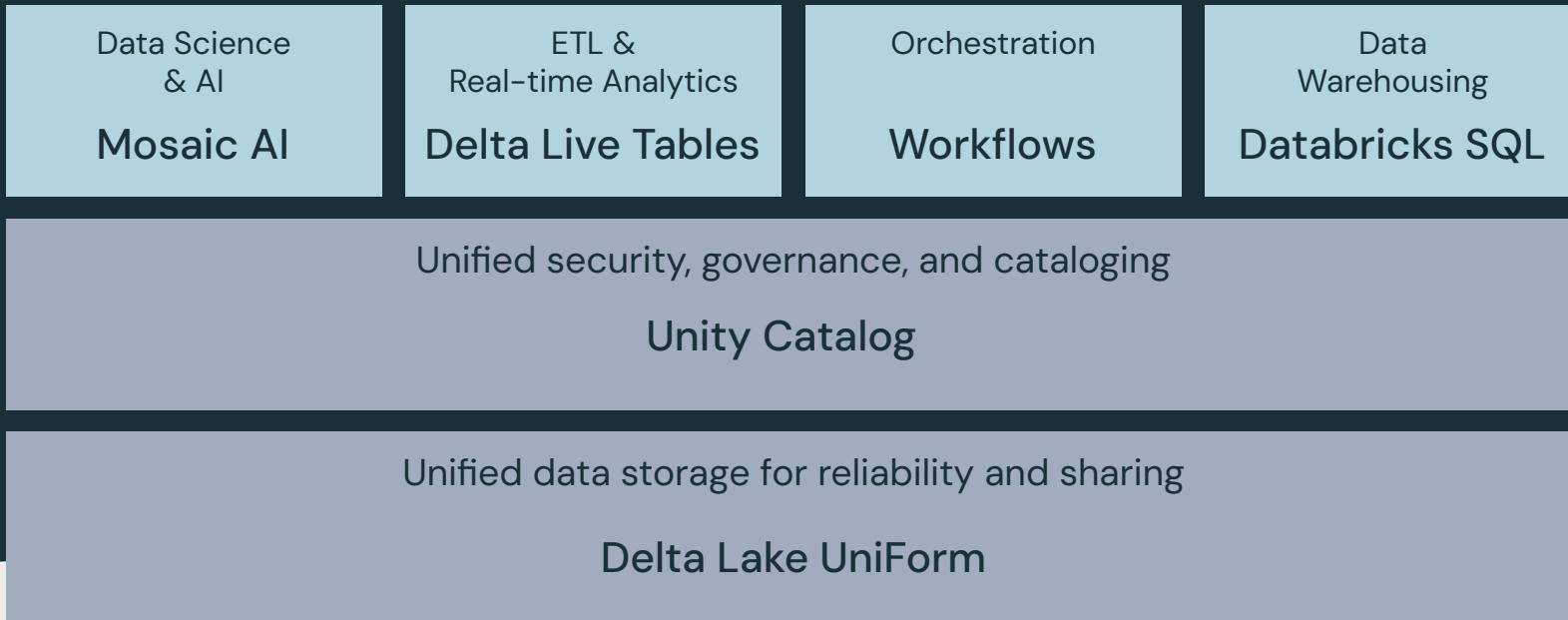
Next steps

Microsoft Fabric Lakehouse is a data architecture platform for storing, managing, and analyzing structured and unstructured data in a single location. It's a flexible and scalable solution that allows organizations to handle large volumes of data using various tools and frameworks to process and analyze that data. It integrates with other data management and analytics tools to provide a comprehensive solution for data engineering and analytics.

The screenshot shows the Microsoft Fabric Lakehouse interface. On the left, there's a sidebar with icons for Home, Create, Browse, Data hub, and Monitoring hub. The 'Lakehouse explorer' section shows a tree view with 'ContosoDailySales' expanded, showing 'Tables' and 'Customer'. The main area displays a table titled 'Customer' with columns: Index, UserId, FirstName, LastName, Sex, Email, Phone, DateOfBirth, and JobTitle. Two rows of data are visible: Row 1 (Index 1) has UserId '3d5AD30A...', FirstName 'Jo', LastName 'Rivers', Sex 'Female', Email 'fergusonkat...', Phone '-10395', DateOfBirth '7/26/1931', and JobTitle 'Dancer'. Row 2 (Index 2) has UserId '810Ce0F27...', FirstName 'Sheryl', LastName 'Lowery', Sex 'Female', Email 'fhoward@e...', Phone '(599)782-0...', DateOfBirth '11/25/2013', and JobTitle 'Copy'. A tooltip for the 'Lakehouse' button in the top right corner says 'Explore your data files and folders'. Another tooltip for the 'SQL endpoint' button says 'Query data using SQL'.

# The Data Lakehouse

An open, unified foundation for all your data



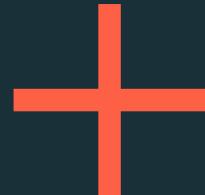
2020  
Databricks pioneered  
the lakehouse  
architecture

Today  
74% of global  
enterprises have  
adopted lakehouse

MIT Technology Review  
Insights, 2023

# Data Lakehouse

An open, unified foundation  
for all your data



# Generative AI

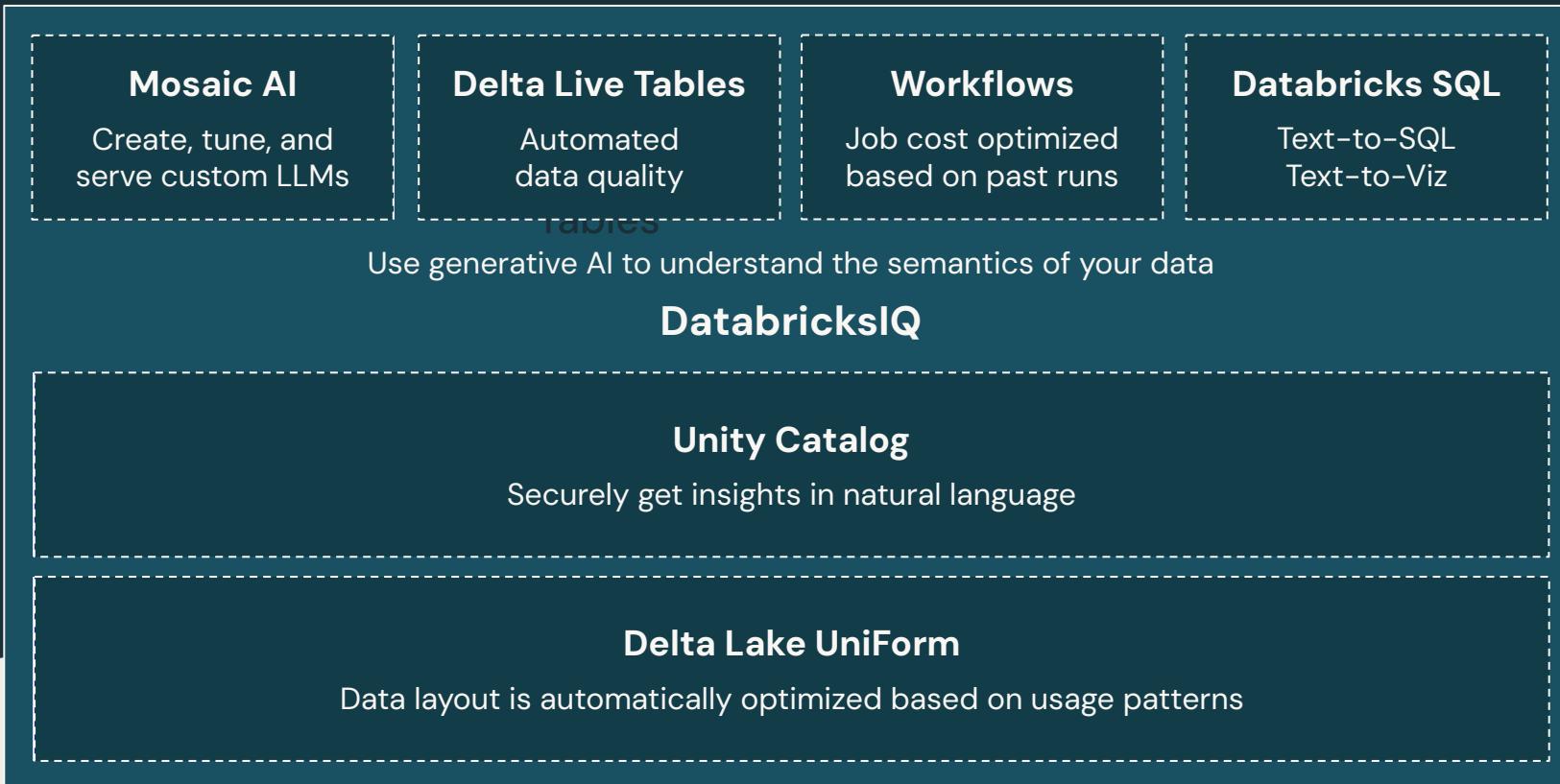
Easily scale and use data and AI



# Data Intelligence Platform

Democratize data + AI across  
your entire organization

# Databricks Data Intelligence Platform



## Open Data Lake

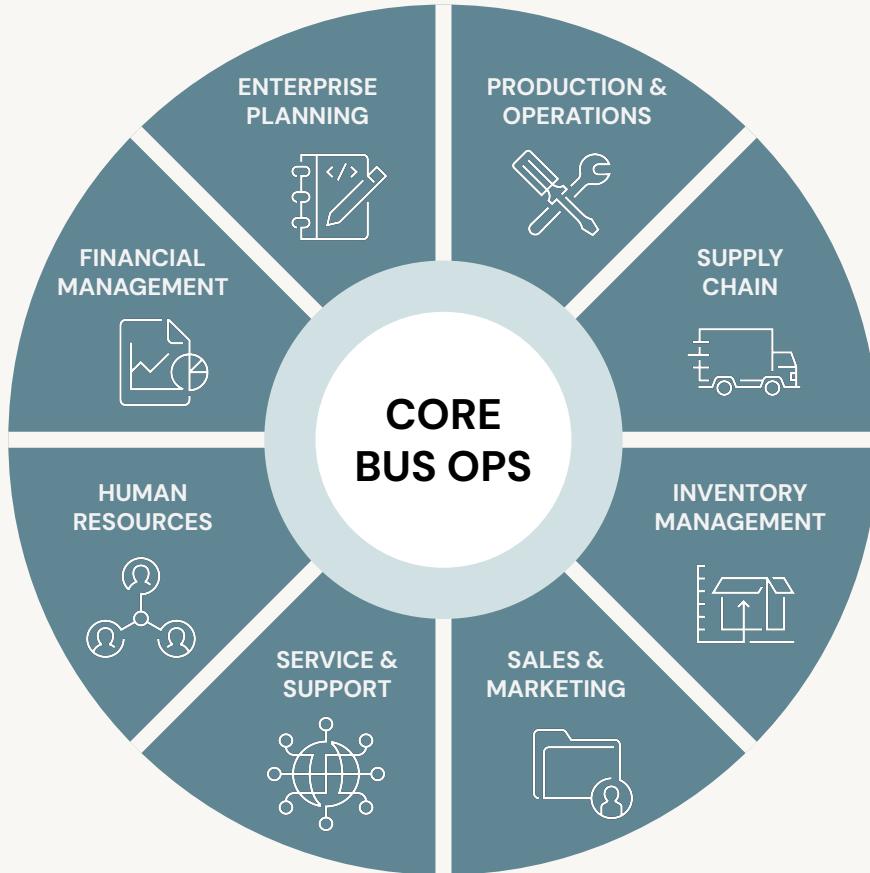
All Raw Data  
(Logs, Texts, Audio, Video, Images)



databricks



# SAP data powers core business operations



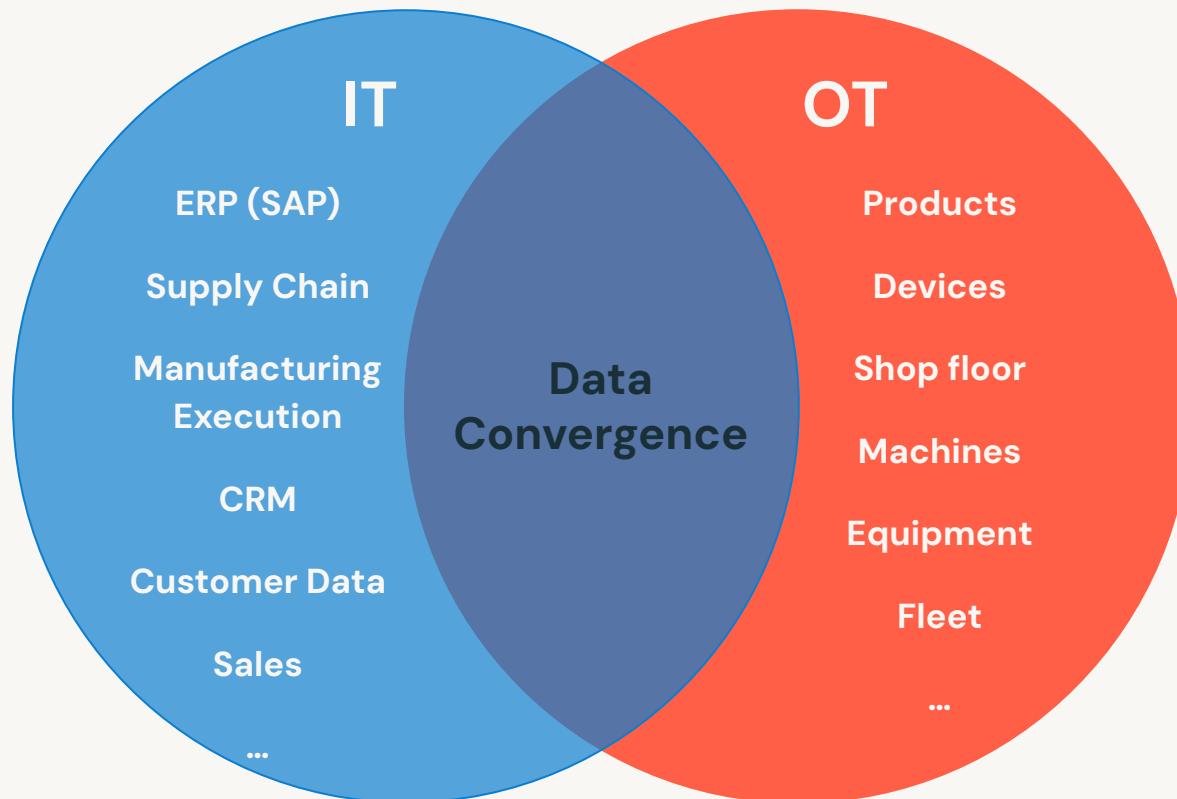
SAP integrates various functions into a **cohesive system, collecting and processing data** across the enterprise.

The **data in SAP is mission critical** – unlocking advanced analytics use cases across supply chain, manufacturing, procurement, finance, and even HR.

**Without ease and speed of SAP data integration, it's hard to optimize end-to-end business processes and improve decision making**



# Organizations need to converge all of their enterprise data to capture its full value



**Resulting Outcomes when effectively leveraging the data**

**Better predict demand**

**Optimize E2E Supply Chain**

**Drive operational & industrial efficiency**

**Enhance workforce productivity**

**Transform buying to service experiences**

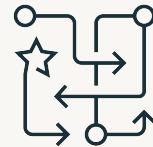


# It has been hard for organizations to easily and effectively manage SAP data and non-SAP data at scale



## Limited access to the all the data due to silos

As the need to join SAP data with other sources for real-time and predictive insights has grown, customers are looking for more efficient ways to accelerate availability of SAP data outside of the core system



## Difficulty connecting data sources

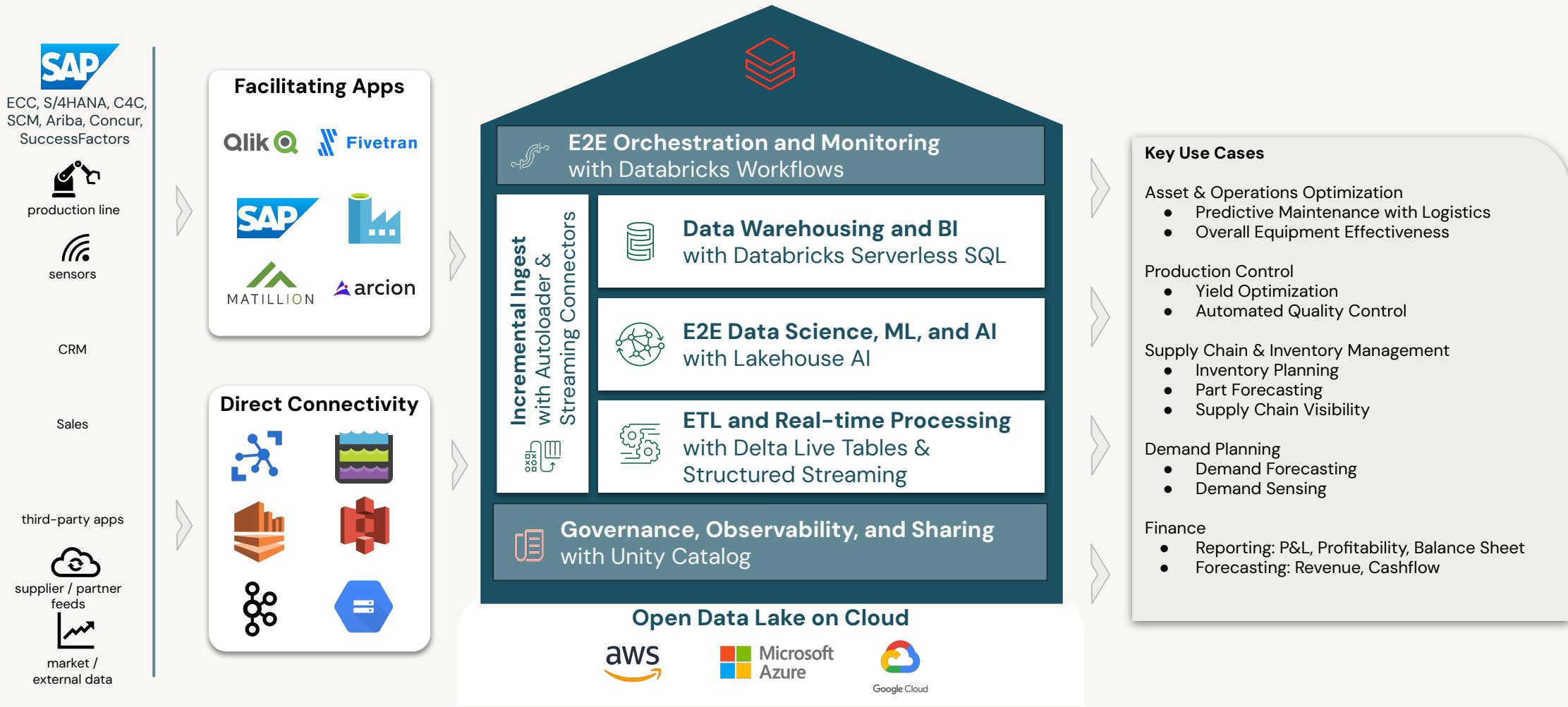
As the need to join SAP data with other sources for real-time and predictive insights has grown, customers are looking for more efficient ways to accelerate availability of SAP data outside of the core system



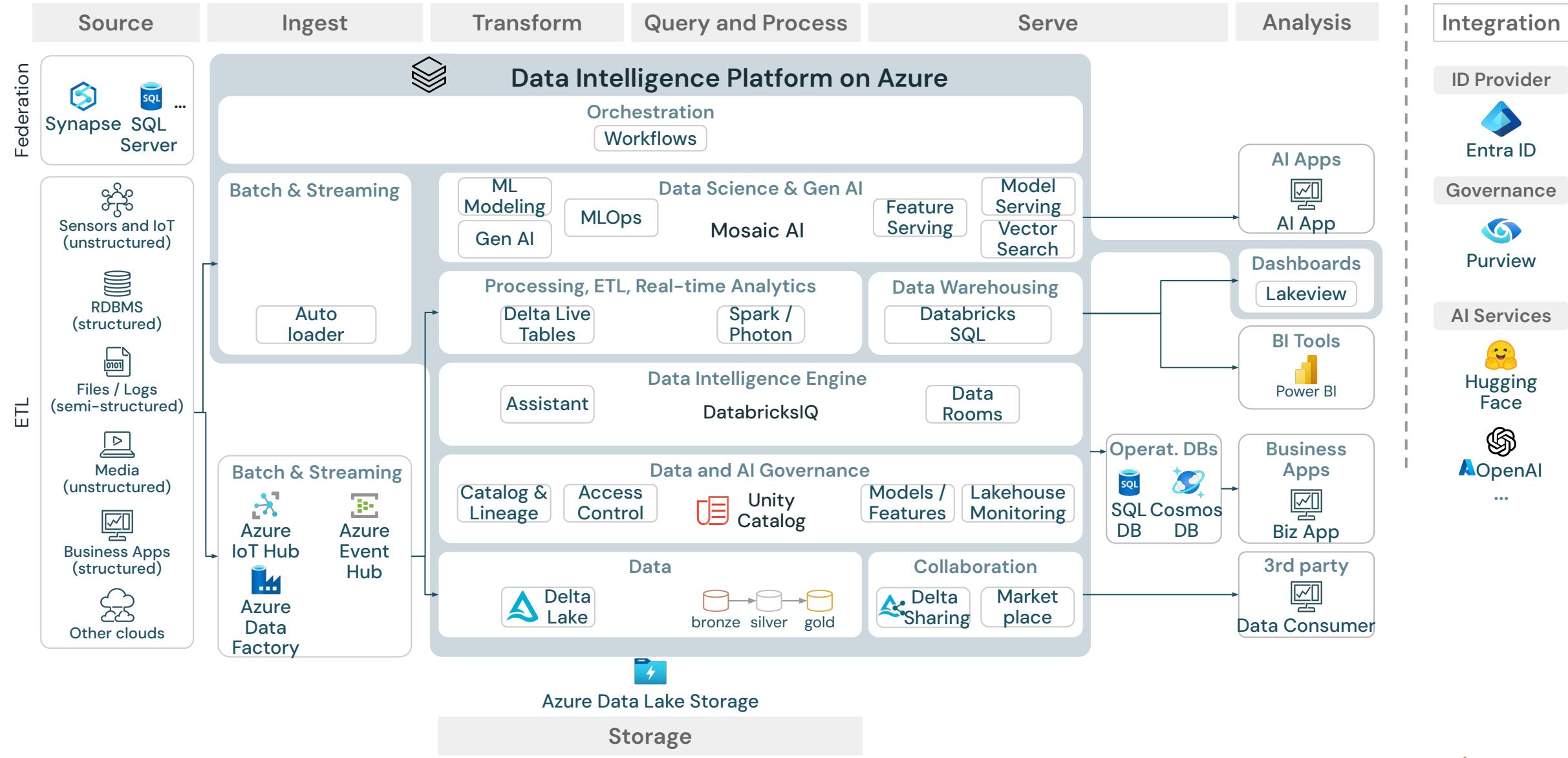
## Limited availability for advanced analytics across all the data

Companies are increasingly looking to implement more advanced analytics at scale and with a broader range of data than is functionally or economically possible in SAP

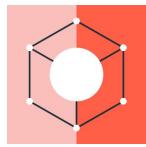
Databricks Lakehouse provides a unified data platform to best manage SAP and non-SAP data



# Databricks Data Intelligence Platform on Azure



# Unlock the full potential of enterprise data with the Databricks Lakehouse



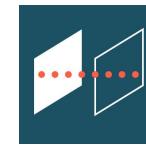
## Simplify access to semantically rich data

Streamline access to semantically rich SAP data and other non-SAP and unstructured data



## Harmonize all enterprise data

Speed time-to-insight by harmonizing SAP data with non-SAP Data (in Lakehouse)



## Deliver predictive insights with AI and ML

Develop more accurate predictions using harmonized data for building and deploying advanced analytics, AI and ML models

at a reduced TCO across the enterprise data technology stack

# Databricks is the only multi-cloud data & AI platform fully supported by SAP Datasphere

Databricks' bi-directional integration with SAP Datasphere, which is SAP's gateway to SAP ERP & LoB application data



**SAP Datasphere** delivers governed, scalable, and seamless **access to SAP data** with business semantics and context intact



*Virtualize or replicate  
(partnership)\**

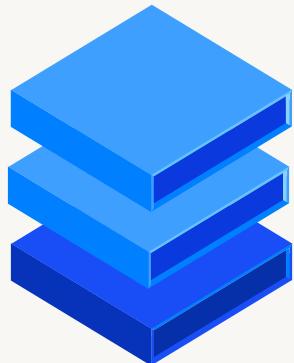


**Databricks** provides an open, **unified data lakehouse platform** for data, analytics and AI uses cases for SAP and non SAP data

**Accelerate** insights with access to semantically rich SAP data

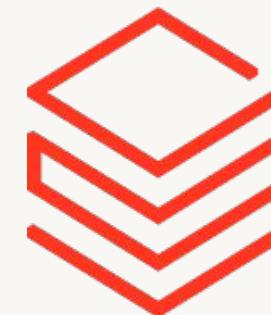
**Unlock** the value of data across the business data fabric

## Accelerate insights with access to semantically rich SAP data



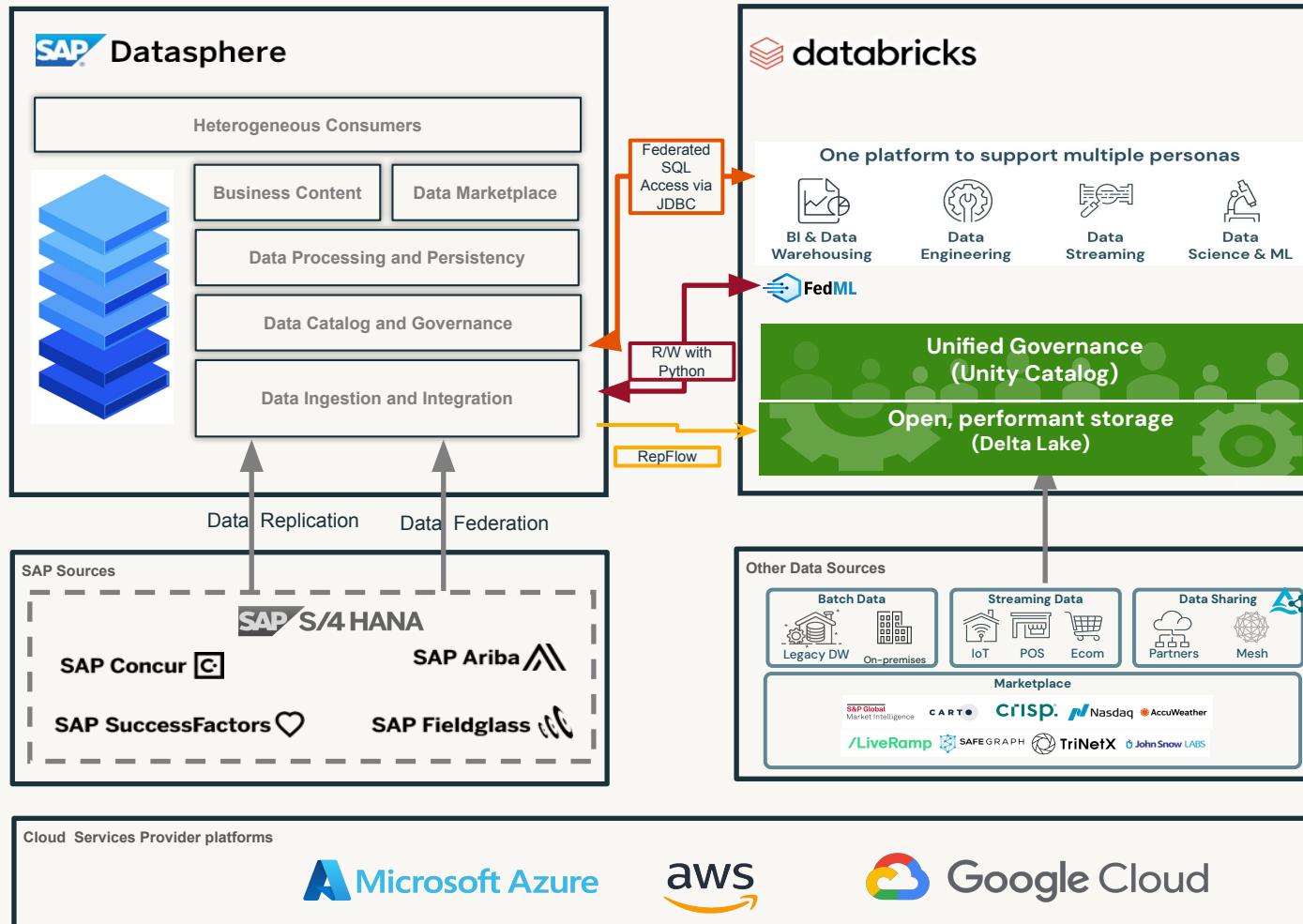
SAP Datasphere delivers governed, scalable, and seamless access to SAP data with business semantics and context intact

## Unlock the value of data across the business data fabric



Open, unified data lakehouse platform for data, analytics and AI uses cases

# Unify Enterprise Data with a Modern Data Stack



## Key Scenarios:

- Advanced analytics & data science on SAP data**  
Use the FedML library from Databricks Notebooks to access Datasphere tables, analyze the data, store analysis results in the Lakehouse or write back to Datasphere.
- ML models on SAP Data**  
Build, train and deploy ML models on enterprise data that resides in SAP Datasphere using FedML. Write the model outputs back to Datasphere if required.
- Lakehouse data federated into Datasphere**  
Use SAP Datasphere's Federation capability to query data in the Lakehouse (Delta tables) and analyze it together with SAP data.
- Advanced analytics with blended SAP/non-SAP data**  
Combine fresh, semantically rich data in SAP Datasphere with the data in the Lakehouse for advanced analytics and ML capabilities in Databricks

<https://blogs.sap.com/2023/05/02/federating-queries-to-databricks-from-sap-datasphere-for-real-time-analytics-in-sap-analytics-cloud/>

<https://community.sap.com/t5/technology-blogs-by-sap/replication-flow-blog-series-part-4-sizing/ba-p/13579486>



# Future Roadmap – TechEd Announcements

SAP's new embedded data lake capabilities will include:

- **An integrated object store**, which provides a simpler way to store large amounts of data, making it easier for businesses to expand their storage as needed.
- **Spark compute**, which facilitates efficient data transformation and processing based on existing SAP Datasphere data integration capabilities.
- **SQL on files functionality**, which lets developers access data in the integrated object store without needing to physically copy the data. This makes it appear in SAP Datasphere data models the same as any other persisted data (that is, data which is permanently stored and readily accessible), streamlining data integration processes, reducing storage redundancy and ensuring data consistency.

Announcements – <https://www.sap.com/events/teched/news-guide/data-and-analytics.html>

\*Availability – Available by the end of Q4 2024

\*check with your SAP rep!

# My Questions/Possibilities

- **An integrated object store:**
  - Can Databricks directly read the Datasphere object store?
- **Spark compute:**
  - Can Datasphere directly read from Databricks Delta sharing?



databricks

