Lakehouse Federation

Learning Objectives

▶ Data ingestion challenges

▶ Lakehouse Federation

Data Ingestion Challenges
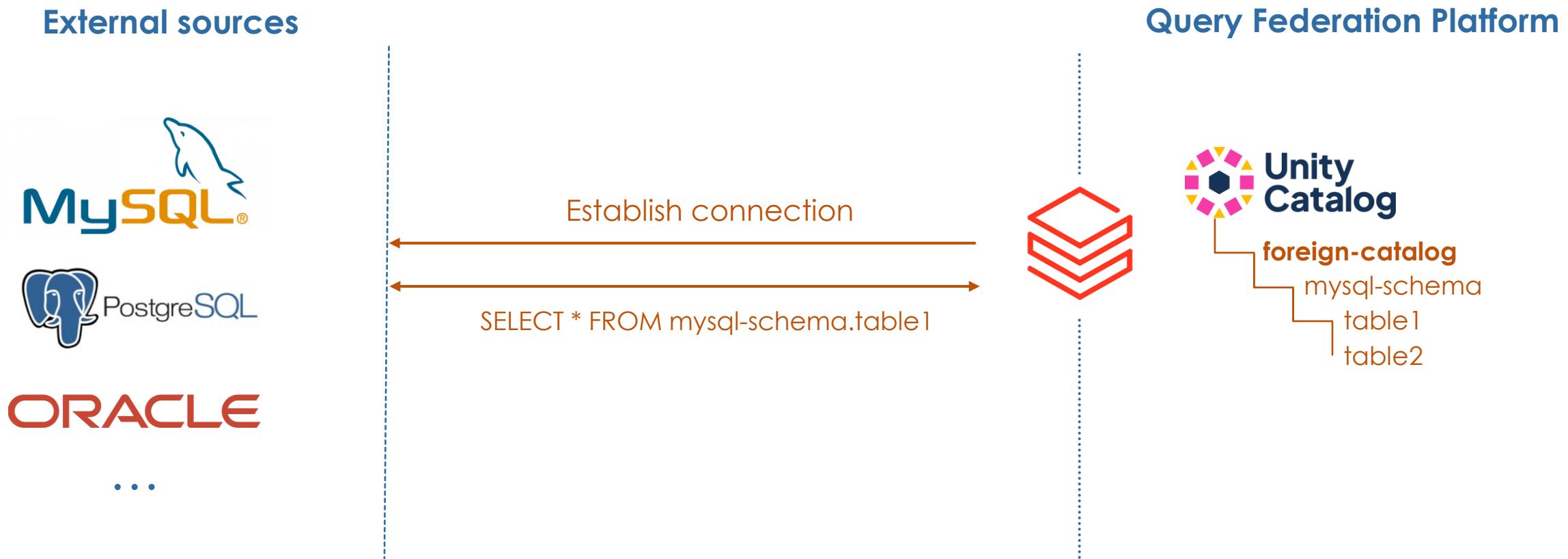
▶ Data ingestion is recommended for most use cases

   ▶ high data volumes

   ▶ low-latency querying

   ▶ third-party API limits.

▶ Ingestion results in duplicate data that may become stale

   ▶ Use Delta Sharing to receive live data from the source

Lakehouse Federation

▶ Run queries against multiple data sources

▶ No data migration

Query Federation

External sources

Query Federation Platform

Establish connection

SELECT * FROM mysql-schema.table1

foreign-catalog
mysql-schema
table1
table2

Use case

▶ Maintain live access to external systems.

▶ Ad hoc reporting or proof-of-concept access to operational data stored in external databases.

Limitations

▶ Complex queries don't benefit from the power of Databricks