



Data Analysis with Databricks SQL



Databricks Academy

2023

Meet your classmates

- Where is everyone joining us from today (city, country)?

Meet your classmates

- How long have you been working with Databricks SQL?

Meet your classmates

- What data analysis tools have you worked with in the past?

Meet your classmates

- What are you hoping to get out of this class?

Getting Started with the Course

Course goals

1 Describe how Databricks SQL:

- Works in the lakehouse architecture
- Implements data security

2 Use Databricks SQL to:

- Ingest data into the Platform
- Use SQL commands specific to Databricks
- Query data
- Create visualizations and dashboards
- Share queries and dashboards with others
- Use automation and integration capabilities

Agenda

Module Name	Duration
Databricks SQL Services and Capabilities	1 hour, 50 min
Data Management in Databricks SQL	1 hour, 40 min
Data Visualization and Dashboarding	3 hours, 30 min

- We will take 10 minute breaks about every hour

Activities used in this course

Lectures

Demos

Labs

Knowledge checks



Technical Environment Overview

The Databricks SQL workspace

- Everyone:
 - is in the same workspace
 - has their own catalog
 - is using the same SQL warehouse
- Only the instructor has administrator privileges in the workspace
 - only a select few tasks in this course require admin privileges
 - you will see these tasks in the slides in order to provide context
 - the labs do not require admin privileges

Databricks SQL Services and Capabilities



Databricks Academy 2023

Databricks SQL Services and Capabilities

Lesson Name	Duration
Lecture: Getting Started with Databricks SQL	15 min
Follow Along Demo: Setting Up a Catalog and Schema	20 mins
Follow Along Demo: Data Importing	20 mins
Follow Along Demo: A Simple (but Quick) Query, Visualization, and Dashboard	15 min
Lecture: Unity Catalog in Databricks SQL	20 min
Lecture: Lakehouse Architecture	20 min
Lecture: Integrations	10 min

Getting Started with Databricks SQL

Learning Objectives

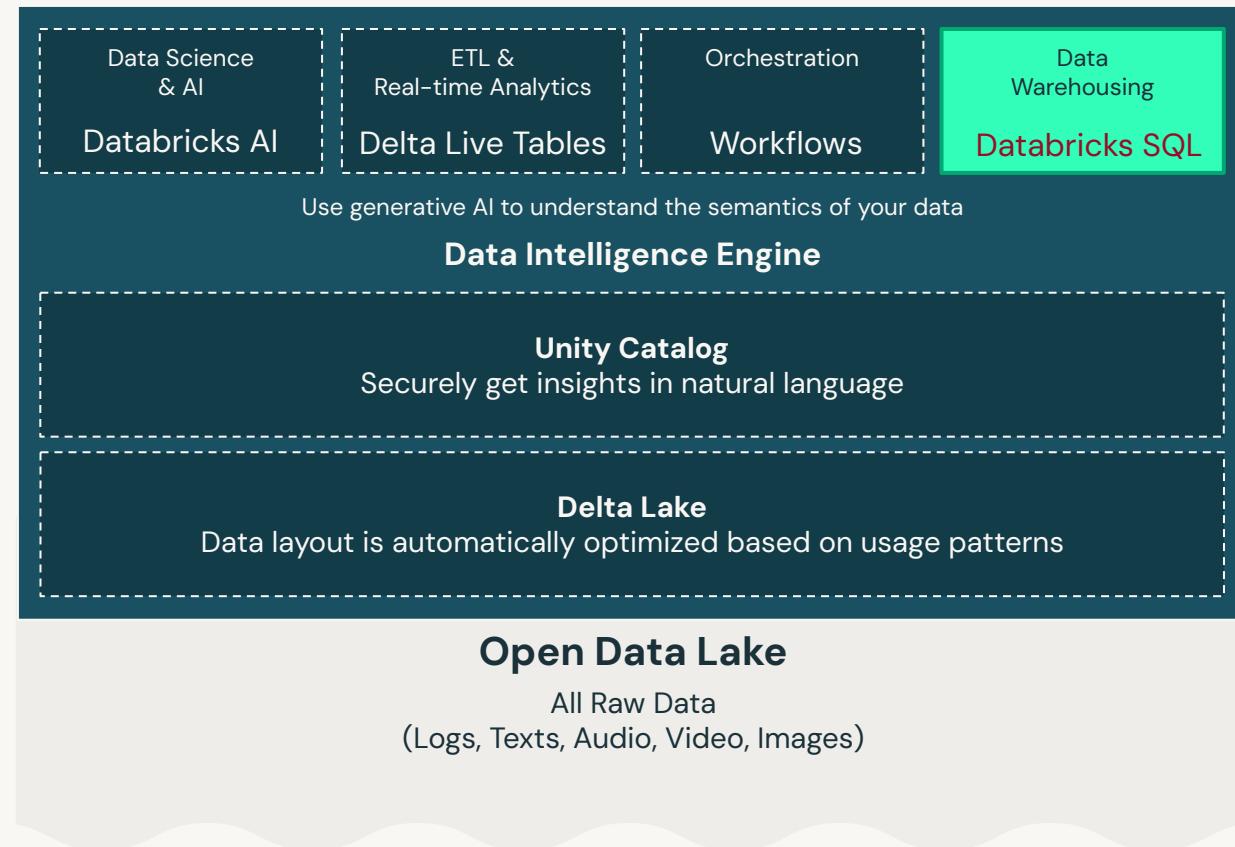
By the end of this lesson, you should be able to:

- 1** Describe what Databricks SQL is.
- 2** Describe the benefits of Databricks SQL.

Databricks SQL

Delivering analytics on the freshest data with data warehouse performance and data lake economics

- Better price / performance than other cloud data warehouses
- Simplify discovery and sharing of new insights
- Connect to familiar BI tools, like Tableau or Power BI
- Simplified administration and governance



Better price / performance

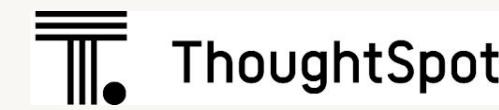
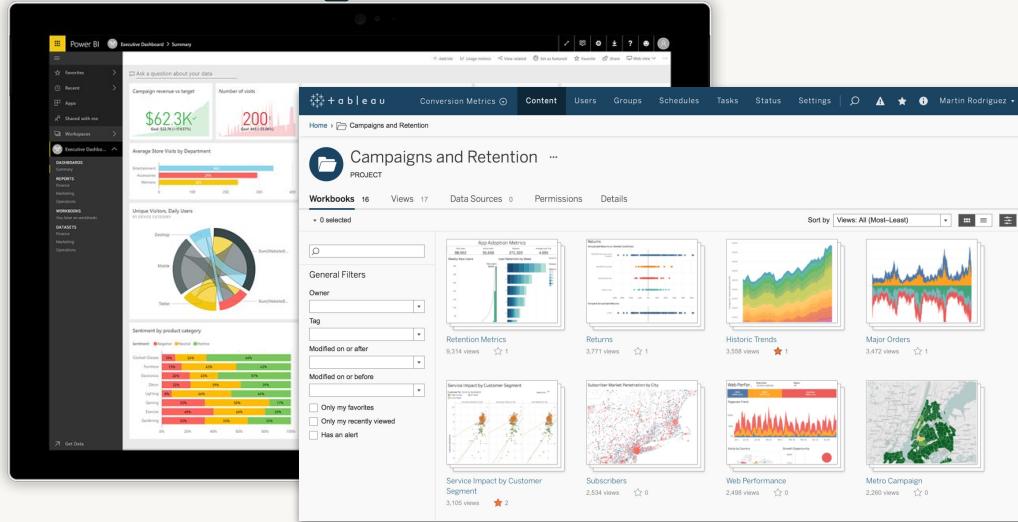
Run SQL queries on your lakehouse and analyze your freshest data with **up to 6x better price/performance** than traditional cloud data warehouses.



Source: Performance Benchmark with Barcelona Supercomputing Center

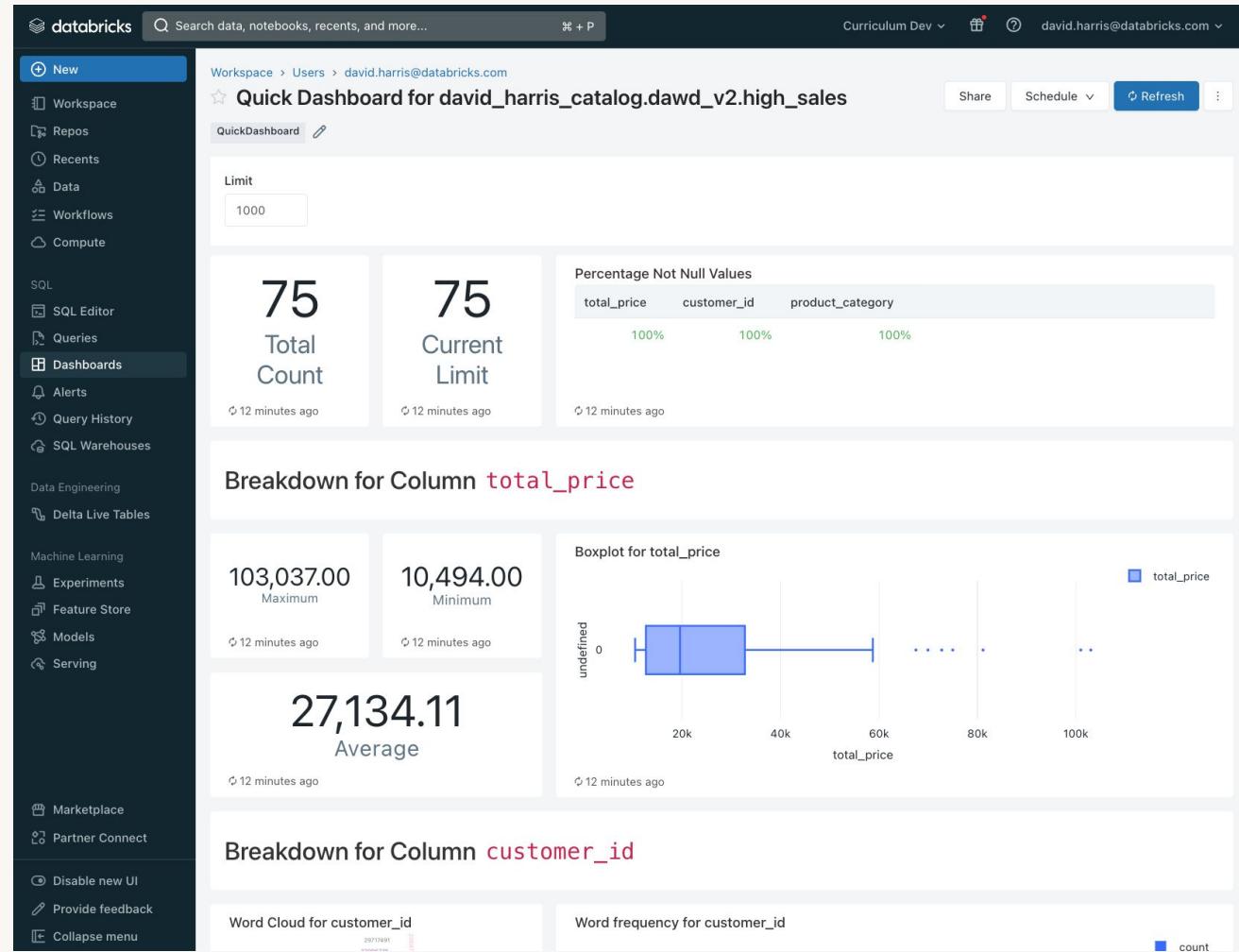
Better together | Broad integration with BI tools

Connect your preferred BI tools with optimized connectors that provide fast performance, low latency, and high user concurrency to your data lake for your existing BI tools.



A new home for data analysts

Enable data analysts to quickly perform ad-hoc and exploratory data analysis, with a new SQL query editor, visualizations and dashboards. Automatic alerts can be triggered for critical changes, allowing to respond to business needs faster.



Simple administration and governance

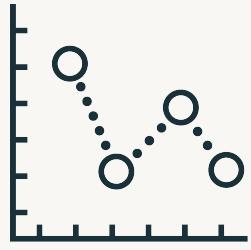
Quickly setup SQL / BI optimized compute with SQL warehouses. Databricks automatically determines instance types and configuration for the best price/performance. Then, easily manage usage, perform quick auditing, and troubleshooting with query history.

The image displays two side-by-side screenshots of the Databricks web interface. Both screenshots show the 'DBAcademy Warehouse' under the 'SQL Warehouses' section. The left screenshot shows the 'Overview' tab, which provides a summary of the warehouse's status (Running), name (DBAcademy Warehouse), type (Serverless), cluster size (2X-Small), and auto-stop settings (After 20 minutes of inactivity). It also lists scaling, channel, and tags information. The right screenshot shows the 'Monitoring' tab, which includes a timeline chart for peak query count and another for running clusters over a 24-hour period. The monitoring tab also displays active sessions (0), status (Running), and clusters (1).

Use Cases



Connect existing BI tools to one source of truth for all your data



Collaboratively explore the latest and freshest data



Build data-enhanced applications

Follow Along Demo: Setting Up a Catalog and Schema



Follow Along Demo

Setting Up a Catalog and Schema

- Get your username for the course
- Create a catalog
- Create a schema

Follow Along Demo: Data Importing

Follow Along Demo

Data Importing

- Upload a .csv file
- Use the Catalog Explorer
- Create a table with data from object store

Follow Along Demo: A Simple (but Quick) Query, Visualization, and Dashboard

Follow Along Demo

A Simple (but Quick) Query, Visualization, and Dashboard

- Query data
- Create a visualization
- Create a dashboard

Knowledge Check

Knowledge check

Think about this question and volunteer an answer

Which of the following statements about Databricks SQL is true? Select one response.

- A. With Databricks SQL, queries deliver up to 2x better price/performance than other cloud data warehouses.
- B. Delta Live Tables can be created in Databricks SQL.
- C. Databricks SQL automatically configures scaling when creating SQL warehouses.
- D. Databricks SQL clusters are powered by Photon.

Knowledge check

Think about this question and volunteer an answer

**Which of the following features is used by Databricks SQL to ensure your data is secure?
Select one response.**

- A. Built-in data governance
- B. Delta sharing
- C. Integration with 3rd party tools
- D. Automatically scalable cloud infrastructure

Knowledge check

Think about this question and volunteer an answer

Which of the following features of Databricks is used for running queries in Databricks SQL?

Select one response.

- A. Dashboards
- B. Job scheduler
- C. SQL Editor
- D. SQL warehouses

Knowledge check

Think about this question and volunteer an answer

What is the primary purpose of Databricks SQL?

- A. To provide better price/performance and simplify discovery for BI tools.
- B. To manage administration and governance of data warehouses.
- C. To support a broad set of BI tools, including Tableau and Power BI.
- D. All of the above.

Unity Catalog in Databricks SQL



Learning Objectives

By the end of this lesson, you should be able to:

- 1 Describe the three-level namespacing system provided by Unity Catalog.
- 2 Describe persistence and scope of catalogs, schemas (databases), tables, and views on Databricks.
- 3 Compare and contrast the behavior of managed and unmanaged tables.

Learning Objectives, cont.

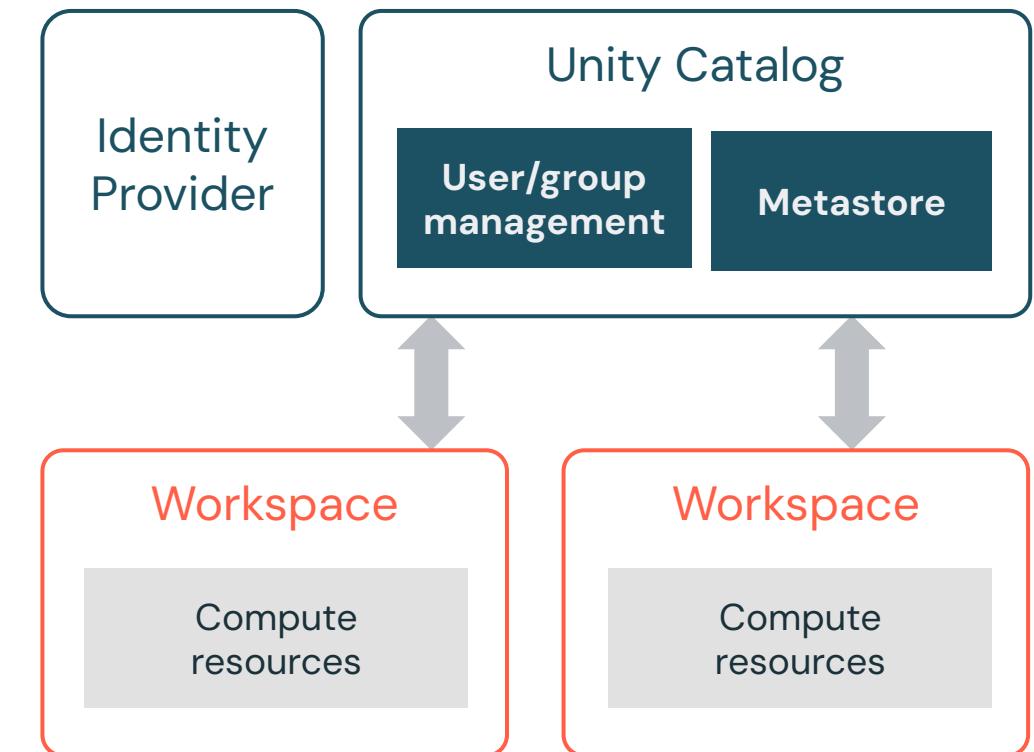
By the end of this lesson, you should be able to:

- 4 Identify the legacy `hive_metastore` as appearing as the default catalog to be compatible with Unity Catalog.
- 5 Describe how and where Unity Catalog stores the data behind its catalogs, schema, and granular data objects.
- 6 Explain the impact of Unity Catalog on existing external storage locations.

Unity Catalog

Architecture

- Implements access control on data
- Access control is always enabled
- Works across multiple workspaces
- Grants permissions to users at the account level



Unity Catalog

Comparison to workspace security model

Workspace security model

Object/privilege/principal access control model

Open by default

Local to workspace

Grants privileges to workspace-level principals

Unity Catalog security model

Object/privilege/principal access control model

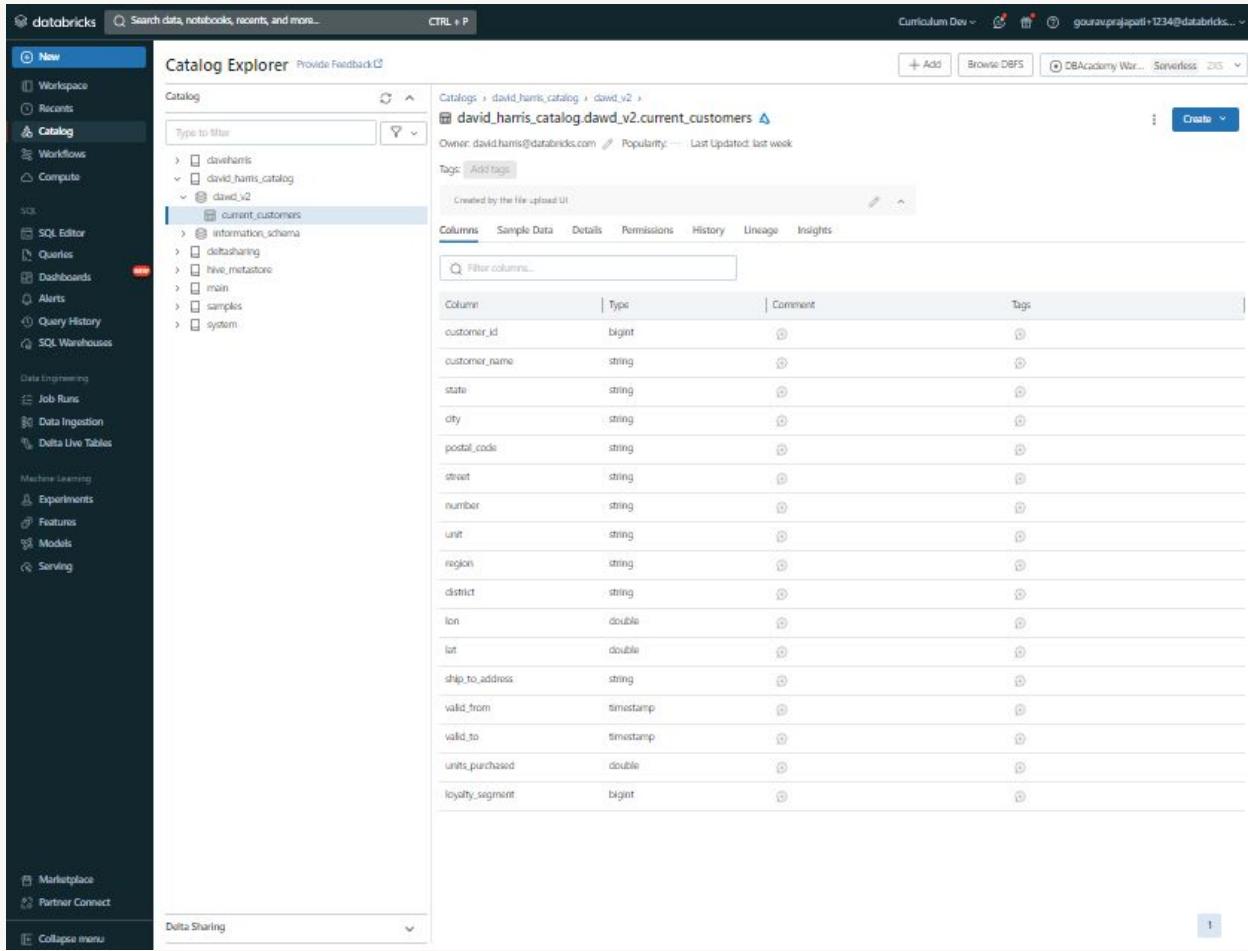
Secure by default

Works across multiple workspaces

Grants privileges to account-level principals

Catalog Explorer UI

Single pane of glass for all of your data



The screenshot shows the Databricks Catalog Explorer interface. On the left is a dark sidebar with navigation links like Workspace, Recents, Catalog, Workflows, Compute, SQL, Data Engineering, Machine Learning, and more. The main area is titled 'Catalog Explorer' and shows a tree view of catalogs: 'davisharris', 'david_harris_catalog', and 'david_v2'. Under 'david_v2', 'current_customers' is selected. Below the tree is a table of columns with their types and comments. At the bottom of the table are 'Sample Data', 'Details', 'Permissions', 'History', 'Lineage', and 'Insights' tabs.

Column	Type	Comment	Tags
customer_id	bigint		
customer_name	string		
state	string		
city	string		
postal_code	string		
street	string		
number	string		
unit	string		
region	string		
district	string		
lon	double		
lat	double		
ship_to_address	string		
valid_from	timestamp		
valid_to	timestamp		
units_purchased	double		
loyalty_segment	bigint		



UI driven access control to simplify secure data permissioning



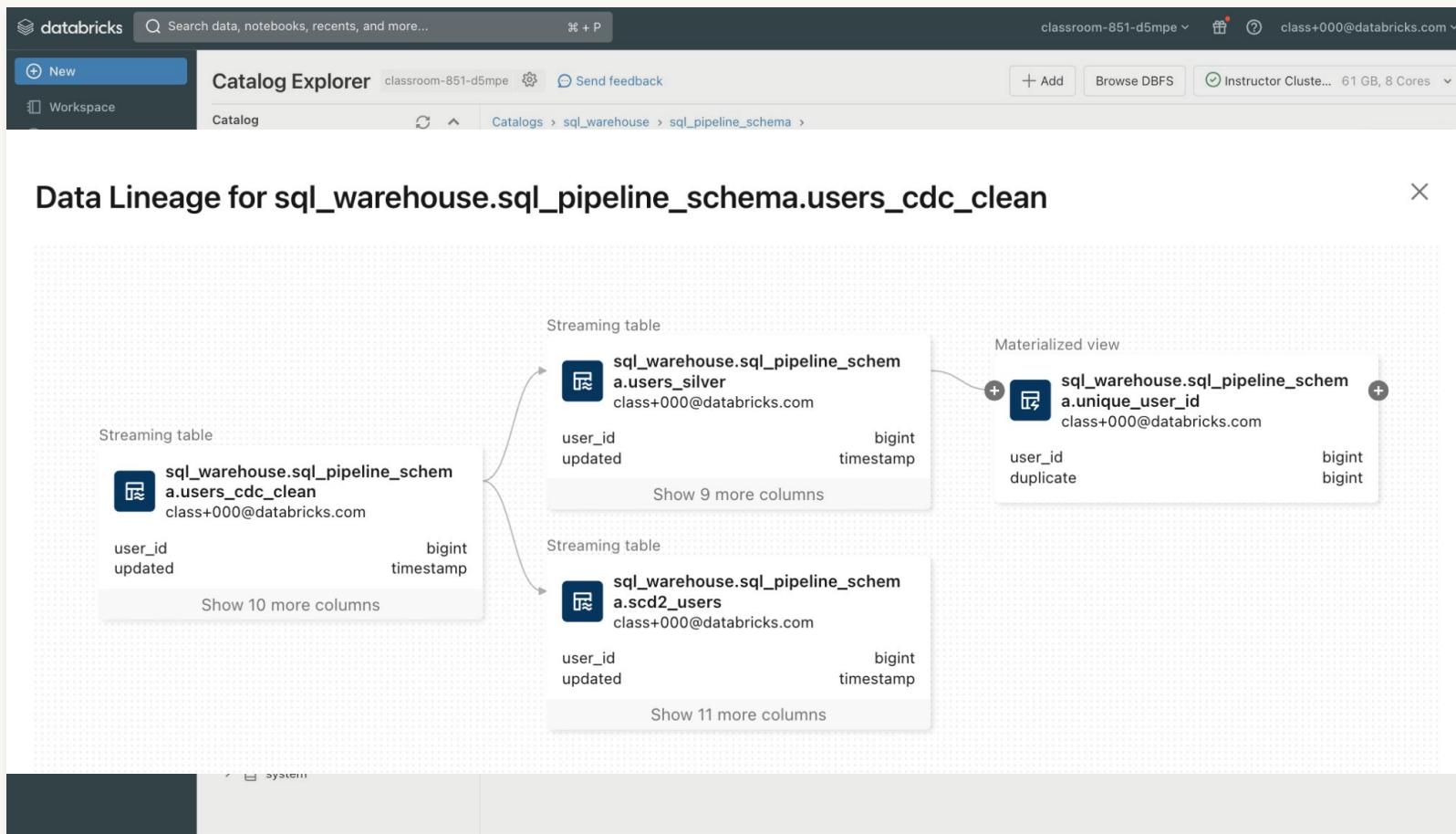
Browse and understand data assets stored in your Lakehouse



Data lineage
End-to-end table & column lineage

Data Lineage

Mapping the flow of data in the lakehouse



Auto-capture runtime data lineage across all languages

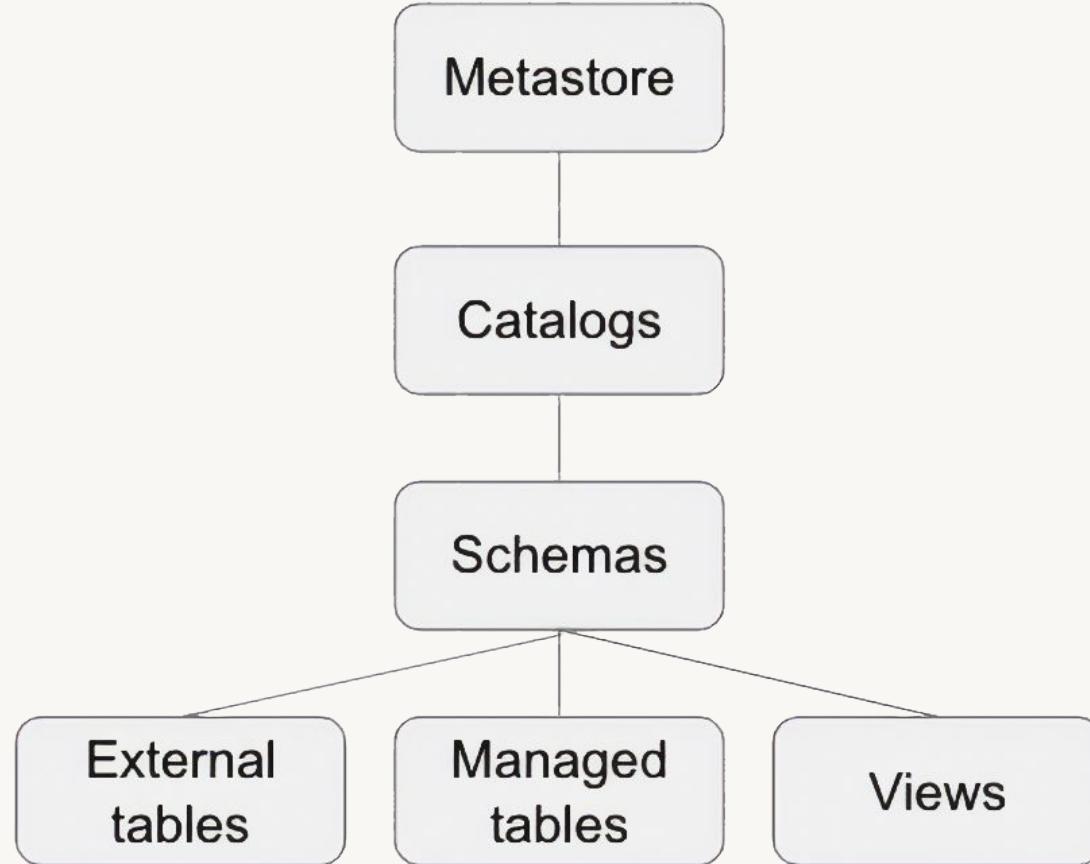


Track lineage down to the **table and column level**



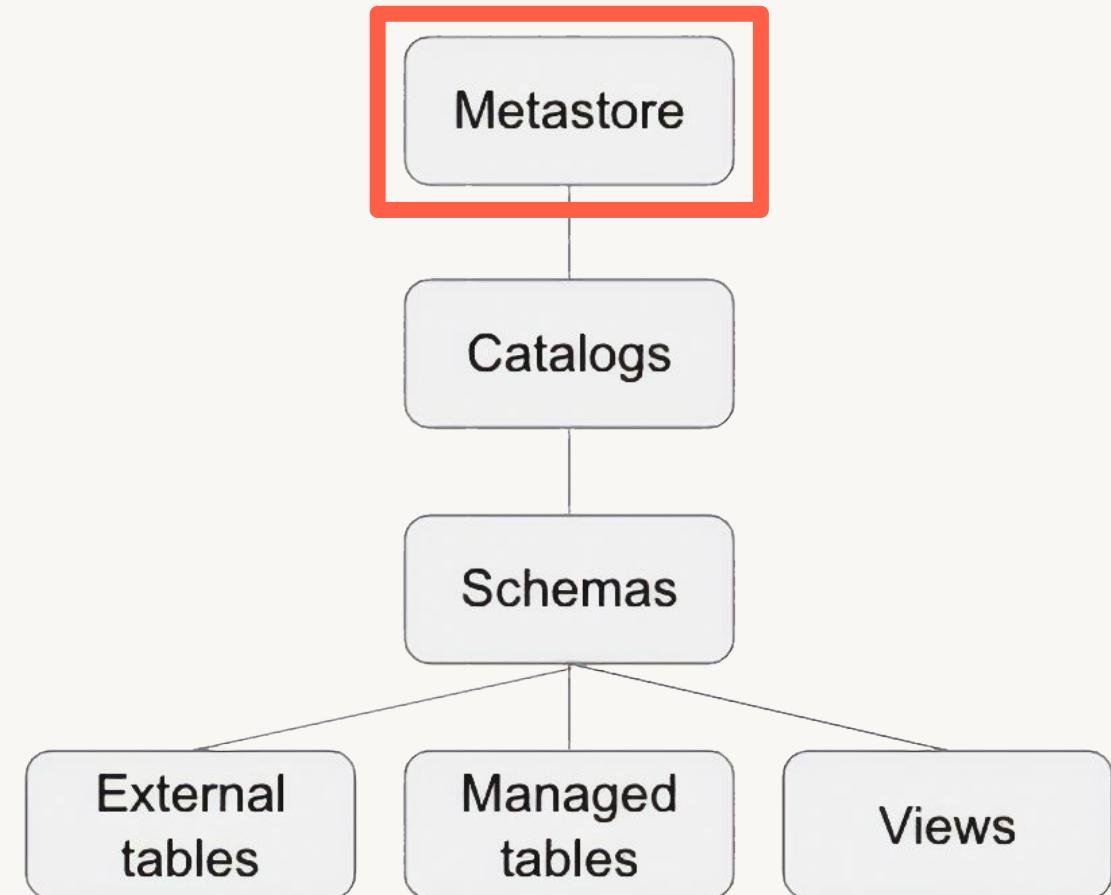
Leverage **common permission model** from Unity Catalog

Object Model



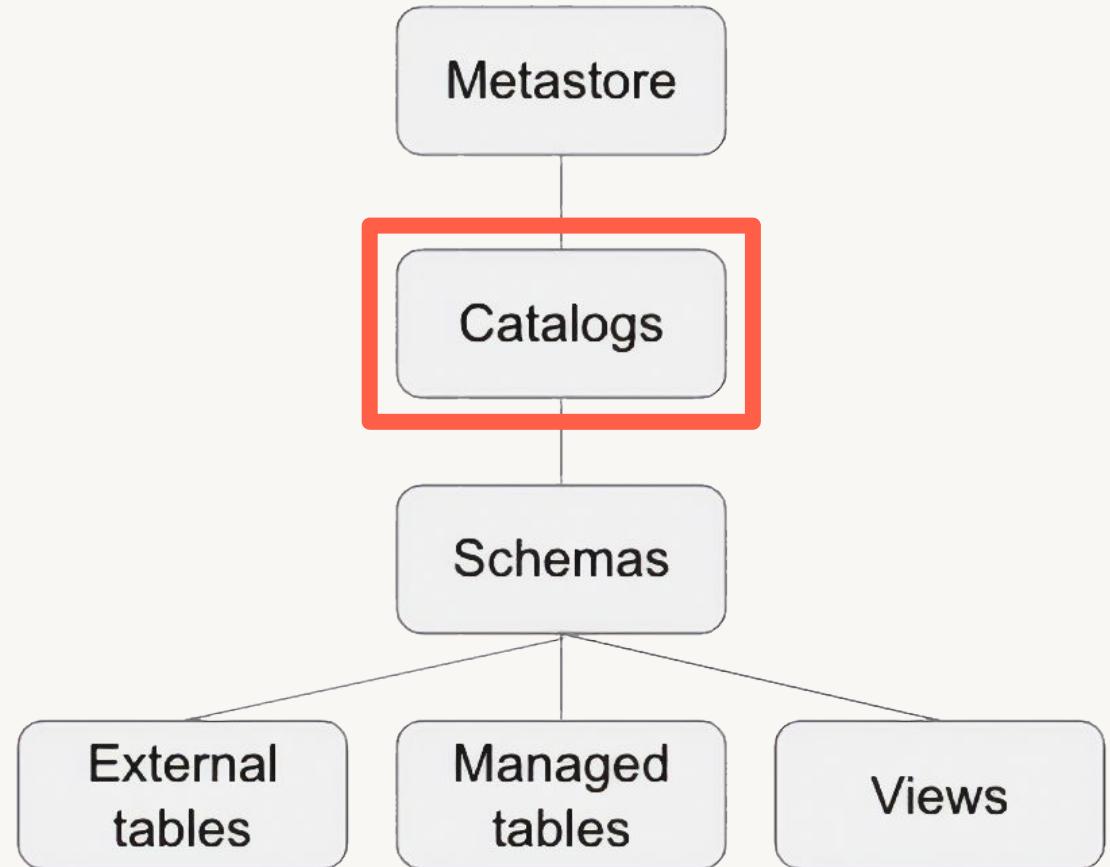
Metastore

- Stores data assets
- Permissions
- Created with default storage location (external object store)
- Metastore Admin



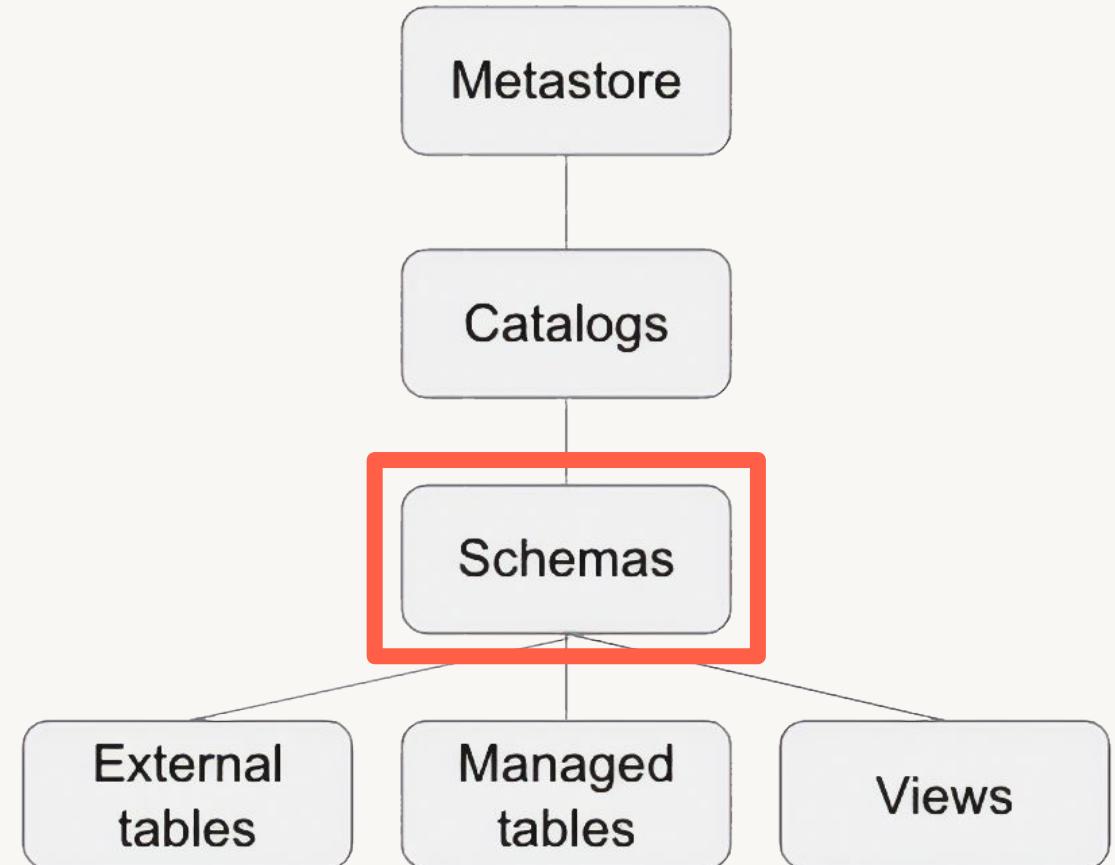
Catalog

- First level of organization
- Users can see all catalogs where USAGE is granted



Schema

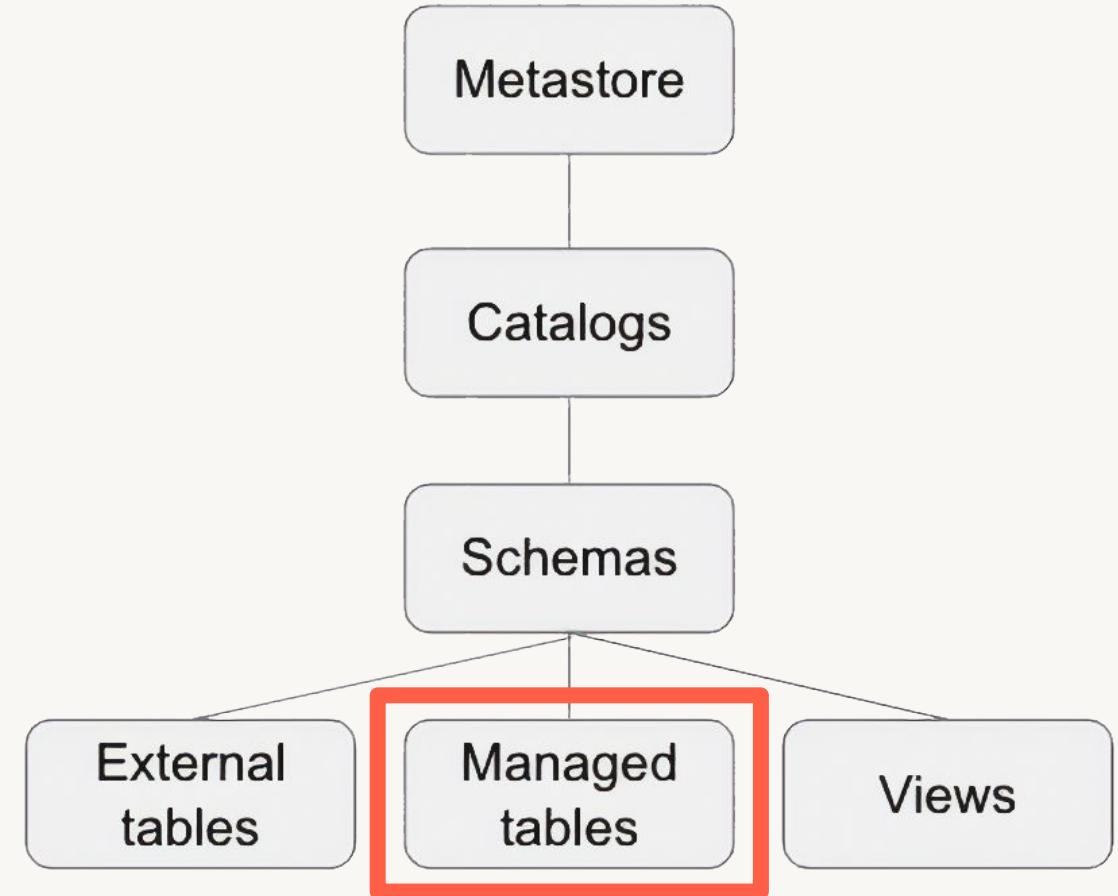
- aka, Database
- Second level of organization
- Users can see all schemas where USAGE is granted on both the schema and the catalog



Managed Table

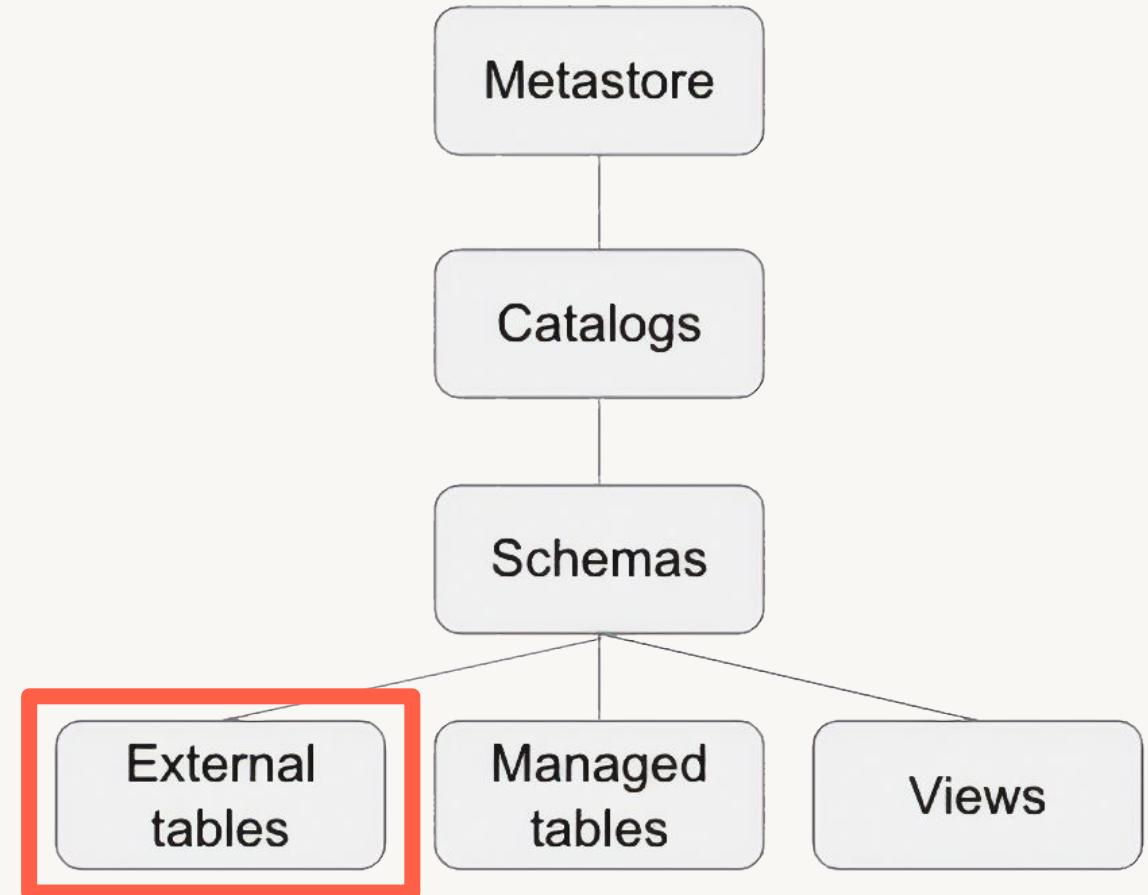
- Third level of organization
- Supported format: Delta
- Data is written to a new directory in the metastore's default location
- Created using CREATE TABLE statement with no LOCATION clause
- Example:

```
CREATE TABLE table1 ...
```



External Table

- Third level of organization
- Data stored in a location outside the managed storage location
- DROP TABLE does not delete data
- Can easily clone a table to a new schema or table name without moving data
- Supported formats:
 - Delta, csv, json, avro, parquet, orc, text



Creating External Tables

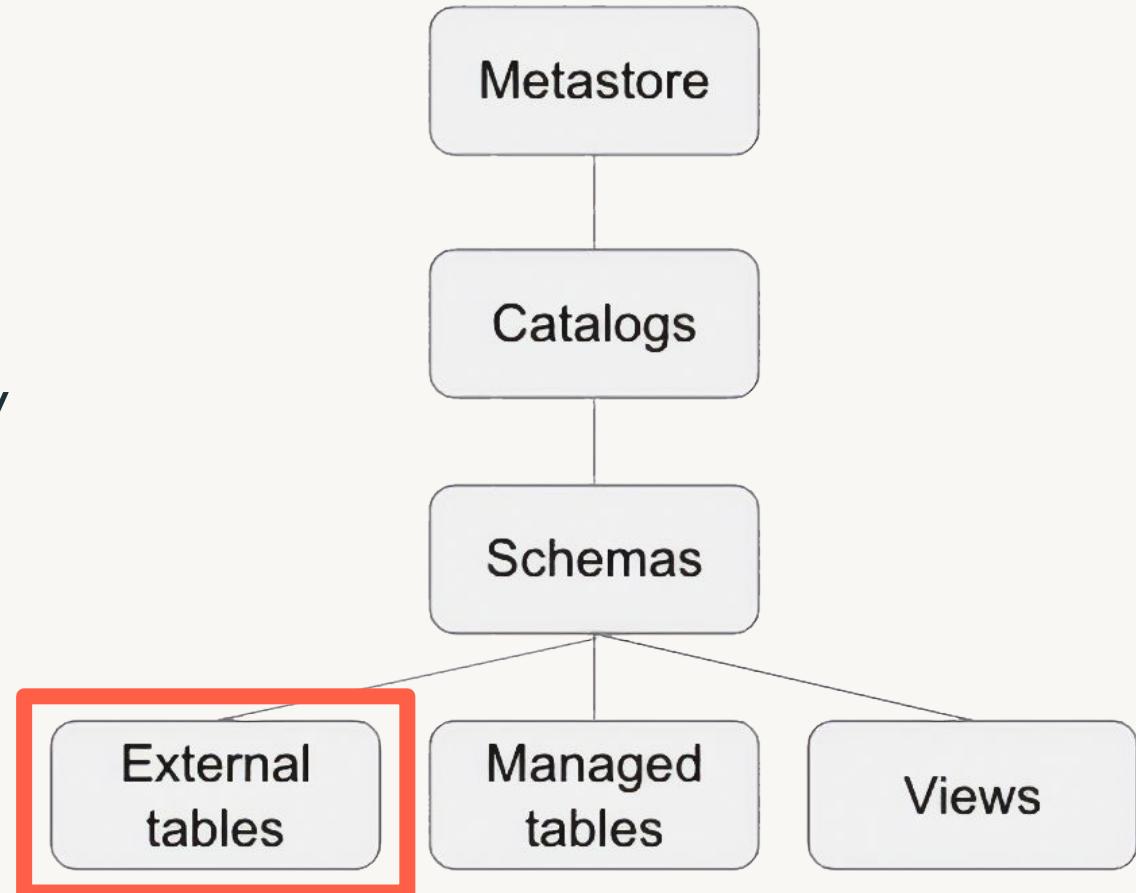
- Two credential types:
 - Storage Credential or External Location
- Use the LOCATION clause
- Example using External Location only

```
CREATE TABLE table2
  LOCATION 's3://<bucket_path>/<table_directory>'
...

```

- Example using Storage Credential

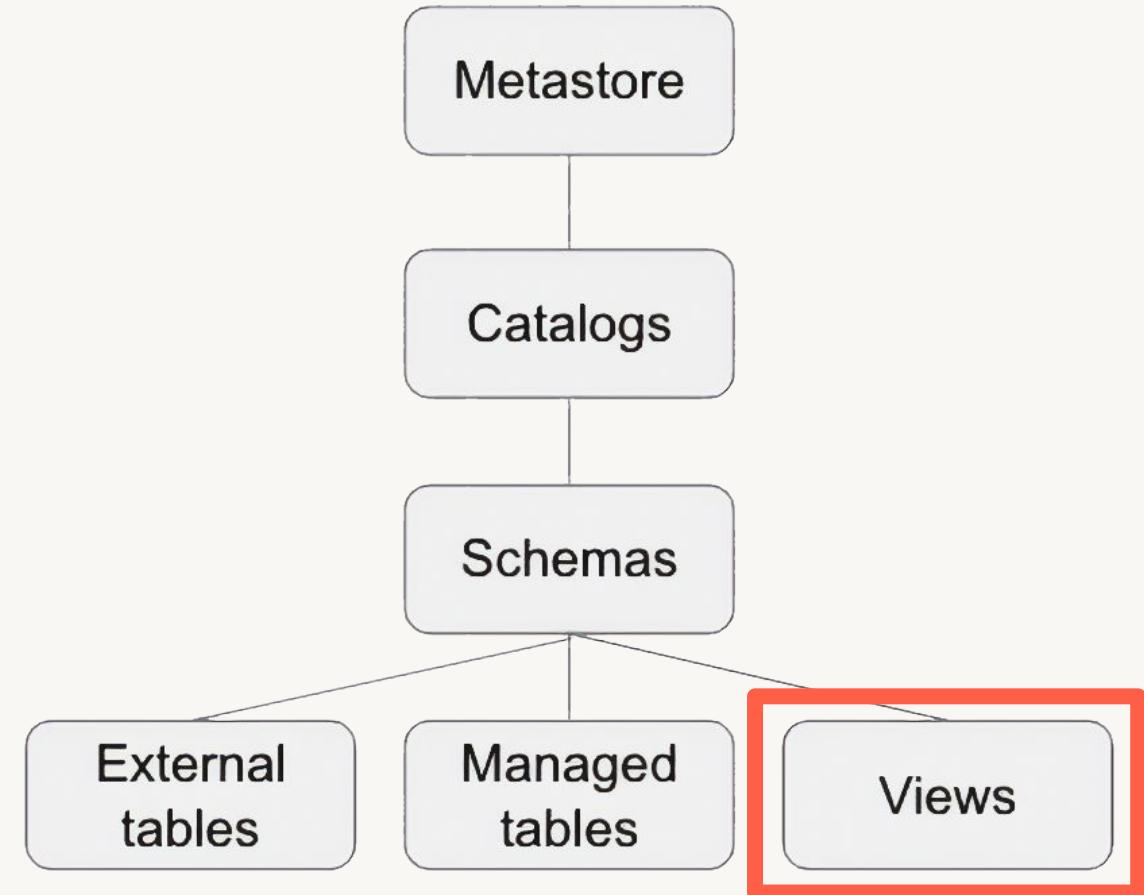
```
CREATE TABLE table2
  LOCATION 's3://<bucket_path>/<table_directory>'
...
  WITH CREDENTIAL <credential-name>;
```



View

- Third level of organization
- Can be composed from tables and views in multiple schemas or catalogs
- Created using CREATE VIEW:

```
CREATE VIEW view1 AS  
  SELECT column1, column2  
    FROM table1 ...
```



Three-Level Namespace Notation

- Data objects must be specified with three elements, depending on granularity required: Catalog, Schema, and Table
- Example:

```
CREATE TABLE main.default.department
(
    deptcode    INT,
    deptname    STRING,
    location    STRING
);
```

- Or, with a USE statement:

```
USE main.default;
SELECT * FROM department;
```

Lakehouse Architecture



Learning Objectives

By the end of this lesson, you should be able to:

- 1 Describe the benefits of using Databricks SQL for in-platform data processing.
- 2 Describe the medallion architecture as a sequential data organization and pipeline system of progressively cleaner data.
- 3 Identify that bronze and silver layers data requires additional processing and cleaning.
- 4 Describe the data in the gold layer of the medallion architecture.

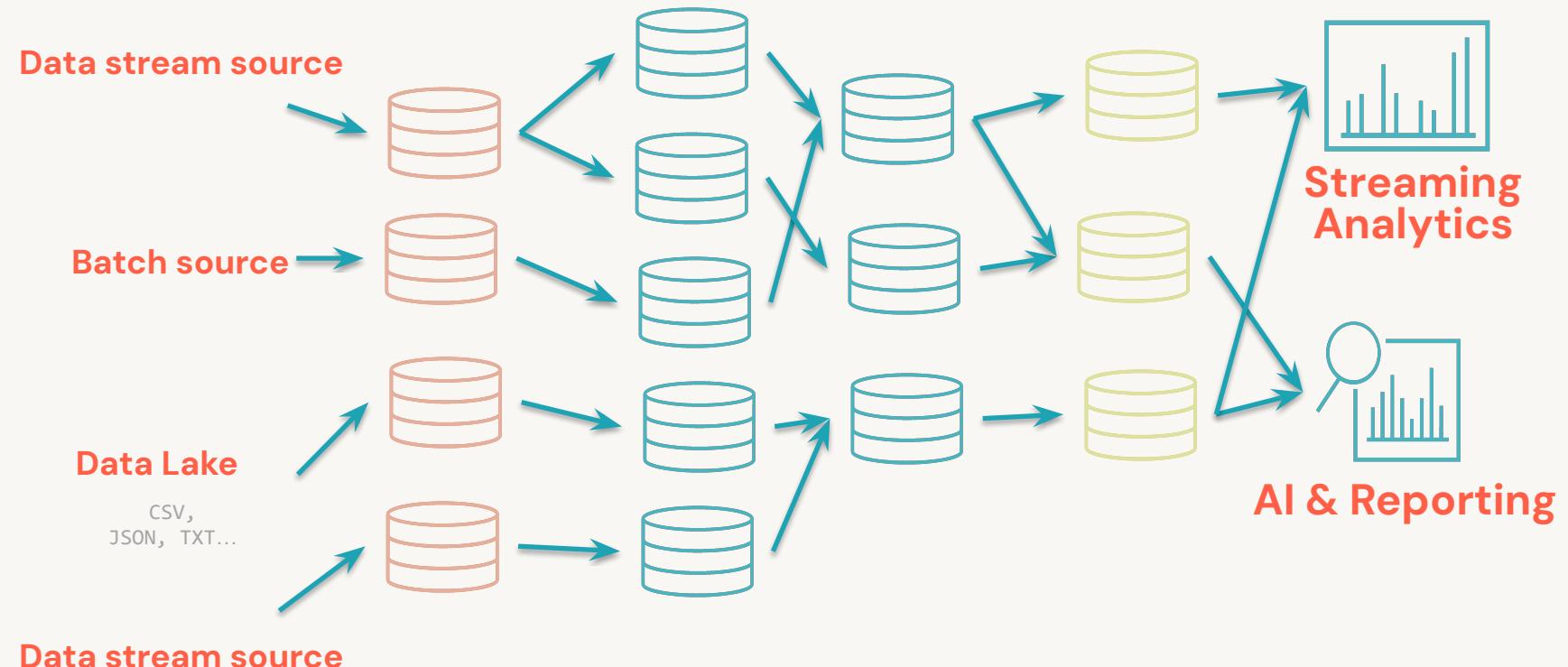
Learning Objectives, cont.

By the end of this lesson, you should be able to:

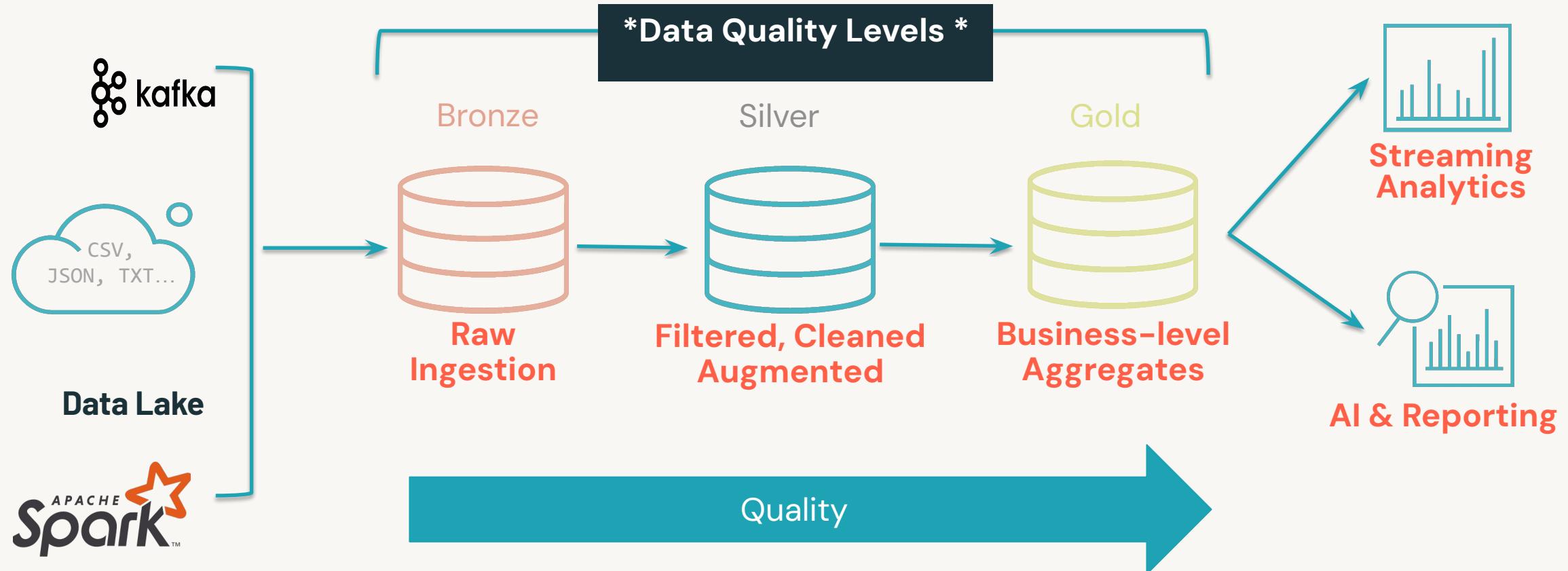
- 5 Describe last-mile ETL workflows fully within the gold layer for specific use cases.
- 6 Identify the gold layer as the most common layer for data analysts using Databricks SQL.
- 7 Describe the benefits of working with streaming data.

The Lakehouse Architecture

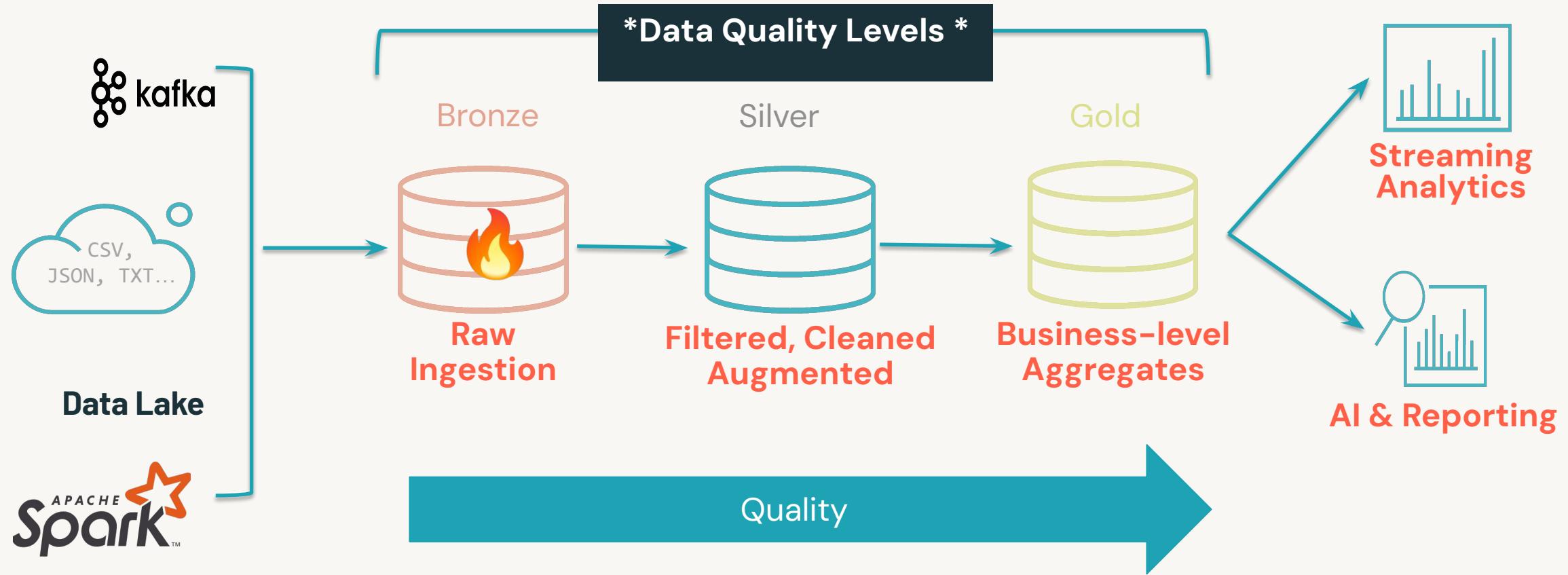
- Full ACID Transaction
- Focus on your data flow, instead of worrying about failures.
- Open Standards, Open Source
- Store petabytes of data without worries of lock-in. Growing community including Presto, Spark and more.
- Powered by 
- Unifies Streaming / Batch. Convert existing jobs with minimal modifications.*



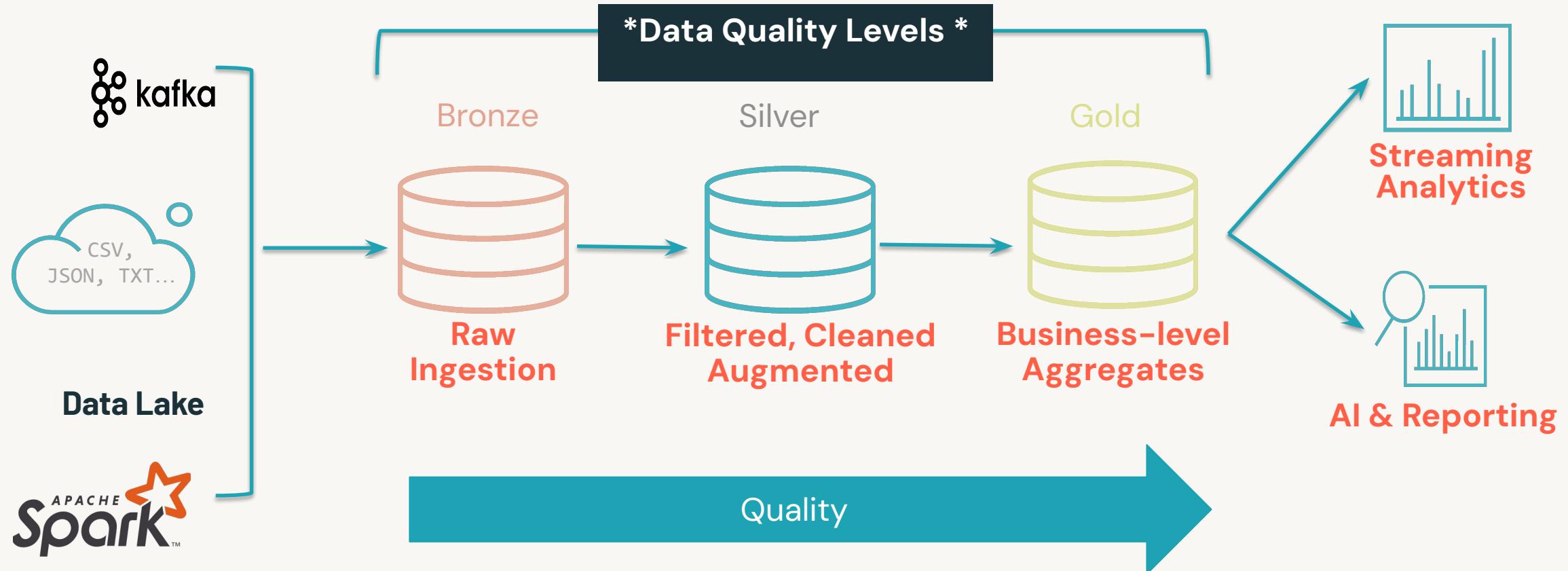
The Delta Lake Architecture



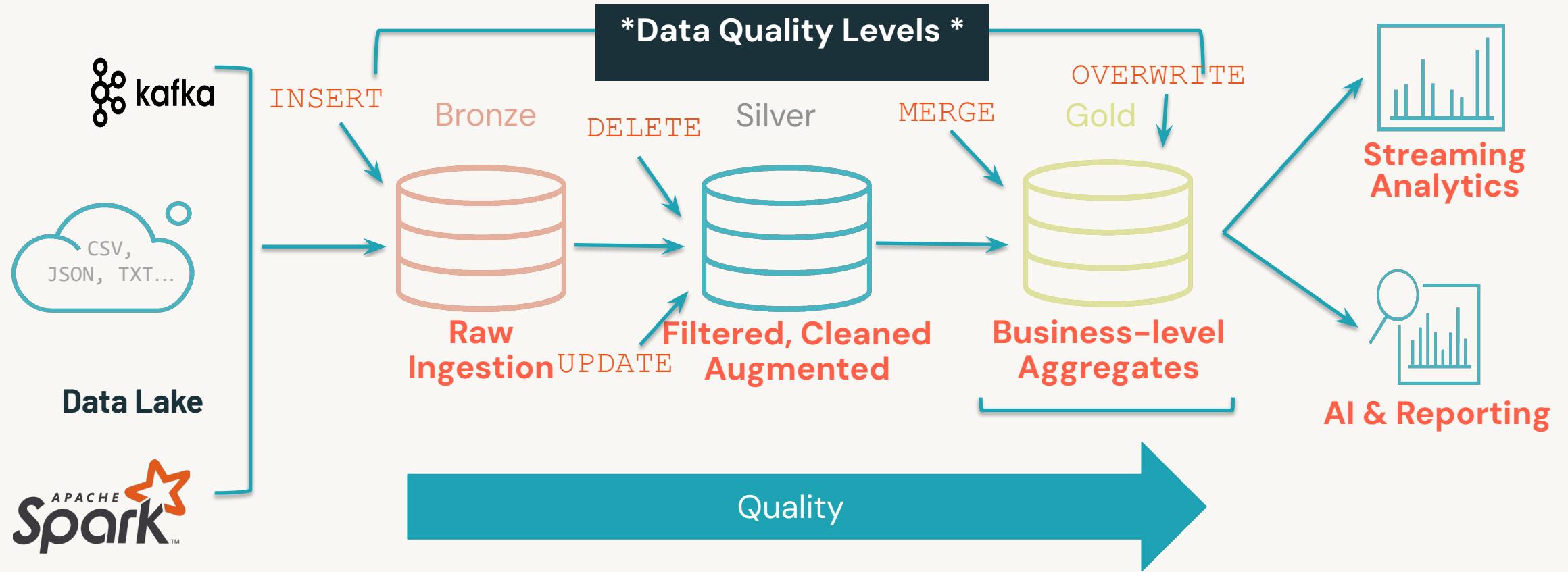
The Delta Lake Architecture



The Delta Lake Architecture



The Delta Lake Architecture



Knowledge Check

Knowledge check

Think about this question and volunteer an answer

Which of the following statements about the lakehouse medallion architecture is true? Select one response.

- A. The data in a single upstream table could be used to generate multiple downstream tables.
- B. The silver layer is for reporting and uses more de-normalized and read-optimized data models with fewer joins.
- C. The gold layer provides a broad view of all key business entities, concepts and transactions.
- D. Only minimal or "just-enough" transformations and data cleansing rules are applied to each layer in the medallion architecture.

Knowledge check

Think about this question and volunteer an answer

Which of the following describes the data quality of the gold layer of data in the lakehouse medallion architecture? Select one response.

- A. The gold layer brings the data from different sources into an Enterprise view.
- B. The gold layer is comprised of clean aggregated data, ready to use in production for a specific use case.
- C. The table structures in the gold layer correspond to the source system table structures "as-is".
- D. The focus of the gold layer is quick Change Data Capture and the ability to provide a historical archive if needed without rereading the data from the source system.

Knowledge check

Think about this question and volunteer an answer

What is the primary purpose of the bronze layer in the "bronze–silver–gold medallion" paradigm in Delta Lake?

- A. To store data in a format suitable for individual business projects or reports.
- B. To perform data cleansing, joining, and enrichment on raw data.
- C. To provide a "single source of truth" for the enterprise across various projects.
- D. To ingest raw data quickly, keeping it in its original format for both current and future projects.

Knowledge check

Think about this question and volunteer an answer

Which of the following statements describes the relationship between the silver and gold layer of data? Select one response.

- A. The gold layer has less clean data than the silver layer.
- B. Project-specific business rules are applied from the silver to gold layer.
- C. Self-service analytics are enabled for the gold layer for ad-hoc reporting in the silver layer.
- D. The gold layer is where we land all the data from external source systems, which are represented by the silver layer.

Integrations

Learning Objectives

By the end of this lesson, you should be able to:

- 1 Identify Databricks SQL as a complementary tool for BI partner tool workflows.
- 2 Identify Databricks SQL as a quick opportunity to create queries, visualizations, and dashboards from within the Lakehouse
- 3 Identify Partner Connect as a tool for implementing simple integrations with a number of other data products.

Databricks Partner Connect

Databricks Partner Connect is a **dedicated ecosystem of integrations** that allows users to **easily connect** with popular **data ingestion, transformation and BI partner products**.

This helps data analysts get useful data into their lakehouse **faster** without the need to manually configure each product so they can get data-driven insights

Databricks Partner Connect

Helps Data Analysts who:

- Struggle to connect to their choice of BI tools
- Struggle to bring data from SaaS apps (Google Analytics, Facebook, etc.) to run SQL queries
- Have to wait on eng/ops to generate data for analysis

Partner Connect Makes it Easy

How do I get the data from SFDC into Delta lake?



What tools can I use to ingest data into Delta?



I heard Fivetran is great! How do I connect it to Databricks?

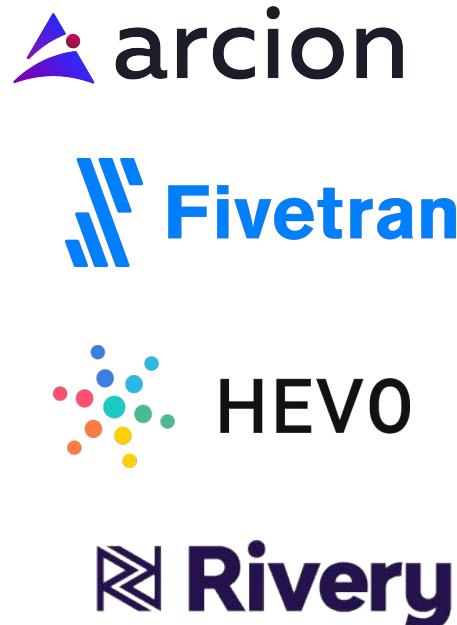


- Many partner integrations take as few as **6 clicks**
- **No** context or page **switches** required
- **Automatically** launches a cluster, calls Partner API to pass on PAT token and the cluster configuration details
- Sets up all the **necessary configs** for an optimized user experience
- Creates **trial account** in the partner product if an account doesn't exist

DATABRICKS
PARTNER
CONNECT

Databricks Partner Connect

Data
Ingestion



BI and
Visualization



Data Prep and
Transformation



Machine
Learning



Databricks Partner Connect

Data
Governance



Data
Quality

Anomalo



Reverse
ETL

hightouch

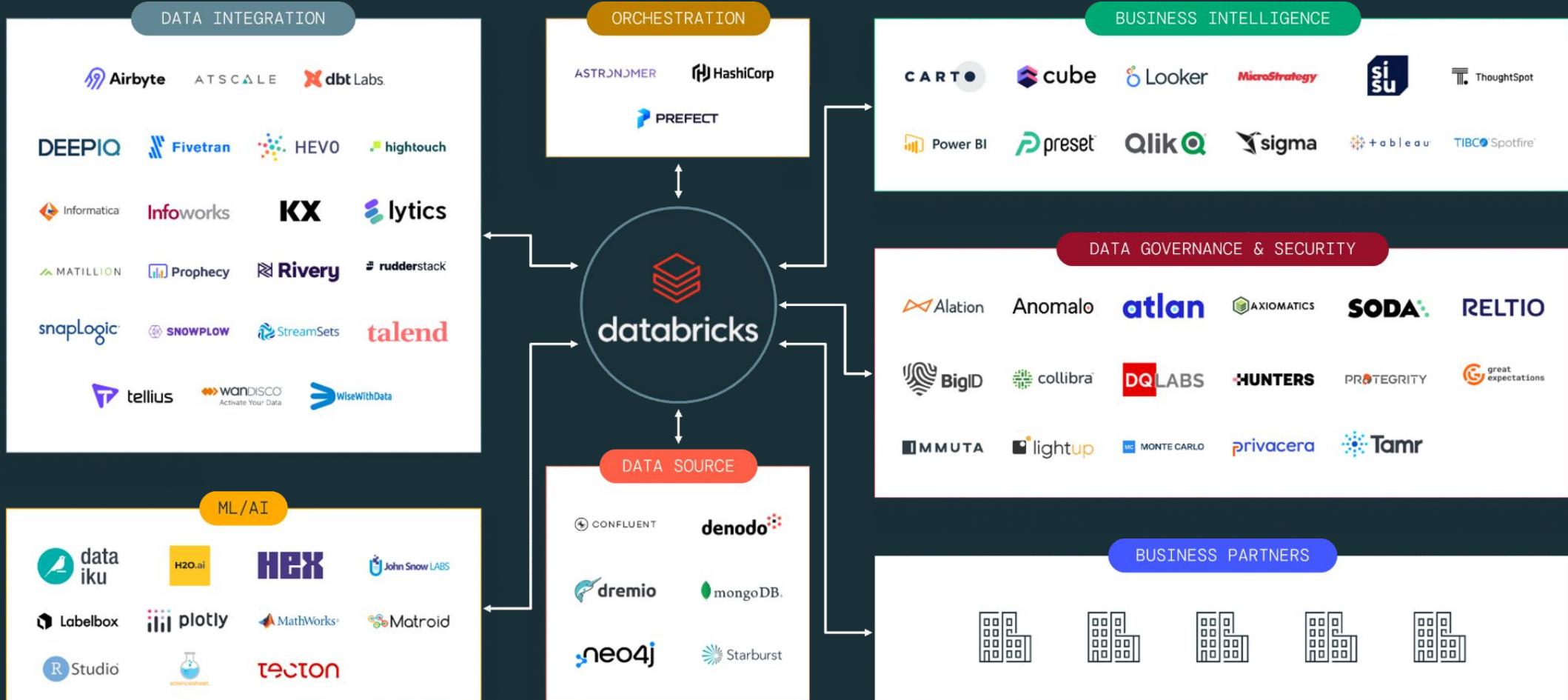
Semantic
Layer

ATSCALE



Built on an open foundation

Easily integrate with the entire data and AI ecosystem



Follow Along Demo: Integrations

Follow Along Demo

Integrations

- Connecting to outside data
- Connecting to BI tools
- Partner Connect

Data Management in Databricks SQL



Databricks Academy 2023

Data Management in Databricks SQL

Lesson Name	Duration
Lecture: Databricks SQL Warehouses	15 min
Follow Along Demo: Delta Lake in Databricks SQL	20 min
Lecture: Data Security	20 min

Databricks SQL Warehouses



Learning Objectives

By the end of this lesson, you should be able to:

- 1 Describe the purpose of Databricks SQL warehouses.
- 2 Compare and contrast Classic, Pro, and Serverless Databricks SQL warehouses.
- 3 Identify Serverless Databricks SQL warehouses as a quick-starting option.

Learning Objectives, cont.

By the end of this lesson, you should be able to:

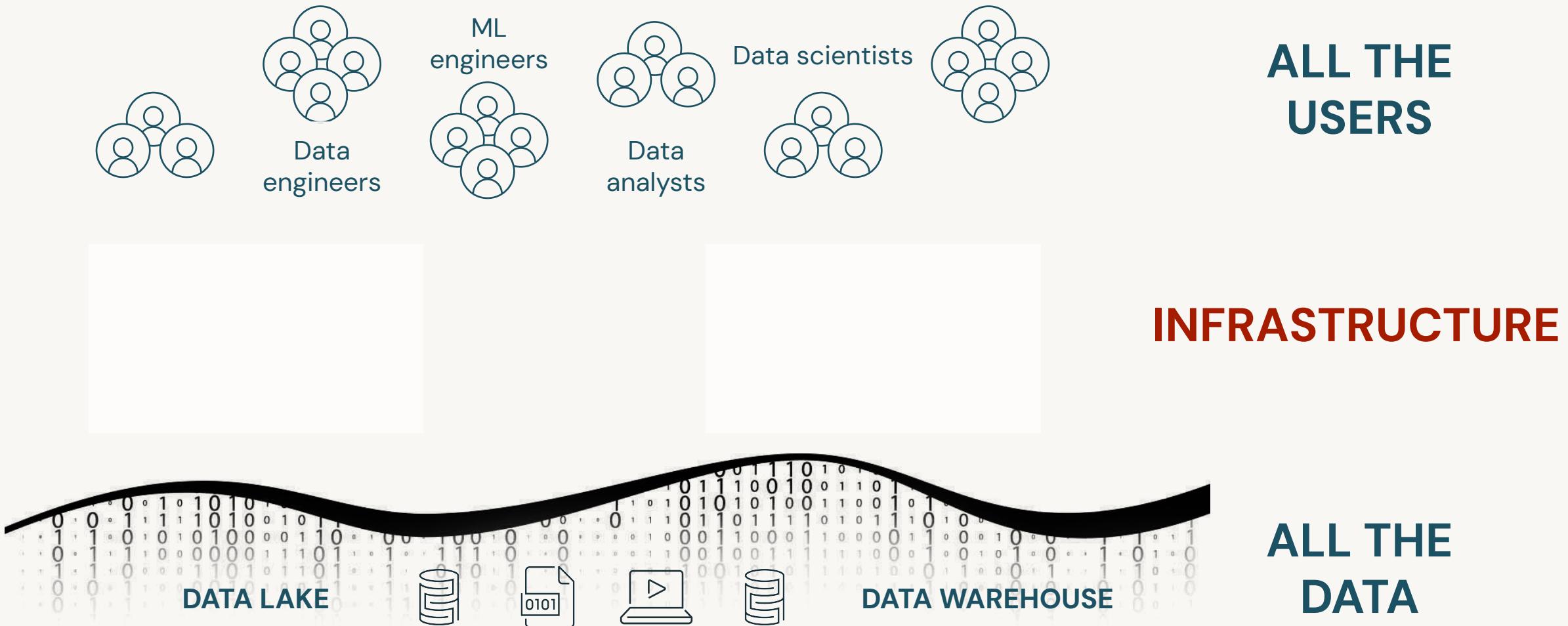
4

Describe basic Databricks SQL warehouse sizing and scaling guidelines in response to slow-running single-user queries and multi-user environments.

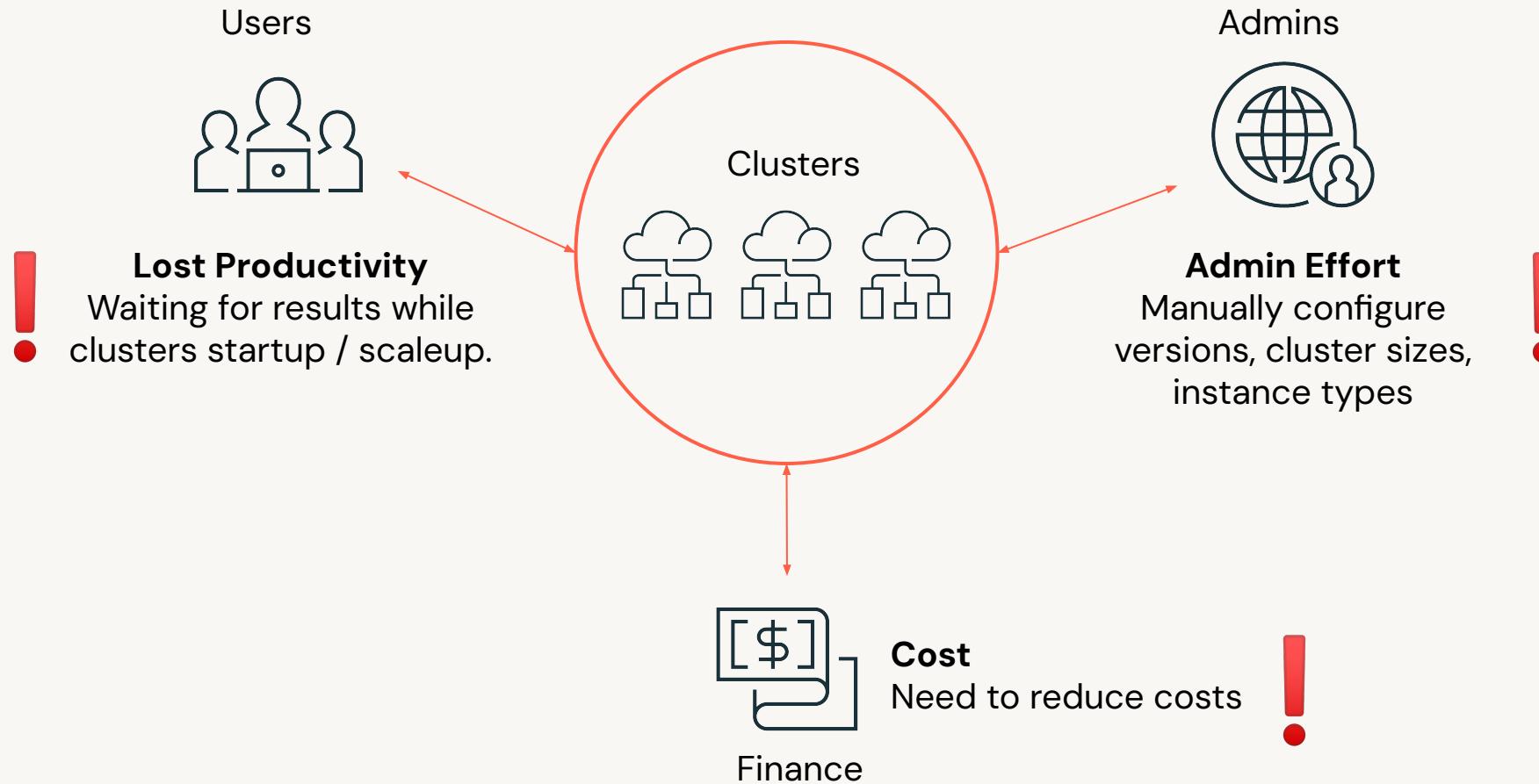
5

Describe the impact of Databricks SQL warehouse permissions on query history availability.

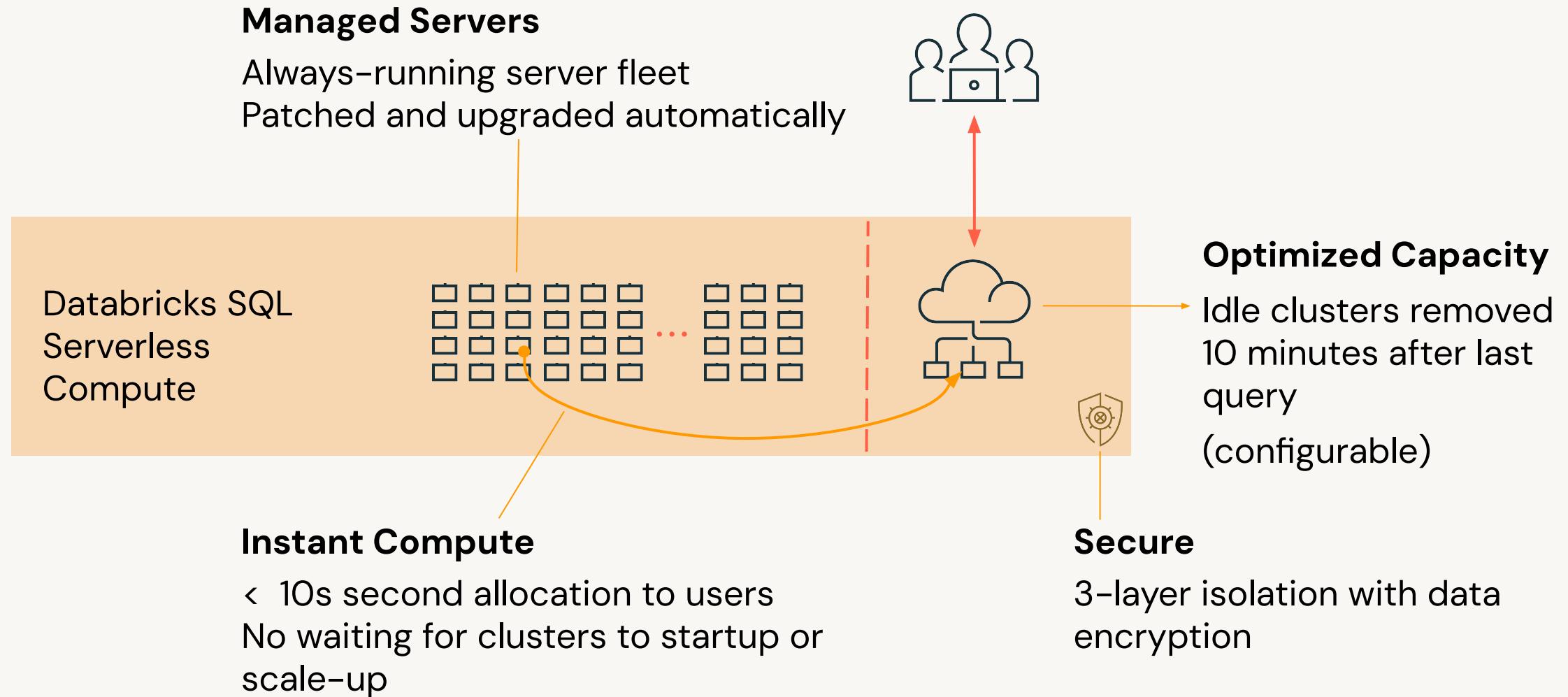
Successful data management platforms rely on efficient infrastructure



Problems with Managing Infrastructure



Databricks SQL Serverless



Industry leading security architecture supporting production workloads

1. Container Isolation

- Hardened container images per industry best practice
- Disable privilege access in the container

2. VM Isolation

- Workloads separated by VM boundaries
- Blocked – reusing VMs among customers

3. Network Isolation

- All nodes egress is blocked except to nodes in same cluster
- Federated access through temporary security tokens
- Ingress traffic from other customers is blocked

Databricks SQL Serverless Benefits



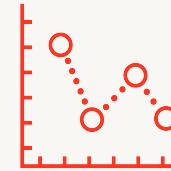
Higher user productivity

- User queries start instantly, no waiting for cluster start-up
- Add more concurrent users with instant cluster scaling



Zero Management

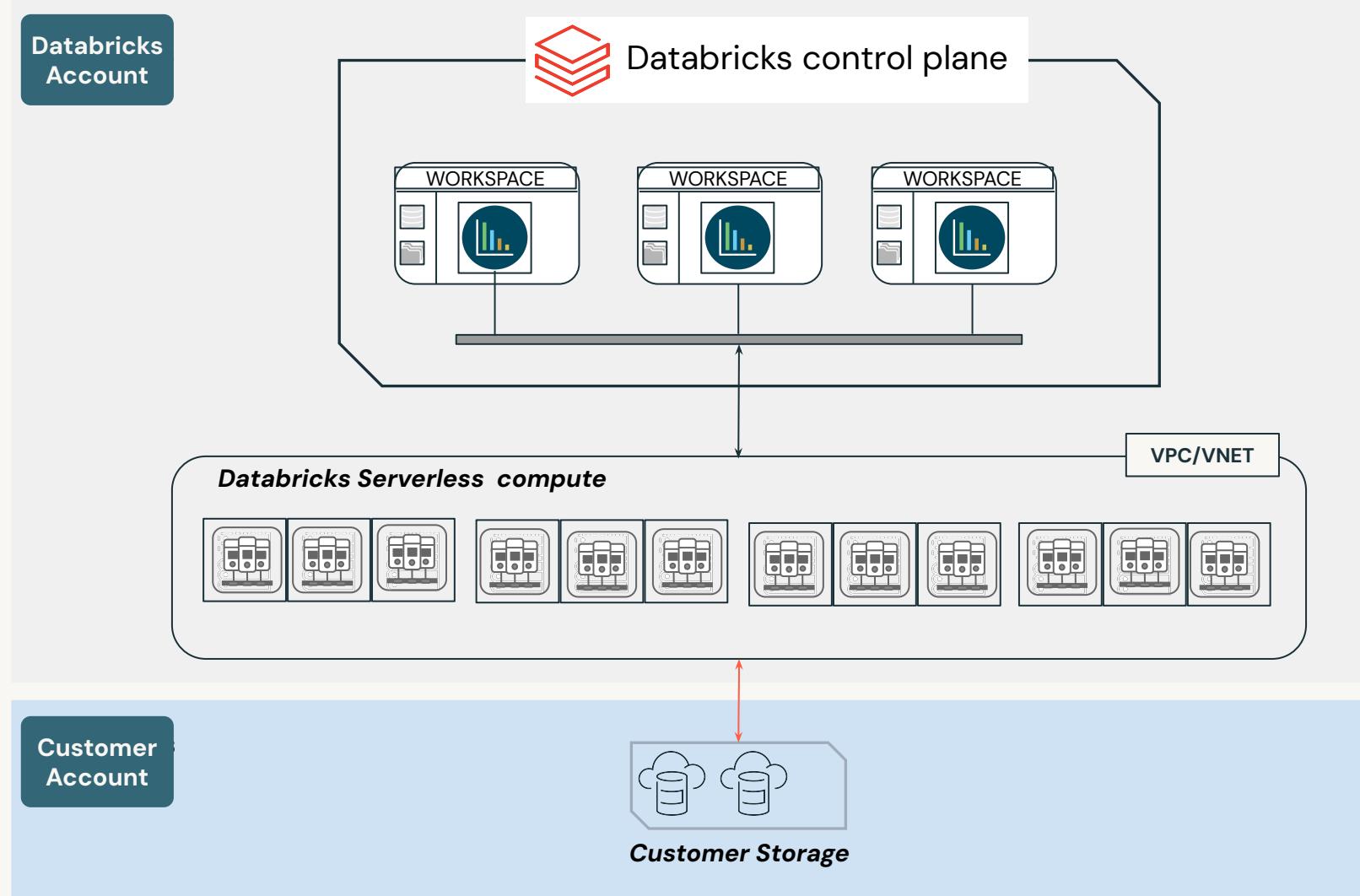
- No configuration
- No performance tuning
- No capacity management
- Automatic upgrades and patching



Lower Cost

- Pay what you consume; eliminate idle cluster time
- No over-provisioning of resources
- Idle capacity removed 10 minutes after last query

Serverless Compute Architecture



Benefits:

- Production ready environment
- Robust security foundation – data isolation and encryption



Warehouse Configuration

AWS

Cluster size	Driver size	Worker count
2X-Small	i3.2xlarge	1
X-Small	i3.2xlarge	2
Small	i3.4xlarge	4
Medium	i3.8xlarge	8
Large	i3.8xlarge	16
X-Large	i3.16xlarge	32
2X-Large	i3.16xlarge	64
3X-Large	i3.16xlarge	128
4X-Large	i3.16xlarge	256

Azure

Cluster size	Driver size	Worker count
2X-Small	Standard_E8ds_v4	1
X-Small	Standard_E8ds_v4	2
Small	Standard_E16ds_v4	4
Medium	Standard_E32ds_v4	8
Large	Standard_E32ds_v4	16
X-Large	Standard_E64ds_v4	32
2X-Large	Standard_E64ds_v4	64
3X-Large	Standard_E64ds_v4	128
4X-Large	Standard_E64ds_v4	256

Warehouse Configuration

GCP

Cluster size	Instance type for driver	Worker count	Total vCPU	Total Persistent Disk SSD (TB)	Total Local SSD (TB)
2X-Small	n2-highmem-8	1	16	1	3
X-Small	n2-highmem-8	2	24	1.5	4.5
Small	n2-highmem-16	4	48	2.5	7.5
Medium	n2-highmem-32	8	96	4.5	15
Large	n2-highmem-32	16	160	8.5	27
X-Large	n2-highmem-64	32	320	16.5	54
2X-Large	n2-highmem-64	64	576	32.5	102
3X-Large	n2-highmem-64	128	1088	64.5	198
4X-Large	n2-highmem-64	256	2112	128.5	390

The instance size of all workers is n2-highmem-8.

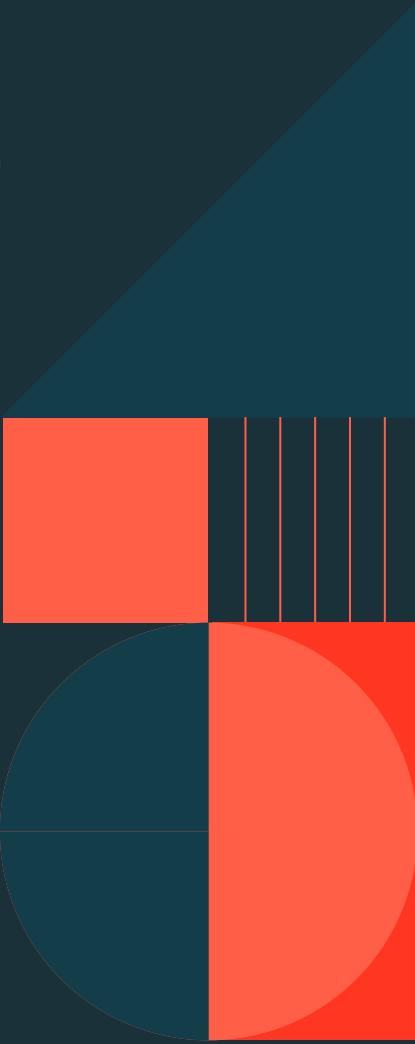
Warehouse Configuration

In the course, SQL Warehouses have the following settings

- Cluster size – 2X-Small
- Scaling – Min: 1, Max 1
- Auto-stop – After ten minutes

Delta Lake

Follow Along Demo: Delta Lake in Databricks SQL



Follow Along Demo

Delta Lake in Databricks SQL

- Create a schema
- Create views that withhold data from unauthorized groups
- Optimize delta tables

Knowledge Check



Knowledge check

Think about this question and volunteer an answer

Which of the following statements describes the purpose of Databricks SQL warehouses? Select one response.

- A. SQL warehouses enable data analysts to find and share dashboards.
- B. SQL warehouses are a declarative framework for building data processing pipelines.
- C. SQL warehouses provide data discovery capabilities across Databricks workspaces.
- D. SQL warehouses allow users to run SQL commands on data objects within Databricks SQL.

Knowledge check

Think about this question and volunteer an answer

What are the benefits of Delta Lake within the Lakehouse Architecture?

- A. Real-time data processing with low latency
- B. Exclusive support for batch processing
- C. ACID transactions, metadata scalability, and storage improvement
- D. Data isolation for multiple software development environments

Knowledge check

Think about this question and volunteer an answer

Which of the following statements about SQL warehouse sizing and scaling is true? Select two responses.

- A. Increasing maximum scaling allows for multiple users to use the same warehouse at the same time.
- B. Scaling is set to a minimum of 1 and a maximum of 1 by default.
- C. The higher the cluster size, the higher the latency in your queries.
- D. The auto-stop feature will restart the warehouse if it remains idle during the auto-stop period.

Knowledge check

Think about this question and volunteer an answer

Which feature of the platform provides users with the ability to quickly connect to third-party tools with simple to implement integrations? Select one response.

- A. SQL Editor
- B. Partner Connect
- C. Workflows
- D. Features

Lab: Lakehouse SQL



Lab

Lakehouse SQL

- Time Travel
- COPY INTO
- MERGE INTO

Data Security

Learning Objectives

By the end of this lesson, you should be able to:

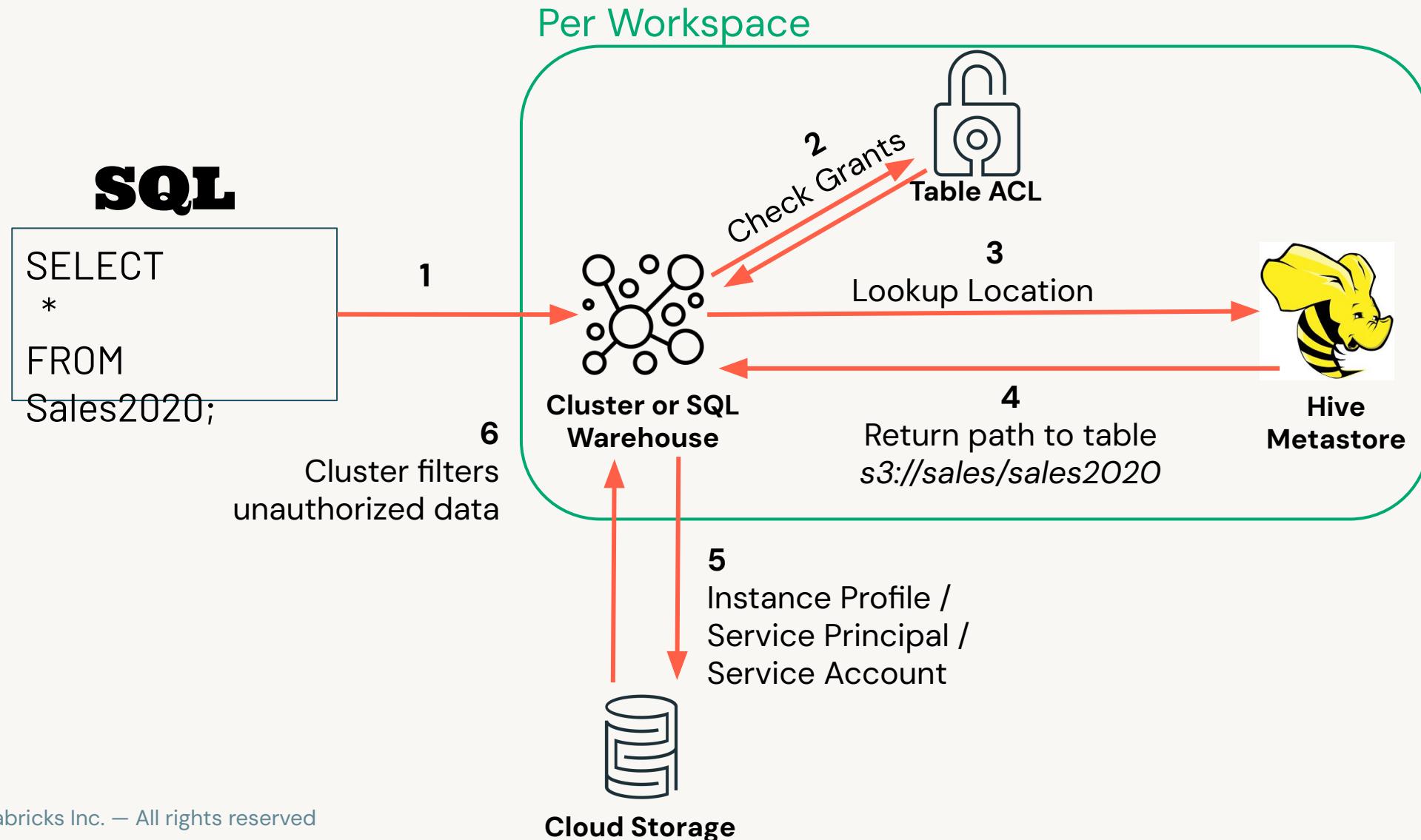
- 1 Describe the different levels of data object access available with Unity Catalog.
- 2 Identify that catalogs, schemas, and tables can all have unique owners.
- 3 Describe how to organize owned data objects for the purposes of security.
- 4 Identify that the creator of a data object becomes the owner of that data object.

Learning Objectives, cont.

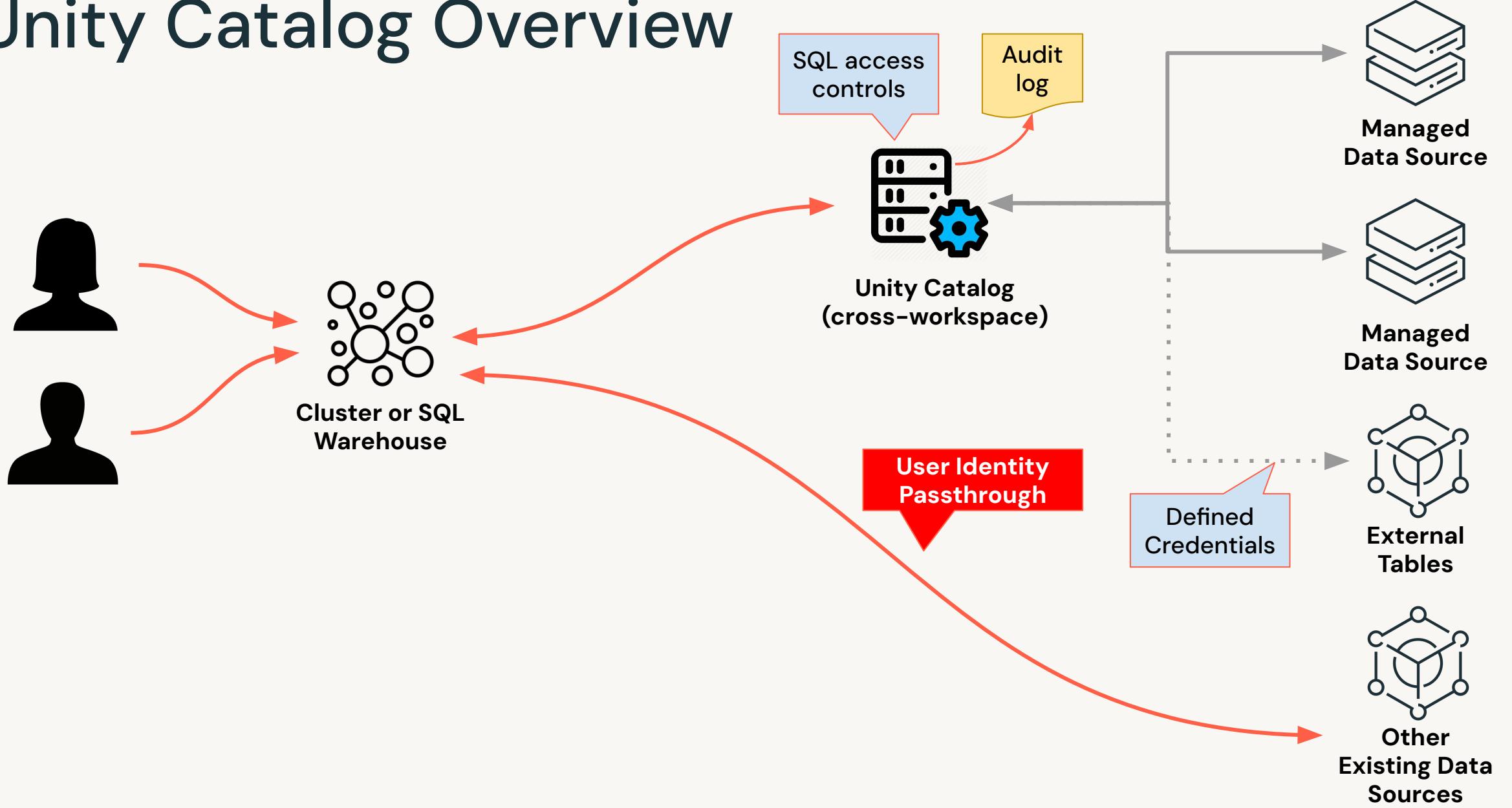
By the end of this lesson, you should be able to:

- 5 Identify the responsibilities of data object ownership.
- 6 Update data object permissions to address user access needs in a variety of common scenarios.
- 7 Identify PII data objects as needing additional, organization-specific considerations.

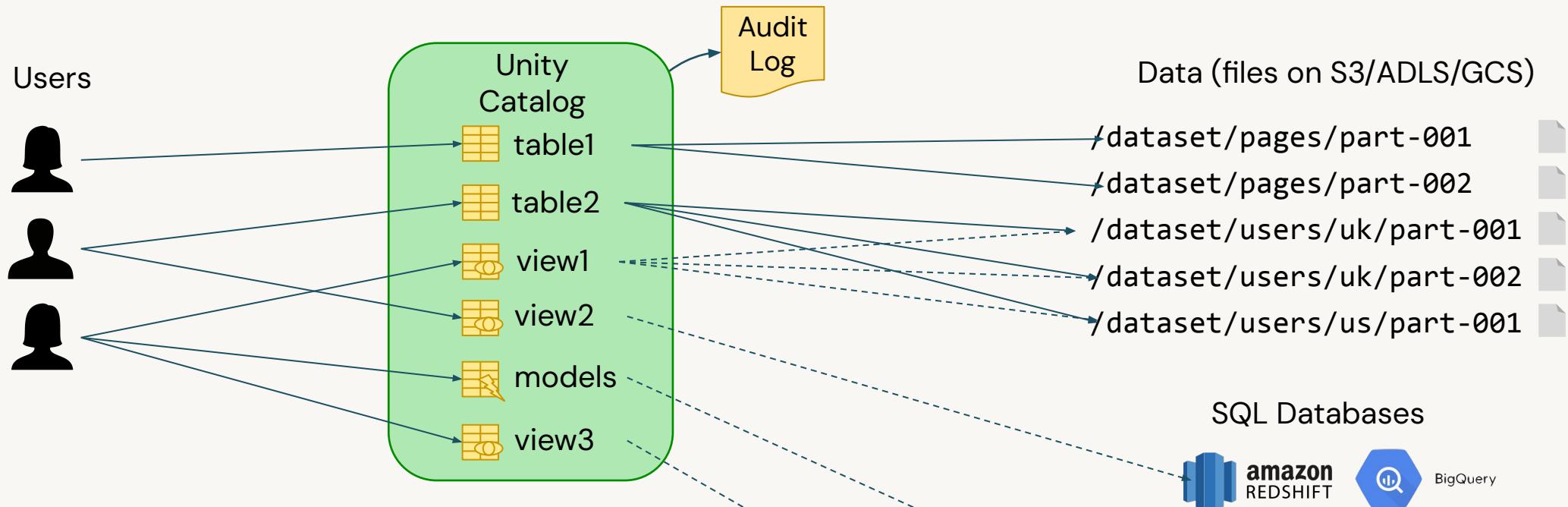
The Life of a Query (Without Unity Catalog)



Unity Catalog Overview



Databricks Unity Catalog



Fine-grained permissions on tables, fields, views: not files

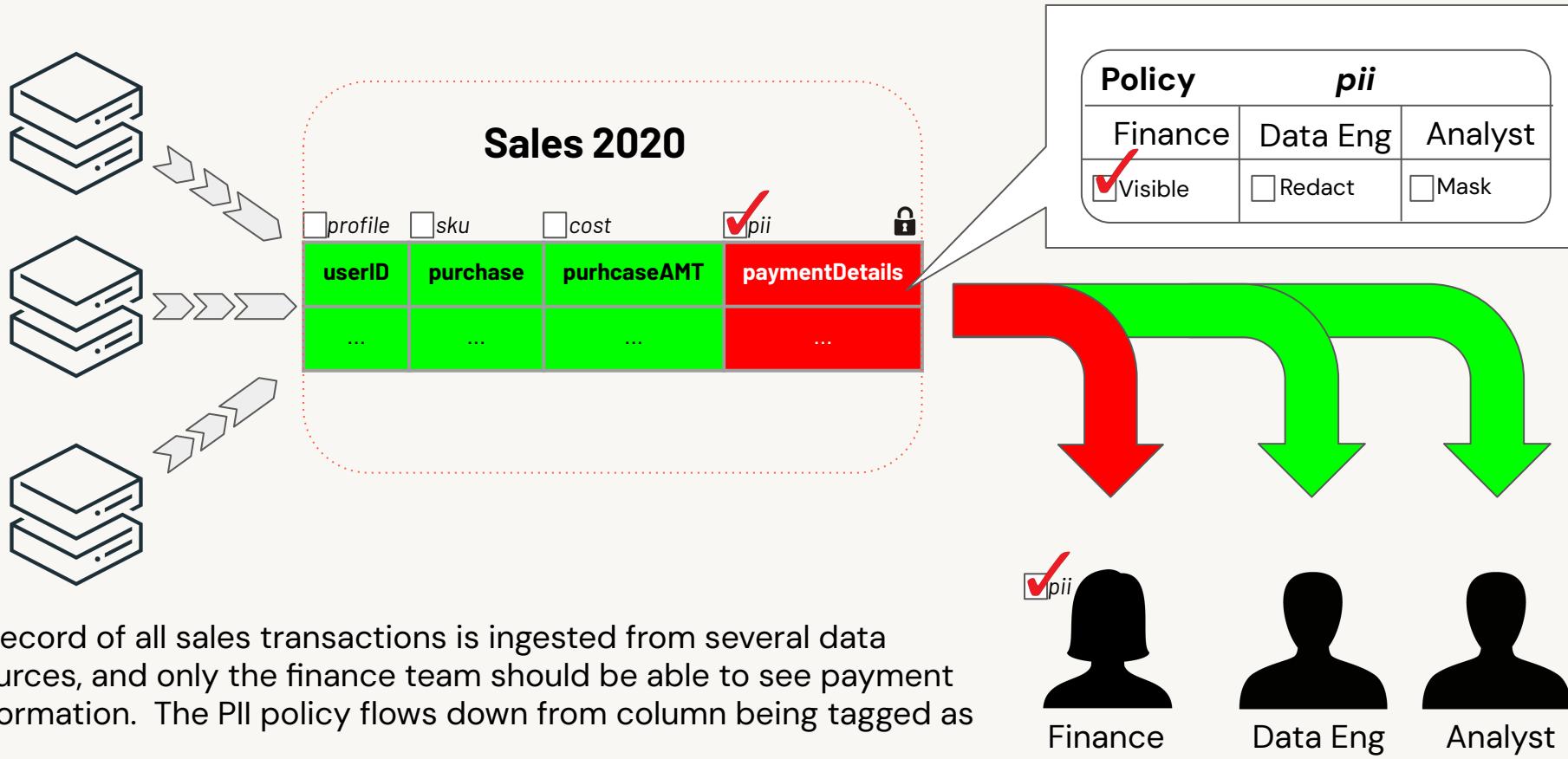
Industry standard interface: ANSI SQL grants

Uniform permission model for all data assets

Centrally audited

Attribute-Based Access Control

True policy driven ABAC that flows from your metadata



Attribute-Based Access Control

```
CREATE ATTRIBUTE pii
```

```
ALTER TABLE iot_events ADD ATTRIBUTE pii ON email  
ALTER TABLE users ADD ATTRIBUTE pii ON phone
```

...

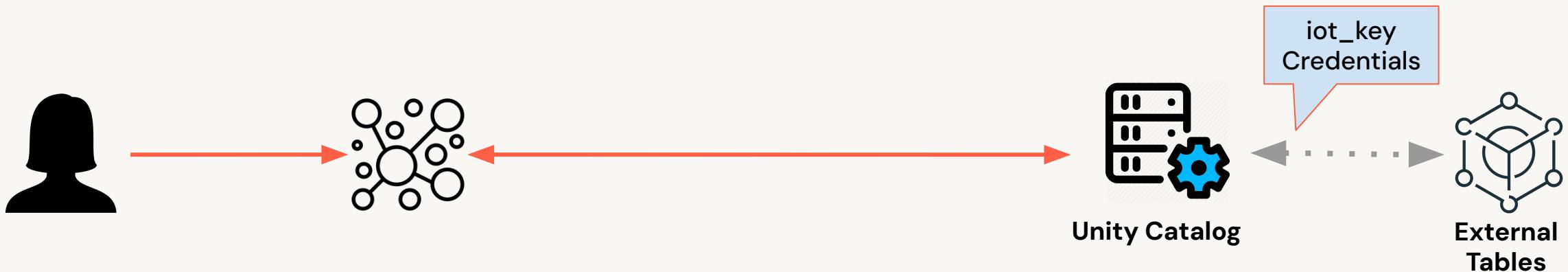
```
GRANT SELECT ON SCHEMA iot_data  
HAVING ATTRIBUTE NOT IN (pii)  
TO product_managers
```

Set permission on all
columns tagged pii
together

Unity Catalog: External Table with Defined Credentials

```
CREATE CREDENTIAL iot_role TYPE AWS_ROLE ...
```

```
CREATE TABLE iot_data LOCATION s3:/...  
WITH CREDENTIAL iot_role
```

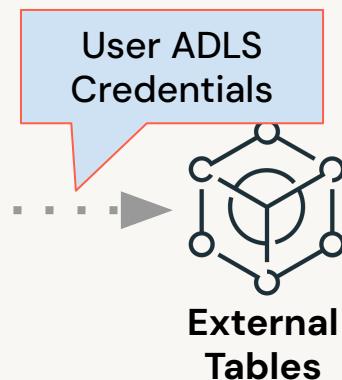
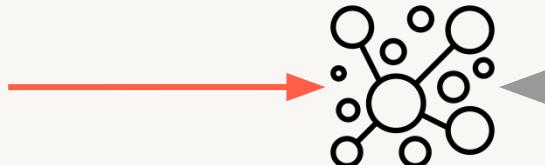


Unity Catalog: External Files with Passthrough

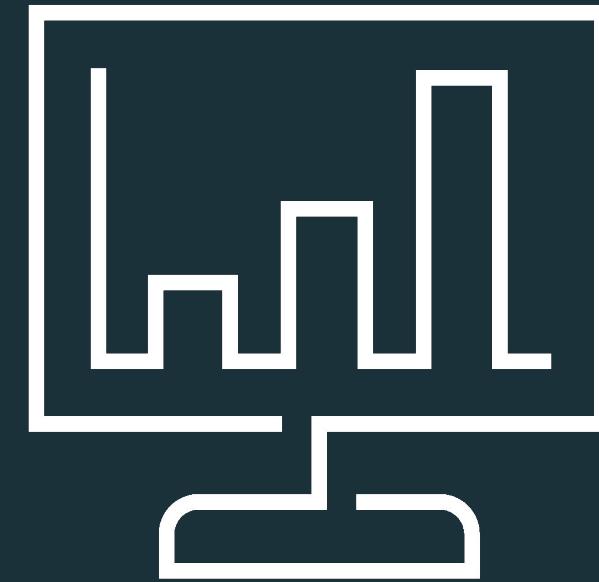
```
SELECT * FROM csv.`adls:/.../myfolder`
```

if a direct file path is specified,
we perform passthrough with
the user's cloud credentials

```
CREATE VIEW v AS SELECT * FROM csv.`adls:/.../myfolder`
```



Data Visualization and Dashboarding



Databricks Academy 2023

Data Visualization and Dashboarding

Lesson Name	Duration
Follow Along Demo: Data Visualizations	40 min
Lab: Data Visualizations	30 min
Follow Along Demo: Dashboarding Basics	40 min
Lab: Dashboards	30 min
Follow Along Demo: Databricks SQL in Production	40 min
Lab: Databricks SQL in Production	30 min

Follow Along Demo: Data Visualizations

Follow Along Demo

Data Visualizations

- Create common visualizations
- Create map visualizations

Knowledge Check

Knowledge check

Think about this question and volunteer an answer

What types of customized visualizations can be created using Databricks SQL?

- A. Pie charts, bar graphs, and line charts
- B. Counter and funnels
- C. Bar charts, combo charts, and geographical maps
- D. All of the above

Knowledge check

Think about this question and volunteer an answer

How can you enable aggregation in a Databricks SQL visualization?

- A. Modify the underlying SQL query to add an aggregation column.
- B. Select the aggregation type directly in the visualization editor.
- C. Use the Aggregation drop-down menu in the Visualization Type options.
- D. Aggregation is not supported in Databricks SQL visualizations.

Knowledge check

Think about this question and volunteer an answer

How can you add a query visualization to a Databricks dashboard?

- A. Select a query and choose the visualization type.
- B. Drag and drop a visualization from the sidebar.
- C. Copy and paste the visualization code into the dashboard.
- D. Query visualizations cannot be added to a Databricks dashboard.

Knowledge check

Think about this question and volunteer an answer

What is the benefit of setting a refresh schedule for a Databricks dashboard?

- A. To change the color palette of visualizations.
- B. To organize and label workspace objects.
- C. To keep the data underlying visualizations up-to-date.
- D. To create query parameters for customization.

Lab: Data Visualizations



Lab

Data Visualizations

- Create a boxplot
- Create a funnel

Follow Along Demo: Dashboarding Basics

Follow Along Demo

Dashboarding Basics

- Create a dashboard
- Parameterize queries
- Organize Databricks SQL assets

Knowledge Check

Knowledge check

Think about this question and volunteer an answer

Which of the following can be added to a query so that the code can be rerun with different variable inputs? Select one response.

- A. User-defined functions
- B. Parameters
- C. Vectors
- D. SQL warehouses

Knowledge check

Think about this question and volunteer an answer

A data analyst needs to create a visualization out of the following query:

```
SELECT order_date  
FROM sales  
WHERE order_date >= to_date('2020-01-01')  
AND order_date <= to_date('2021-01-01');
```

Which of the following visualization types is best suited to depict the results of this query? Select one response.

- A. Funnel
- B. Stacked bar chart
- C. Bar chart
- D. Boxplot

Knowledge check

Think about this question and volunteer an answer

A team of stakeholders needs to be notified of changes in a dashboard's statistics on a daily basis. Which of the following actions can be taken to ensure they always have the newest information? Select one response.

- A. A refresh schedule can be configured and stakeholders can be subscribed to the dashboard's output.
- B. A trigger alert can be created for the dashboard and stakeholders can be added to the alert notification list.
- C. A webhook can be created and shared with stakeholders.
- D. None of the above

Knowledge check

Think about this question and volunteer an answer

Which of the following data visualizations displays a single number by default? Select one response.

- A. Bar chart
- B. Counter
- C. Map – markers
- D. Funnel

Lab: Dashboards

Lab

Dashboards

- Create your own dashboard

Follow Along Demo: Databricks SQL in Production

Follow Along Demo

Databricks SQL in Production

- Automation in Databricks SQL
- Sharing Databricks SQL assets
- Creating gold-level tables

Knowledge Check

Knowledge check

Think about this question and volunteer an answer

Which of the following automations are available in Databricks SQL? Select one response.

- A. Query refresh schedules
- B. Dashboard refresh schedules
- C. Alerts
- D. All of the above

Knowledge check

Think about this question and volunteer an answer

What is the purpose of Alerts in Databricks SQL?

- A. To automatically execute SQL queries.
- B. To organize queries within a folder structure.
- C. To trigger notifications based on specific conditions in scheduled queries.
- D. To share dashboards with other team members.

Knowledge check

Think about this question and volunteer an answer

What is the purpose of configuring a refresh schedule for a query in Databricks SQL?.

- A. To automatically pull new data into a table.
- B. To create a new table based on specified criteria.
- C. To manually execute queries on-demand.
- D. To edit existing data in the database.

Knowledge check

Think about this question and volunteer an answer

What level of permissions is the owner of a query granted on their query? Select one response.

- A. Can View
- B. Can Run
- C. Can Edit
- D. Can Manage

Lab: Databricks SQL in Production

Lab

Databricks SQL in Production

- Create automations
- Share a dashboard

Summary and Next Steps

Earn a Databricks certification!

Certification helps you gain industry recognition, competitive differentiation, greater productivity, and results.

- This course helps you prepare for the **Databricks Certified Data Analyst Associate exam**
- Recommended Self-Paced Courses
 - Ingesting Data for Databricks SQL
 - Integrating BI Tools with Databricks SQL
- Please see the Databricks Academy for additional prep materials



For more information visit:
databricks.com/learn/certification

Thank you!



databricks