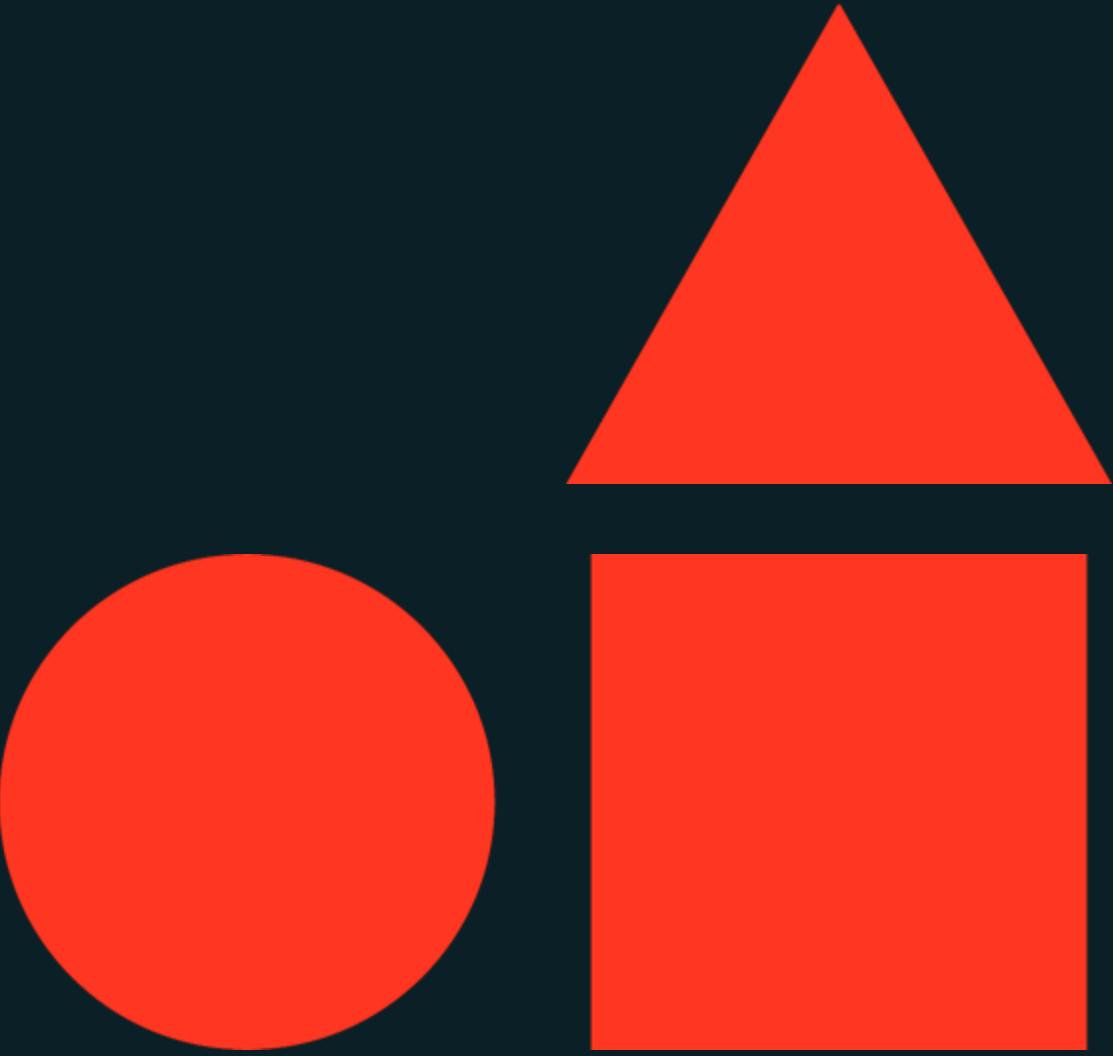




# Get Started with Databricks for Data Warehousing



---

**Databricks Academy**  
**August 2025**



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Agenda

Databricks Overview	Time	Lecture	Demo	Lab
<b>Databricks Data Intelligence Platform</b>	8 mins	✓		
<b>Databricks Workspace Walkthrough</b>	15 mins		✓	



# Agenda

Using Databricks for Data Warehousing	Time	Lecture	Demo	Lab
<b>Introduction to Data Warehousing with Databricks</b>	10 mins	✓		
<b>Databricks SQL Warehouses</b>	5 mins	✓		
<b>Delta Lake Overview</b>	8 mins	✓		
<b>Using Delta Lake features with Databricks SQL</b>	5 mins		✓	



# Agenda

Data Ingestion and Transformation	Time	Lecture	Demo	Lab
Delta Lake UniForm	5 mins	✓		
Data Ingestion Techniques Overview	5 mins	✓		
Data Ingestion Techniques	10 mins		✓	
Data Transformation	8 mins	✓		
Exploring Data Transformation in Databricks	10 mins		✓	



# Agenda

Data Orchestration and Querying Capabilities	Time	Lecture	Demo	Lab
Orchestration in Databricks	5 mins	✓		
Setting Up and Managing Serverless Lakeflow Jobs	15 mins		✓	
Databricks Querying Capabilities	5 mins	✓		



# Agenda

Data Presentation	Time	Lecture	Demo	Lab
Introduction to AI/BI	5 mins	✓		
Creating a Dashboard in Databricks	5 mins		✓	
Creating Genie Spaces	5 mins		✓	
Data Warehousing Comprehensive Lab	30 mins			✓





# Lab Exercise Environment

## Technical Details

- Your lab environment is provided by Vocareum.
- It will open in a new tab.
- It has been configured with the permissions and resources required to accomplish the tasks outlined in the lab exercise.
- Third party cookies must be enabled in your browser for Vocareum's user experience to work properly.
- Make sure to enable pop ups!



# Course Learning Objectives

- Describe the available compute options for workloads performed on the Databricks Data Intelligence Platform.
- List the products and features Databricks offers for different data-centric needs within the Databricks Platform.
- Navigate the Databricks Workspace UI.
- Use Databricks to complete common data warehousing tasks.
- Explain the purpose of Delta Lake for data warehousing purposes.
- Describe and apply various techniques for ingesting and transforming data in Databricks.
- Explain how Databricks support data orchestration needs within the platform.
- Describe the impact of data warehousing work on Databricks AI/BI needs.
- Identify the AI supported features in the Databricks platform.





# Databricks Overview

---

Get Started with Databricks for Data Warehousing



# Learning Objectives

- List the four key product lines available from Databricks.
- Describe how each product line serves its respective audience.
- Navigate the Databricks Workspace UI to locate key features and functionalities.



# Agenda

Databricks Overview	Time	Lecture	Demo	Lab
<b>Databricks Data Intelligence Platform</b>	8 mins	✓		
<b>Databricks Workspace Walkthrough</b>	15 mins		✓	





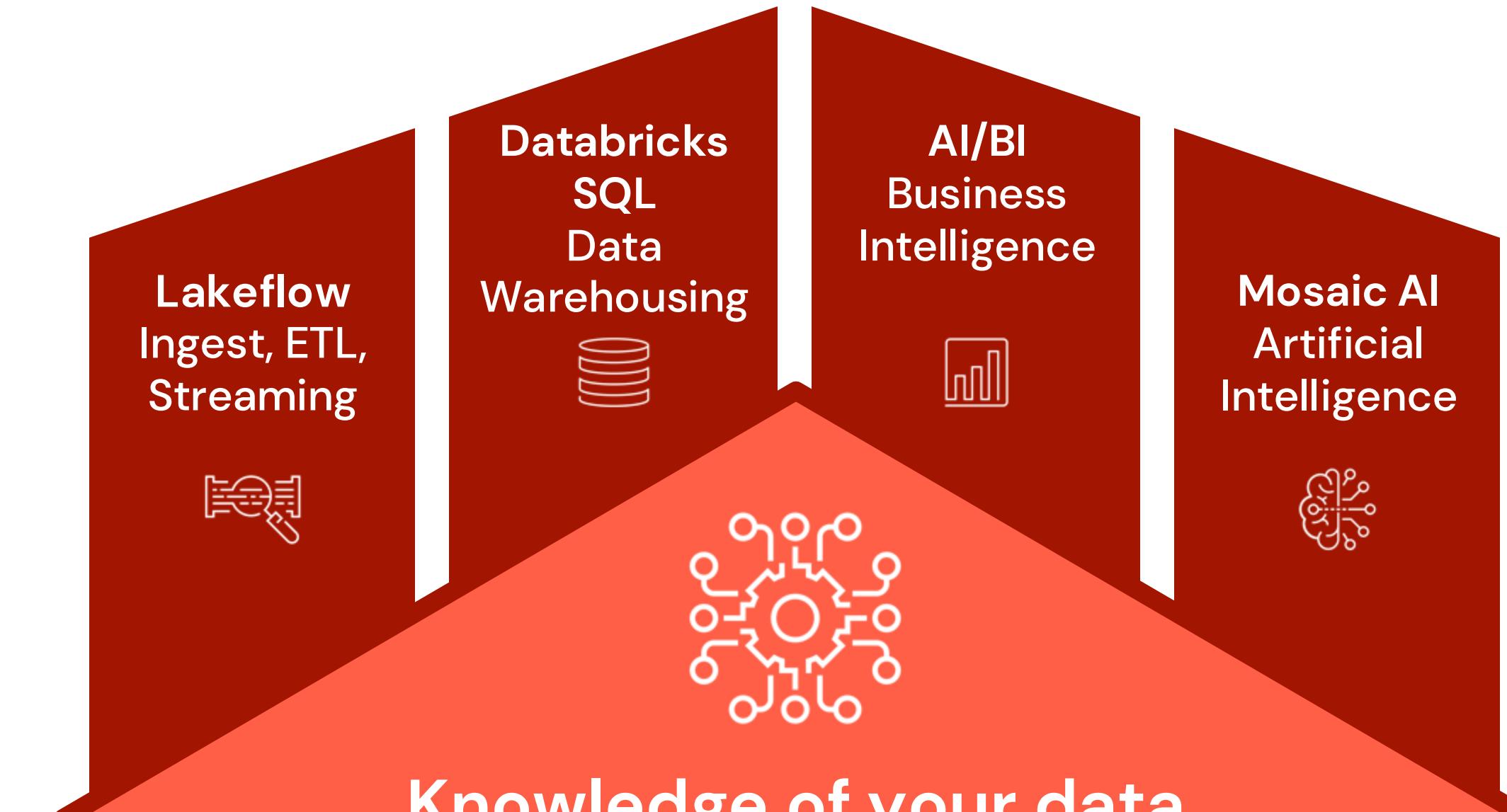
Databricks Overview

LECTURE

# Databricks Data Intelligence Platform



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).



**Knowledge of your data**

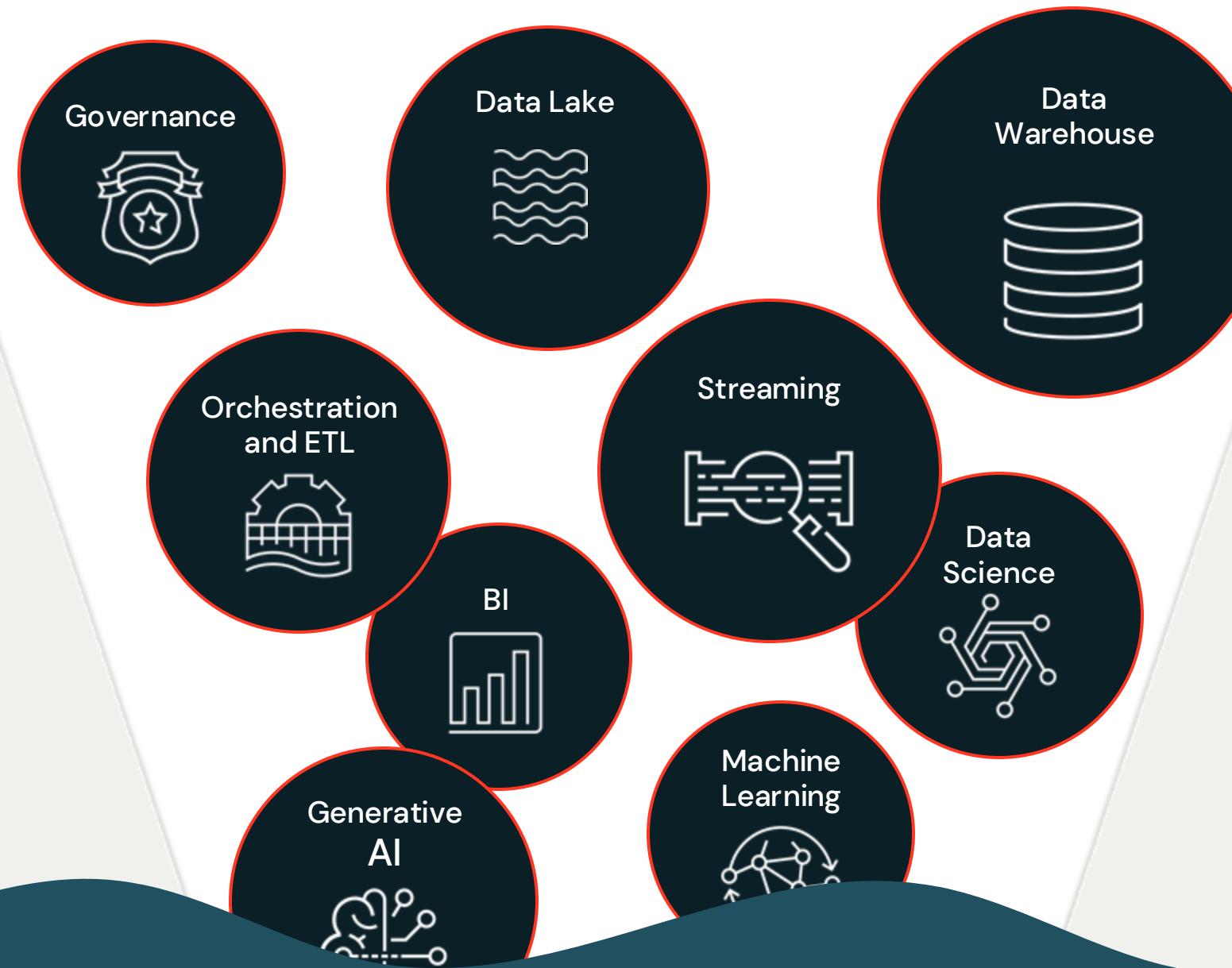
 **Unity Catalog**

Your data stored in an open, broadly accessible lakehouse format

Databricks  
simplifies the complexity in your data  
estate



# Unify and own your data



Adopt an open source Lakehouse  
format of choice

Break free from lock-in  
with open formats

Reduce unnecessary costs  
from multiple copies of data



Iceberg



Delta Lake

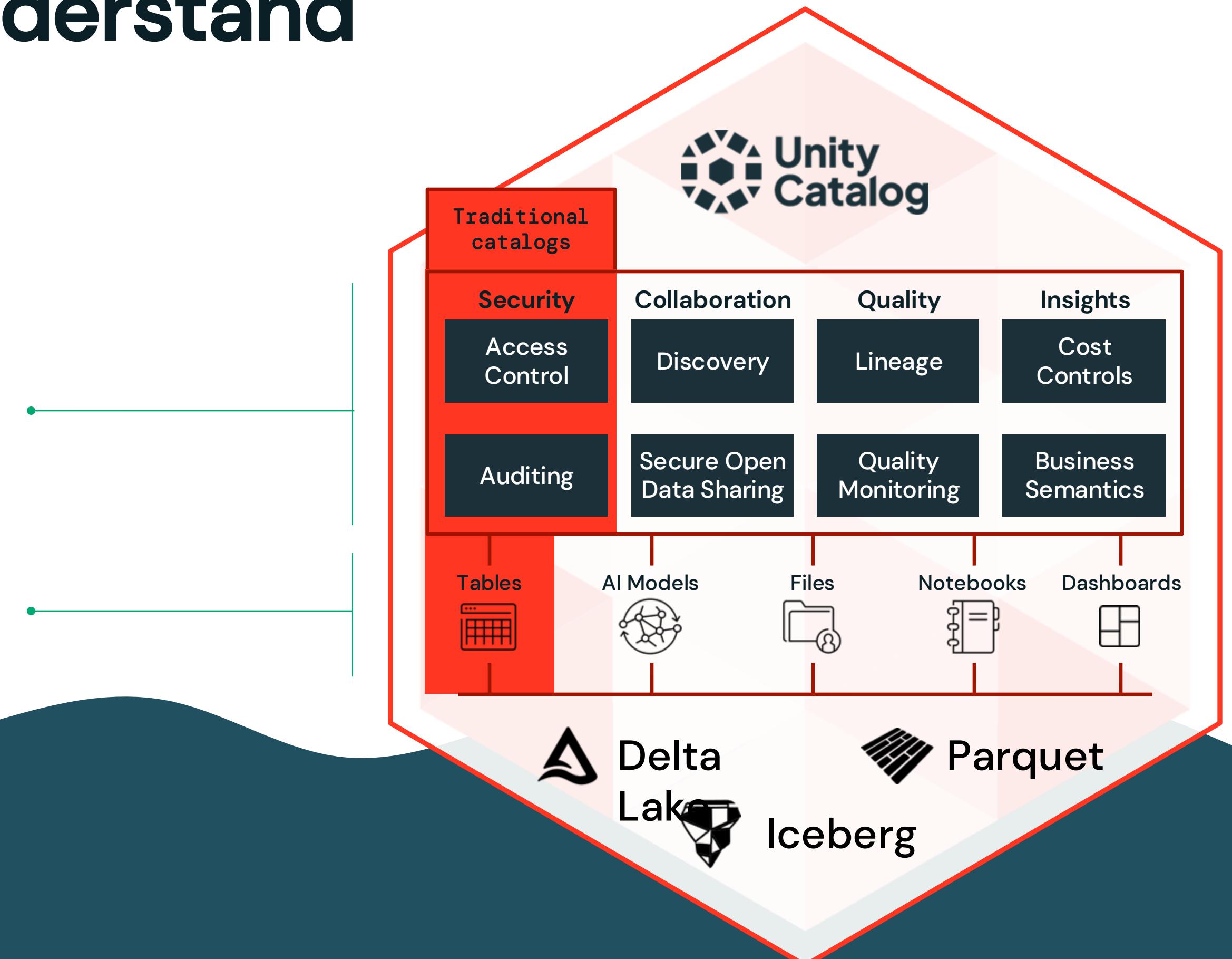


Parquet

# Protect and understand your data

Unified capabilities  
for every use case

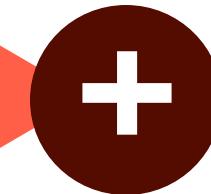
Unified governance  
for all assets





# Data Intelligence Platform

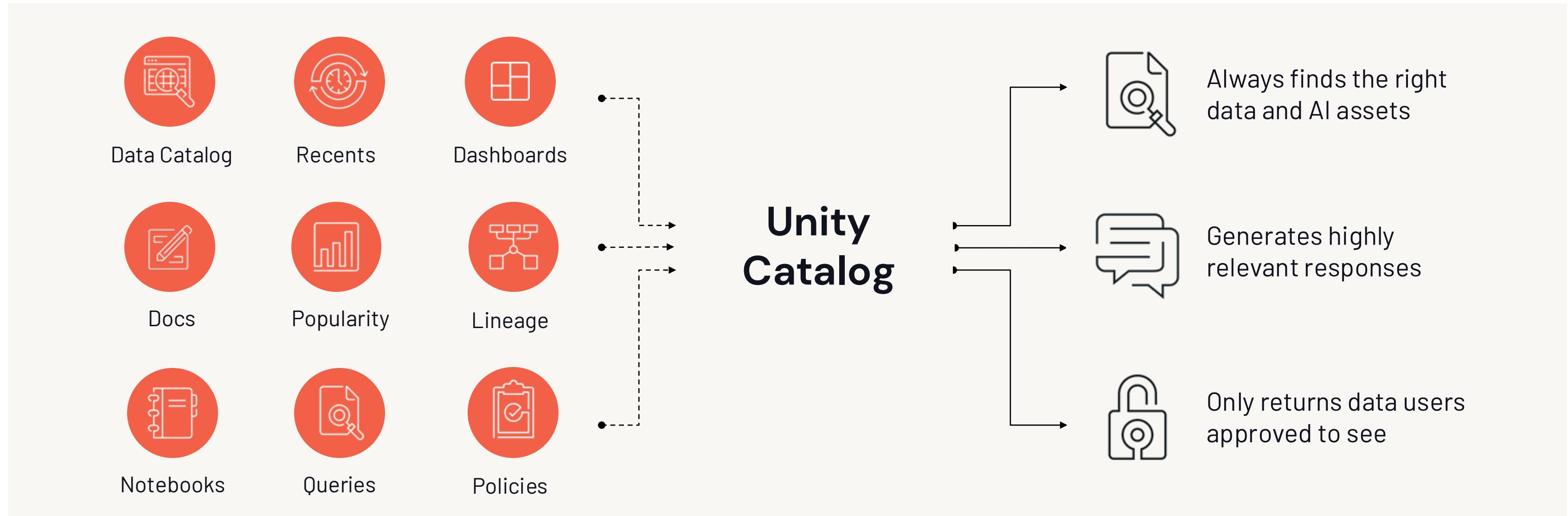
AI tuned to your  
business



Unified data  
and governance



# Data Intelligence delivers the unique context of your business



# Boosts productivity for all users

Data Scientists | Data Engineers | SQL Analysts | Business Users

For  
Data  
Teams

The screenshot shows a Databricks Notebook interface. On the left, the sidebar includes options like Workspace, Catalog, Workflows, Compute, SQL, and Data Engineering. The main area displays a notebook titled "Untitled Notebook 2024-01-26 12:27:52". A prominent feature is the "Databricks Assistant" panel, which provides AI-powered suggestions for code, queries, and answers. Below it, there's a code editor with a snippet of Python code and a "Run cell" button.

Assistant

The screenshot shows the Databricks Catalog Explorer. The sidebar lists catalogs such as "richardt\_demos", "chicago\_data", and "rh\_demo\_more\_bedbug\_catalog". The main area focuses on the "food\_inspections" table within the "chicago\_data" catalog. It displays detailed metadata including columns, sample data, and permissions. An AI-generated comment about the table's purpose is shown at the top right.

Metadata

The screenshot shows the Databricks search interface. It features a "Get started" section with links for importing data, creating notebooks, and running SQL queries. Below it is a "Pick up where you left off" section displaying recent dashboards and notebooks. The sidebar on the left shows a list of popular dashboards.

Search

For  
Business  
Teams

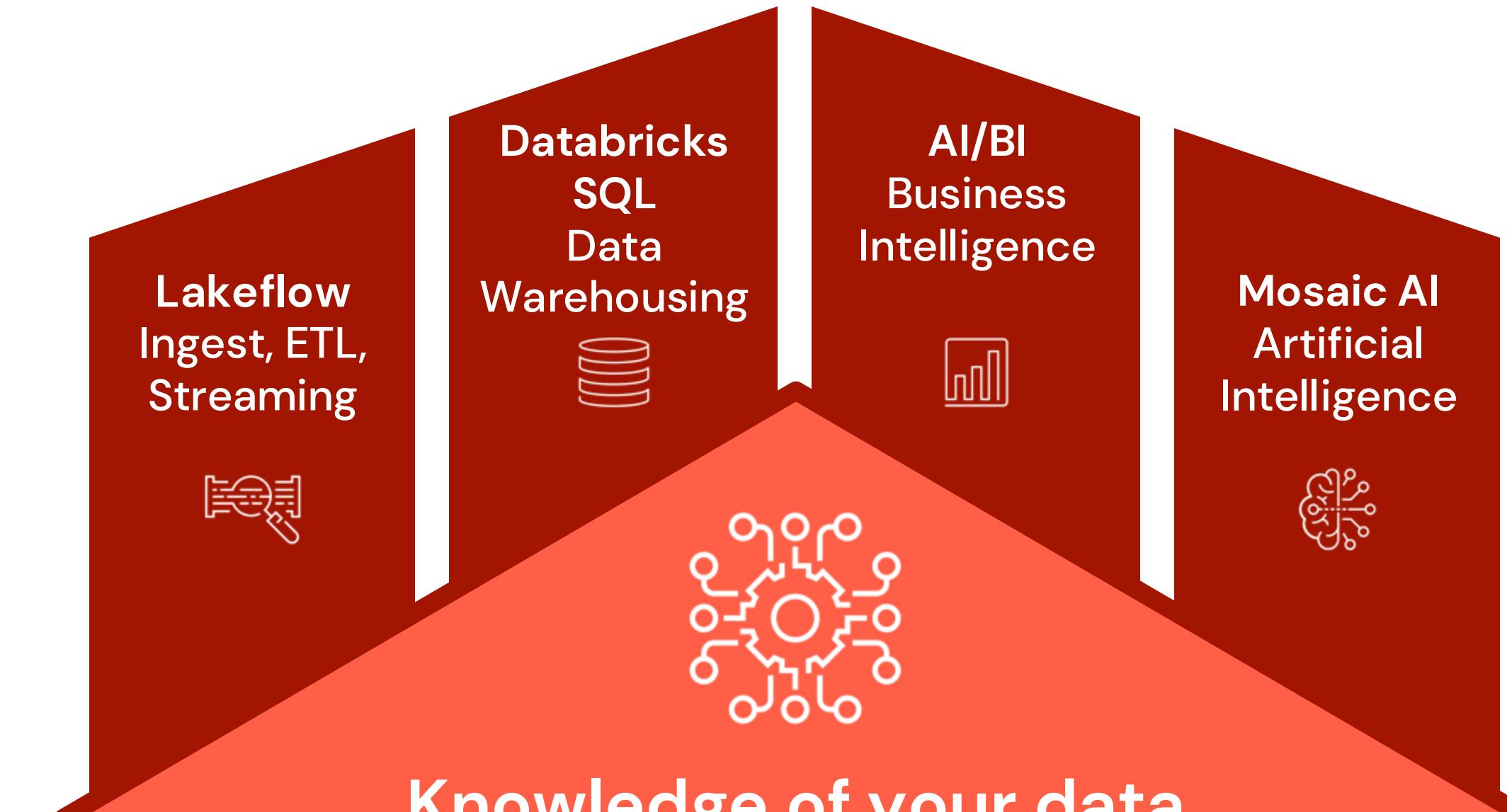
The screenshot shows the Databricks Canvas interface. It displays a chart titled "Electric Vehicles and Hybrids in Washington State" showing the number of EV and hybrid models by car maker from 2016 to 2023. To the right, there's a visualization editor with fields for "Dataset" (set to "ev vehicles"), "Visualization" (set to "Bar"), and "X axis" and "Y axis". A sidebar on the right contains AI-generated suggestions for charts and visualizations.

AI/BI Dashboards

The screenshot shows the Databricks AI/BI Genie interface. It features a chat window titled "Acme Services Product Usage" with a message from the AI asking for example questions. Below the chat is a section titled "Key product performance" with a bar chart. At the bottom, there are buttons for "Give me 5 example questions" and "Explain the data set".

AI/BI Genie





**Your data stored in an open, broadly accessible lakehouse format**

# Achieving data intelligence begins with getting data in



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).



# Lakeflow

Reliable & automated  
dataflow from  
systems-of-record



Knowledge of your data



Your data stored in an open, broadly  
accessible lakehouse format

# Get the right data without the hassle



## databricks Lakeflow

### Connect



Easily connect key data sources

ORACLE  
DATABASE

PostgreSQL

MySQL

Microsoft  
SQL Server

salesforce

workday.

Google  
Analytics

servicenow

Google  
Ads

ORACLE  
NETSUITE

SharePoint

Dynamics 365

### DLT



Reliable data pipelines made easy

### Jobs



Unified orchestration for analytics and AI

©

Databricks 2020. All rights reserved. Apache, Apache Spark, Spark, the Spark logo, Apache Iceberg, iceberg, and the Apache

Iceberg logo are trademarks of the [Apache Software Foundation](#).

# What is data intelligence for analytics and BI?





# Databricks SQL

The intelligent enterprise  
data warehouse



Your data stored in an open, broadly  
accessible lakehouse format



# Complete data warehousing capabilities

## Lower migration cost

Serverless Warehouses	Views/Temp Views	ANSI SQL	Data Quality Monitoring	Monitor Permissions
1P/3P Connectors	SQL Scripting	SQL UDFs	Table ACLs	Billing System Tables
CDC Support	H3 GeoSpatial	Session Vars	Lakehouse Federation	Internet On/Off
Streaming Tables	Spatial SQL (ST_)	SQL Alerts	HMS Federation	Queries System Table
Materialized Views	Foreign/Primary Keys	Dashboards	Table Lineage	Warehouses Sys Table
Autoloader for Ingest	Lateral Col Alias	Dashboard Sharing	Entity Relation Diagram	WH Events Sys Table
Scheduled Workflows	Named Arguments	Publish to PowerBI	Row/Col Security	Query Duration Limits
1p/3p Orchestration	HyperLogLog	Publish to Tableau	OAuth	Foreign/Primary Keys
SQL Editor	Array Functions	Cloud Fetch	ABAC	Query History
Python/Go Connectors	Identifier Clause	Python UDFs	Marketplace	Query Profiling
SQL Rest API	Variant Data Type	Row Level Concurrency	100K+ User Support	Warehouse Monitoring

Foundational Functionality

Governance and Administration

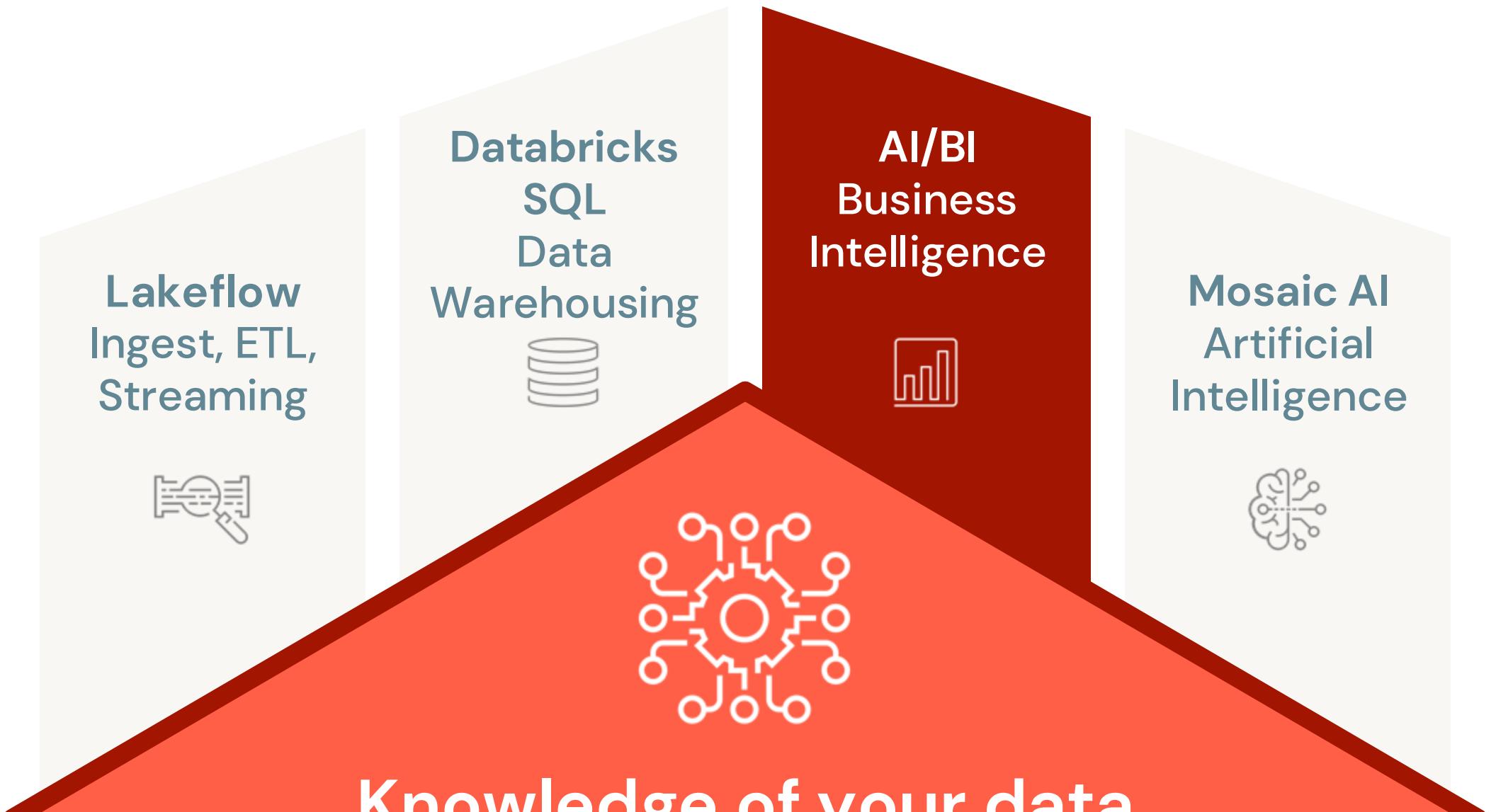


© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).



# AI/BI

Intelligent analytics  
for real-world data



Your data stored in an open, broadly  
accessible lakehouse format

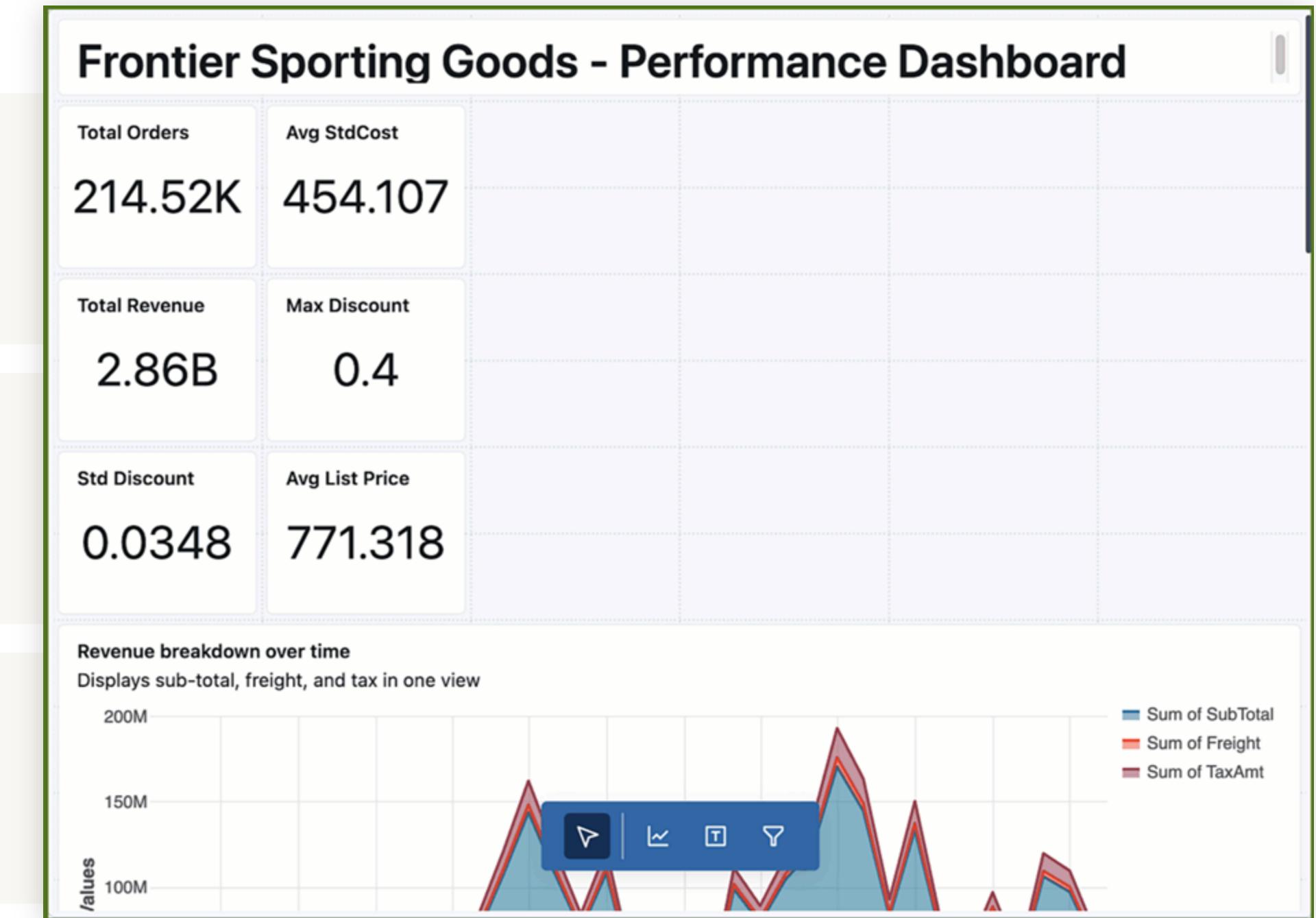
# Dashboards to analyze and visualize data

Get useful results with AI-powered understanding

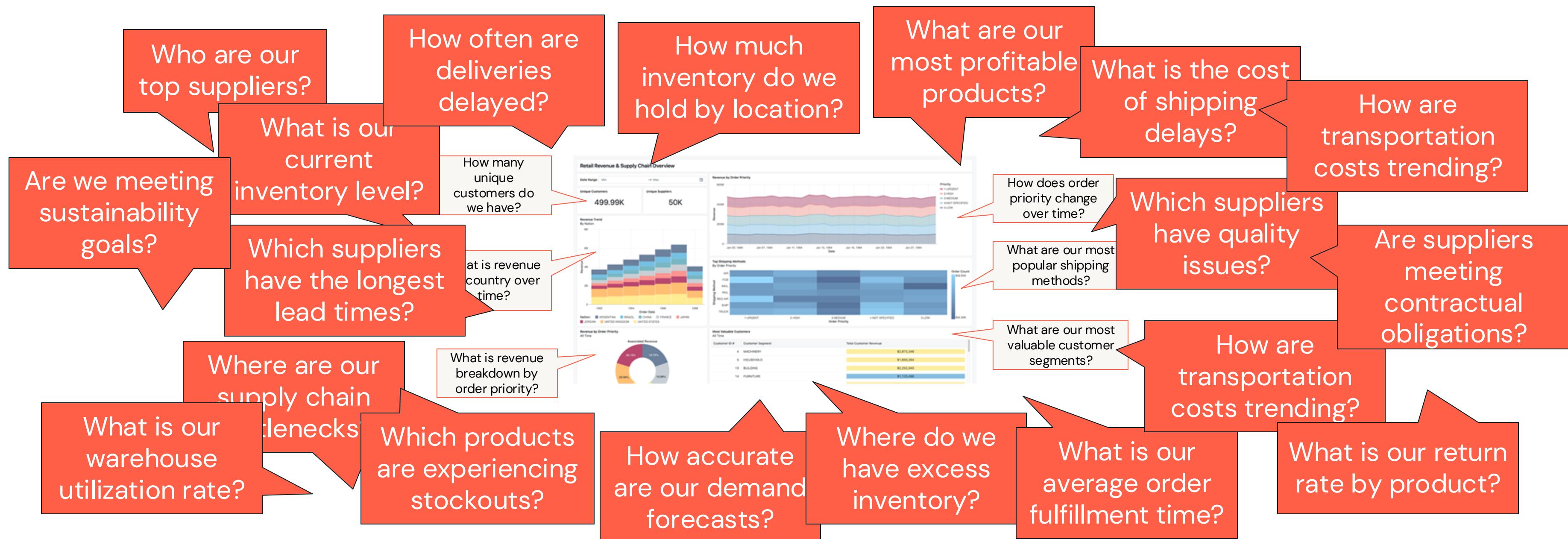
Scale BI to everyone with no additional cost

Blazing fast results

Secure by design with Unity



# But, you can't anticipate every question users will ask



# Genie lets you go beyond dashboards

Talk to your data to explore any question

The screenshot shows the Genie interface. At the top, there's a navigation bar with buttons for '+ New chat', 'History', 'Configure', 'Monitoring', 'Share', and more. Below the navigation bar, the title 'Sales Data' is displayed. On the left side, there's a sidebar with a 'Sales Data' icon and a list of sample questions: 'Explain the data set', 'What is the average number of units purchased by customers?', 'What are the most common loyalty segments among customers?', and 'How many customers are there in each state?'. Below the sidebar is a large input field labeled 'Ask your question...' with a placeholder 'Ask your question...'. At the bottom of the input field, a note says 'Always review the accuracy of responses.'

**Query your enterprise data directly with open Q&A**

**Maintain context between sessions to follow-up on research**

**Get responses in natural language, data tables, and visualizations**



# What is data intelligence for AI applications?





# Mosaic AI

Create domain-specific  
agentic applications



Your data stored in an open, broadly  
accessible lakehouse format



## General Intelligence

Large models  
trained on the entire web  
leveraging scaling laws

vs

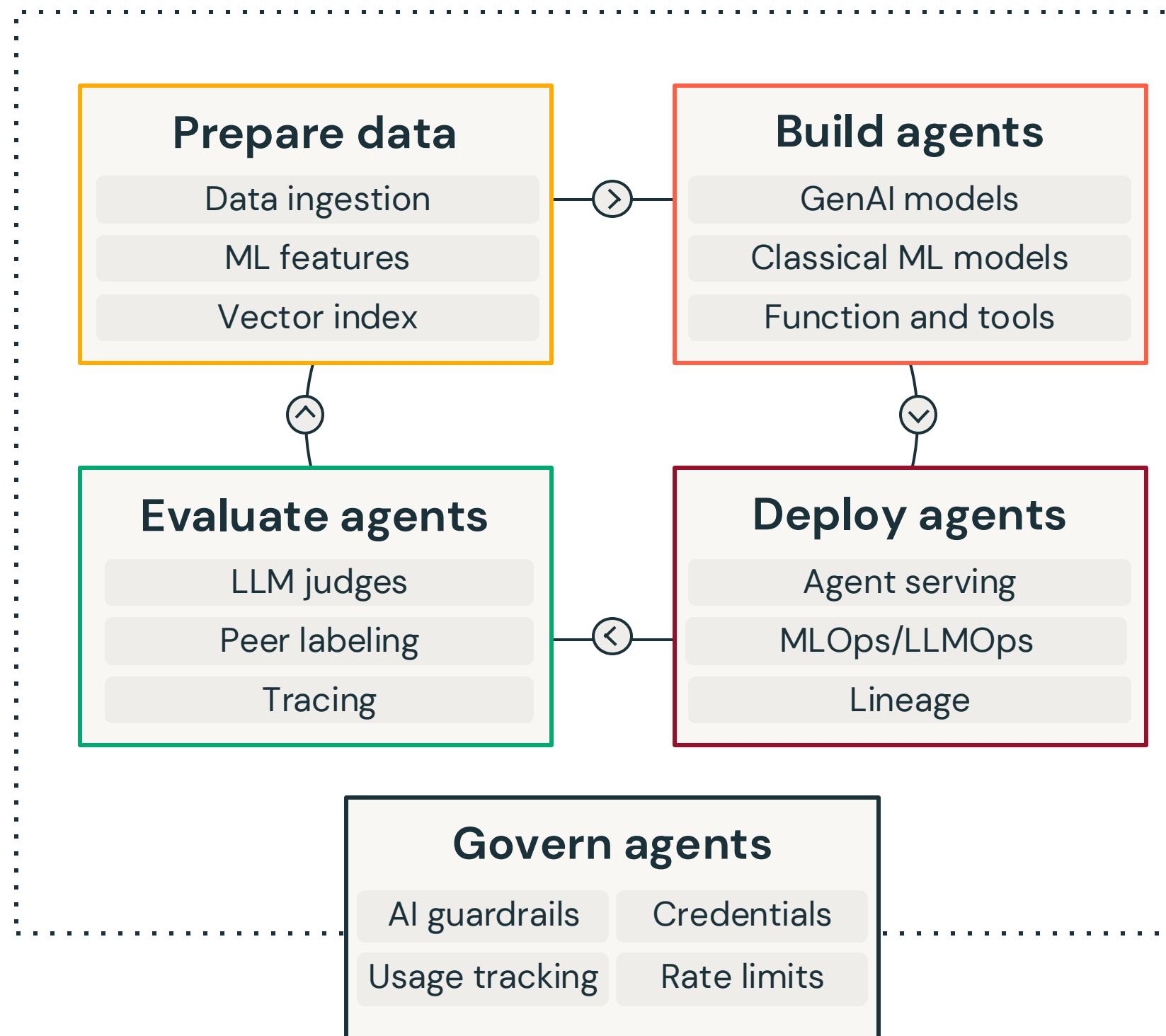
## Data Intelligence

AI agents that reason  
on your data and solve  
domain-specific problems



# Get accurate, domain-specific results

Mosaic AI is the complete agent platform

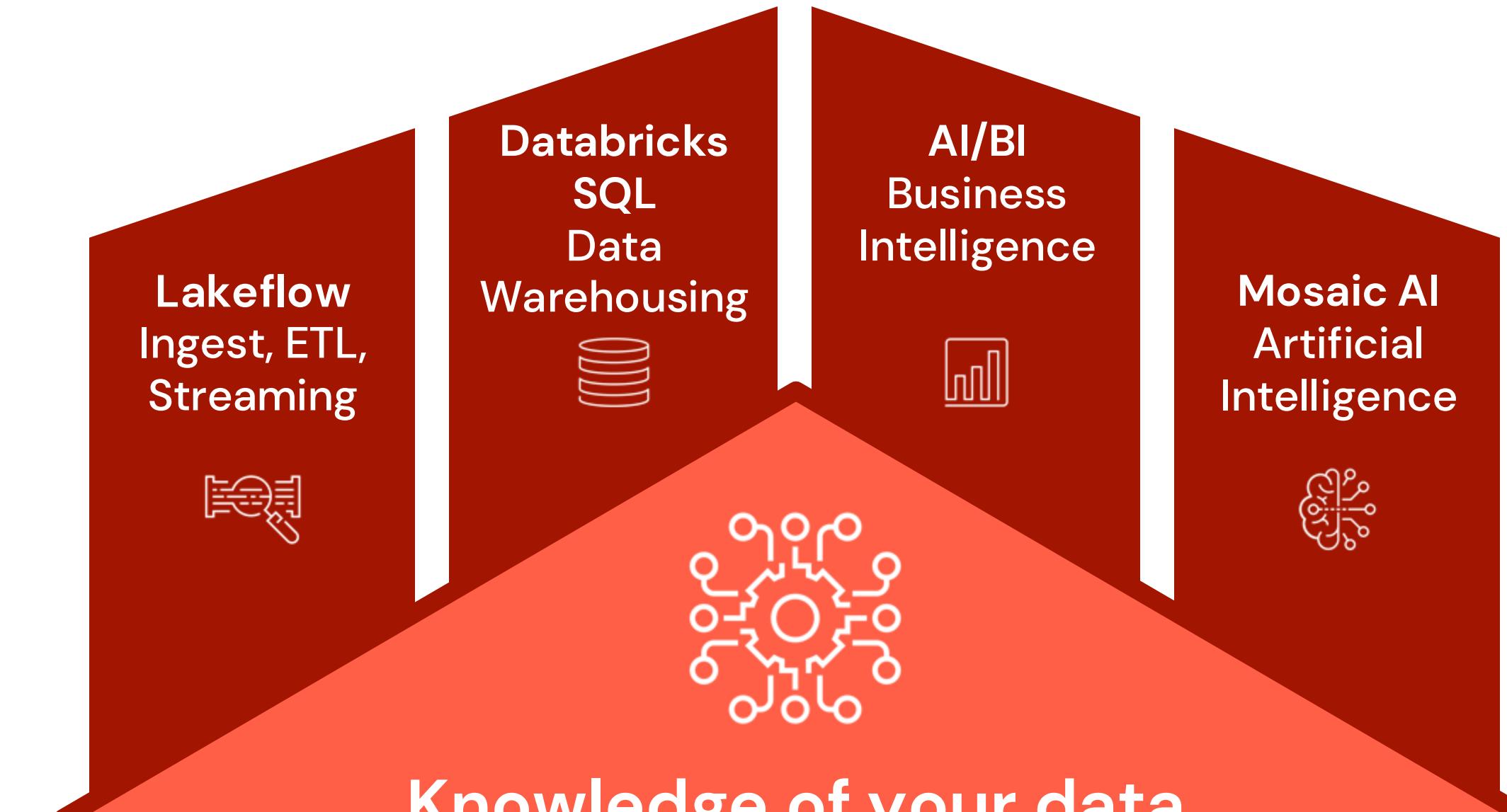


Agents that reason across every enterprise system

Support for all existing and future AI models

Build trust with guardrails, evaluation, and monitoring





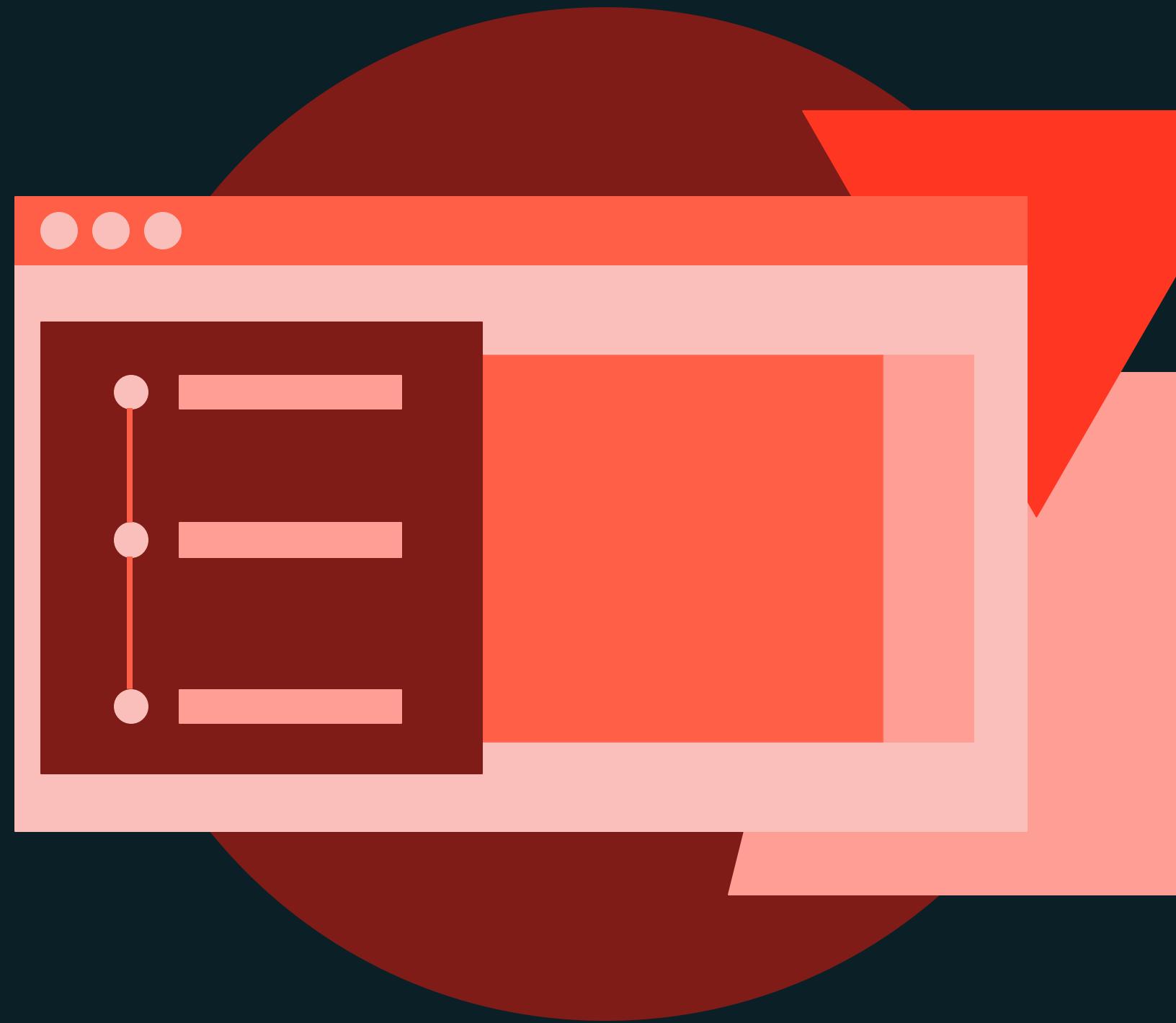
Your data stored in an open, broadly accessible lakehouse format



Databricks Overview

**DEMONSTRATION**

# Databricks Workspace Walkthrough



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Follow-along Instructions

**Estimated Time: 15 minutes**

For this demonstration, the instructor will take you on a guided tour of the platform. If you have access to the Vocareum lab environment, feel free to follow-along to explore the workspace and get oriented with where future labs will be completed.

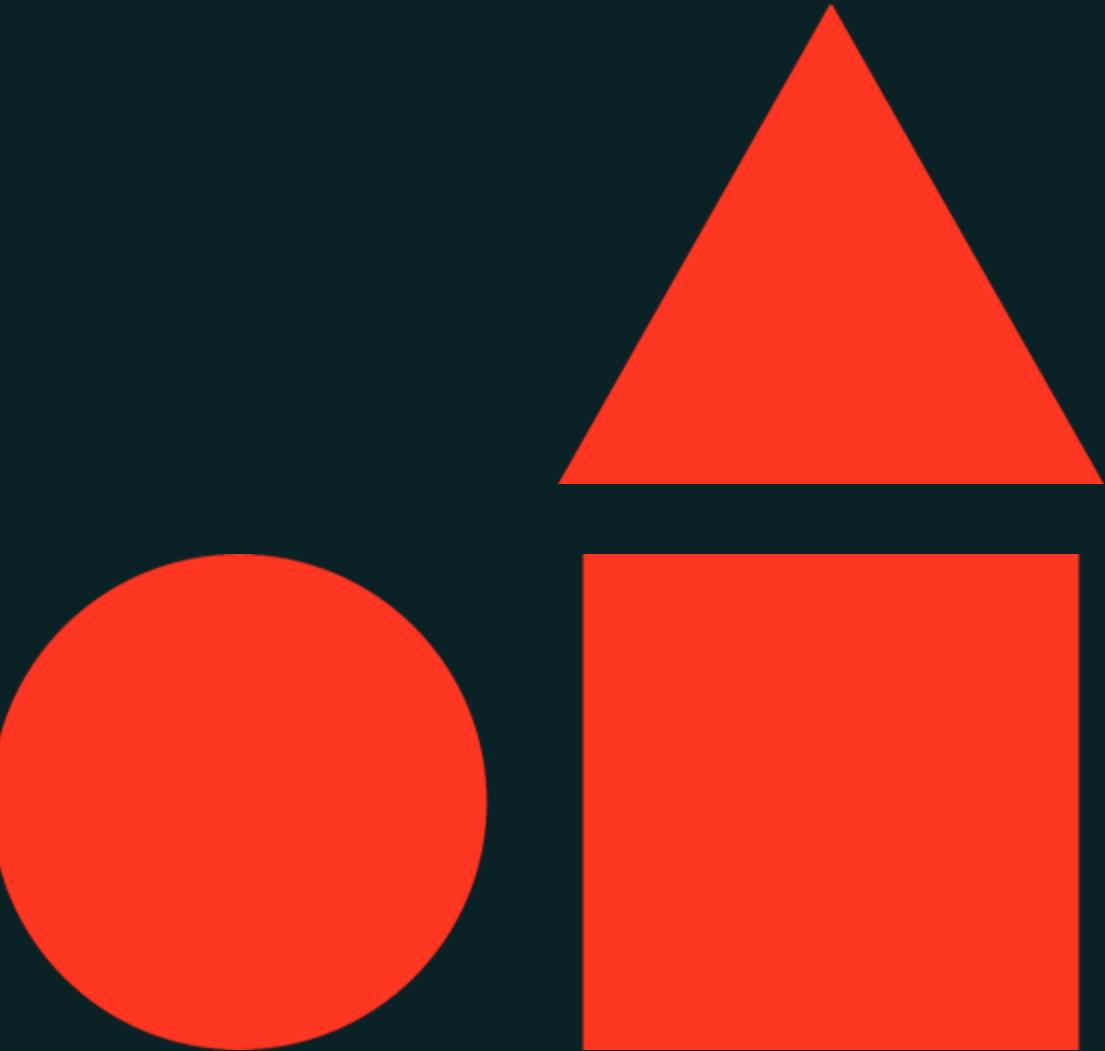




# Using Databricks for Data Warehousing

---

Get Started with Databricks for Data Warehousing



# Objectives

- Use Databricks to complete common data warehousing tasks.
- Explain the purpose of Delta Lake for data warehousing purposes.



# Agenda

Using Databricks for Data Warehousing	Time	Lecture	Demo	Lab
<b>Introduction to Data Warehousing with Databricks</b>	10 mins	✓		
<b>Databricks SQL Warehouses</b>	5 mins	✓		
<b>Delta Lake Overview</b>	8 mins	✓		
<b>Using Delta Lake features with Databricks SQL</b>	5 mins		✓	





Using Databricks for Data Warehousing

LECTURE

# Introduction to Data Warehousing with Databricks



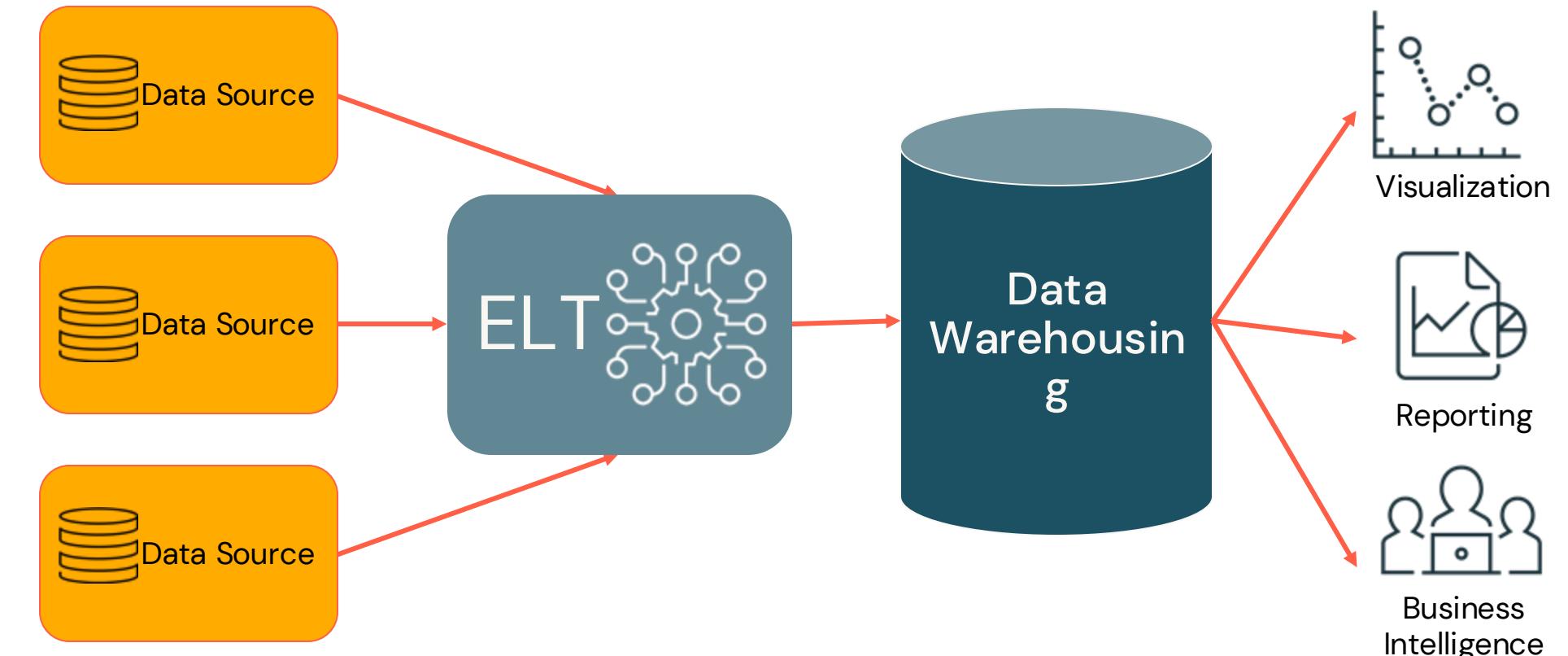
© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Introduction to Data Warehousing

## A Centralized Solution for Reporting, Analytics, and Insights

A data warehouse serves as the foundation for reporting, analytics, and BI, helping organizations make informed decisions.

- **Centralized Storage:** Integrates data from multiple sources (databases, logs, etc.).
- **ETL Process:** Data is extracted, cleaned, and organized for analysis.
- **Optimized for Insights:** Powers visualization, reporting, and BI tools.



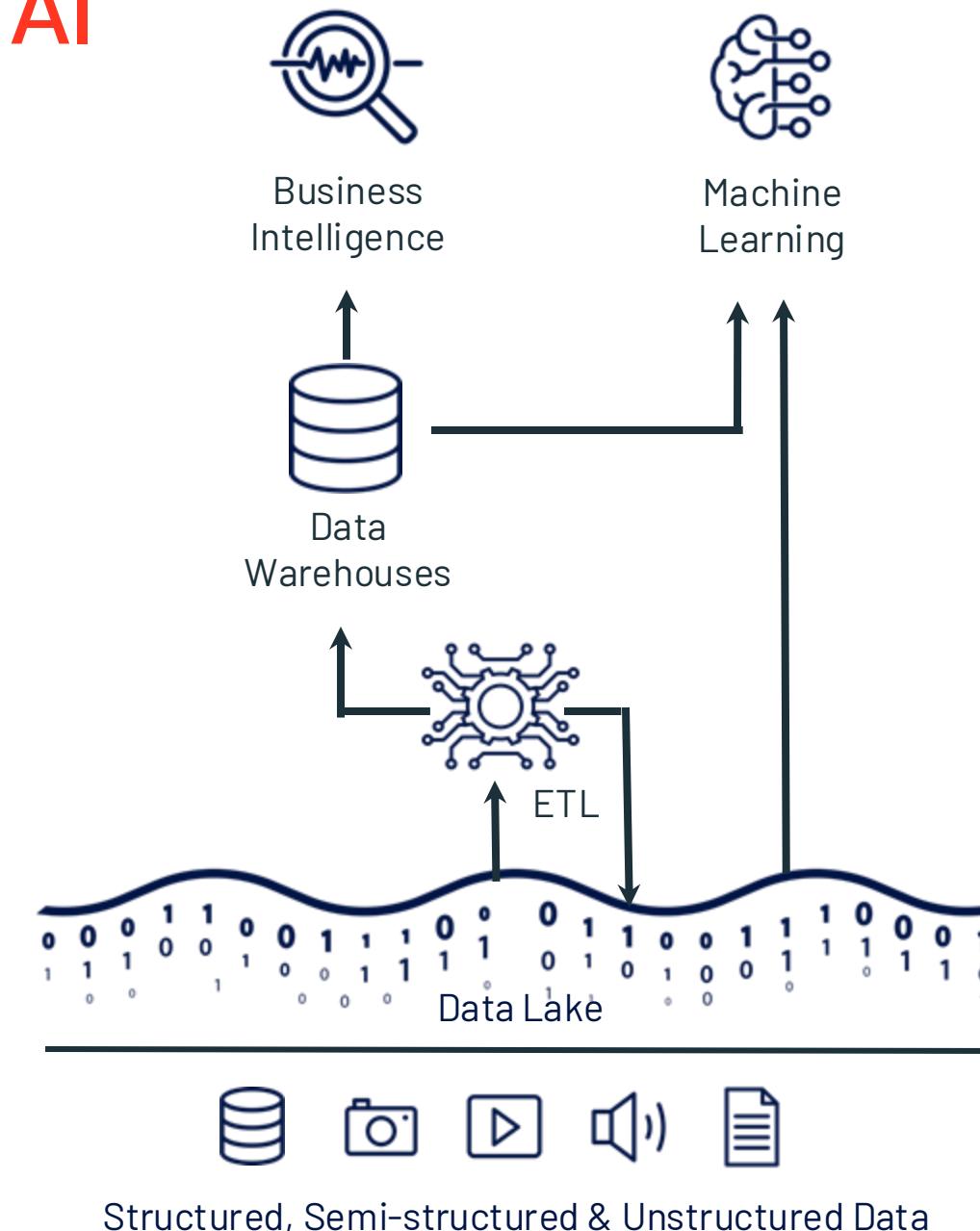
*Data warehousing professionals ensure the ETL process runs smoothly and data is centralized, reliable, and optimized for reporting and insights.*



# Why Data Warehousing Matters

Unified data powers analytics, reporting, and AI

- **Consolidates Data from Disparate Sources**
- **Improves Data Quality and Consistency**
- **Enables Reporting, Analytics, and BI**
- **Supports Better Decision-Making**



Unified data enables faster insights and better decisions, driving organizational success.



# Challenges in Data Warehousing

Data warehousing can be complex and expensive.



Users need specialized technical skills



Complex infrastructure



Data silos and fragmented systems



Issues with data governance and compliance



Slow performance at scale



Inconsistent data quality and access

- Users must have specialized knowledge to operate and manage traditional data warehouses.
- Talent is hard to obtain and retain as the industry is highly competitive.
- Expensive to train up these skills to have them lost to other companies.



# Challenges in Data Warehousing

Data warehousing can be complex and expensive.



Users need specialized  
technical skills



Complex  
infrastructure



Data silos and fragmented  
systems



Issues with data  
governance and  
compliance



Slow performance  
at scale



Inconsistent data quality  
and access

- Redundant tooling within the data warehouse make it difficult to achieve consistency
- Systems and processes become inefficient over time and need to be frequently adjusted to maintain performance requirements
- Maintenance difficulties arise with easily outdated foundational components



# Challenges in Data Warehousing

Data warehousing can be complex and expensive.



Users need specialized technical skills



Complex infrastructure



Data silos and fragmented systems



Issues with data governance and compliance



Slow performance at scale



Inconsistent data quality and access

- Data is spread across multiple systems in the ecosystem, fragmenting management
- Difficulty in consolidating insights and identifying accurate and clean data
- Delayed decision-making as processes are lengthy and heavily manual



# Challenges in Data Warehousing

Data warehousing can be complex and expensive.



Users need specialized  
technical skills



Complex  
infrastructure



Data silos and fragmented  
systems



Issues with data  
governance and  
compliance



Slow performance  
at scale



Inconsistent data quality  
and access

- Lack of centralized access controls in traditional platforms
- Challenges in ensuring compliance with regulations (GDPR, HIPAA)
- Limited capabilities to track data lineage and challenges with tackling quality issues



# Challenges in Data Warehousing

Data warehousing can be complex and expensive.



Users need specialized technical skills



Complex infrastructure



Data silos and fragmented systems



Issues with data governance and compliance



Slow performance at scale



Inconsistent data quality and access

- Runaway increase in cost for cloud compute as the data warehouses scale
- Degraded performance as the amount of data increases from small initial testing
- Lack of access in traditional on-premise data warehouses to the latest technologies that impact performance capabilities



# Challenges in Data Warehousing

Data warehousing can be complex and expensive.



Users need specialized technical skills



Complex infrastructure



Data silos and fragmented systems



Issues with data governance and compliance



Slow performance at scale



Inconsistent data quality and access

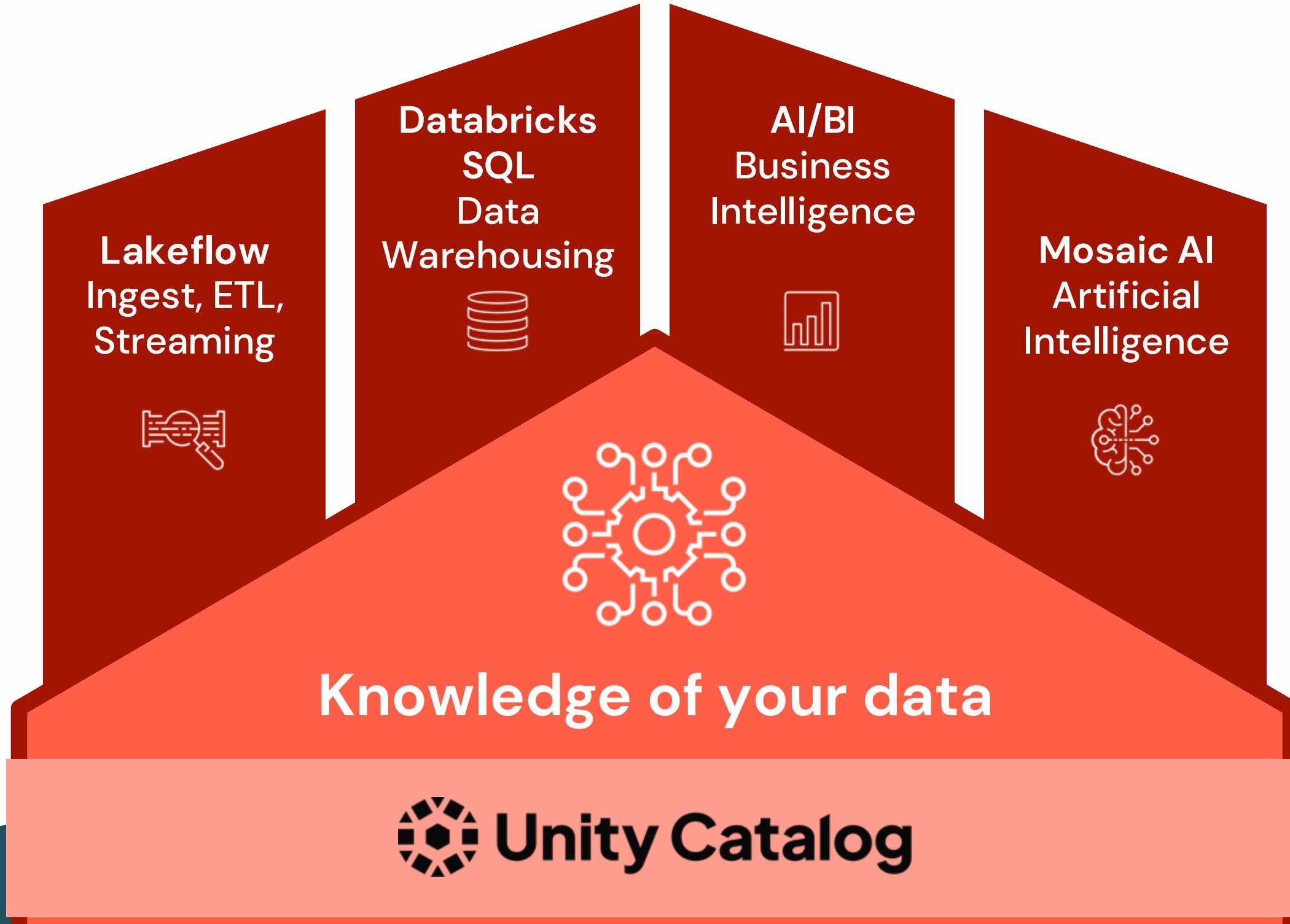
- Inconsistent, incomplete, or outdated data persists within the environment
- Limited access for business users leads to manual overhead to process requests for access
- Complex processes requiring manual oversight to clean and validate data persist regardless of the volume of data in the warehouse



**With Databricks, data warehousing  
becomes simpler, faster, and smarter  
– enabling you to overcome  
traditional challenges.**



# Databricks Data Intelligence Platform



Your data stored in an open, broadly  
accessible lakehouse format



# What Makes Databricks Different?

## Databricks Data Warehousing: Simplified, Centralized, and Powerful



### Unified Data Management:

- Combines structured and semi-structured data using Delta Tables for centralized storage.



### Transform with Delta:

- Leverages Databricks' compute for ETL and real-time transformations.



### Powered by Unity Catalog:

- Ensures governance, metadata management, and seamless collaboration across teams.



### Optimized Outputs:

- Enables data marts, machine learning models, and applications for actionable insights.



# What Makes Databricks Different?

## Databricks Data Warehousing: Simplified, Centralized, and Powerful



### Unified Data Management:

- Combines structured and semi-structured data using Delta Tables for centralized storage.



### Transform with Delta:

- Leverages Databricks' compute for ETL and real-time transformations.



### Powered by Unity Catalog:

- Ensures governance, metadata management, and seamless collaboration across teams.



### Optimized Outputs:

- Enables data marts, machine learning models, and applications for actionable insights.



# What Makes Databricks Different?

## Databricks Data Warehousing: Simplified, Centralized, and Powerful



### Unified Data Management:

- Combines structured and semi-structured data using Delta Tables for centralized storage.



### Transform with Delta:

- Leverages Databricks' compute for ETL and real-time transformations.



### Powered by Unity Catalog:

- Ensures governance, metadata management, and seamless collaboration across teams.



### Optimized Outputs:

- Enables data marts, machine learning models, and applications for actionable insights.



# What Makes Databricks Different?

## Databricks Data Warehousing: Simplified, Centralized, and Powerful



### Unified Data Management:

- Combines structured and semi-structured data using Delta Tables for centralized storage.



### Transform with Delta:

- Leverages Databricks' compute for ETL and real-time transformations.



### Powered by Unity Catalog:

- Ensures governance, metadata management, and seamless collaboration across teams.



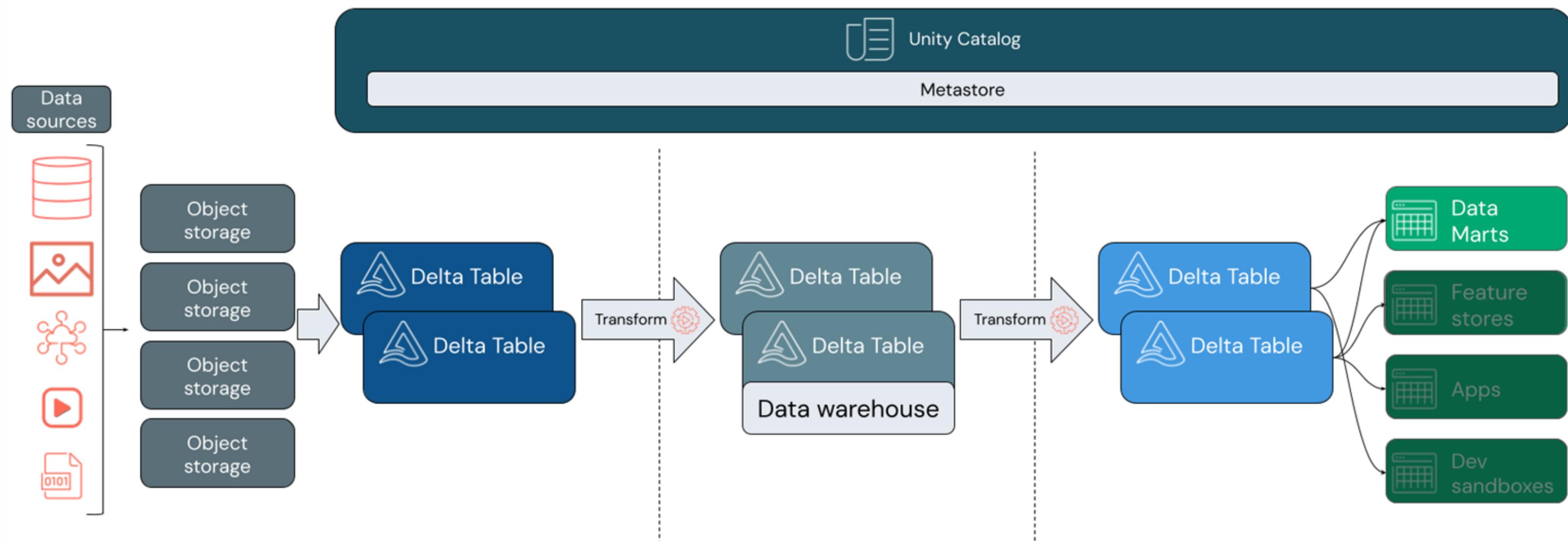
### Optimized Outputs:

- Enables data marts, machine learning models, and applications for actionable insights.



# Databricks Approach to Data Warehousing

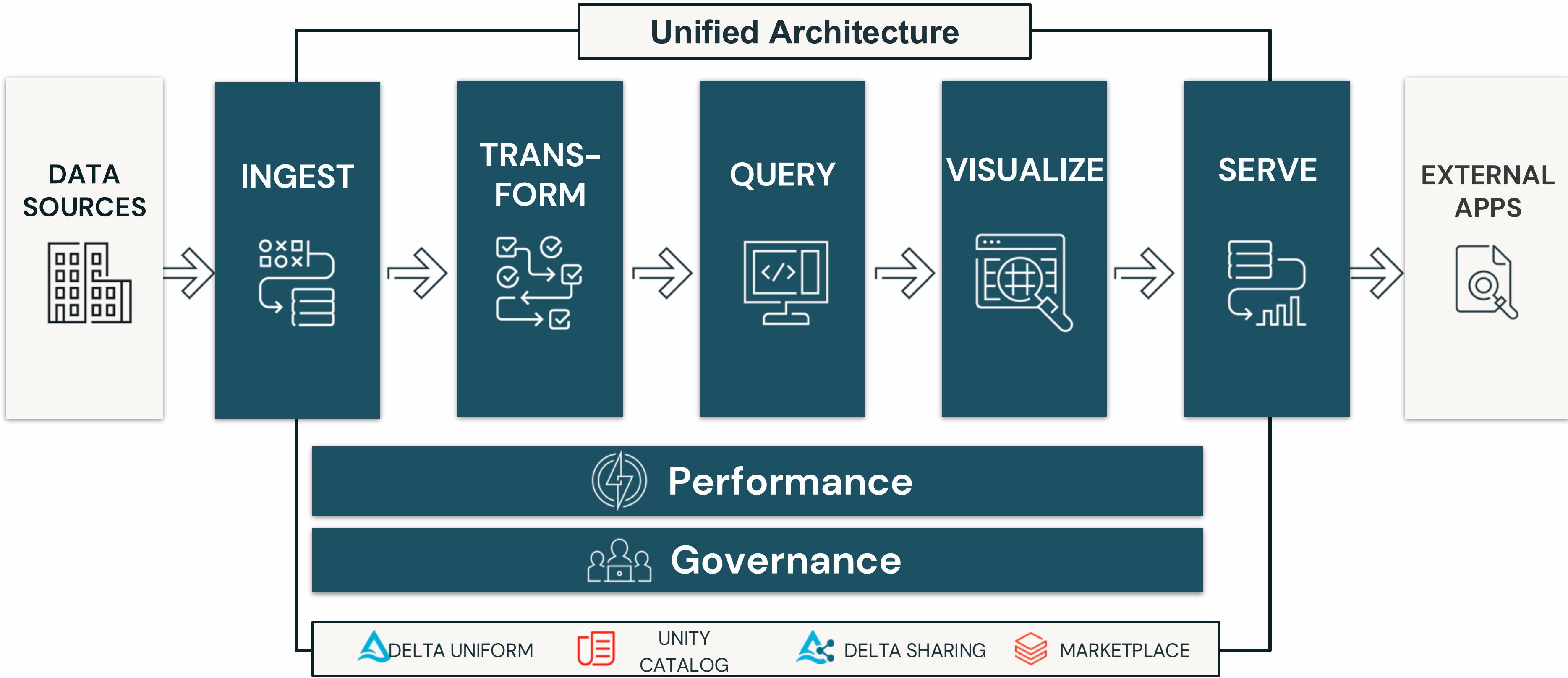
From raw data to insights: Leveraging Delta Tables and Unity Catalog.



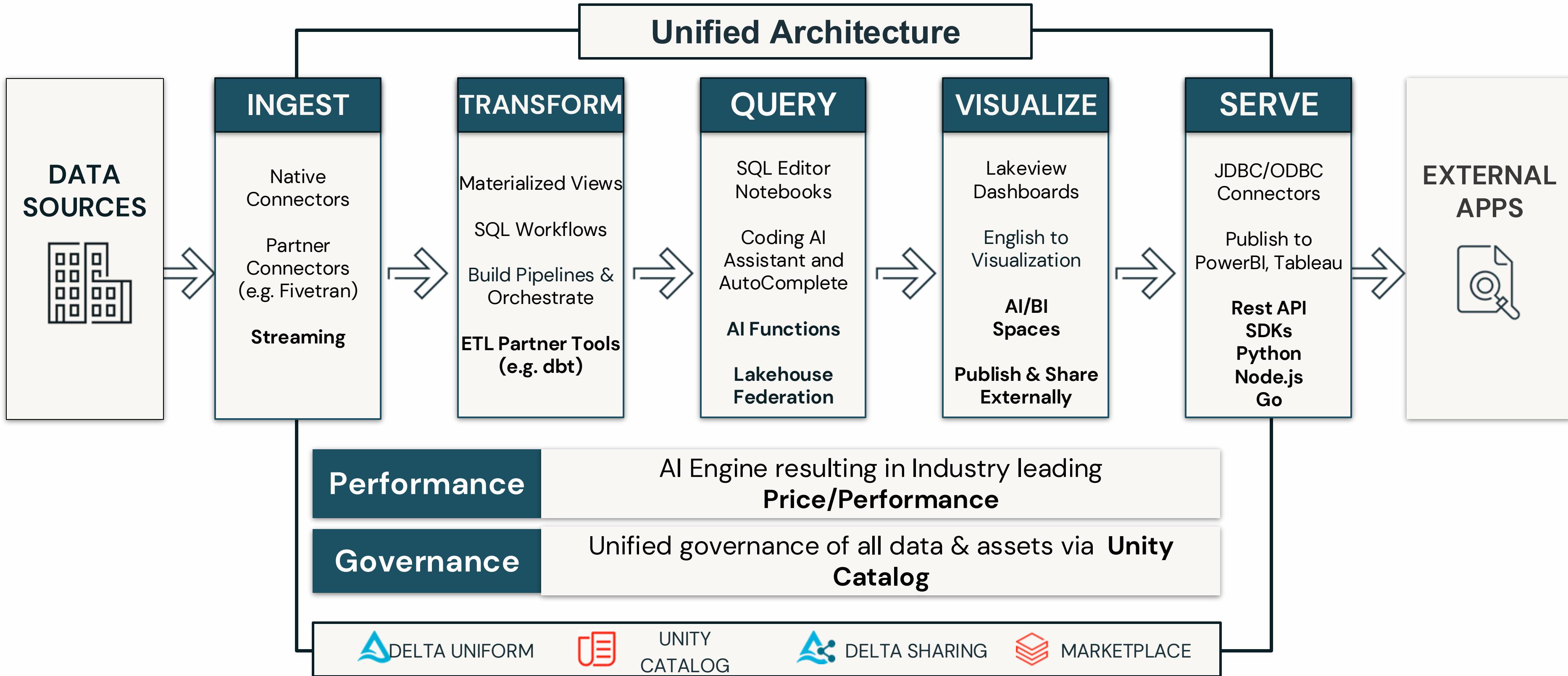
*Databricks combines unified storage, real-time transformations, and governance to enable smarter data-driven insights.*



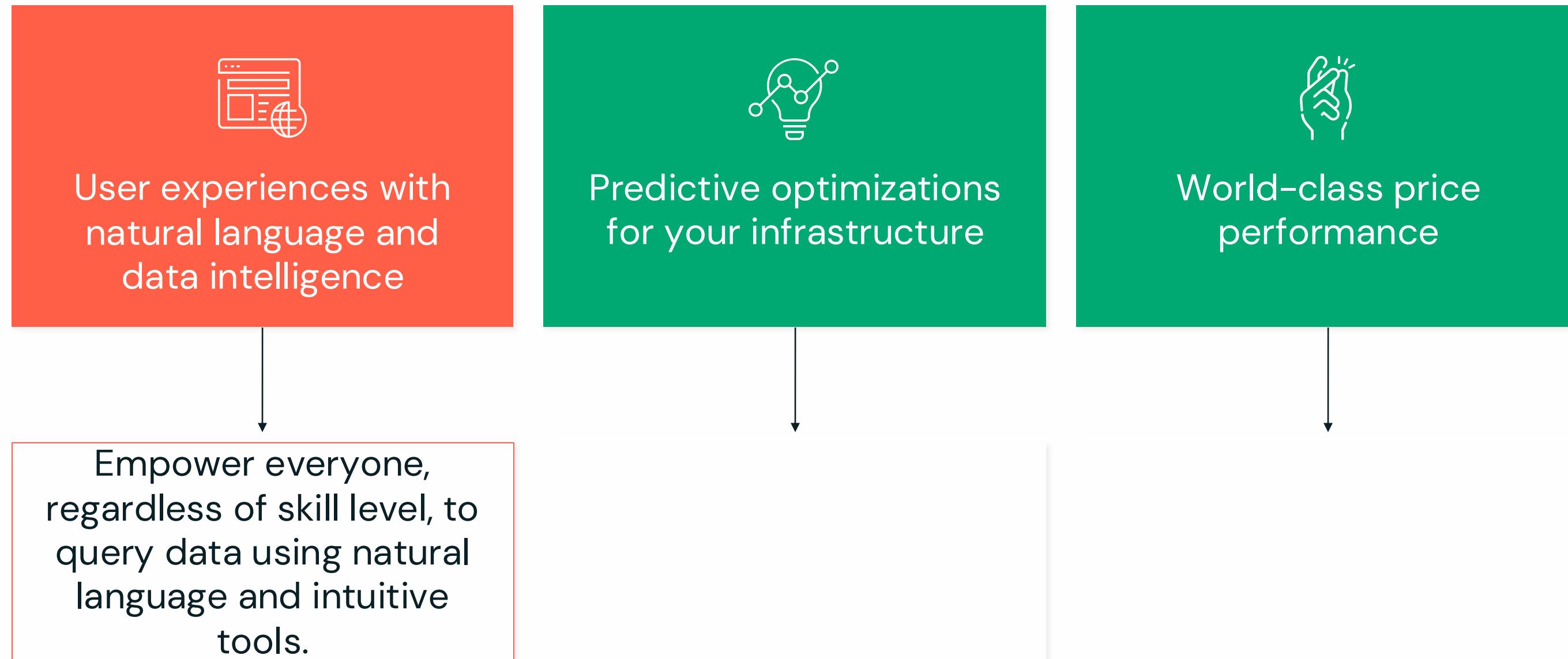
# Complete Data Warehousing Solution



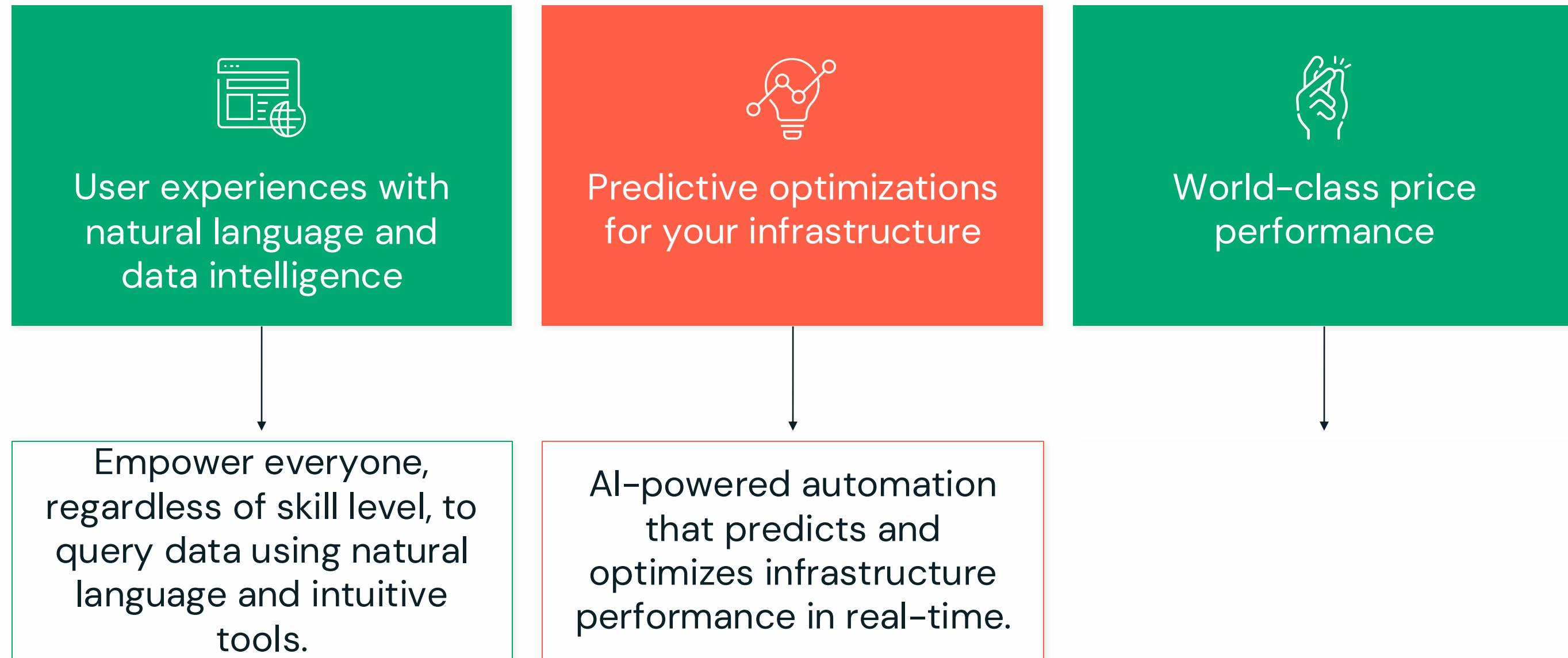
# Complete Data Warehousing Solution



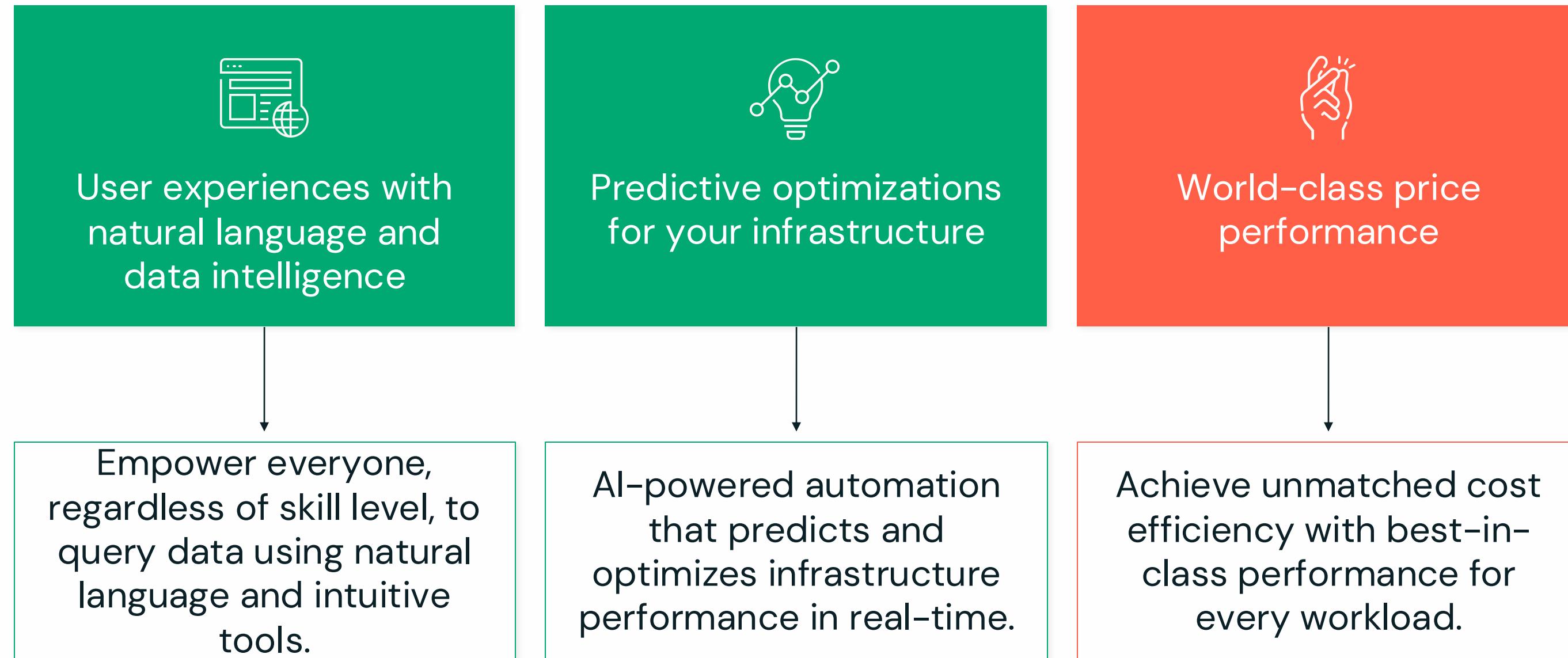
# The Future of Data Warehousing: Smarter, Faster, and AI-Powered



# The Future of Data Warehousing: Smarter, Faster, and AI-Powered

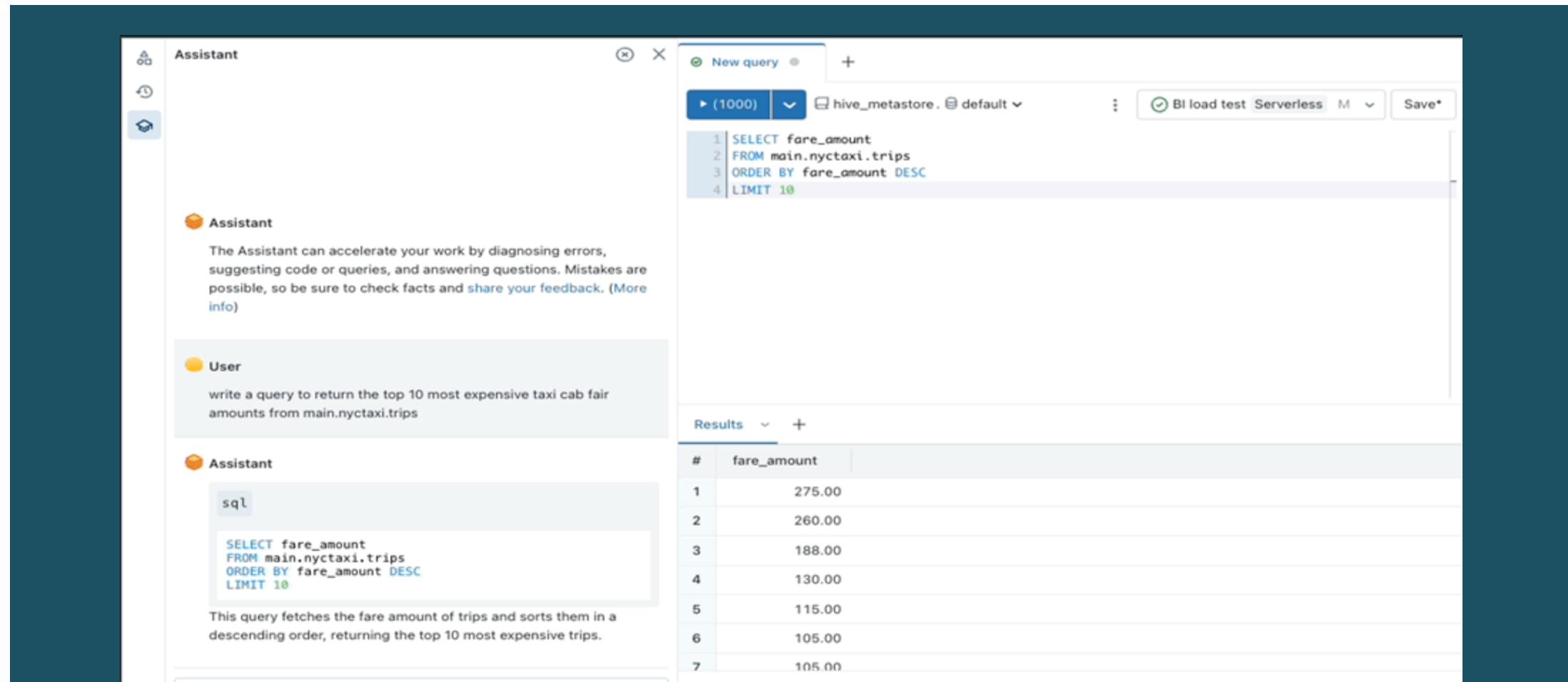


# The Future of Data Warehousing: Smarter, Faster, and AI-Powered



# Tools for Intelligent Data Warehousing

Powered by Intelligence Engine



The screenshot shows the Databricks Unified SQL Editor interface. At the top, there's a navigation bar with 'Assistant' and other icons. Below it is a search bar with '(1000)' and a dropdown for 'hive\_metastore.default'. To the right of the search bar are buttons for 'BI load test', 'Serverless', and 'Save\*'. The main area has a code editor with the following SQL query:

```
1 SELECT fare_amount
2 FROM main.nyctaxi.trips
3 ORDER BY fare_amount DESC
4 LIMIT 10
```

Below the code editor is a 'User' section with a prompt: 'write a query to return the top 10 most expensive taxi cab fair amounts from main.nyctaxi.trips'. An 'Assistant' section below it provides a suggested query:

```
sql
SELECT fare_amount
FROM main.nyctaxi.trips
ORDER BY fare_amount DESC
LIMIT 10
```

A note explains: 'This query fetches the fare amount of trips and sorts them in a descending order, returning the top 10 most expensive trips.' At the bottom of the interface, there's a message input field 'Enter your message' and a status bar indicating '357 ms | 10 rows returned' and 'Refreshed just now'.

## Unified SQL Editor

Data Science / Engineers



# Tools for Intelligent Data Warehousing

Powered by Intelligence Engine



**Sales Opportunity Overview**

Deals created between Mar 12, 2024 and Jun 12, 2024

Avg days to close: 29.347 | Avg probability of close: 73.854 | Avg active opp size: 82.29K | Avg cust annual rev: 246.78M

Percentage of opportunities by stage:

- 5. Closed Won: 60.76%
- 1. Discovery: 13.54%
- 3. Validation: 7.84%
- 4. Procure: 7.80%

Opportunity by market region size:

Region	Sum of opportunity amount
AMER	~6.5M
APAC	~5.8M
EMEA	~5.2M
LATAM	~4.5M

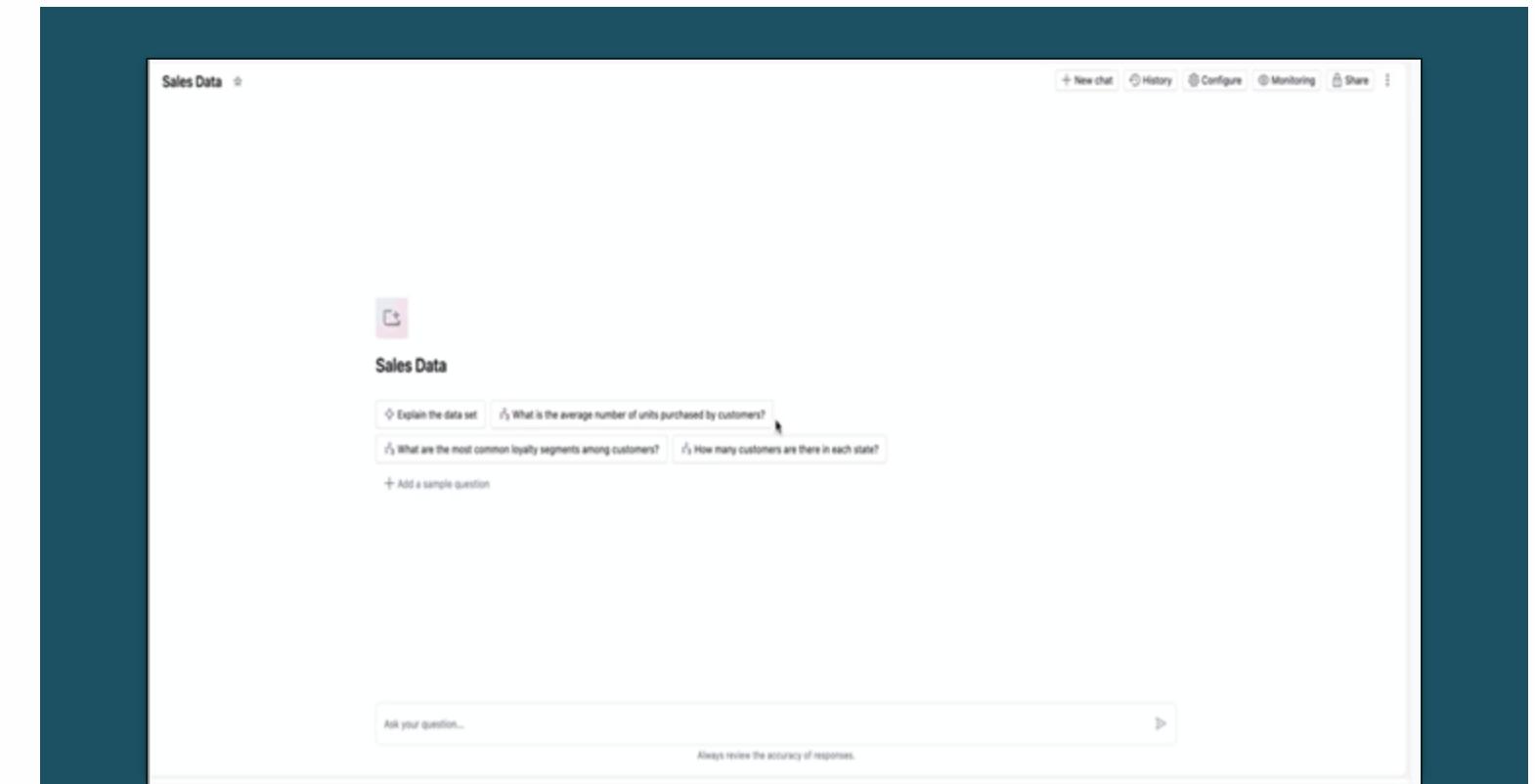
Opportunity amount by customer annual revenue:

Customer annual revenue (\$USD)

Scatter plot showing opportunity amount vs customer annual revenue.

AI/BI Dashboards

Analysts



**Sales Data**

New chat History Configure Monitoring Share

Sales Data

Explain the data set

What is the average number of units purchased by customers?

What are the most common loyalty segments among customers?

How many customers are there in each state?

Add a sample question

Ask your question... Always review the accuracy of responses.

AI/BI Genie

Business Users



# Tools for Intelligent Data Warehousing

Powered by Intelligence Engine

The diagram illustrates the Databricks Partner Connect ecosystem, structured into several layers:

- Top Layer (Partner for everything on top):** Data Ingestion (Fivetran, Airbyte, Rivery, CONFLUENT, arcion, Qlik), Data Pipelines (MATILLION, dbt, Informatica, Prophecy, PRIVACERA, Alation), Data Governance (Collibra, MMUTA), BI & Dashboards (ThoughtSpot, tableau, Looker, Qlik), Data Science (PyCharm, Power BI, R Studio, jupyter, HEX), and Machine Learning (H2O.ai, Labelbox, Azure Machine Learning, Amazon SageMaker, John Snow LABS).
- Middle Layer (Data Processing Engines):** Photon, ML Runtime.
- Foundation Layer (Cloud Data Lake):** Delta Lake, Unity Catalog, Cloud Data Lake (All structured and unstructured data) supported by Microsoft Azure, AWS, and Google Cloud.
- Bottom Layer (Lakehouse foundation):** Own the lakehouse foundation.

## Partner Connect

Data Science / Engineers / Analysts / Business Users



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).



Using Databricks for Data Warehousing

LECTURE

# Databricks SQL Warehouses



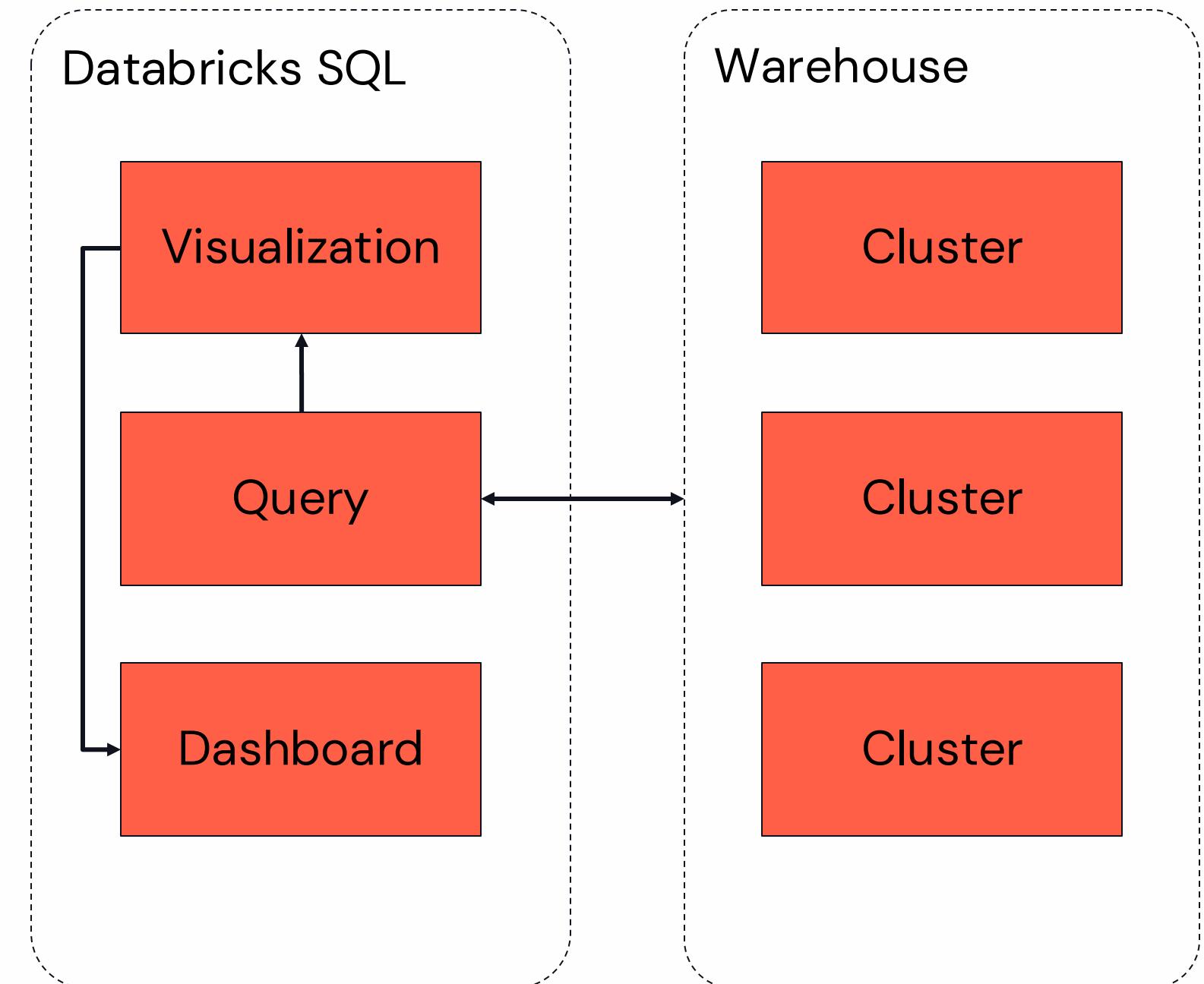
# SQL Warehouses

## Overview

SQL Warehouse in Databricks is a scalable compute resource optimized for executing SQL queries, analyzing data, creating visualizations, and supporting business intelligence (BI) workflows.

### Key Features:

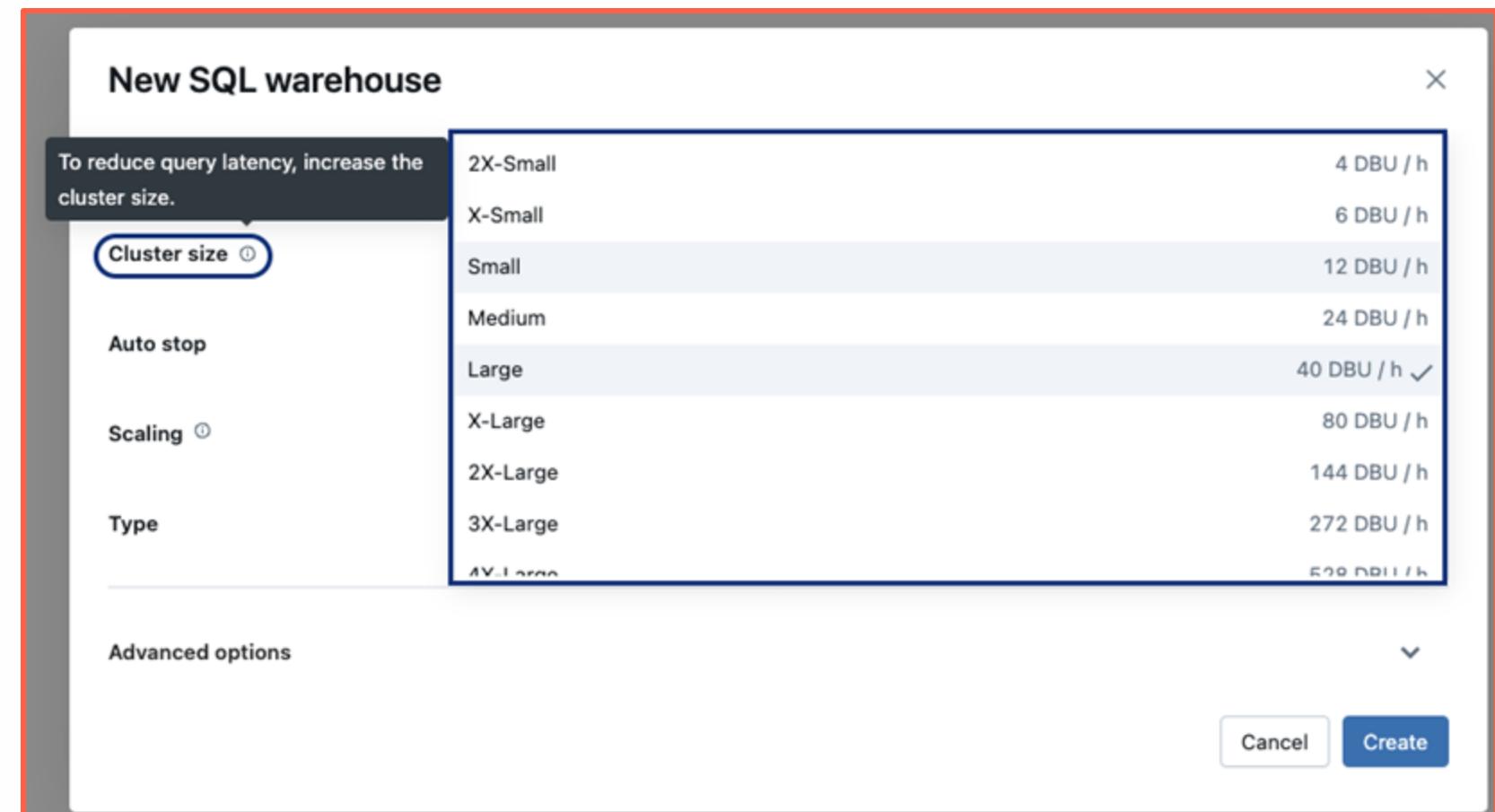
- Scalable compute for distributed SQL workloads.
- Optimized for data queries and BI workflows.
- Supports integration with dashboards and analytics tools.



# SQL Warehouses

## Cluster Size:

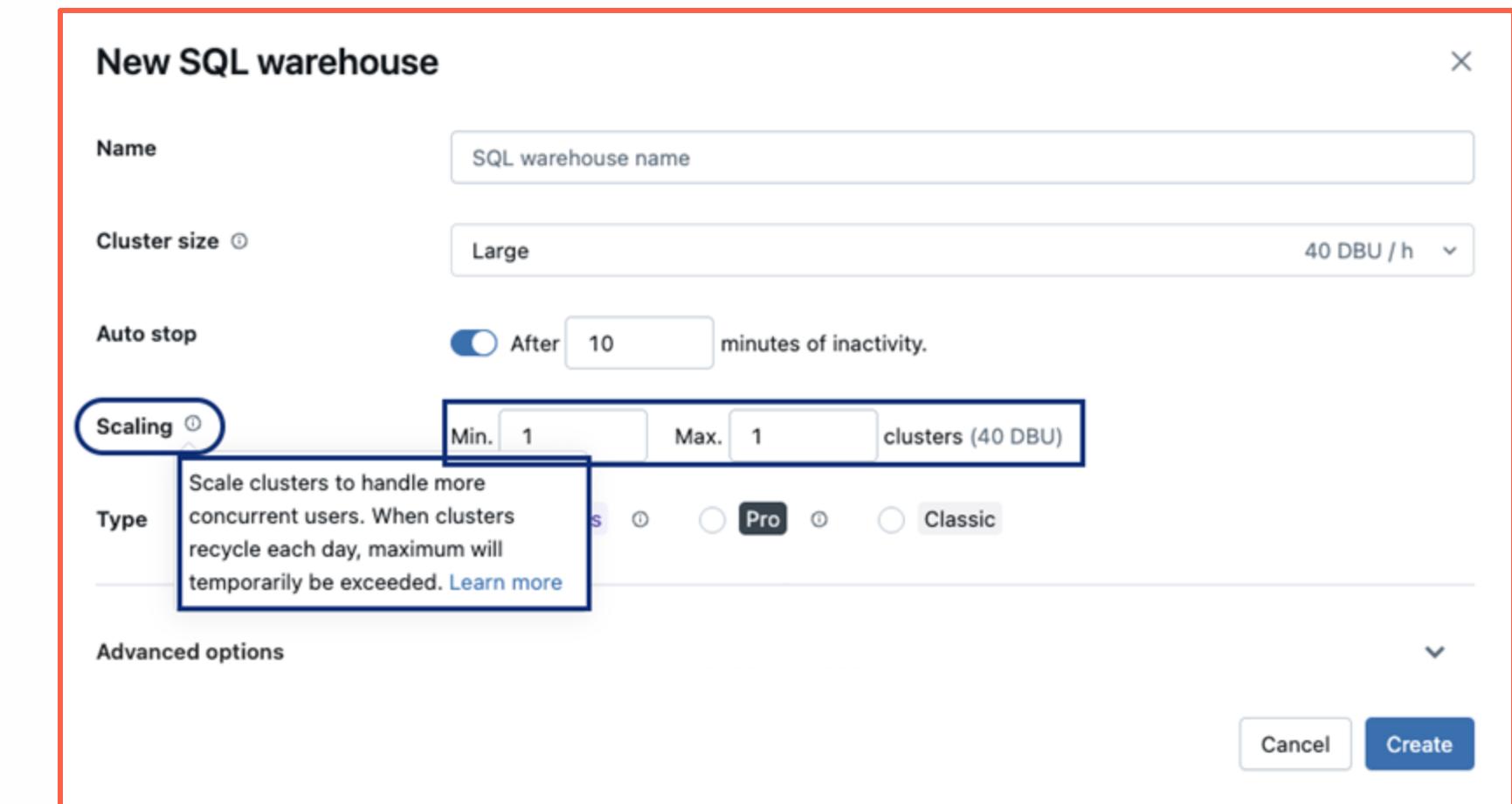
- Refers to the computational power allocated to a SQL Warehouse.
- Expressed in “**t-shirt sizes**” such as **Small**, **Medium**, **Large**, and **X-Large**.
- Each size defines the number of compute nodes and capacity for parallel processing.
- **Example:**
  - **Small:** Suitable for light queries.
  - **X-Large:** Ideal for heavy, high-concurrency workloads.



# SQL Warehouses

## Scaling:

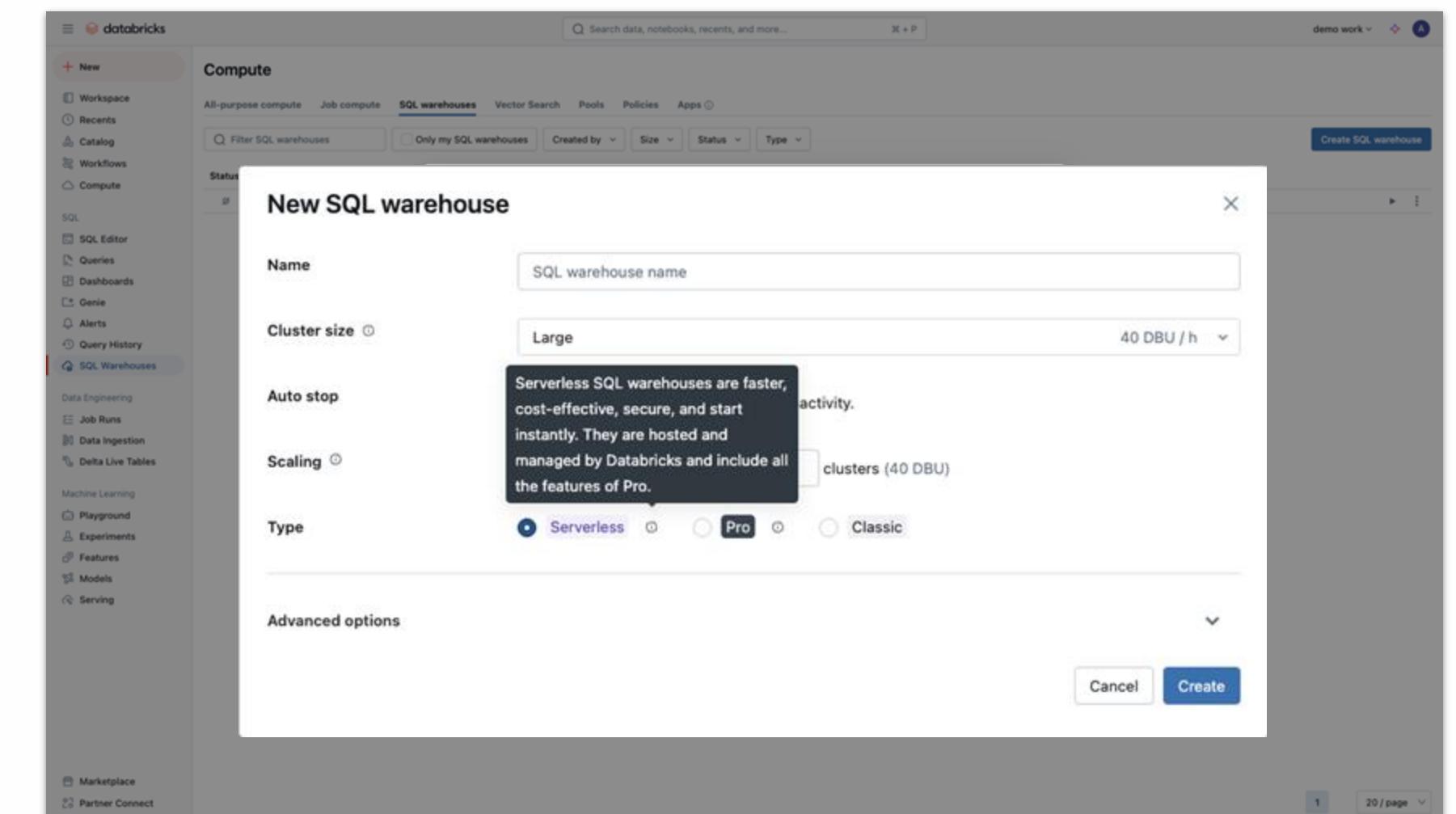
- Ensures SQL Warehouses adapt to varying workloads by dynamically allocating resources.
- **Users can configure:**
  - **Minimum clusters:** Base capacity for consistent performance.
  - **Maximum clusters:** Auto-scale limit to handle peak demand.
- **Benefits:**
  - **Cost Efficiency:** Only pay for resources used.
  - **Performance Optimization:** Prevents resource bottlenecks.



# Serverless Compute for Databricks SQL

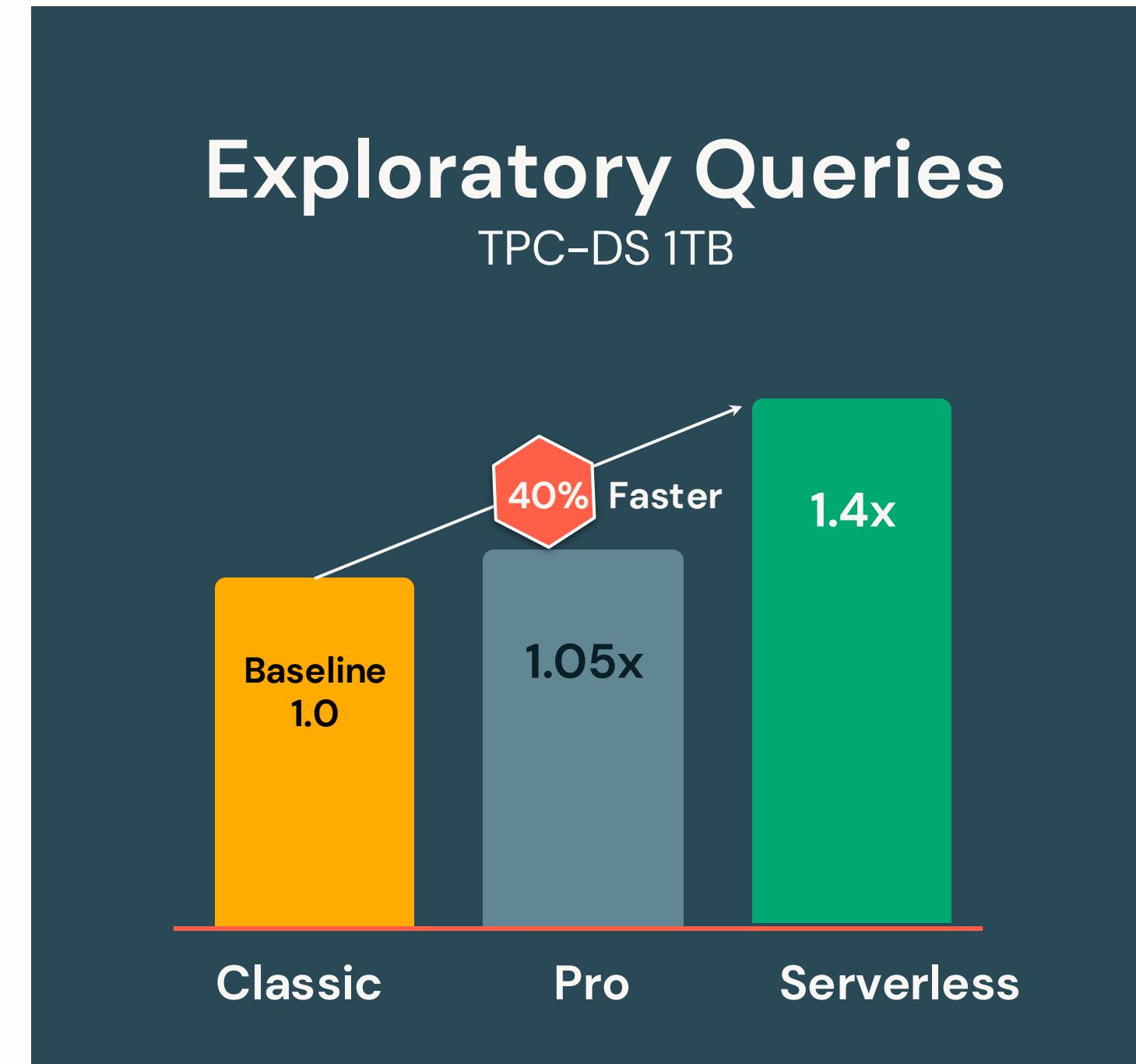
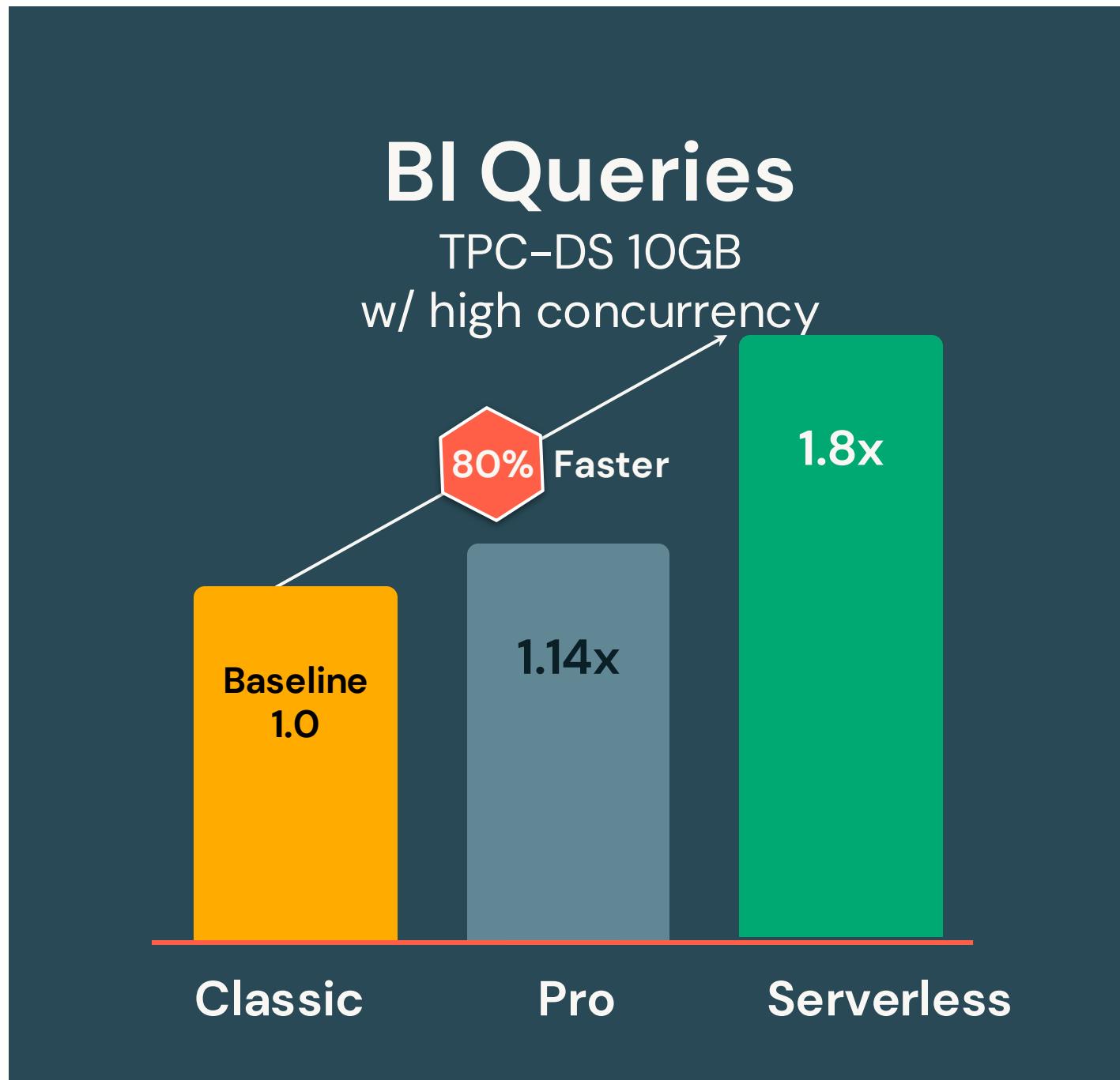
Instant, elastic SQL compute decoupled from storage

- Quickly setup **instant, elastic SQL warehouse** – decoupled from storage – Powered by Photon
- Automatically determines instance types and configuration for best price/perf (up to 12x)
- **High concurrency** built-in, automatic load balancing
- Intelligent **workload management** and faster reads from cloud storage
- Instant startup, greater availability, and **40% average lower overall costs with serverless**



# Serverless SQL has the Best Performance!

Let's see how they perform in real-world scenarios:





Using Databricks for Data Warehousing

LECTURE

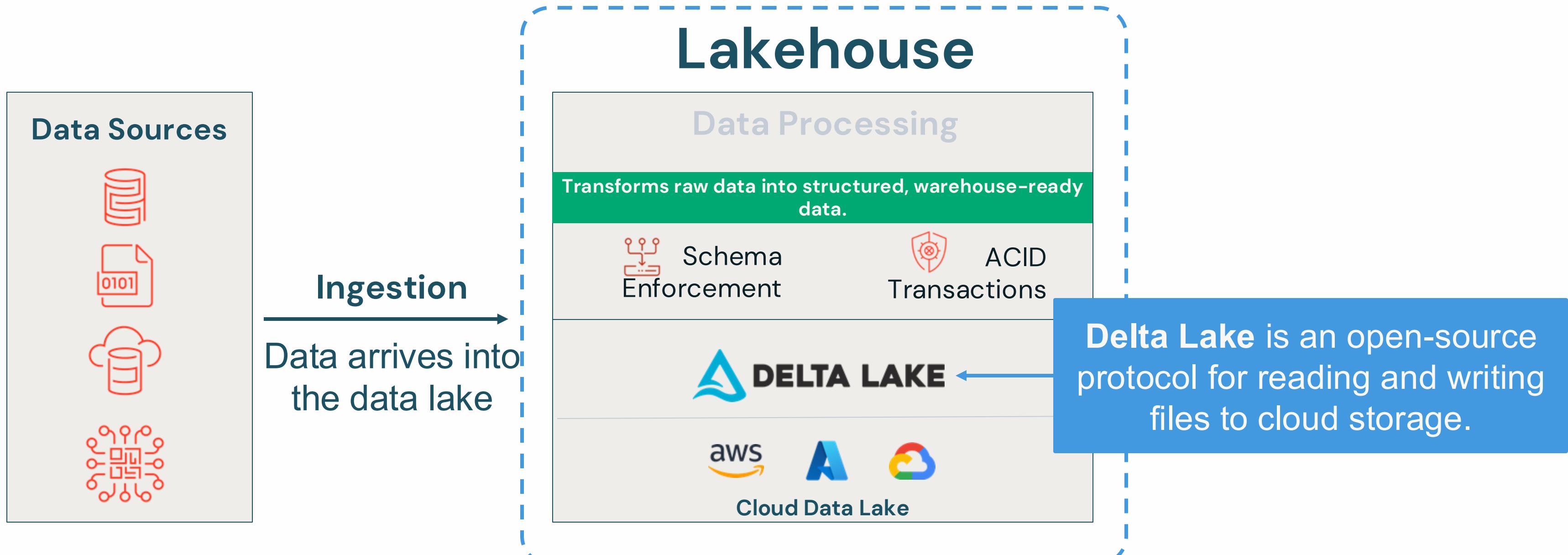
# Delta Lake Overview



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Delta Lake Overview

Delta Lake: Foundation for Modern Data Warehousing.



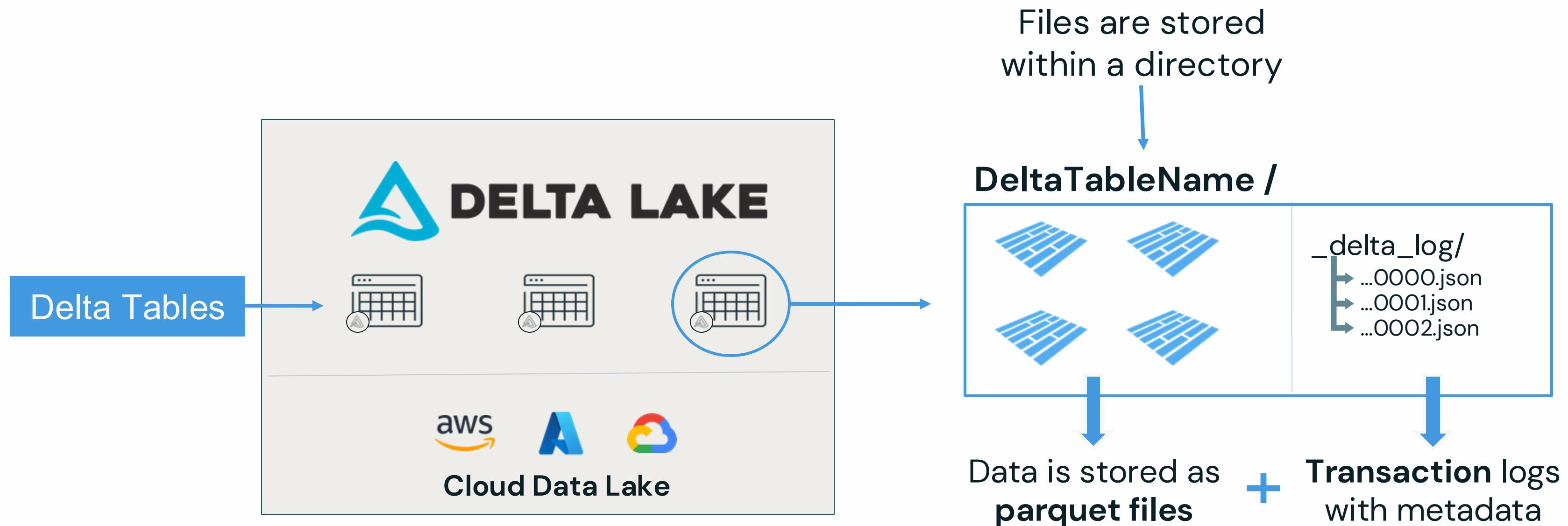
*Delta Lake ensures reliability, performance, and scalability for modern data warehousing workflows.*



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Delta Lake Overview

## Delta Tables: The Backbone of Reliable Data Warehousing



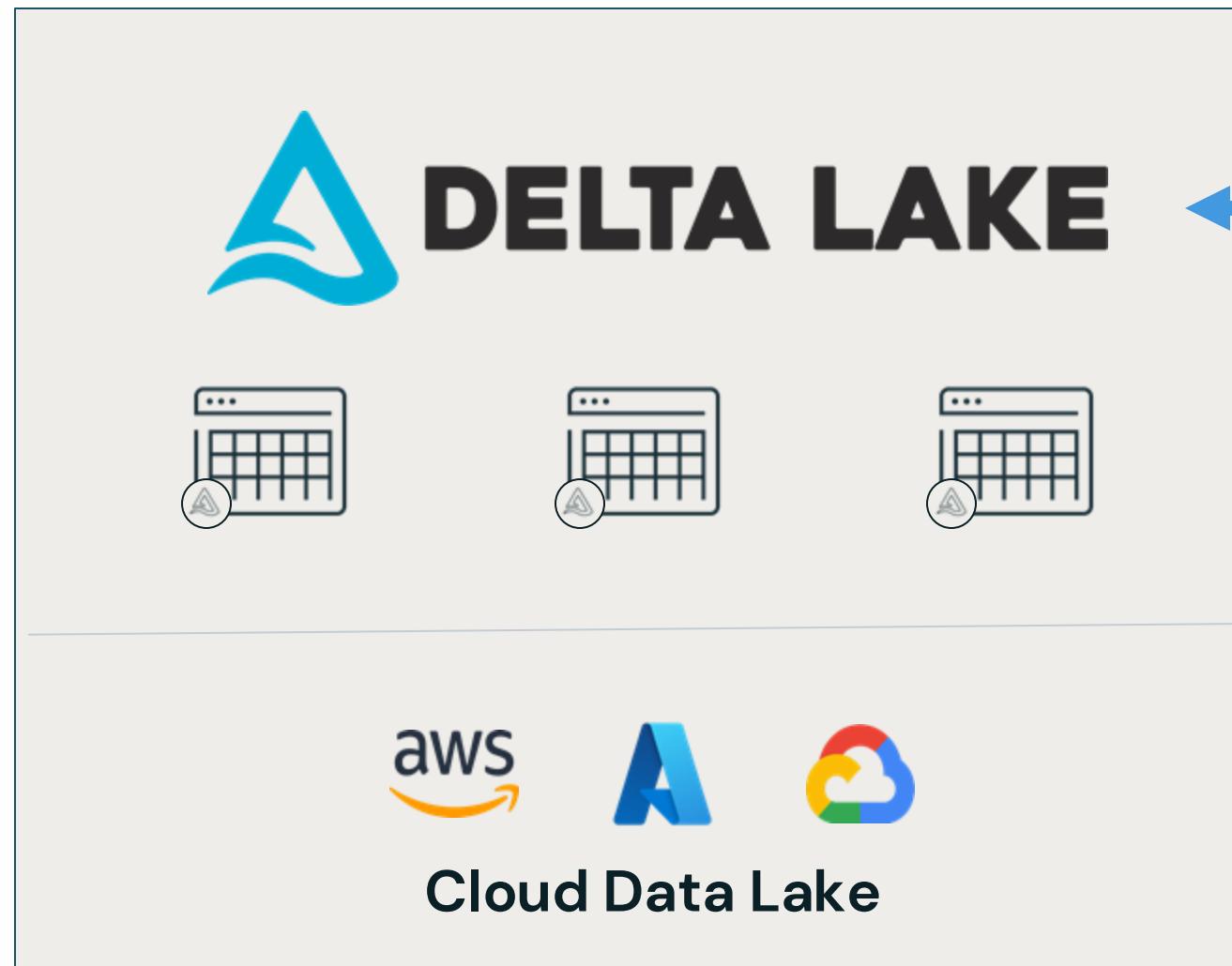
*Delta Tables support both batch and streaming data, making them flexible for real-time and historical analytics in a warehouse.*



© Databricks 2020. All rights reserved. Apache, Apache Spark, Spark, the Spark logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Delta Lake Overview

Delta Lake: Default Format for Databricks Warehousing



**Delta Lake** is the default format for  
tables created in Databricks

**SQL** CREATE TABLE orders ~~USING DELTA~~

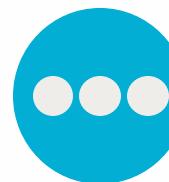
**Python** df.write.format("delta")

*Delta Tables power Serverless SQL Warehouses, providing instant scalability and optimized performance for data warehousing workflows.*



# Delta Lake Overview

## Key Features

 ACID Transactions DML Operations Time Travel Schema Evolution and Enforcement Many more!

# Delta Lake Overview

## Key Features



ACID Transactions



DML Operations



Time Travel



Schema Evolution and  
Enforcement



Many more!

Delta Lake provides ACID transactional guarantees at the table level to data stored in **cloud-based object storage**.

**Atomicity** – Entire transaction completes

**Consistency** – Data follows rules or it will be rolled back

**Isolation** – One transaction completed before the start of another

**Durability** – Data is saved in a persistent state once completed



# Delta Lake Overview

## Key Features



ACID Transactions



DML Operations



Time Travel



Schema Evolution and  
Enforcement



Many more!

## Data Manipulation Language (DML)

- **INSERT** – Add new rows
- **UPDATE** – Update existing values
- **DELETE** – Delete rows
- **MERGE** – Selects records from a source table and performs multiple DML operations on a target table



# Delta Lake Overview

## Key Features



ACID Transactions



DML Operations



Time Travel



Schema Evolution and  
Enforcement



Many more!

Delta Tables keep a **transaction log** for each **version (writes)** of the table.

- Historical Querying
- Snapshot Isolation
- Auditing



# Delta Lake Overview

## Key Features



ACID Transactions



DML Operations



Time Travel



Schema Evolution and  
Enforcement



Many more!

### Schema Evolution

- Automatically **adjusts** the schema of your Delta table as your data changes



# Delta Lake Overview

## Key Features



ACID Transactions



DML Operations



Time Travel



Schema Evolution and  
Enforcement



Many more!

### Schema Evolution

- Automatically **adjusts** the schema of your Delta table as your data changes

### Schema Enforcement

- Ensures that any data written to the Delta table **matches** the table's defined schema



# Delta Lake Overview

## Key Features



ACID Transactions



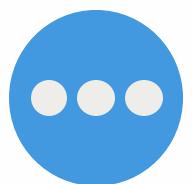
DML Operations



Time Travel



Schema Evolution and  
Enforcement



Many more!

- Unified Batch and Streaming
- Performance
- Scalable Metadata
- Optimization
- **Delta Lake is Open Source!**

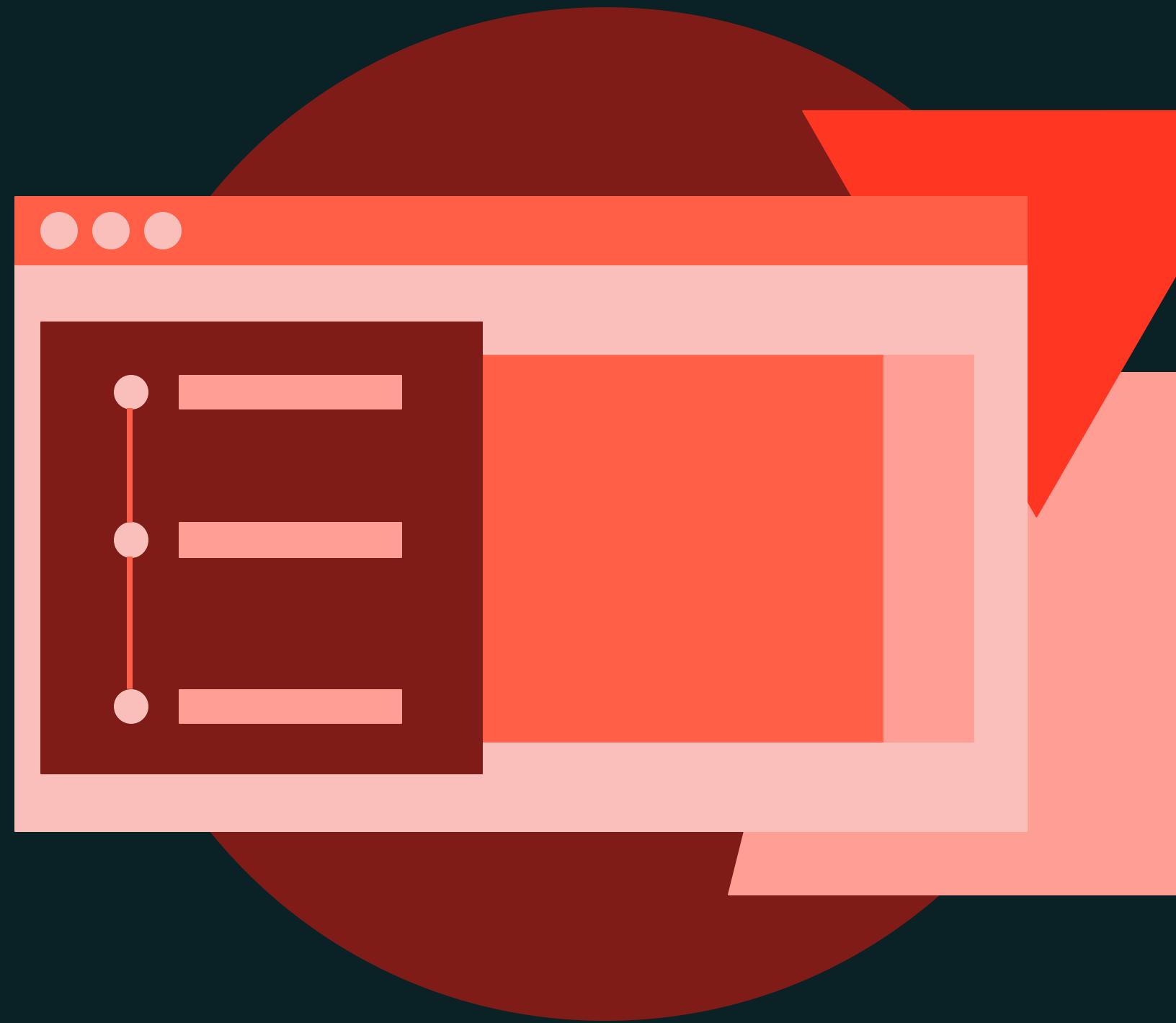




Using Databricks for Data Warehousing

## DEMONSTRATION

# Using Delta Lake features with Databricks SQL



# High Level Steps

- Explore key features of Delta Lake such as Time Travel, Version History, and metadata management.
- Use SQL commands like **DESCRIBE EXTENDED**, **DESCRIBE HISTORY**, **VERSION AS OF**, and **RESTORE TABLE**.
- Understand how Delta Lake's features are beneficial for managing data in data warehousing scenarios.
- Work with the **sales\_orders** table in the <schema\_name> schema.





# Data Ingestion and Transformation

---

Get Started with Databricks for Data Warehousing



# Objectives

- Describe and apply various techniques for ingesting and transforming data in Databricks.



# Agenda

Data Ingestion and Transformation	Time	Lecture	Demo	Lab
Delta Lake UniForm	5 mins	✓		
Data Ingestion Techniques Overview	5 mins	✓		
Data Ingestion Techniques	10 mins		✓	
Data Transformation	8 mins	✓		
Exploring Data Transformation in Databricks	10 mins		✓	





Data Ingestion and Transformation

LECTURE

# Delta Lake UniForm



# Delta Lake with UniForm

Rise of Open Table Formats

Which open table format to pick?



Delta Lake



Hudi



Iceberg

Delta Lake with UniForm simplifies data management by bridging the gap between different table formats, empowering data warehouses with unified access and control.

# Introducing Delta Lake with Uniform (**Delta Universal Format**)



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Delta Lake with UniForm

## Open Table Formats Overview



### Apache Hudi

#### Metadata

Used for transactional source of truth, concurrency control, etc.



#### Metadata



#### Parquet

#### Connector Ecosystem



© Databricks 2025. All rights reserved. The Apache logo and the Apache Hadoop logo are trademarks of the Apache Software Foundation.



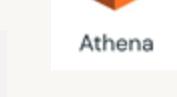
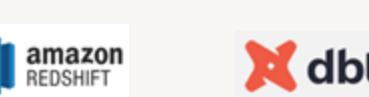
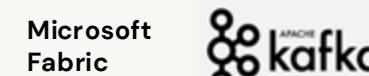
### Delta Lake



#### Metadata



#### Parquet



### Apache Iceberg



#### Metadata



#### Parquet



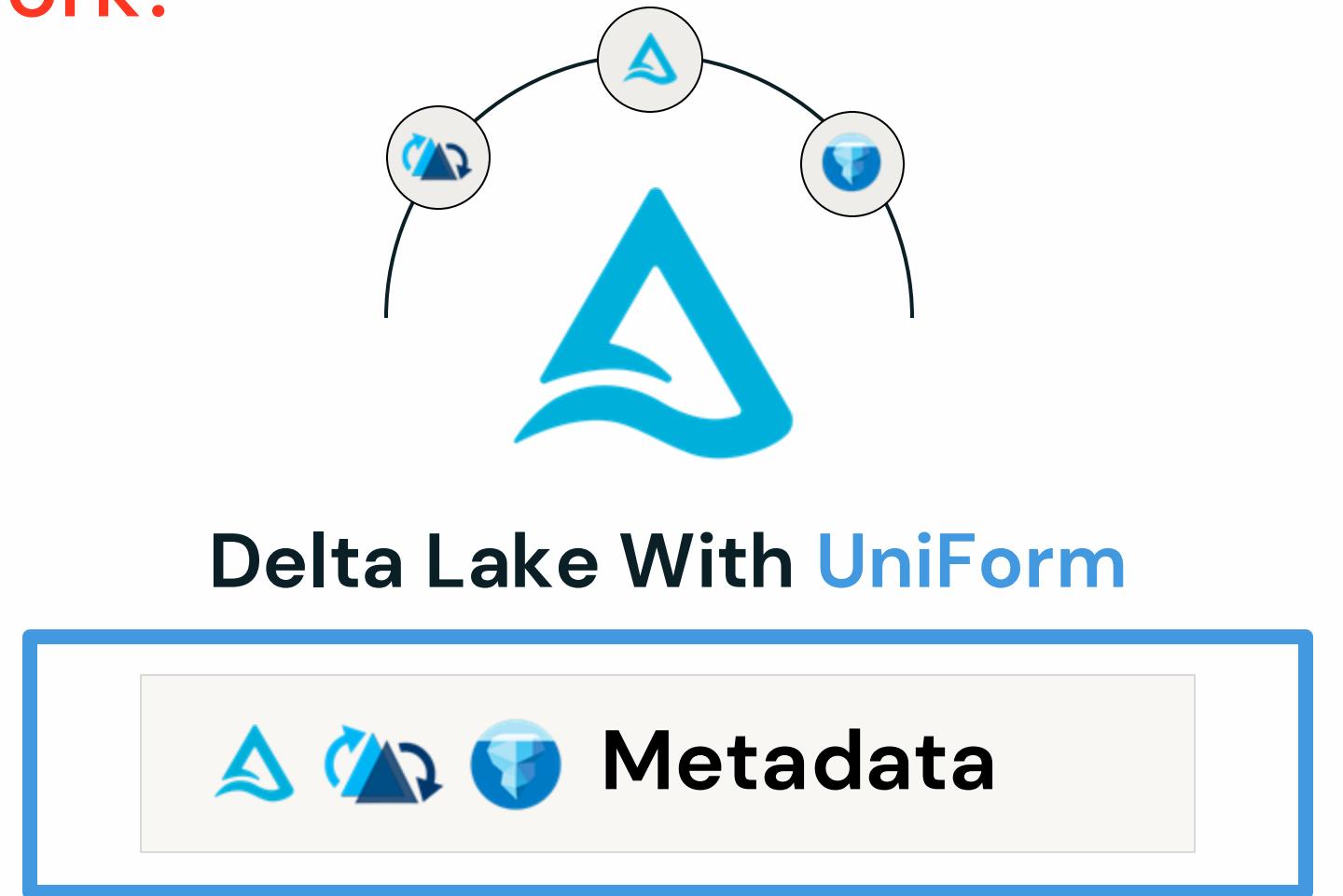
Spark, the  
Apache  
Hadoop  
and the Apa

Apache  
Fabric  
Amazon  
Redshift  
Apache  
kafka  
dbt  
Athena

Microsoft  
Fabric  
Apache  
kafka  
dbt  
Athena

# Delta Lake with UniForm

How does it work?

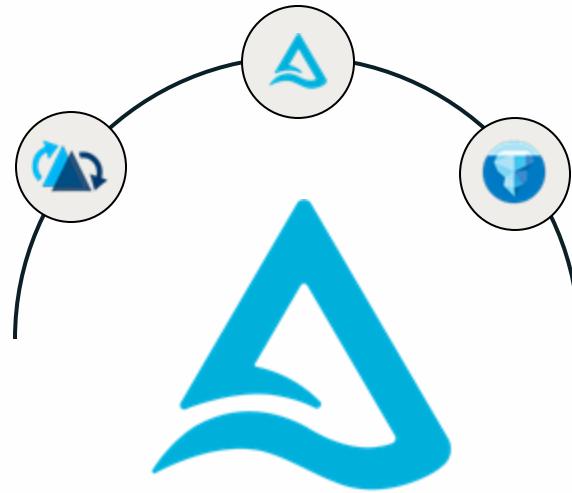


- ✓ **Metadata automatically generated to make Delta accessible as Iceberg/Hudi readers**



# Delta Lake with UniForm

How does it work?



Delta Lake With **UniForm**

## Metadata

Used for transactional source of truth, concurrency control, etc.

## Data

All formats use Parquet!



**Metadata**



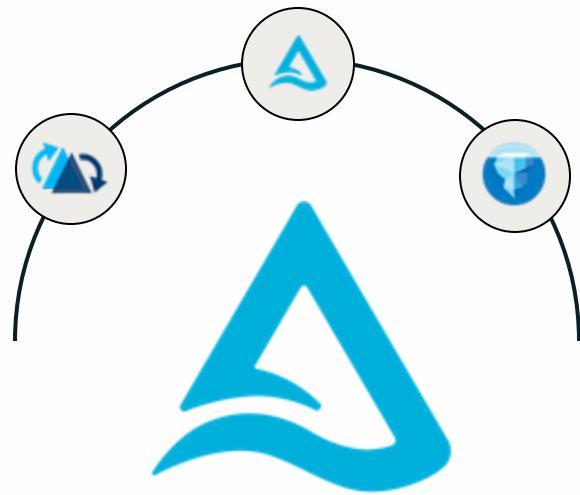
**Parquet**

- ✓ Metadata automatically generated to make Delta accessible as Iceberg/Hudi readers
- ✓ Parquet files remain the same



# Delta Lake with UniForm

How does it work?



## Delta Lake With UniForm

### Metadata

Used for transactional source of truth, concurrency control, etc.

### Data

All formats use Parquet!

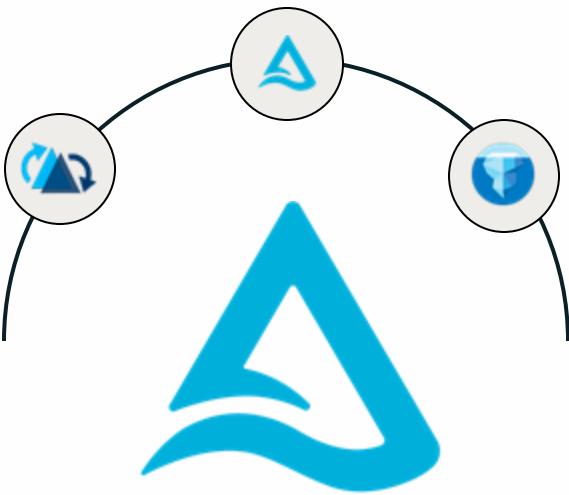


- ✓ Metadata automatically generated to make Delta accessible as Iceberg/Hudi readers
- ✓ Parquet files remain the same
- ✓ Metadata is co-located with data



# Delta Lake with UniForm

How does it work?



Delta Lake With [UniForm](#)

## Metadata

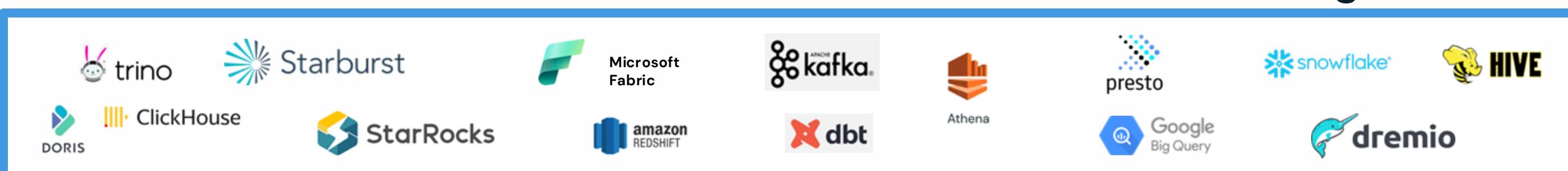
Used for transactional source of truth, concurrency control, etc.

## Data

All formats use Parquet!



## Connector Ecosystem



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).



Data Ingestion and Transformation

LECTURE

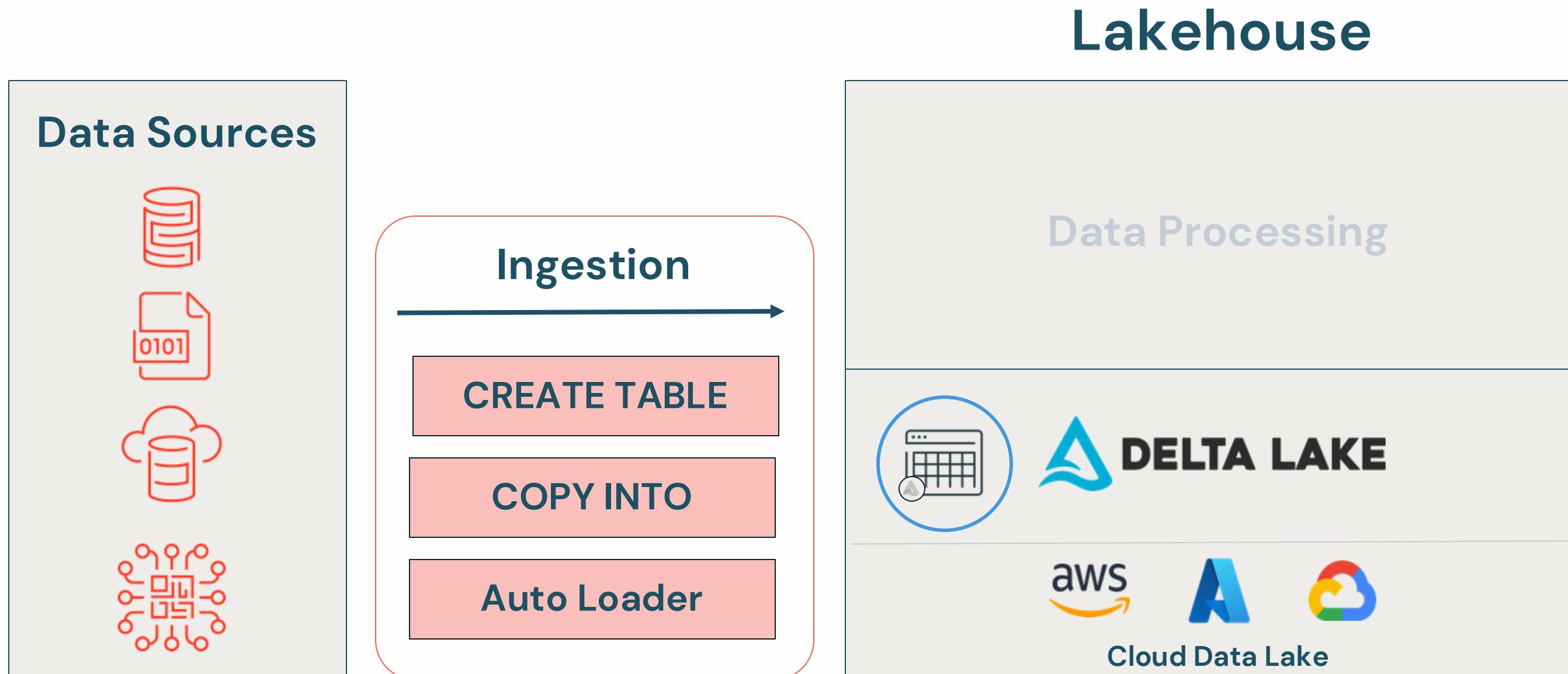
# Data Ingestion Techniques Overview



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Delta Lake Data Ingestion Techniques

## Overview



# Delta Lake Data Ingestion Techniques

## Overview

**CREATE TABLE**

**Upload UI**

**COPY INTO**

**Auto Loader**



# Delta Lake Data Ingestion Techniques

## Overview

CREATE TABLE

Upload UI

COPY INTO

Auto Loader

```
CREATE TABLE mydeltatable  
USING DELTA  
AS  
<query>
```

USING DELTA is optional.

- CREATE TABLE AS (CTAS) creates a table by selecting data from an existing table or data source
- Default table format is **Delta**



# Delta Lake Data Ingestion Techniques

## Overview

CREATE TABLE

Upload UI

COPY INTO

Auto Loader

- The UPLOAD UI provides a **point-and-click interface** to upload files and create tables
- Allows you to upload CSV, TSV, JSON, Avro, Parquet, or text files
- The UI also enables you to upload files directly to a Unity Catalog volume



# Delta Lake Data Ingestion Techniques

## Overview

CREATE TABLE

Upload UI

COPY INTO

Auto Loader

```
COPY INTO mydeltatable  
FROM 'your-path'  
FILE_FORMAT = 'format'  
FILE_OPTIONS = ('format-options')
```

- Loads files from a **file location** into a Delta table
- Supports **various file formats** and cloud storage locations
- Automatically can handle **schema changes**.
- **Idempotent** operation—Already-loaded files in the source are skipped



# Delta Lake Data Ingestion Techniques

## Overview

CREATE TABLE

Upload UI

COPY INTO

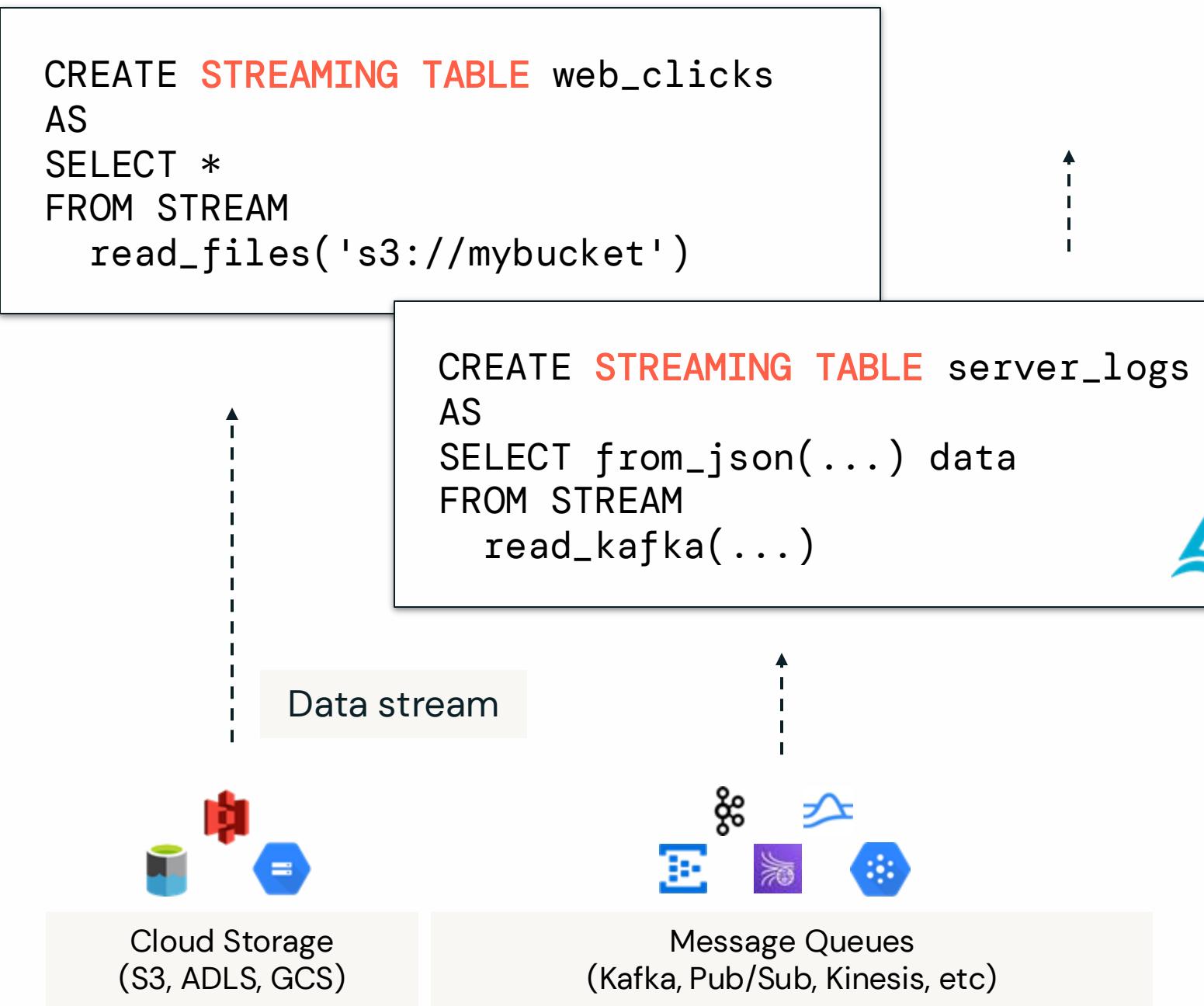
Auto Loader

- Auto Loader **incrementally (streaming) processes** new data files as they arrive in cloud storage
- Automatically **infers schemas** and accommodates **schema changes**
- Includes a **rescue data** column for data that does not adhere to the schema



# Streaming Tables in Delta Lake

Real-Time and Incremental Data Processing Made Simple



## Benefits:

1. **Unlock real-time use cases.** Ability to support real-time analytics/BI, machine learning and operational use cases with streaming data.
2. **Better scalability.** More efficiently handle high volumes of data via incremental processing vs large batches.
3. **Enable more practitioners.** Simple SQL syntax makes data streaming accessible to all data engineers and analysts.





# Databricks LakeFlow Connect

Native ingestion connectors for the Data Intelligence Platform

Efficient ingestion pipelines

Simple setup and maintenance

Unified orchestration, observability, and governance

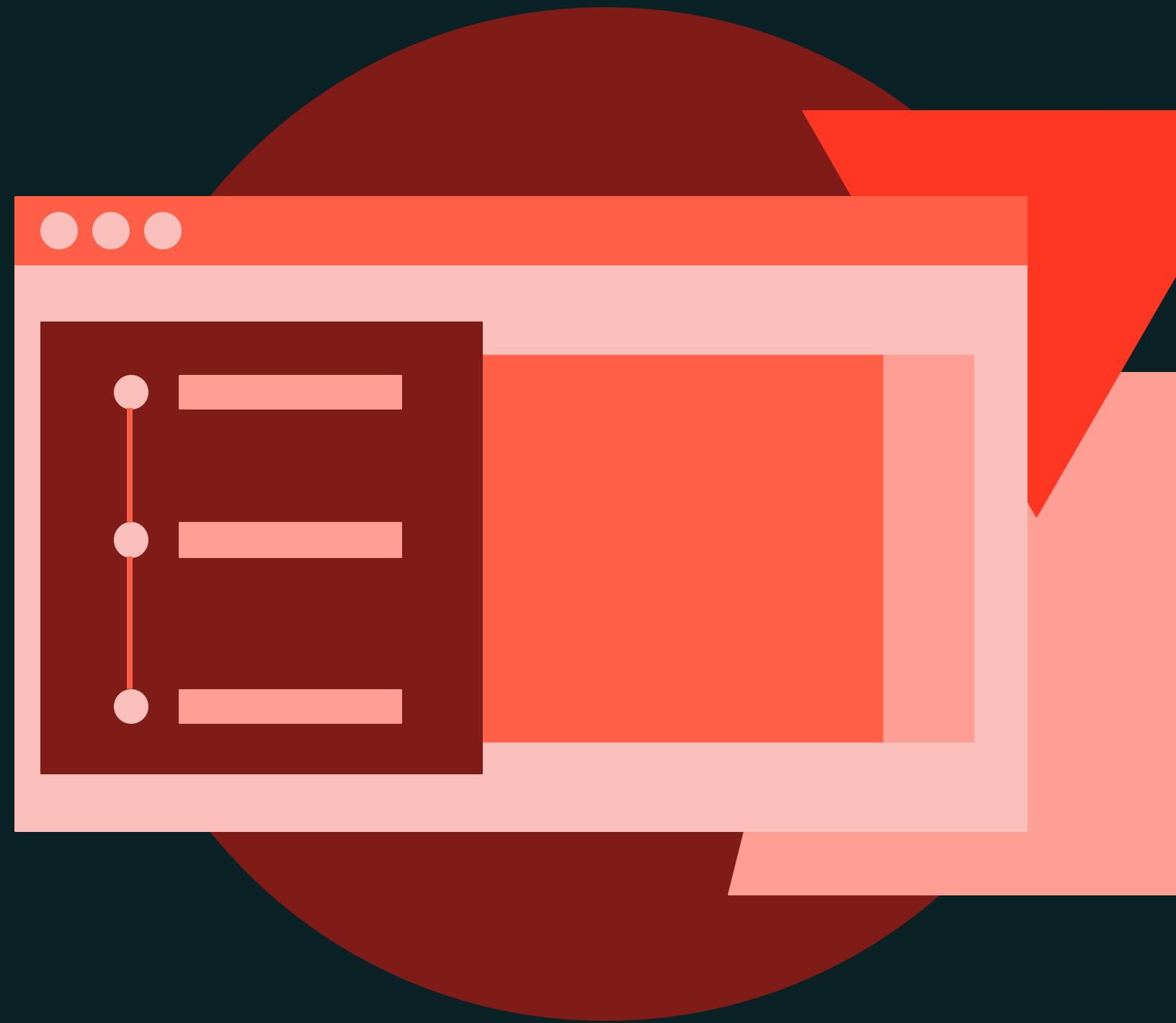




Data Ingestion and Transformation

**DEMONSTRATION**

# Data Ingestion Techniques



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# High Level Steps

- Ingest data into Delta Lake using CREATE TABLE AS SELECT.
- Incrementally load data using COPY INTO.
- Use the Databricks Upload UI for data ingestion.
- Set up real-time ingestion pipelines using Auto Loader.
- Understand the role of Lakeflow Connect in data ingestion workflows.





Data Ingestion and Transformation

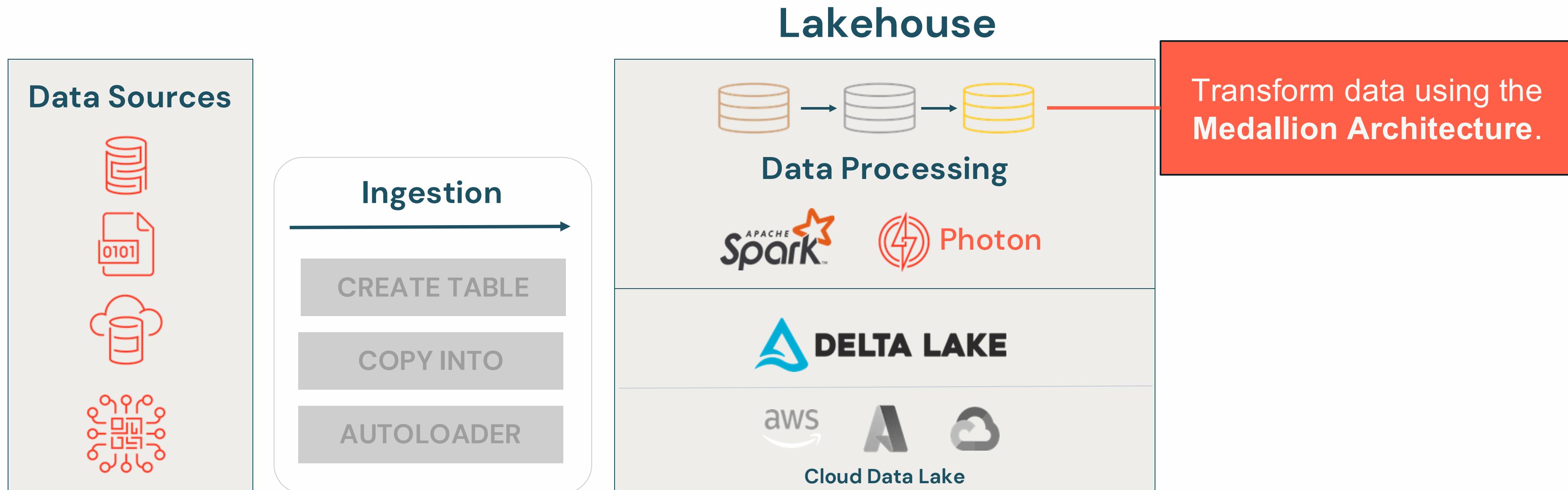
LECTURE

# Data Transformation



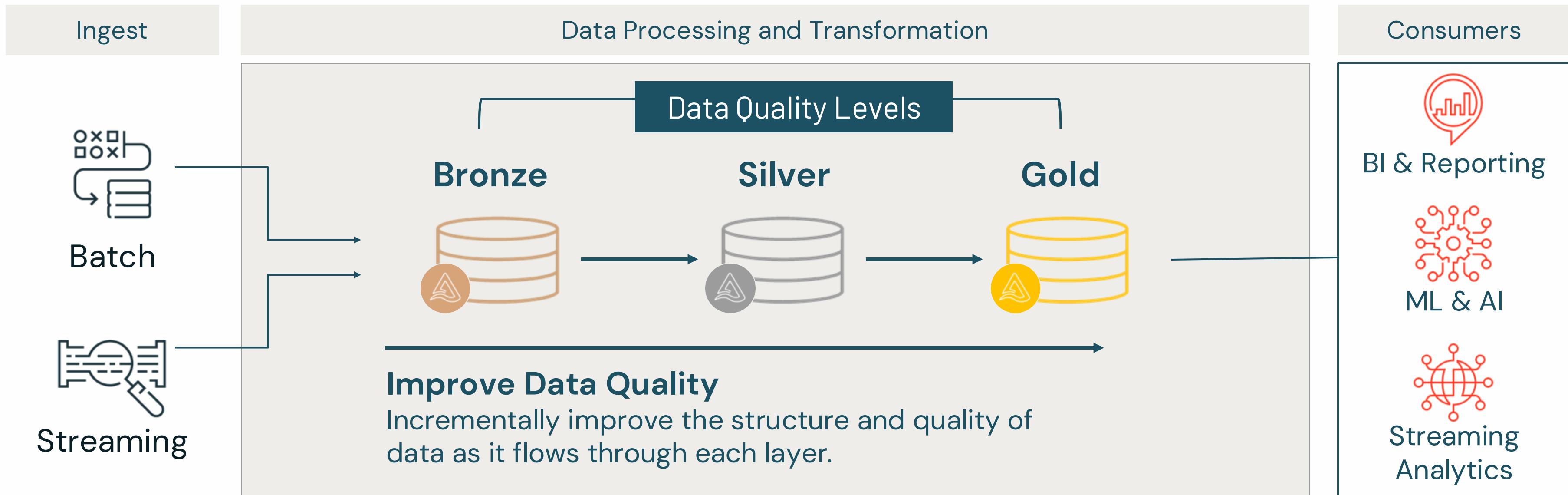
# Data Transformation Overview

## Medallion Architecture (Multi Hop)



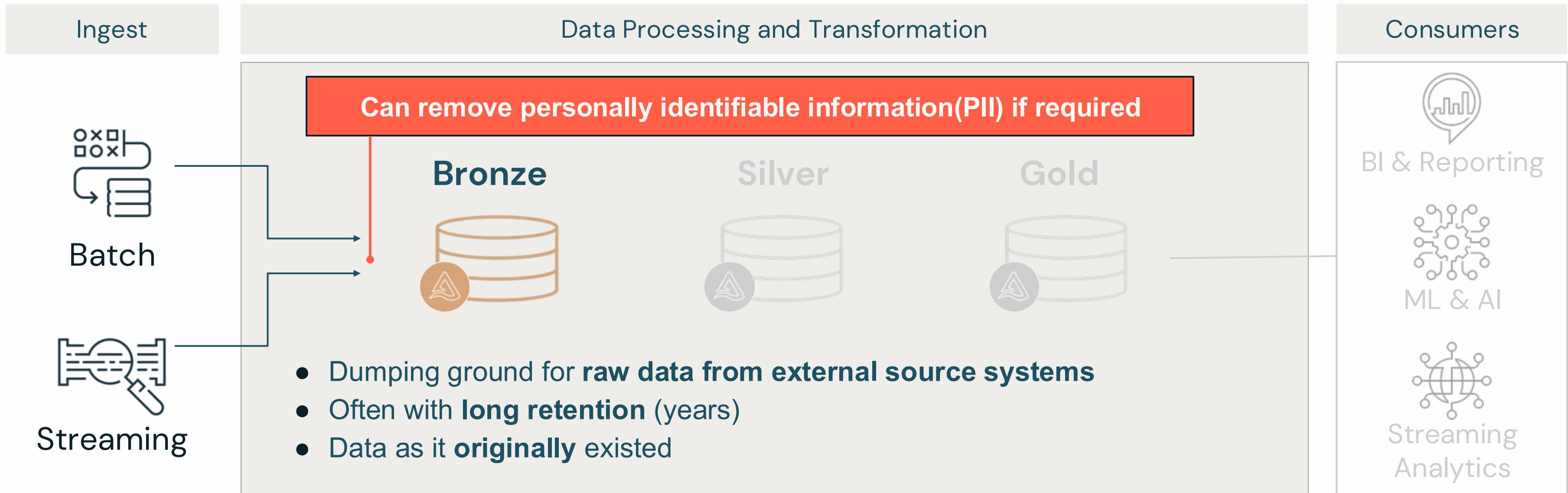
# Data Transformation Overview

## Medallion Architecture (Multi Hop)



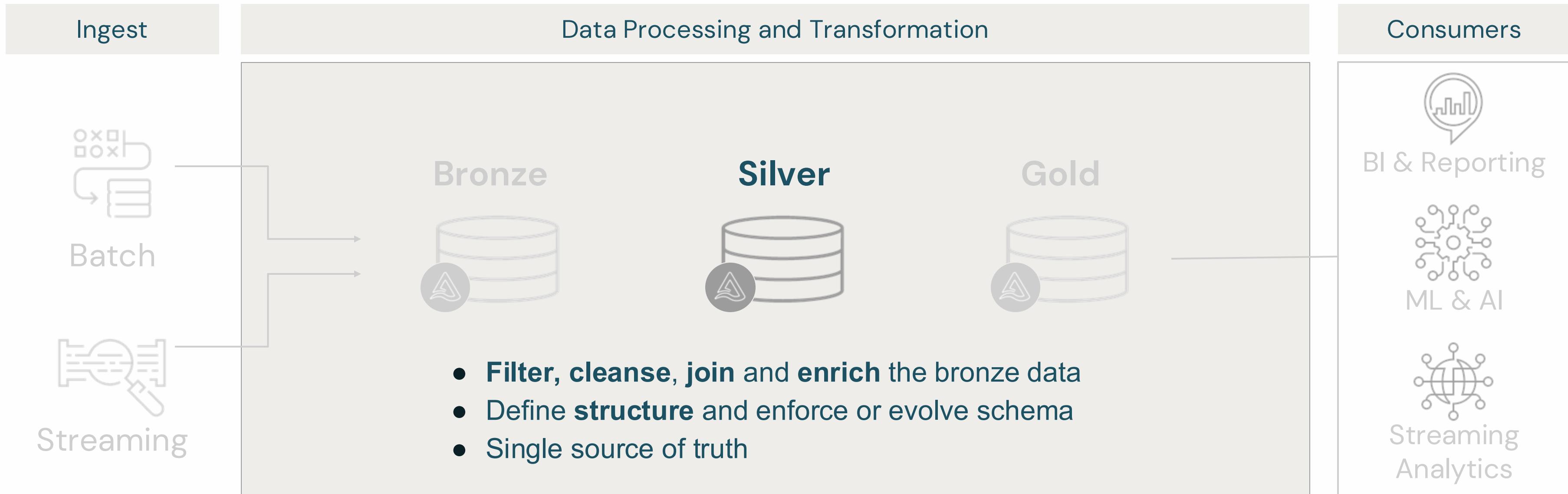
# Data Transformation Overview

## Medallion Architecture (Multi Hop)



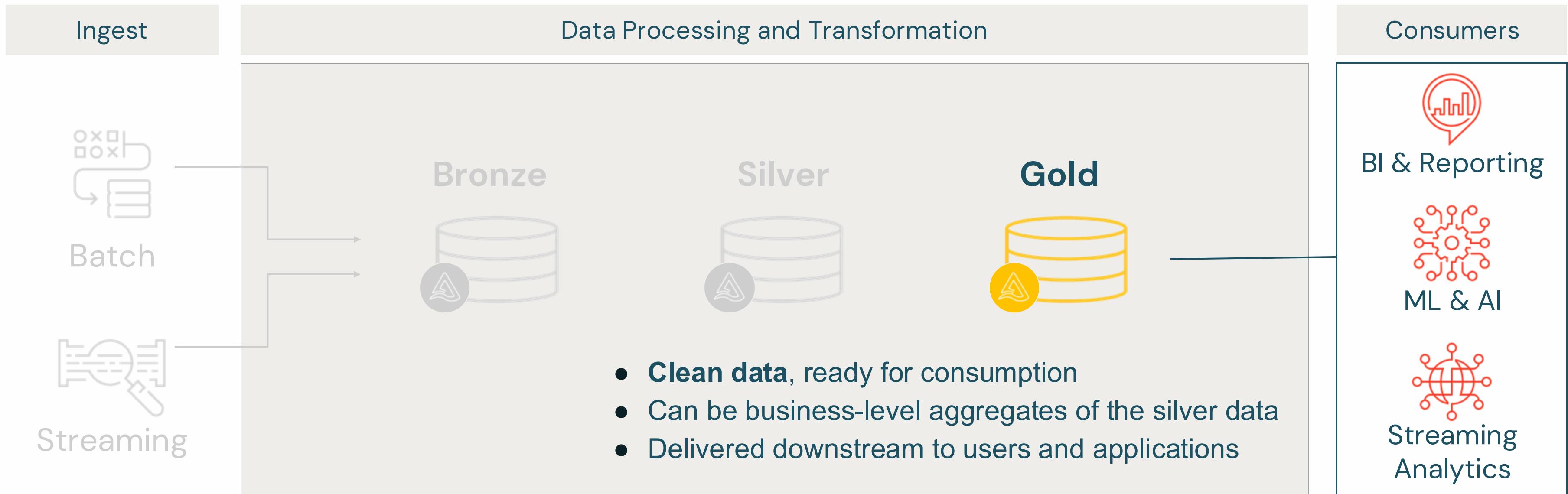
# Data Transformation Overview

## Medallion Architecture (Multi Hop)



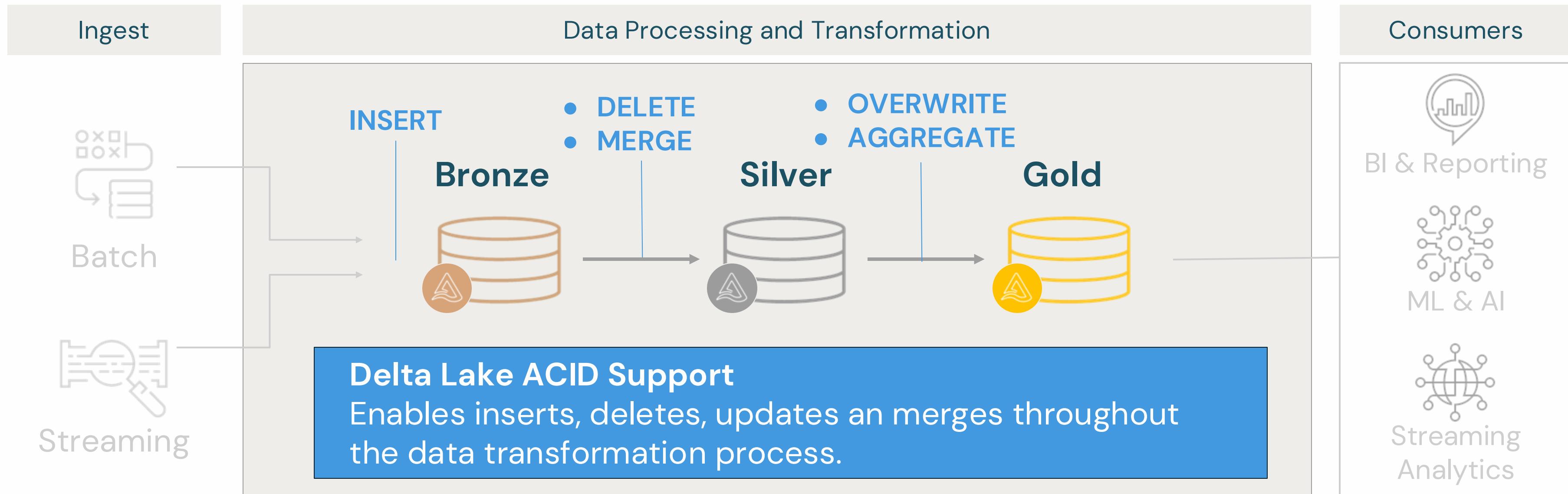
# Data Transformation Overview

## Medallion Architecture (Multi Hop)



# Data Transformation Overview

## Medallion Architecture (Multi Hop)

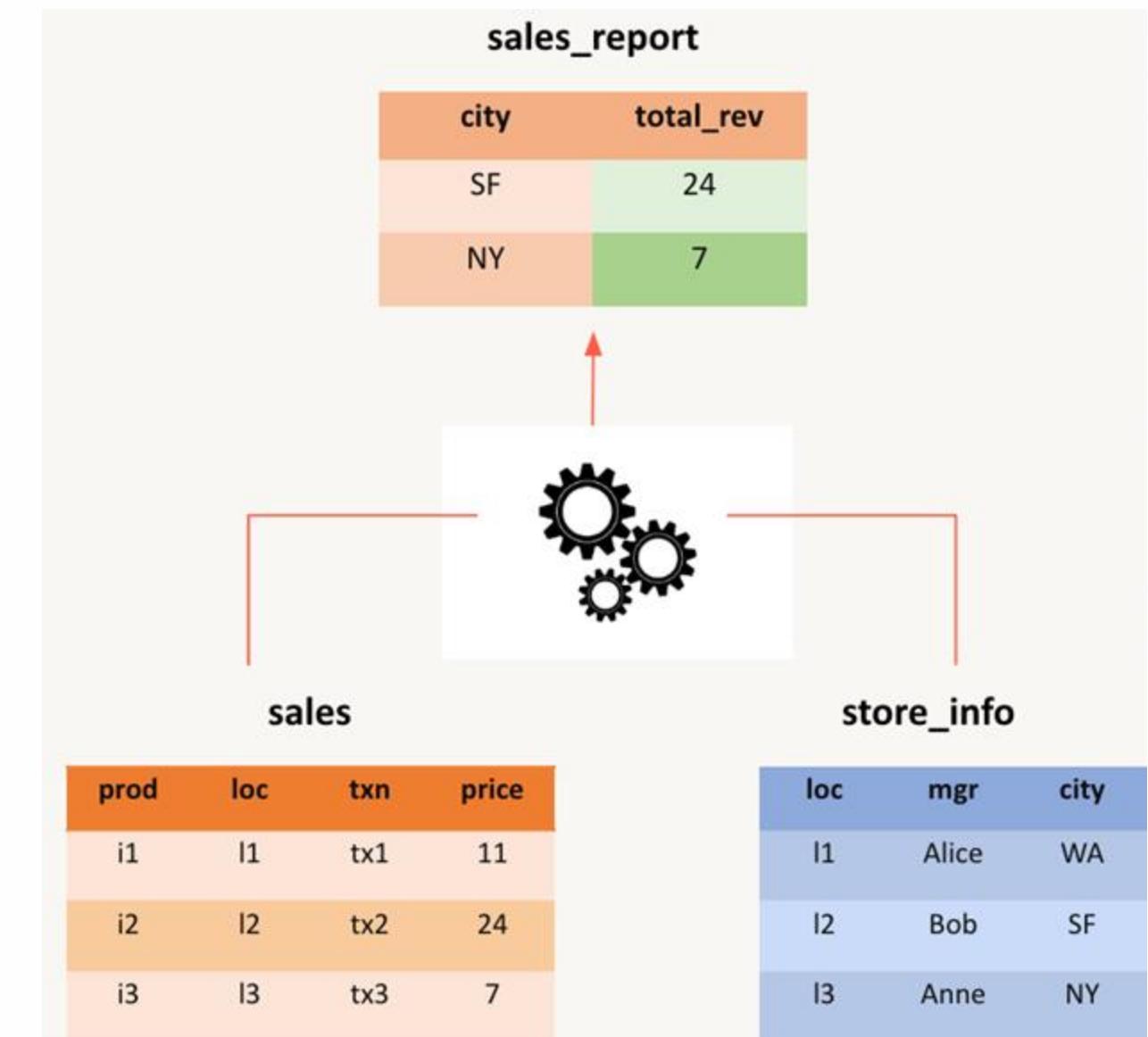


# Materialized Views

Speed up queries with pre-computed results

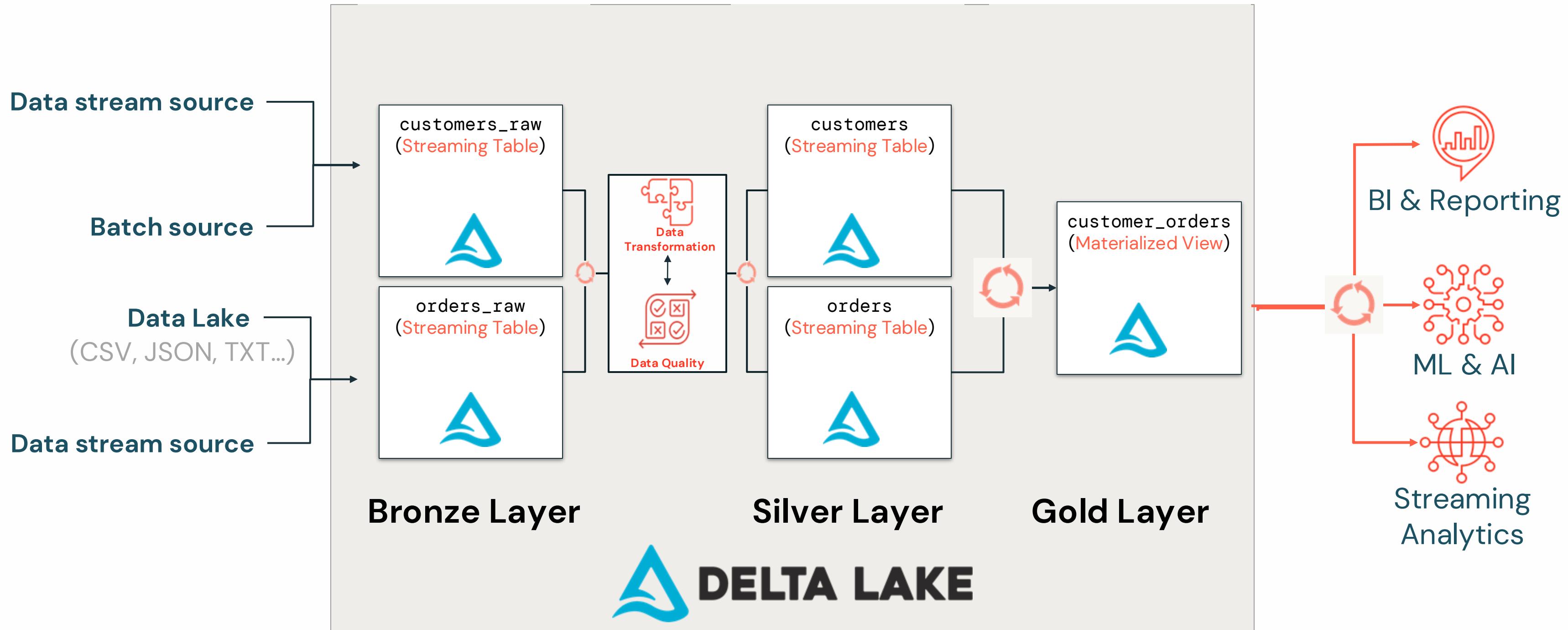
Accelerate end-user queries and reduce infrastructure costs with efficient, incremental computation

- Accelerate **BI dashboards** and **ETL queries**
- **Streaming**: build MVs on top of live tables
- **Easy ELT**: Simplify reporting by cleaning, enriching, denormalizing the base tables
- **Data Sharing & Access Control**: control what info can be seen by internal and external users and organizations



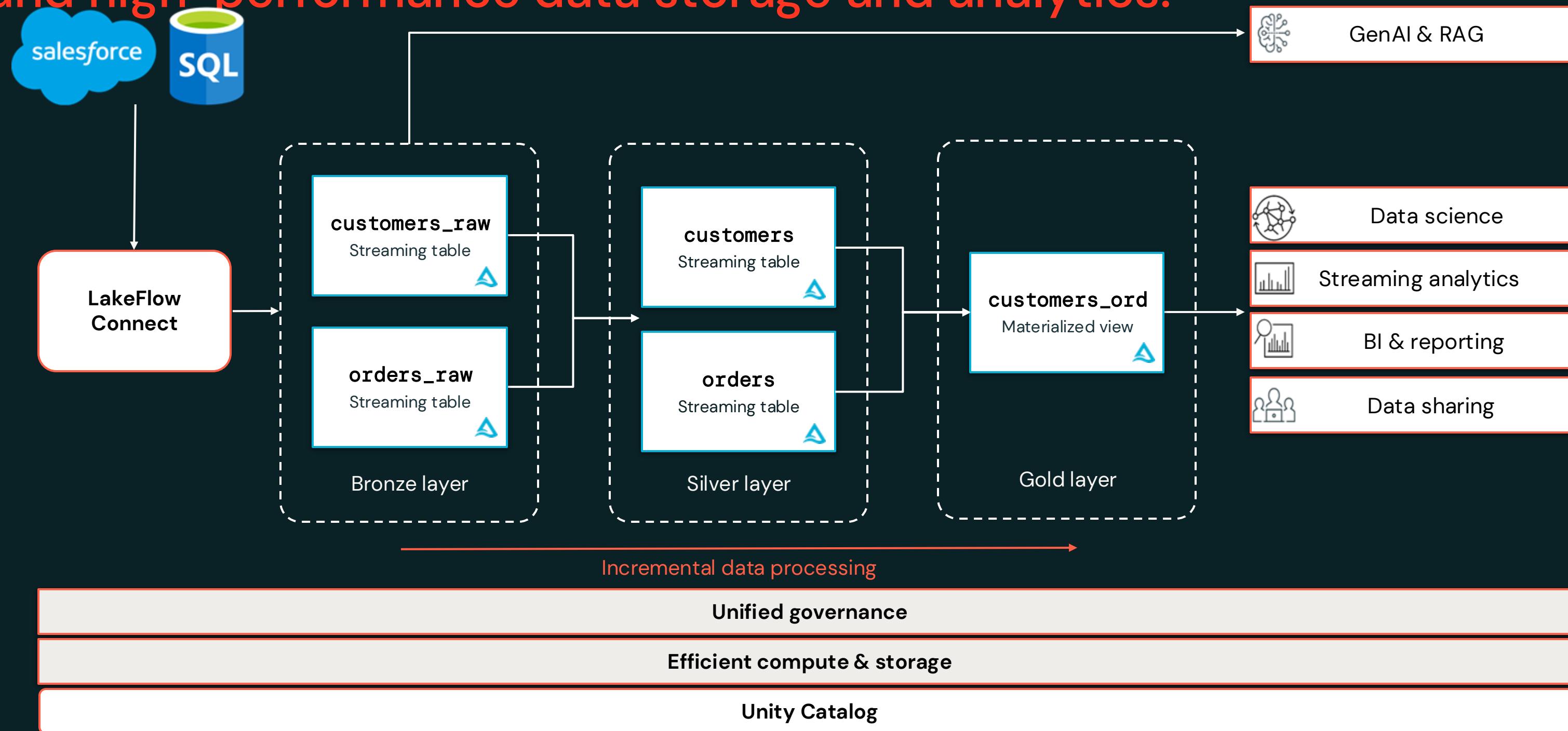
# Constructing a Data Transformation Pipeline

Real World Architecture is Typically More Complex



# LakeFlow Connect in Action

Medallion Architecture powered by Delta Lake ensures reliable, scalable, and high-performance data storage and analytics.



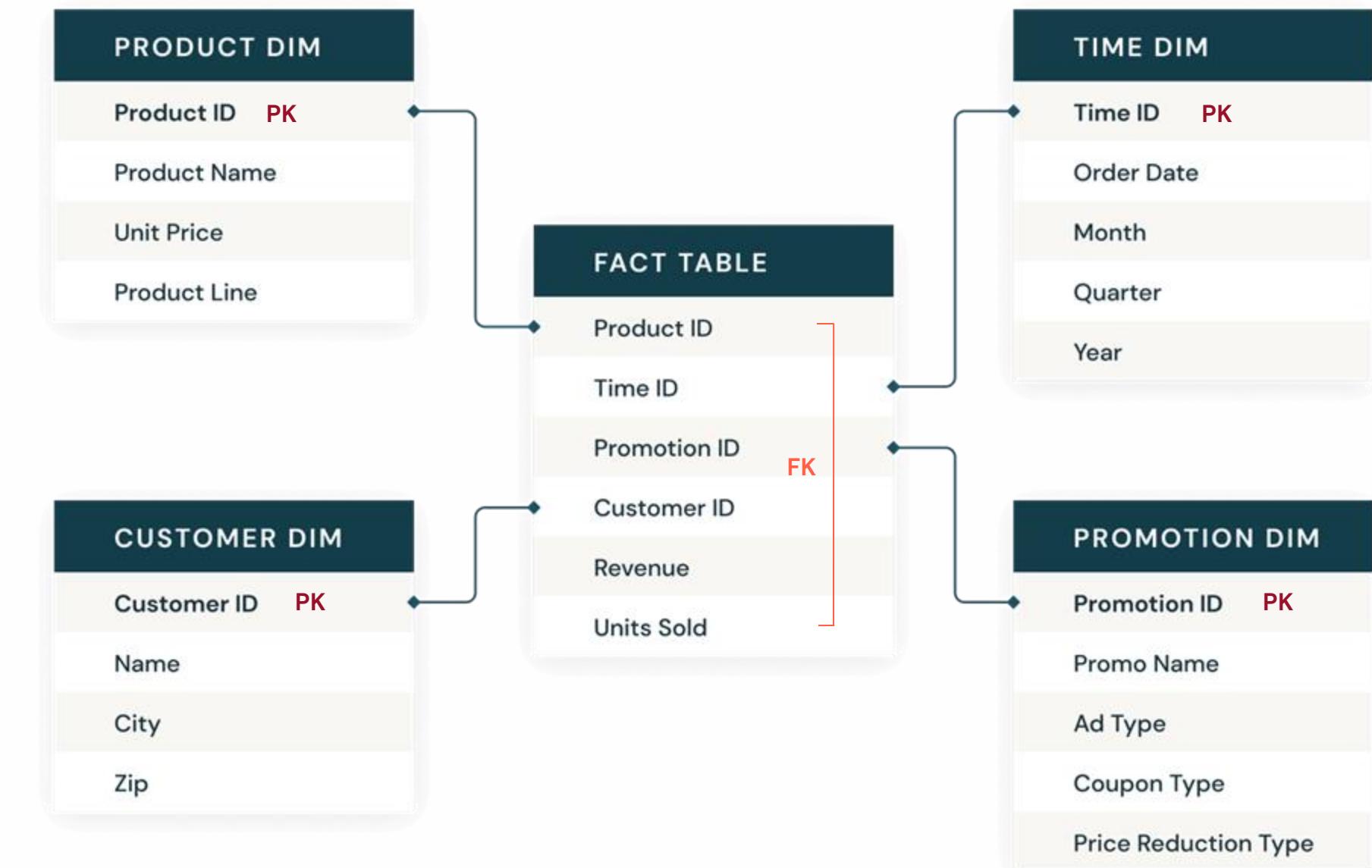
# Overview of Data Modeling with Databricks

## Introduction to Data Modeling in Databricks

Data Modeling enables structured, high-quality, and consistent data to drive accurate analysis and informed decision-making.

- **Support for Various Modeling Approaches:**
  - Design scalable models using Star Schema, Snowflake Schema, or Data Vault.
- **Primary Key/Foreign Key Constraints:**
  - Ensure referential integrity in Delta tables for reliable relationships.
- **Surrogate Keys:**
  - Enable efficient data management with auto-incrementing IDs.

Star schema



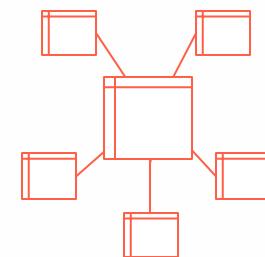
# Modern Data Modeling Approaches

## Key Modeling Approaches:



### Top-Down Approach (Inmon):

- Relational Modeling with normalized data as the core of the Data Warehousing
- Data marts (often dimensional or denormalized models)



### Bottom-Up Approach (Kimball):

- Dimensional Modeling
- Denormalized data model, built as a star or snowflake schema: central fact tables surrounded by dimension tables



### Data Vault (Linstedt):

The Data Vault model is based on three basic entity types:

- **Hubs** separate core business concepts
- **Links** store relationships between business concepts
- **Satellites** store the attributes of a business concepts or relationships

#### Unique Databricks Features:

- Seamless integration with Delta Lake for schema management.
- Identity columns for generating Surrogate Keys.

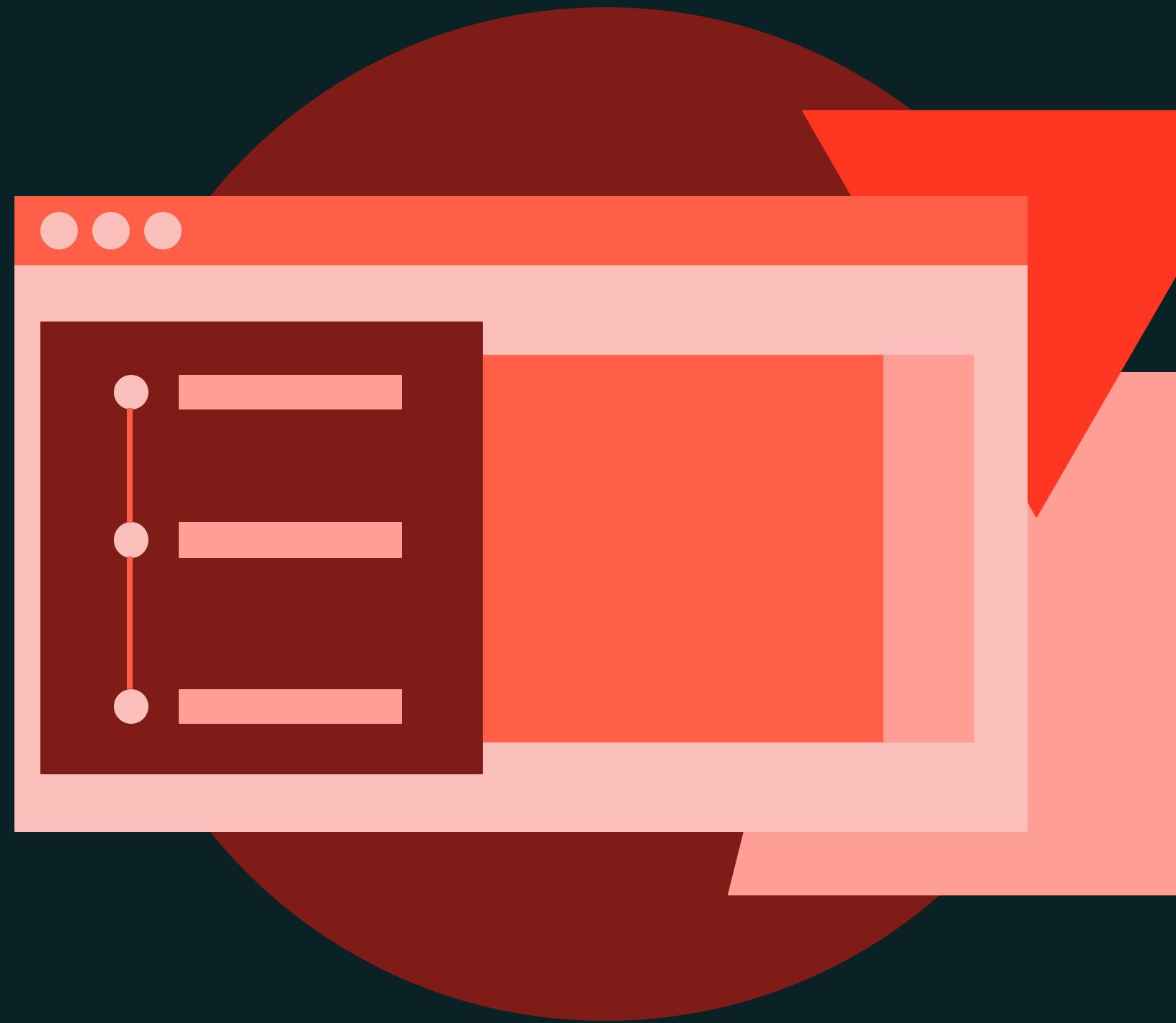




Data Ingestion and Transformation

**DEMONSTRATION**

# Exploring Data Transformation in Databricks



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# High Level Steps

- Create a multi-hop transformation pipeline using the Medallion Architecture.
- Use Streaming Tables and Materialized Views to transform raw data into meaningful insights.
- Demonstrate how to move data from the Bronze layer to the Silver and Gold layers while maintaining data quality.

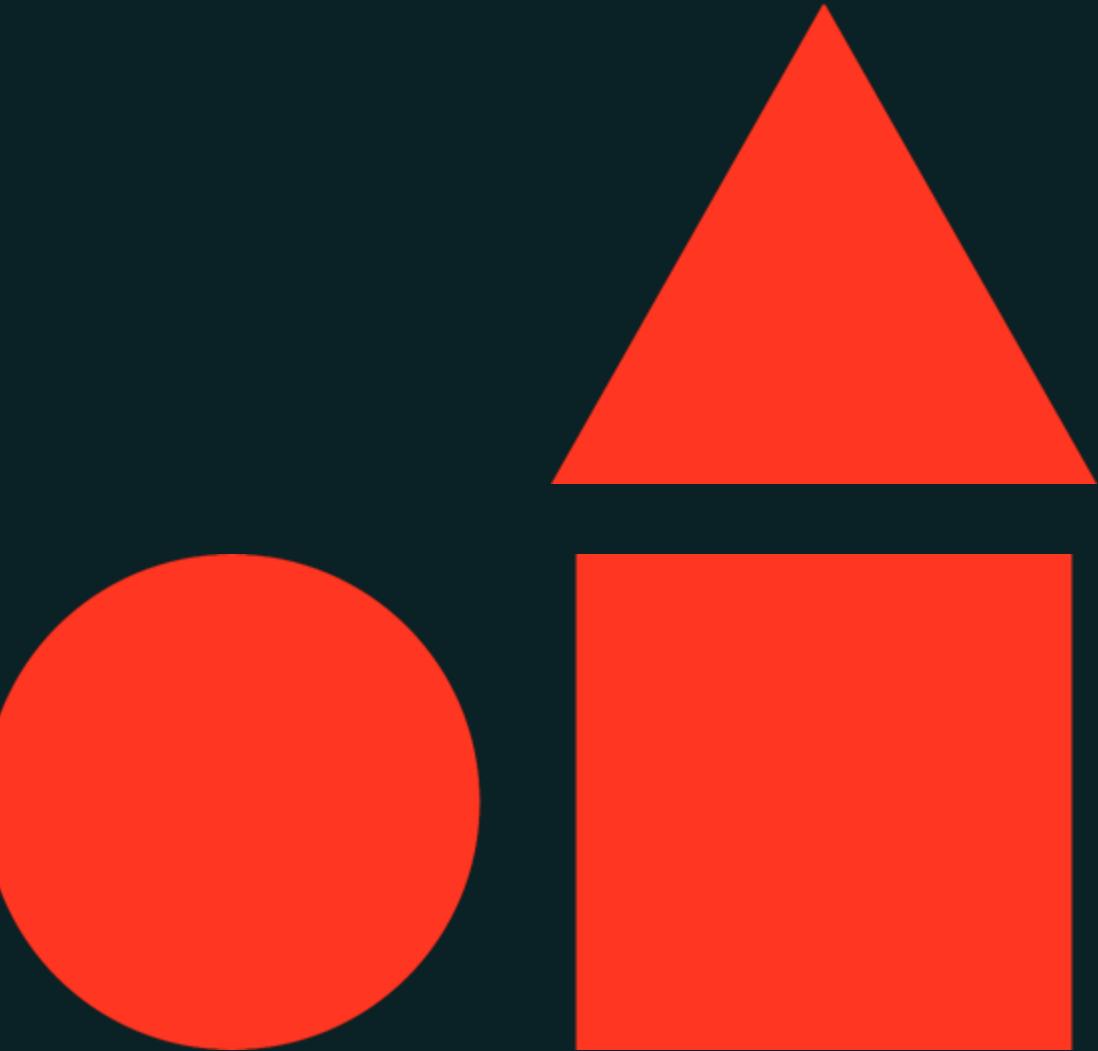




# Data Orchestration and Querying Capabilities

---

Get Started with Databricks for Data Warehousing



# Objectives

- Explain how Databricks support data orchestration needs within the platform.



# Agenda

Data Orchestration and Querying Capabilities	Time	Lecture	Demo	Lab
Orchestration in Databricks	5 mins	✓		
Setting Up and Managing Serverless Lakeflow Jobs	15 mins		✓	
Databricks Querying Capabilities	5 mins	✓		





Data Orchestration and Querying Capabilities

LECTURE

# Orchestration in Databricks

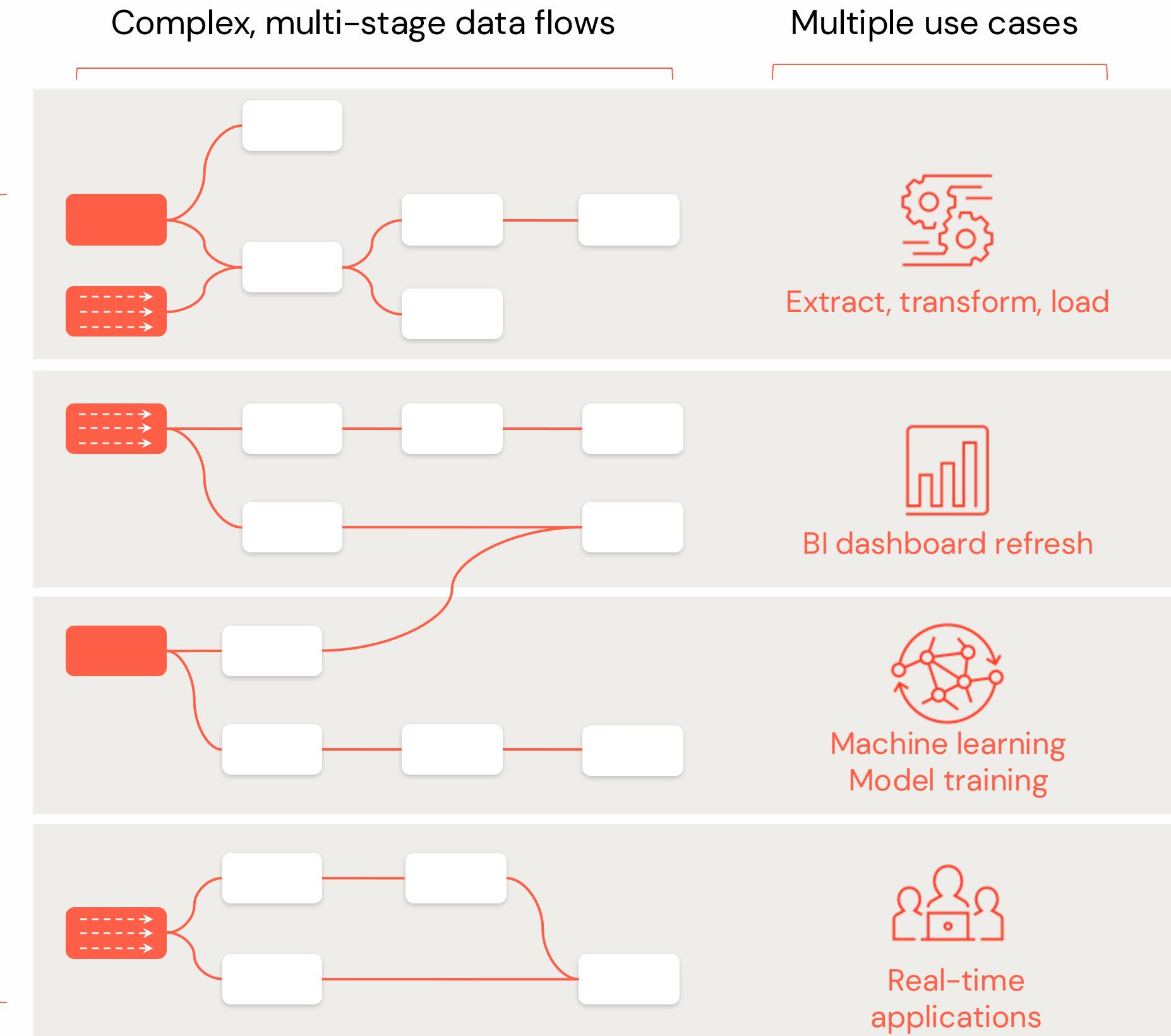


# Modern Data Warehousing Requires Modern Orchestration

**Orchestrating processes across all data, analytics and AI use cases is business critical**

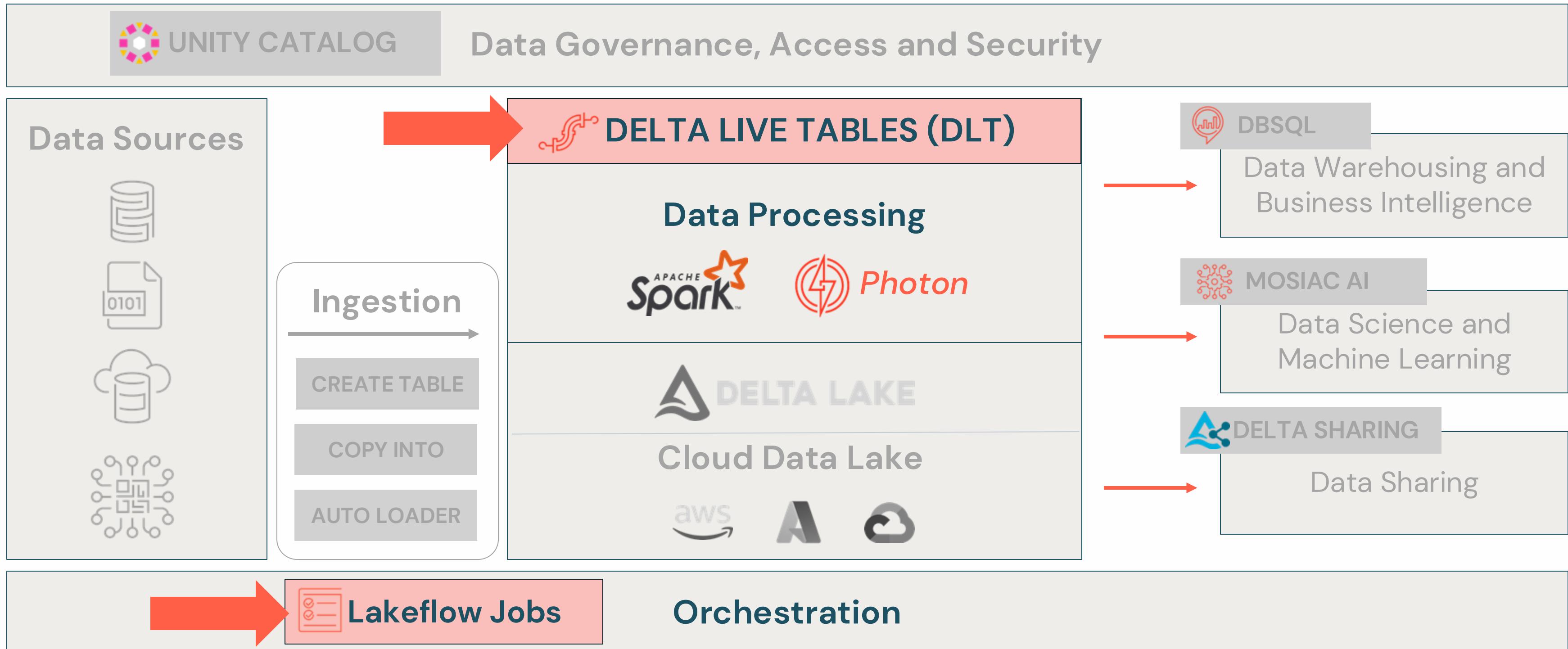
- Handles complex, multi-stage data flows.
- Manages multiple dependencies in Lakeflow Jobs.
- Supports diverse use cases, including:
  - Extract, transform, load (ETL).
  - BI dashboard refresh.
  - Machine learning model training.
  - Real-time applications.

Multiple data sources & triggers



# Unified Orchestration Using Databricks

## Overview



# Databricks Lakeflow Jobs

**Reliable orchestration for data, analytics, and AI**



**Orchestrate  
Anything  
Anywhere**



**Simple  
Lakeflow Jobs  
Authoring**



**Deep Platform  
Integration**



**Proven  
Reliability**

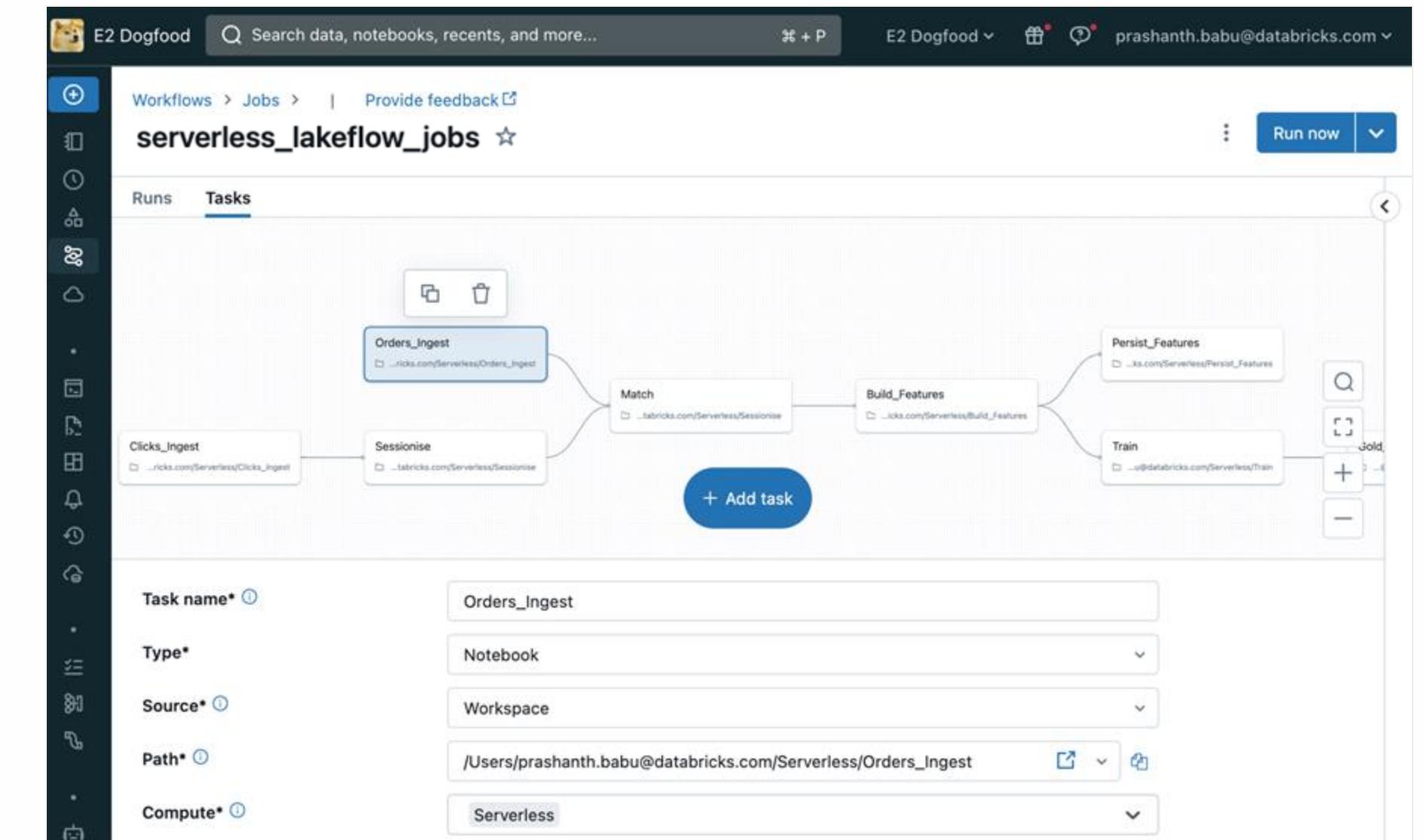


**Fully  
Managed**



# Serverless Lakeflow Jobs

- **Fully Managed Infrastructure**, eliminates the human efforts to handle compute, providing fast analysis & cost savings.
- **Advanced Autoscaler** ensures efficient resource management
- **Elastic scaling** up and down is achieved using a warm pool of machines.



# LakeFlow

The evolution of data warehousing on Databricks

A unified data warehousing solution  
powered by Data Intelligence

DLT

LakeFlow  
Jobs



LakeFlow

Connect

Pipelines

Jobs

Powered by  
**Data  
Intelligence**

Governance with  
**Unity  
Catalog**

Efficiency with  
**Serverless  
Compute**

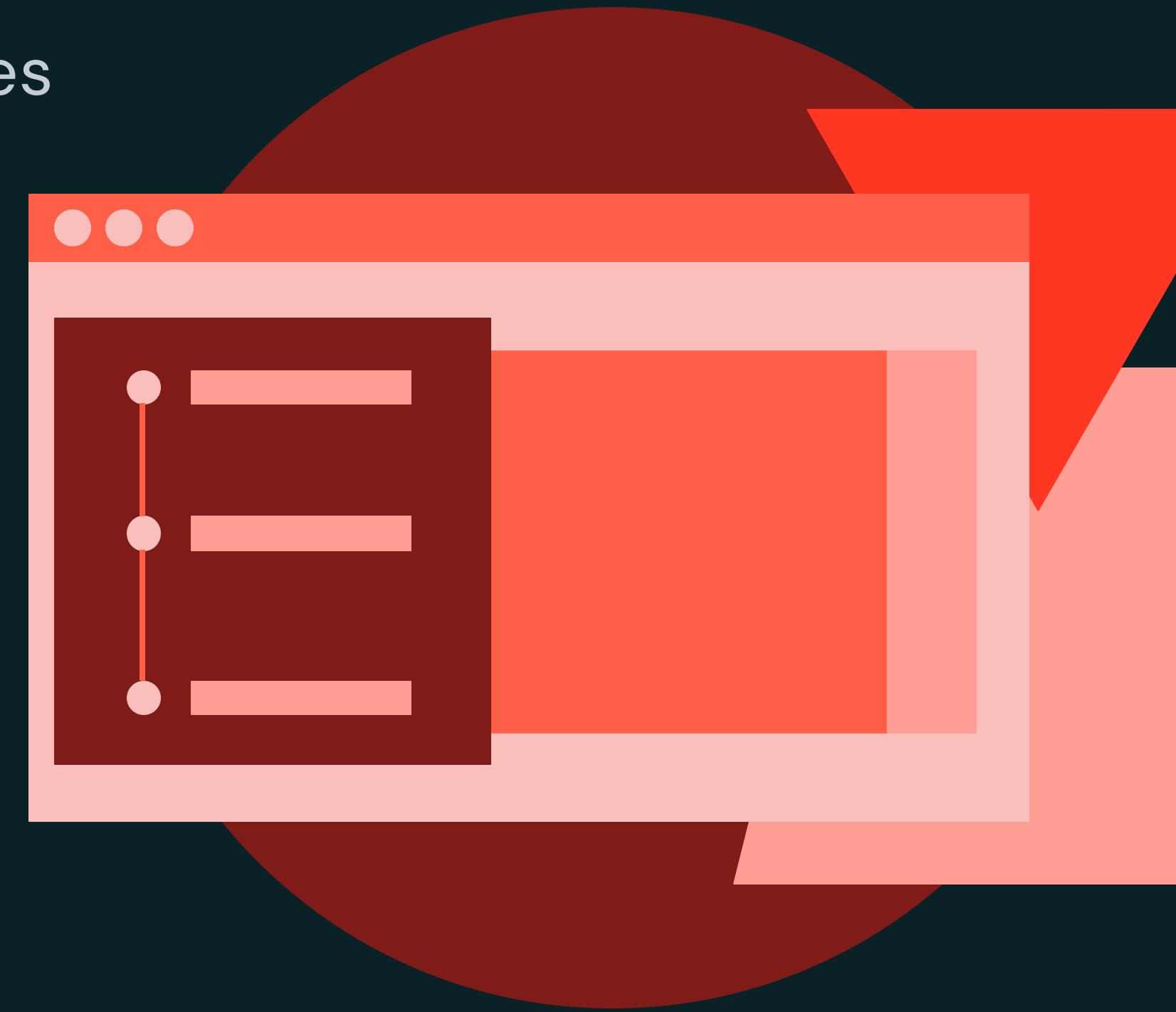




Data Orchestration and Querying Capabilities

## DEMONSTRATION

# Setting Up and Managing Serverless Lakeflow Jobs



# High Level Steps

- Create a Lakeflow Job with SQL tasks for a Medallion Architecture pipeline.
- Set up task dependencies and implement conditional logic for lakeflow jobs control.
- Use the Lakeflow Job UI for monitoring and data lineage visualization to trace data transformations and dependencies.
- Configure error handling and retries for tasks.
- Schedule Lakeflow Job using manual triggers.
- Set up notifications for monitoring and analyze the execution history





Data Orchestration and Querying Capabilities

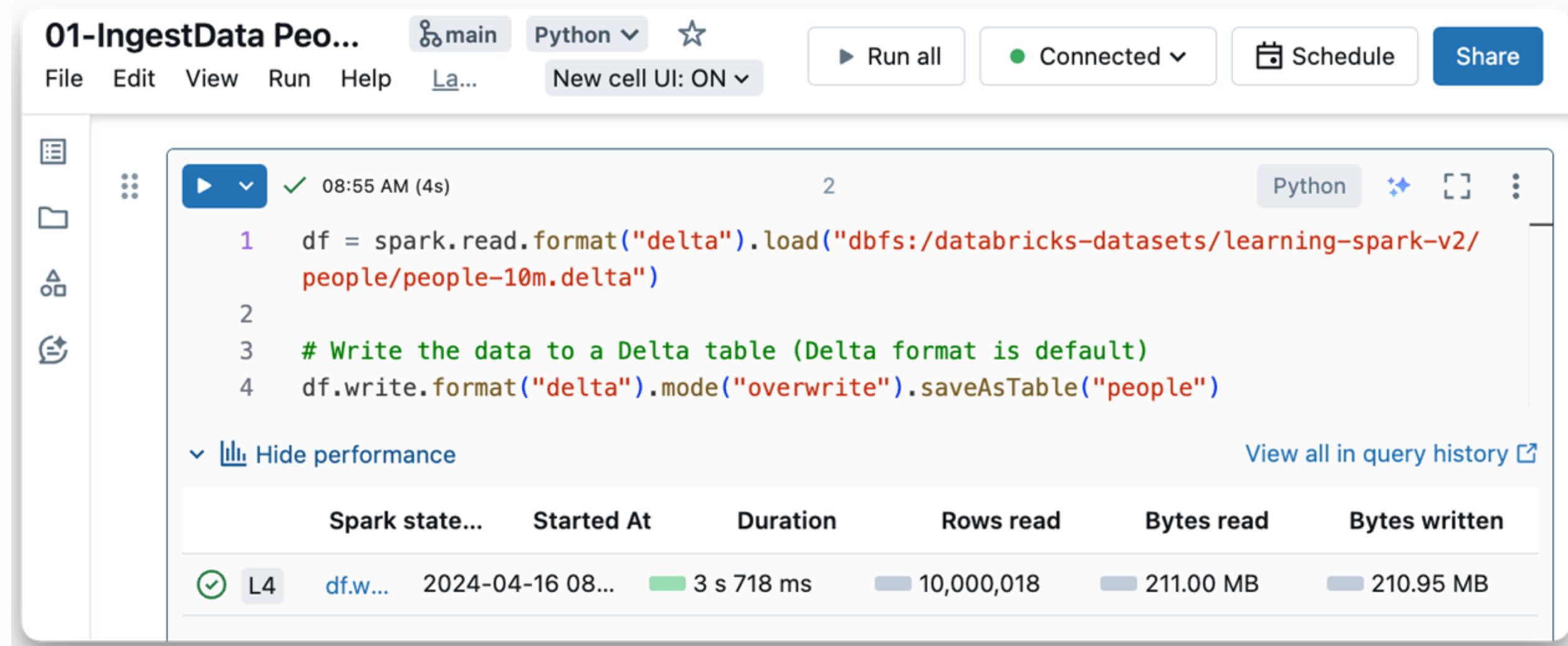
LECTURE

# Databricks Querying Capabilities



# Serverless Compute for Notebooks

With *Environments* for dependency management



```
df = spark.read.format("delta").load("dbfs:/databricks-datasets/learning-spark-v2/people/people-10m.delta")
# Write the data to a Delta table (Delta format is default)
df.write.format("delta").mode("overwrite").saveAsTable("people")
```

Spark state...	Started At	Duration	Rows read	Bytes read	Bytes written
L4	df.w...	2024-04-16 08...	3 s 718 ms	10,000,018	211.00 MB

## Fast and easy to use

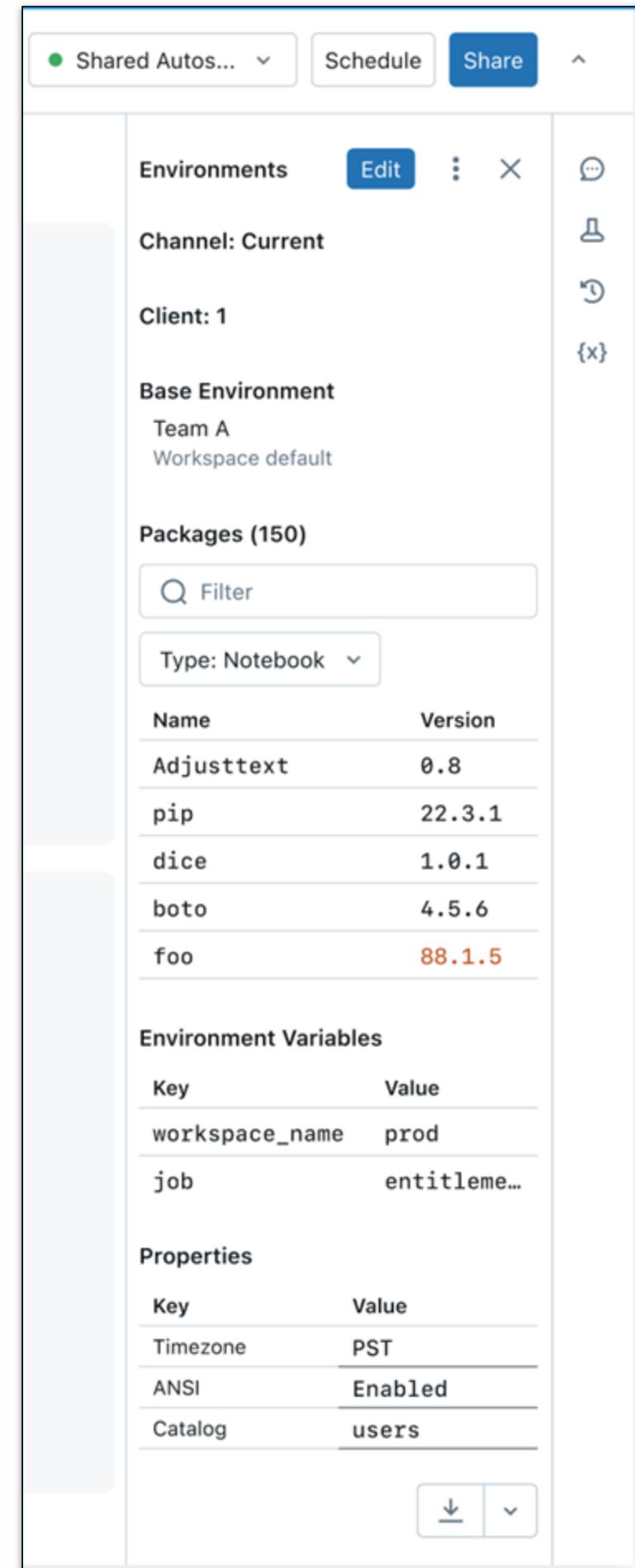
- Starts in seconds

## Portable, reusable environments

Automatically cached

## Base Python and package versions

Customizable by users or admins



Shared Autos... Schedule Share

Environments Edit Channel: Current Client: 1 Base Environment Team A Workspace default

Packages (150)

Name	Version
Adjusttext	0.8
pip	22.3.1
dice	1.0.1
boto	4.5.6
foo	88.1.5

Environment Variables

Key	Value
workspace_name	prod
job	entitlementme...

Properties

Key	Value
Timezone	PST
ANSI	Enabled
Catalog	users



# Query Insights

## Extended to SQL Queries & DataFrames on Serverless Compute

- **Query History**

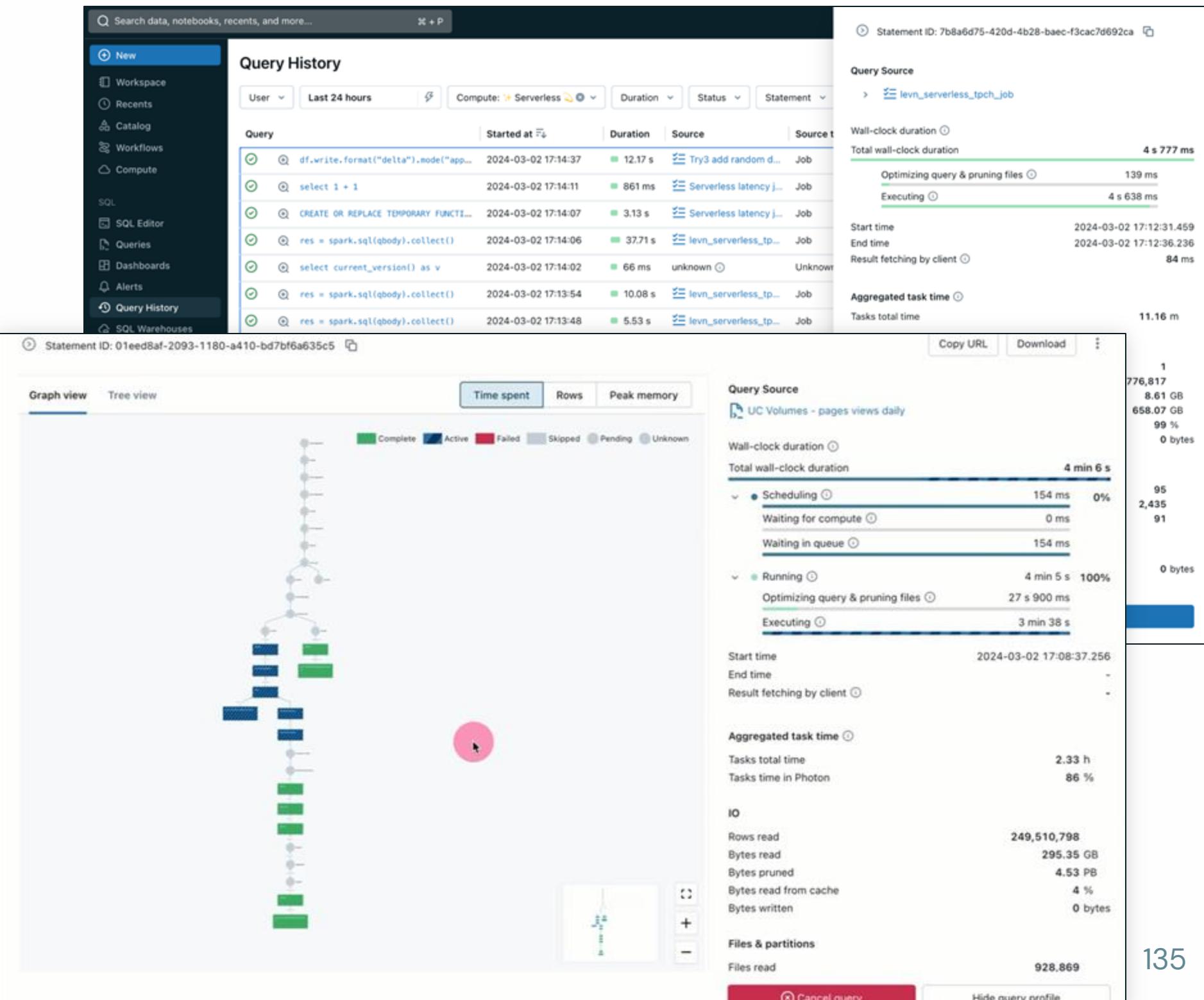
- Searchable and filterable overview of all current and past statements + cross-linking to source object that generated the query

- **Query Profile**

- Deep-dive into execution details to identify performance bottlenecks and optimization opportunities

- **Live Query Profile – coming soon**

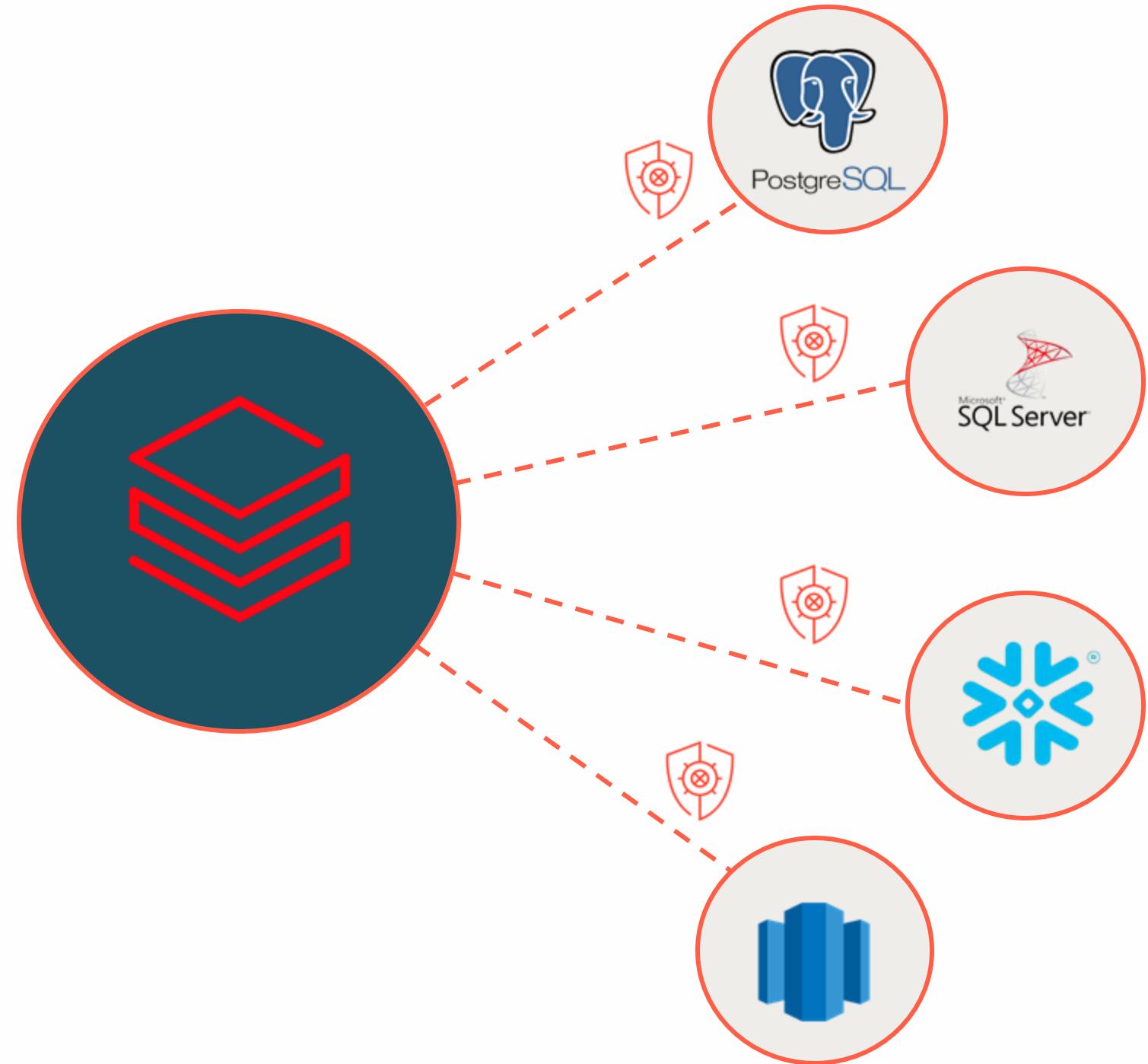
- Performance insights updated live as the query executes to enable immediate action



# Query Federation

Discover, query, and govern your data wherever it lives

- Connect data sources to Unity Catalog and set policies across them
  - Table, Row, Col, tags, etc.
- Efficient optimization & caching across sources



# AI Functions for DBSQL

Access Proprietary and Open Source LLMs directly within DBSQL

- Enable analysts to classify data with LLMs
- Extract actionable insights using AI language models.
- Generate product descriptions, summaries, and more by simply adding the ai\_query function.

```
SELECT
  sku_id,
  product_name,
  ai_query(
    "my-external-openai-chat",
    "You are a marketing expert for a winter
holiday promotion targeting GenZ. Generate a
promotional text in 30 words mentioning a 50%
discount for product: " || product_name
  )
FROM
  uc_catalog.schema.retail_products
WHERE
  inventory > 2 * forecasted_sales
```



# Autocomplete Code or Queries

- Generate code from comments inside a notebook cell or SQL editor

The screenshot shows the Databricks Assistant Demo - Code Completion interface. The left sidebar includes options like New, Workspace, Recents, Data, Workflows, Compute, SQL, SQL Editor, Queries, Dashboards, Alerts, Query History, and SQL Warehouses. The main area is titled "Databricks Assistant Demo - Code Completion" and shows a Python notebook cell labeled "Cmd 1". The cell contains the following code:

```
1 # Write code to reverse a string. Then display then the reversed string.
```

Below the cell, there are instructions: "Shift+Enter to run", "Shift+Ctrl+Enter to run selected text", and "Option+Shift+Space to suggest code". The top right of the interface shows user status (JJ), interrupt button, shared status (13.x Shared), scheduling button, and share button.

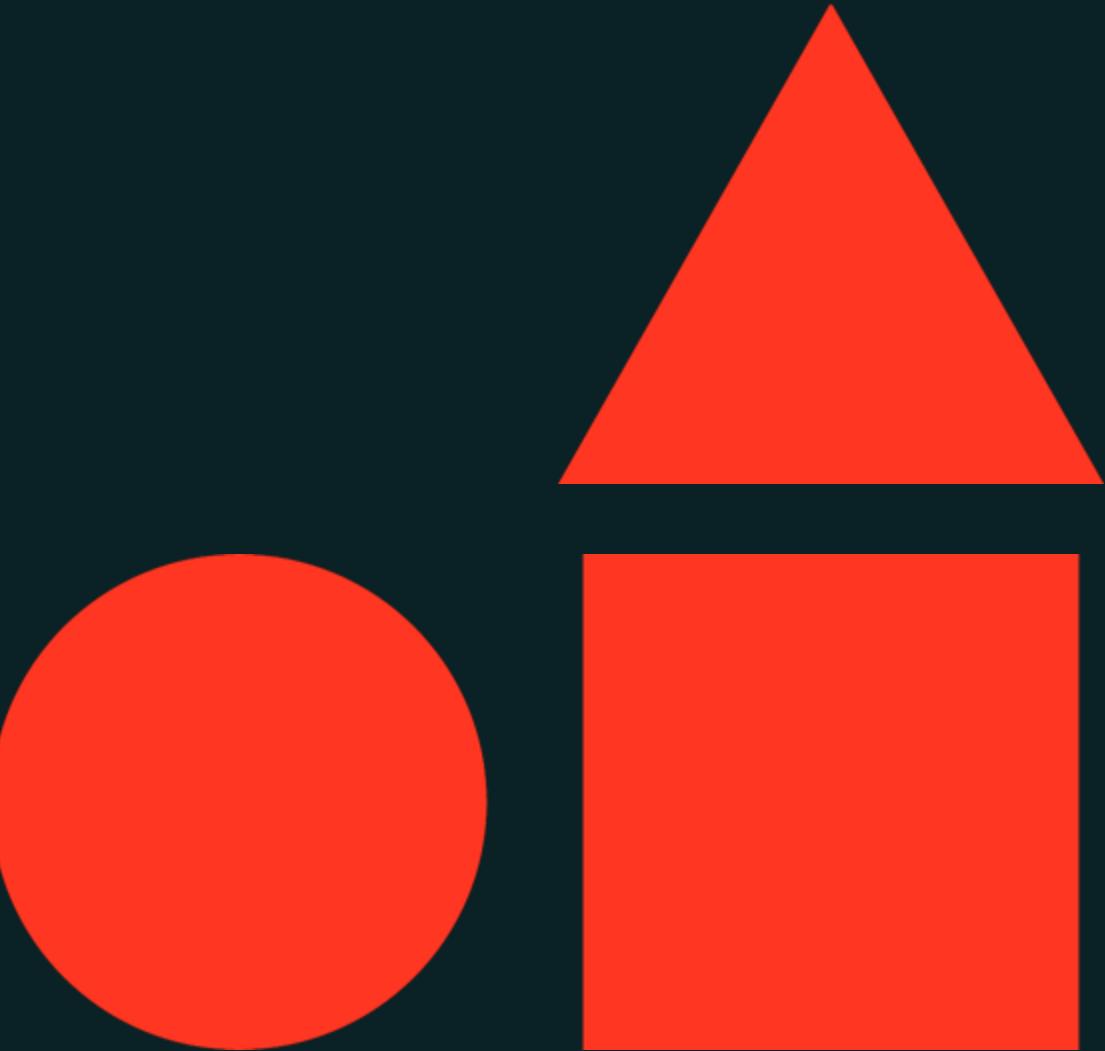




# Data Presentation

---

Get Started with Databricks for Data Warehousing



# Objectives

- Describe the impact of data warehousing work on Databricks AI/BI needs.
- Identify the AI supported features in the Databricks platform.



# Agenda

Data Presentation	Time	Lecture	Demo	Lab
Introduction to AI/BI	5 mins	✓		
Creating a Dashboard in Databricks	5 mins		✓	
Creating Genie Spaces	10 mins		✓	
Data Warehousing Comprehensive Lab	30 mins			✓





Data Presentation

LECTURE

# Introduction to AI/BI



# Infusing AI in our user experience

AI-generated docs for easier search and discoverability

This screenshot shows the Databricks Data Catalog interface. On the left, there's a sidebar with navigation links like 'Data', 'Catalog', 'Workflows', etc. In the center, a table named 'deposits' is selected. A modal window titled 'deposits' provides a detailed description: 'This table provides a comprehensive record of all deposit transactions.' It also lists columns: 'type' (string), 'amount' (int), 'date' (date), and 'email' (string). At the bottom of the modal, there are 'Edit' and 'Accept' buttons.

This screenshot shows the Databricks SQL Editor. On the left, a sidebar lists various notebooks and workspace items. In the main area, there's a 'New query' tab with a sample query: 'SELECT fare\_amount FROM main.nyctaxi.trips ORDER BY fare\_amount DESC LIMIT 10'. Below this, a 'User' section asks 'write a query to return the top 10 most expensive taxi cab fair amounts from main.nyctaxi.trips'. To the right, an 'Assistant' section shows the results of the query, listing fare amounts from 275.00 down to 105.00. The results table has a header row 'fare\_amount'.

AI Assistant for quick creation of visualizations

This screenshot shows the Databricks E2 Dogfood interface. On the left, a sidebar includes 'Compute', 'ML', 'Dashboards', and 'Jobs'. The main area is a 'Canvas' where a new dashboard is being created. A message at the top says 'Select a widget'. At the bottom, there are several small icons for different visualization types.

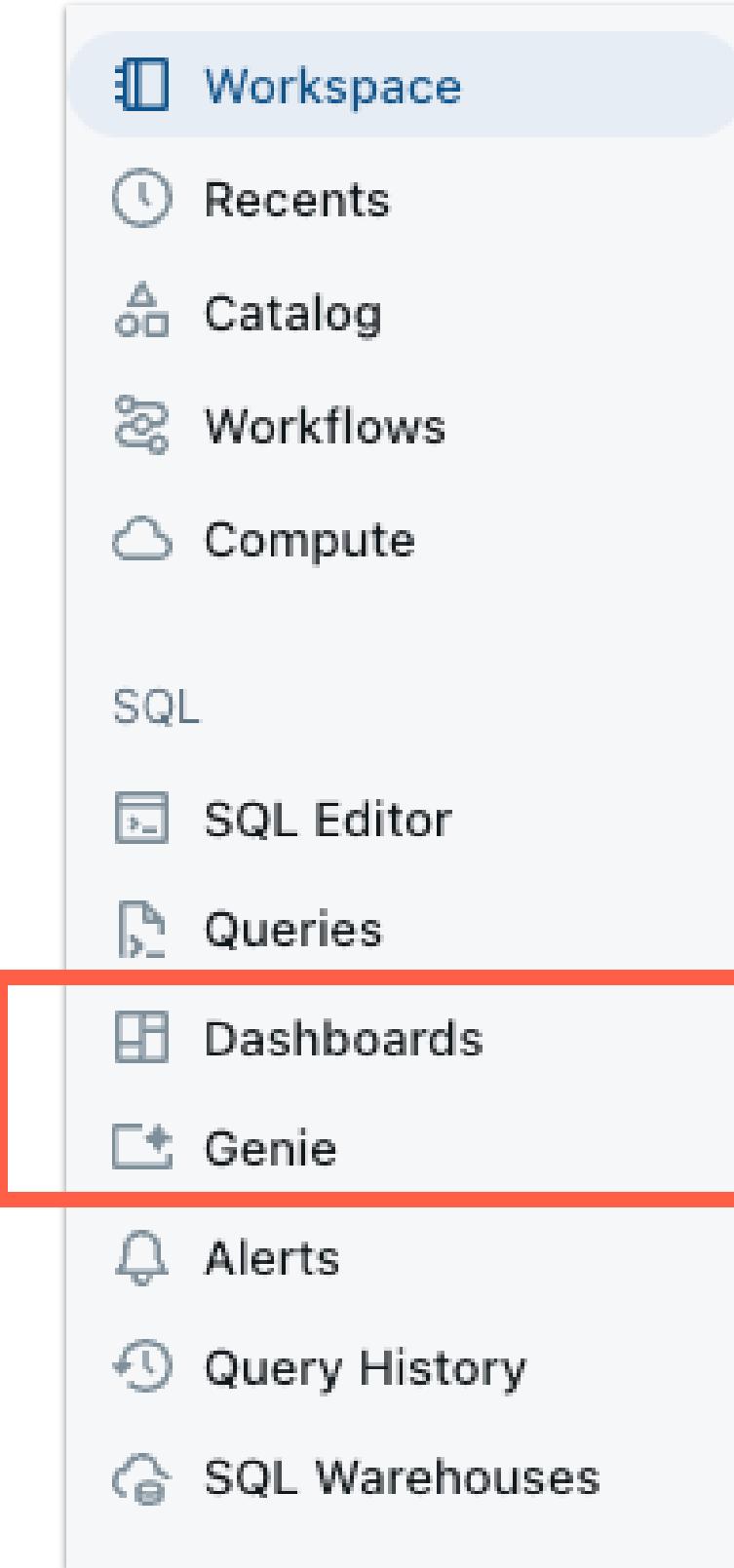
This screenshot shows the Databricks Genie Spaces interface. On the left, a sidebar lists 'EMEA Sales Analysis Data' and other workspace items. In the center, a 'New chat' window is open with the message 'Conversation history is currently disabled.' Below it, there are buttons for 'Example questions', 'Explain data set', and 'Surprise me'. At the bottom, a text input field contains the question 'Name all the unique stores with their footfall'.

AI Assistant for quick code fixes, queries and descriptions

AI-powered Genie Spaces to quickly understand your data using natural language



# Databrick's AI/BI Offerings



Located under the SQL header in the navigation:

## AI/BI Dashboards

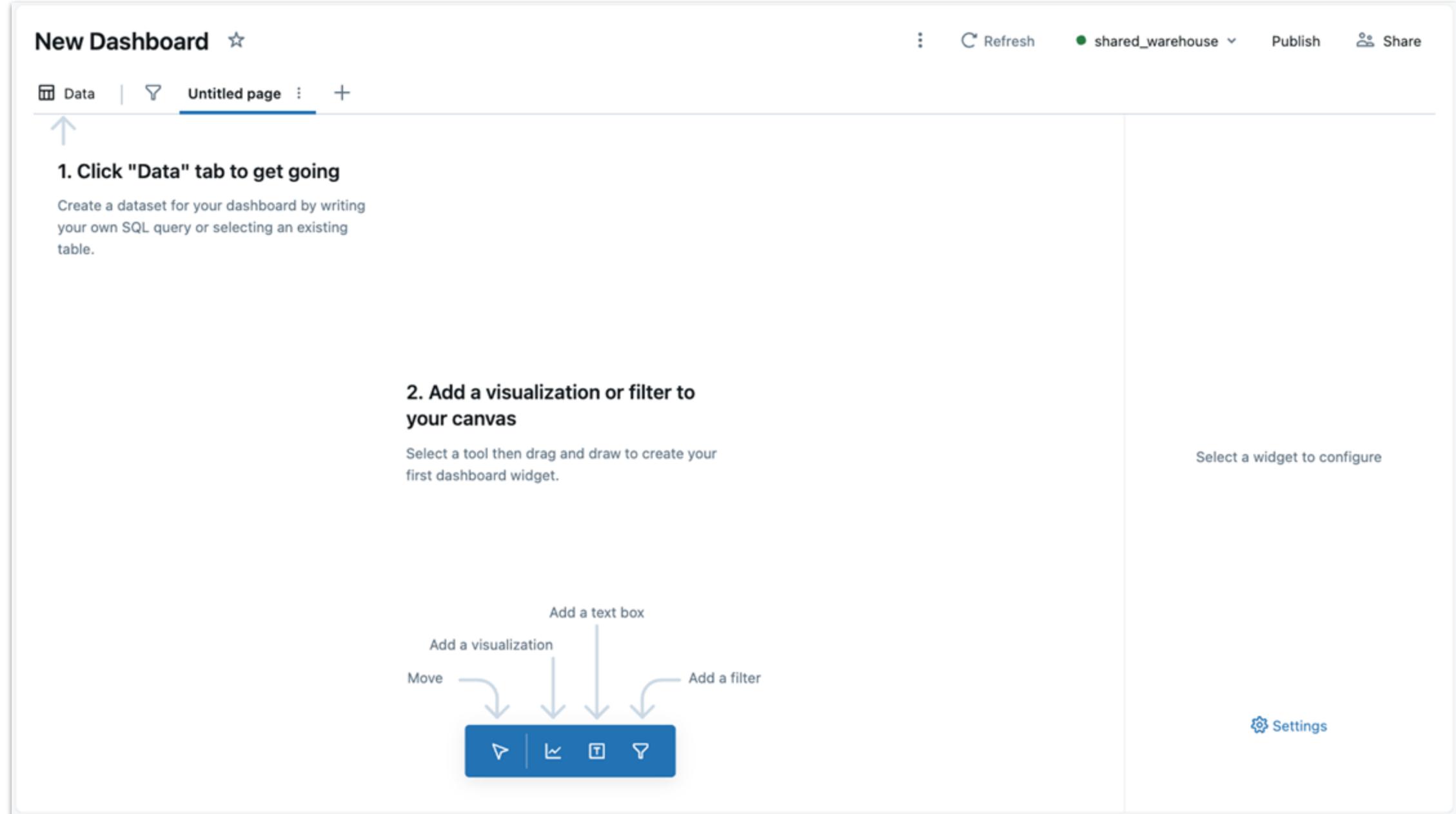
- Presentation environment for visualizations, titles, filters, and other features traditionally found in presentation software.

## AI/BI Genie

- Unique space associated to a specific curated data set for asking questions using natural language to gain further insight or alternative visualizations without needing a data analyst to oversee the development.



# AI/BI Dashboards



- Support for technical and non-technical users.
- Simple, easy to user interface.
- All-in-one sharing manages the permissions for the data, queries, and dashboard all from one interface.



# AI/BI Genie

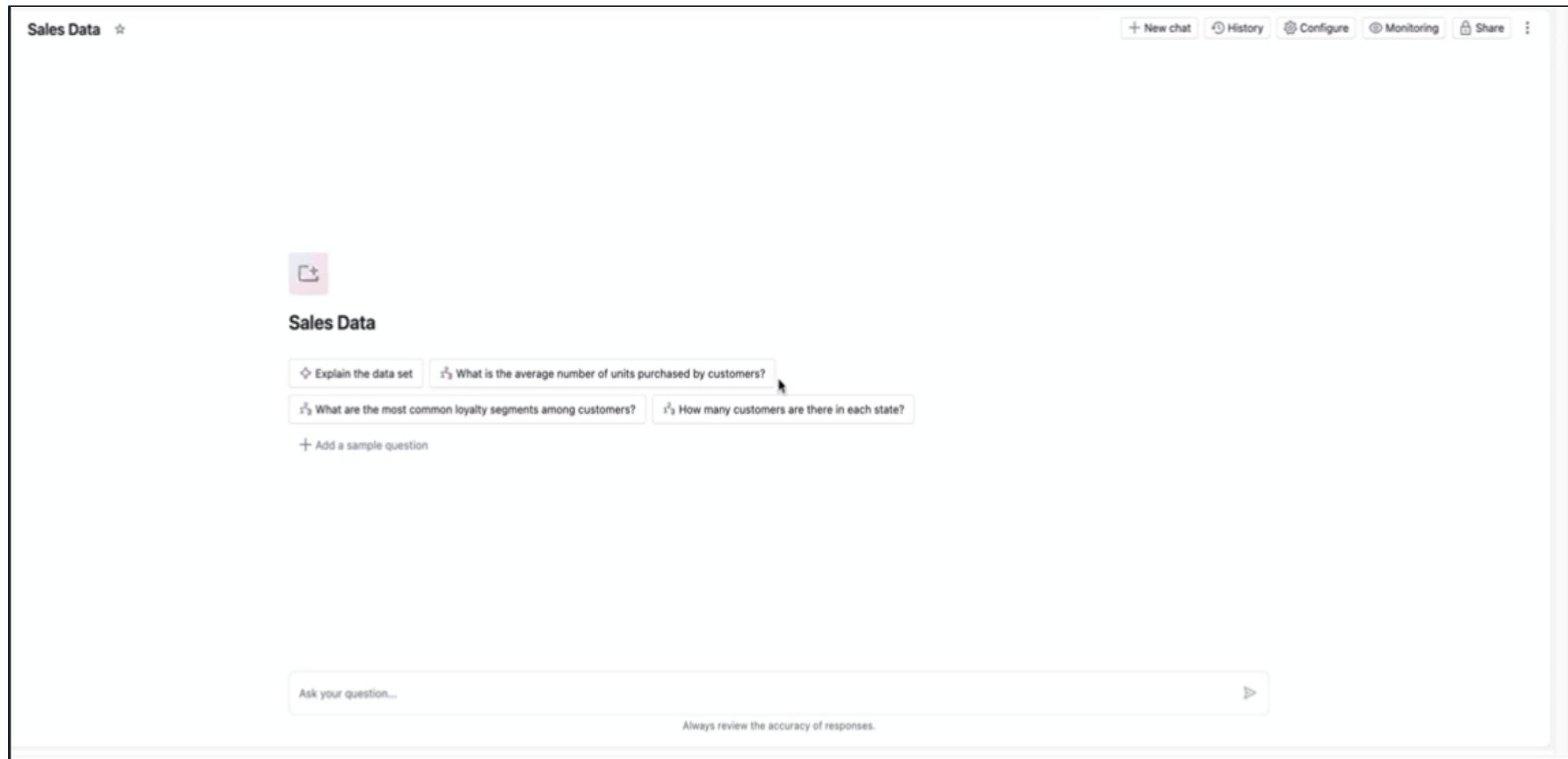
- Easily created straight from a Dashboard.
- Simple UI for creation outside of Dashboards.
- Supports follow-up discussions directly with the data.

The screenshot shows the Databricks AI/BI Genie interface. On the left, a 'New Space' dialog box is open, displaying various sample questions related to a dataset named 'customers'. On the right, the main workspace shows a context menu for a dashboard. The menu includes options like 'Open draft Genie space' (which is highlighted with a red box), 'Preview', 'Clone', 'Move', 'Move to trash', 'Export dashboard', 'Replace dashboard', and 'Send feedback'.



# AI/BI Genie Spaces

Ask questions of your data in natural language using your terms



- Supports non technical users with a no-code solution to gaining data insights using natural language.
- Provides a self-service access point to the data for follow-up questions without needing to involve a data analytics specialist.

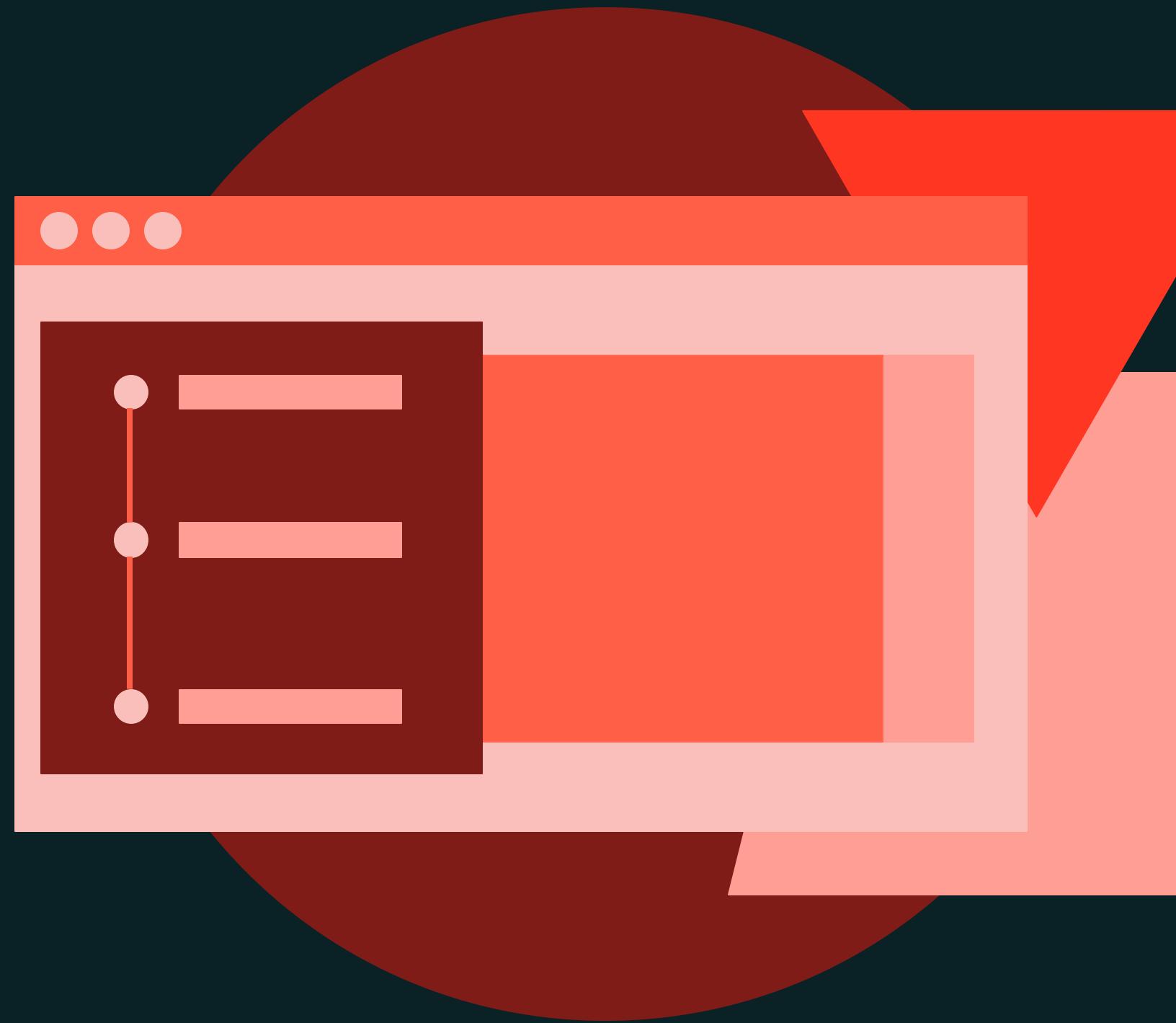




Data Presentation

**DEMONSTRATION**

# Creating a Dashboard in Databricks



# Follow-along Instructions

**Estimated Time: 5 minutes**

For this demonstration, the instructor will guide you through the Dashboards interface, including adding data sets both with the UI and through a query, creating visualizations, and adding features like filters to the dashboard. This demo closes with a brief introduction to sharing your dashboards with others in your business. If you have access to the Vocareum lab environment, feel free to follow-along to explore the workspace and get oriented with where future labs will be completed.

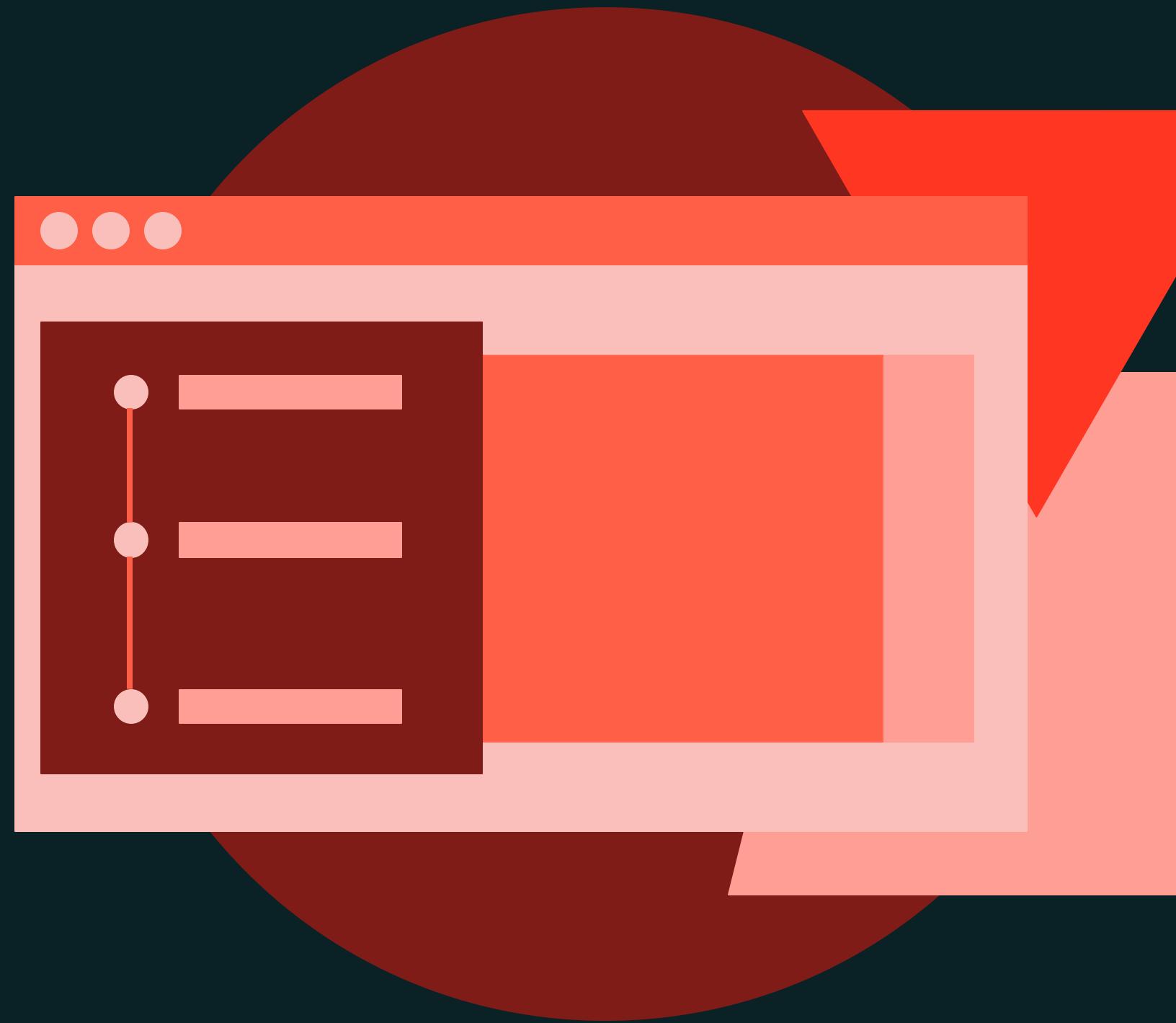




Data Presentation

**DEMONSTRATION**

# Creating AI/BI Genie Spaces



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# Demonstration Details

**Estimated Time: 5 minutes**

For this demonstration, the instructor will walk you through creating Genie Spaces both from a dashboard but also from scratch and discuss the capabilities of this tool.



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).

# High Level Steps

- Understand Databricks AI/BI Genie Spaces:
- Create Genie Spaces:
- Leverage Genie Space Features:
- Create Genie Spaces from Dashboards:
- Understand Genie's Monitoring and Preview Features:

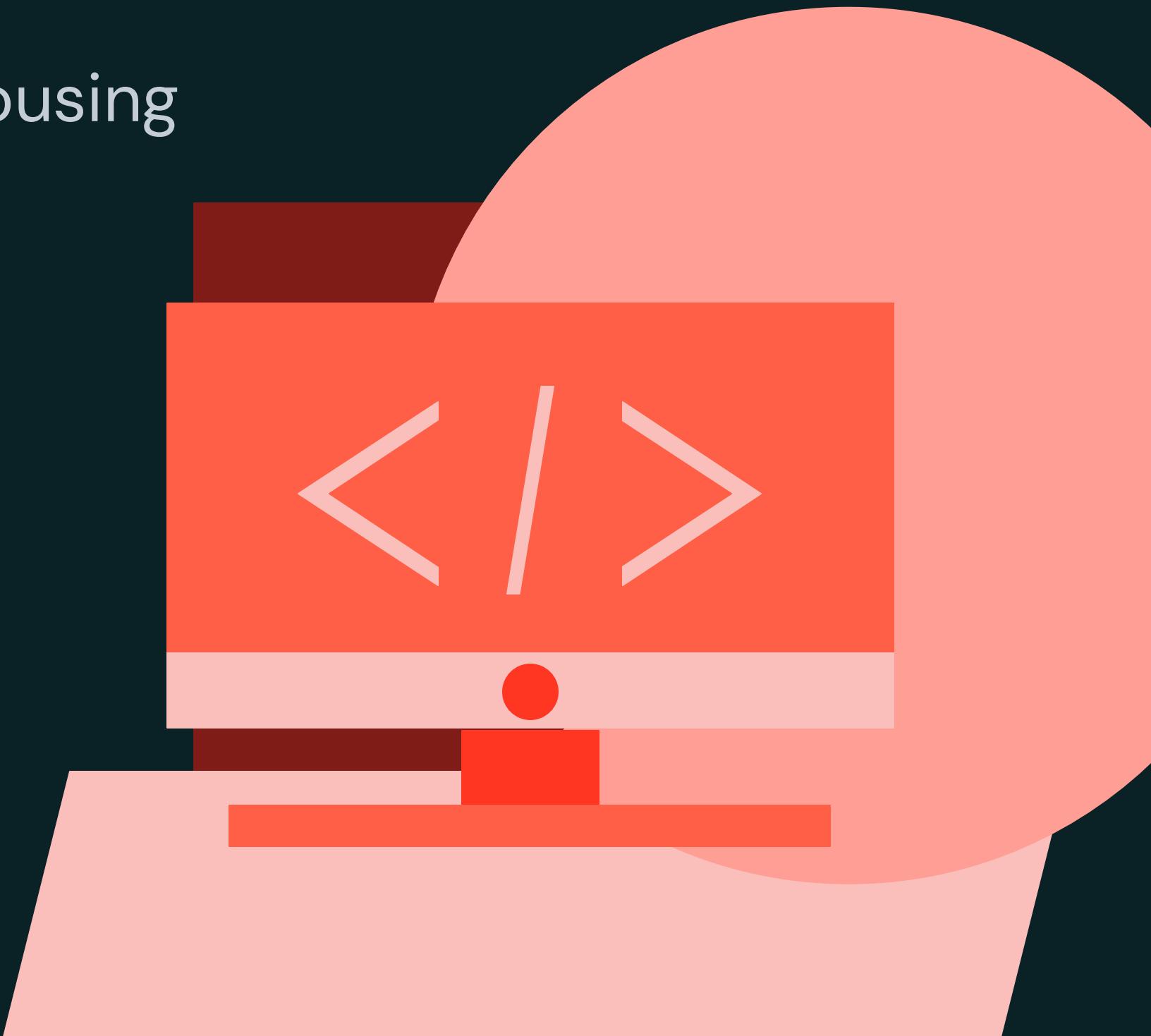




Get Started with Databricks for Data Warehousing

**LAB EXERCISE**

# Data Warehousing Comprehensive Lab



# Lab Exercise Goals

- What the learner is to accomplish within the lab environment as part of completing the lab.
- Explain how it relates to the lecture or demonstration preceding the exercise.
- Announce whether the lab will be completed with the instructor or if it's an independent exercise to be completed in the following X minutes.

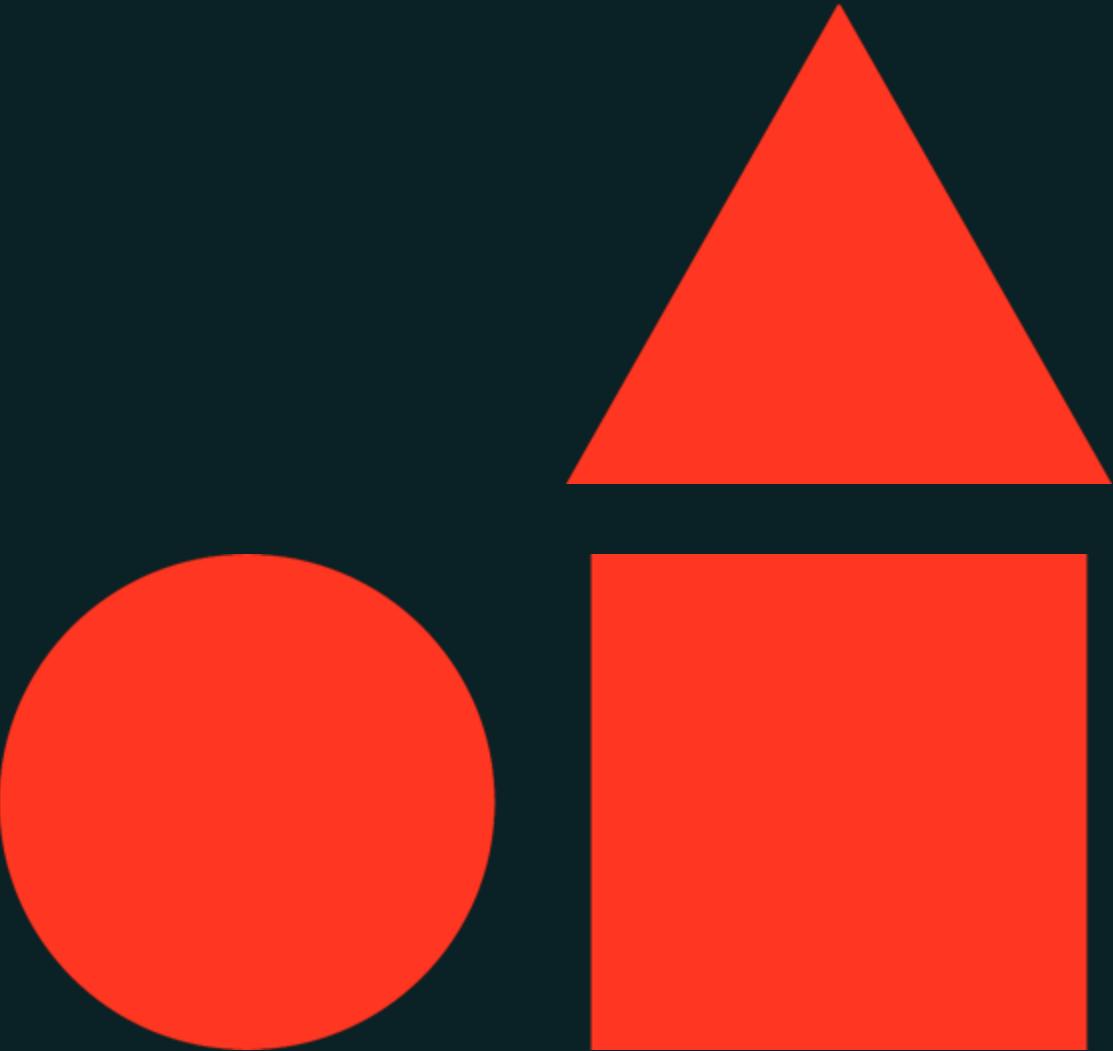




# Summary and Next Steps

---

[Get Started with Databricks for Data Warehousing](#)



# Course Learning Objective Recap

- Describe the available compute options for workloads performed on the Databricks Data Intelligence Platform.
- List the products and features Databricks offers for different data-centric needs within the Databricks Platform.
- Navigate the Databricks Workspace UI.
- Use Databricks to complete common data warehousing tasks.
- Explain the purpose of Delta Lake for data warehousing purposes.
- Describe and apply various techniques for ingesting and transforming data in Databricks.
- Explain how Databricks support data orchestration needs within the platform.
- Describe the impact of data warehousing work on Databricks AI/BI needs.
- Identify the AI supported features in the Databricks platform.





# databricks



© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the [Apache Software Foundation](#).