SQL in the Data Intelligence Platform

INTRODUCTION TO DATABRICKS SQL

Kevin Barlow
Data Manager

datacamp

Instructor Introduction

Kevin Barlow

DATA MANAGER

Who am I?

Professional in the data analytics industry with over a decade of working with different analytical tools.

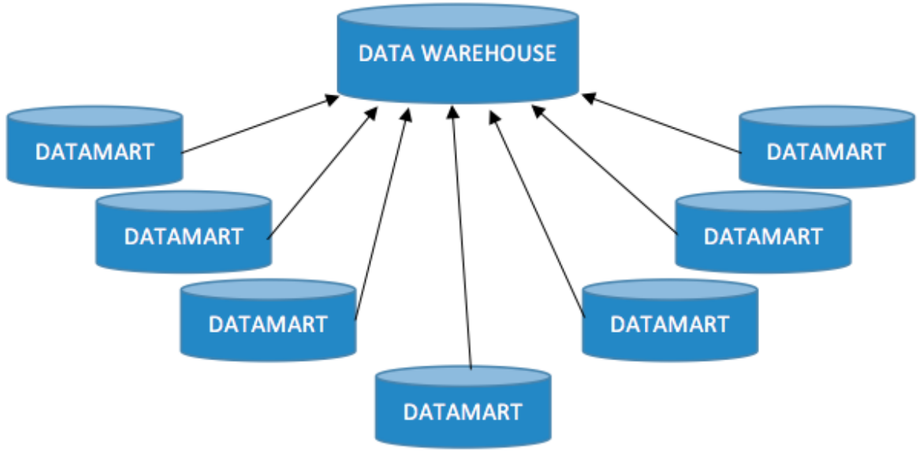
Worked with companies of all sizes to deliver analytics solutions.

Extensive knowledge of the Databricks platform.

Motivation

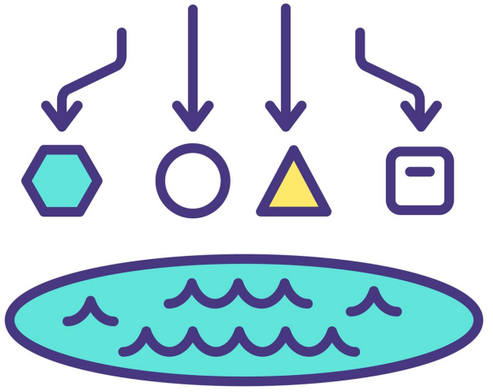
Data Warehouses

- Great for SQL workloads
- Typically expensive
- Proprietary technologies
- Limited capabilities and integrations

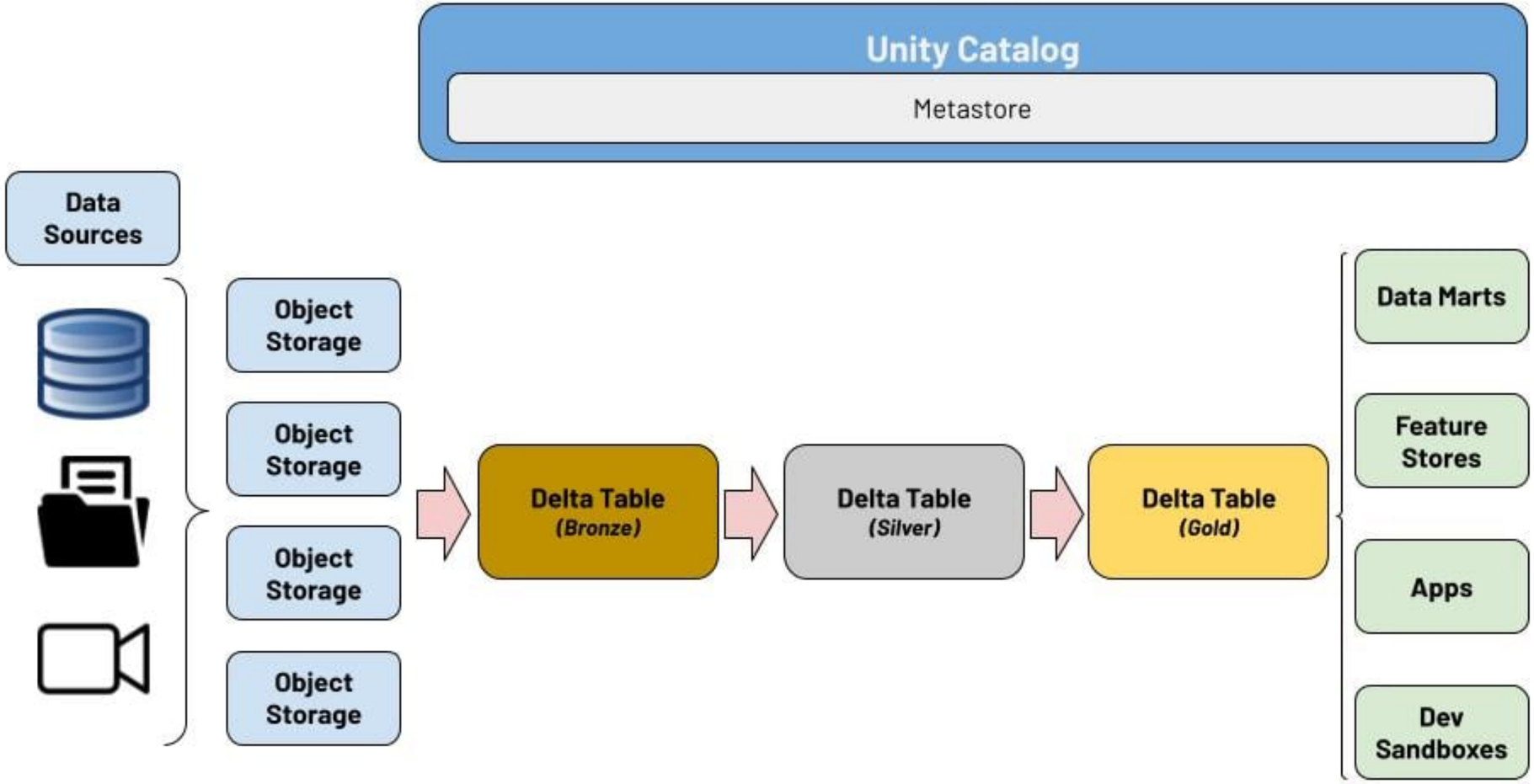


Data Lakes

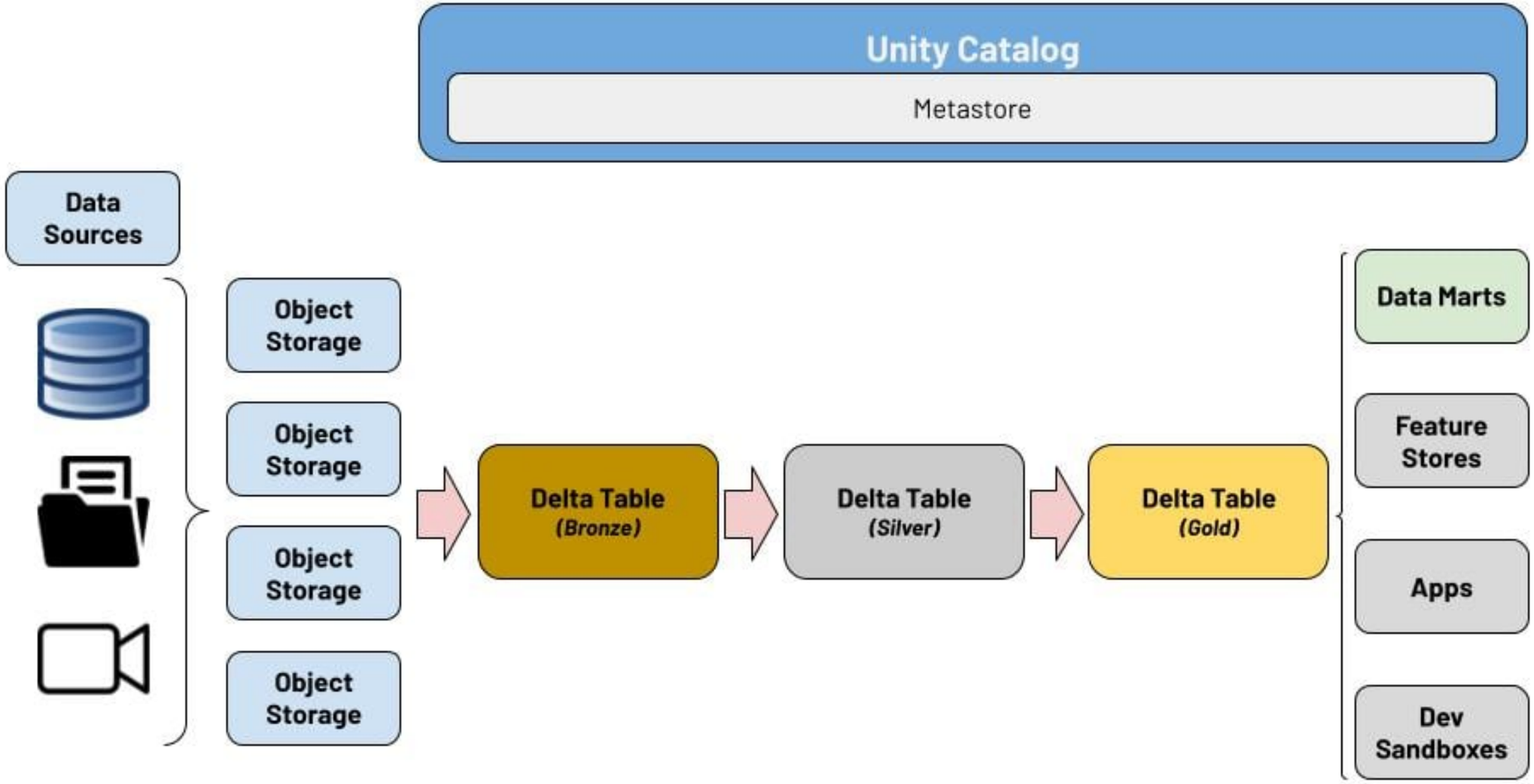
- Great for non-SQL workloads
- Cost effective, lackluster performance
- Open-source technologies
- Unlimited capabilities



Data warehousing in the Lakehouse

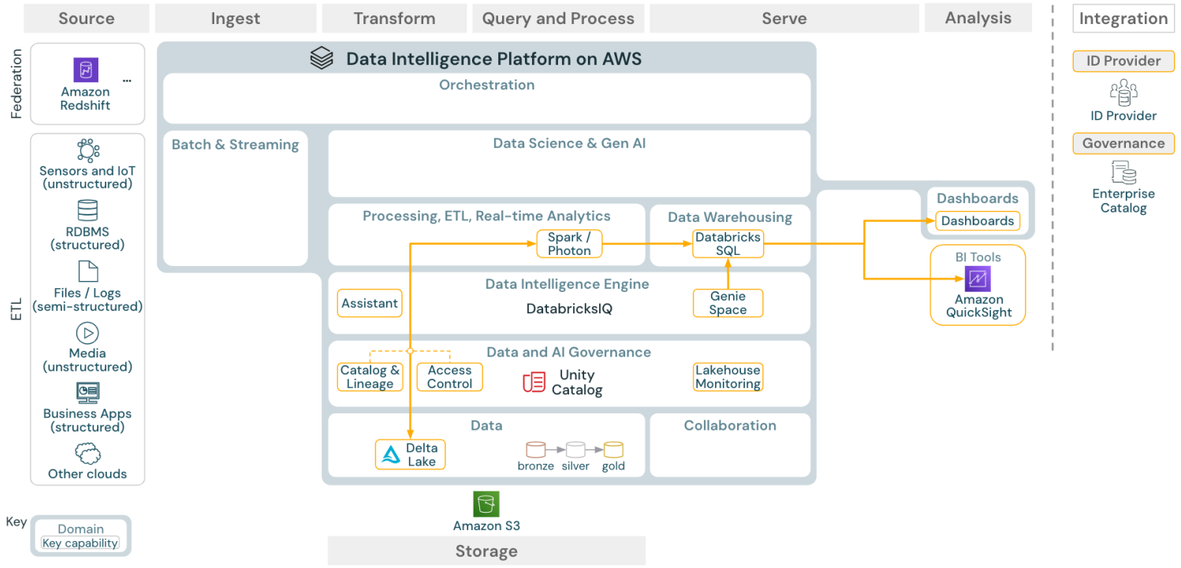


Data warehousing in the Lakehouse



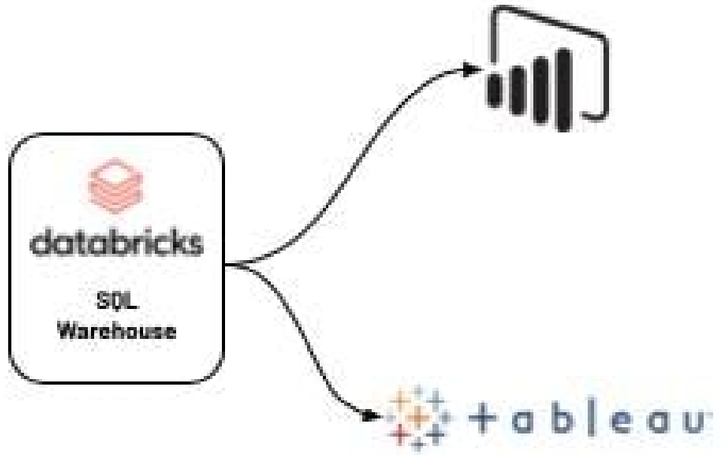
Benefits

- Single architecture for all workloads
- Flexibility and ownership with data
- Open-source technologies
  - Delta
  - ANSI SQL
- Cost effective solution



Business Intelligence ecosystem

- Integrate directly with your BI tool of choice
  - Partner Connect
  - Databricks Connect
  - JDBC / ODBC
- Performance and scalability
- Keep users where they are



Let's practice!

INTRODUCTION TO DATABRICKS SQL

Exploring Databricks SQL

INTRODUCTION TO DATABRICKS SQL

Kevin Barlow
Data Manager

datacamp

empty

INTRODUCTION TO DATABRICKS SQL

Databricks SQL key assets

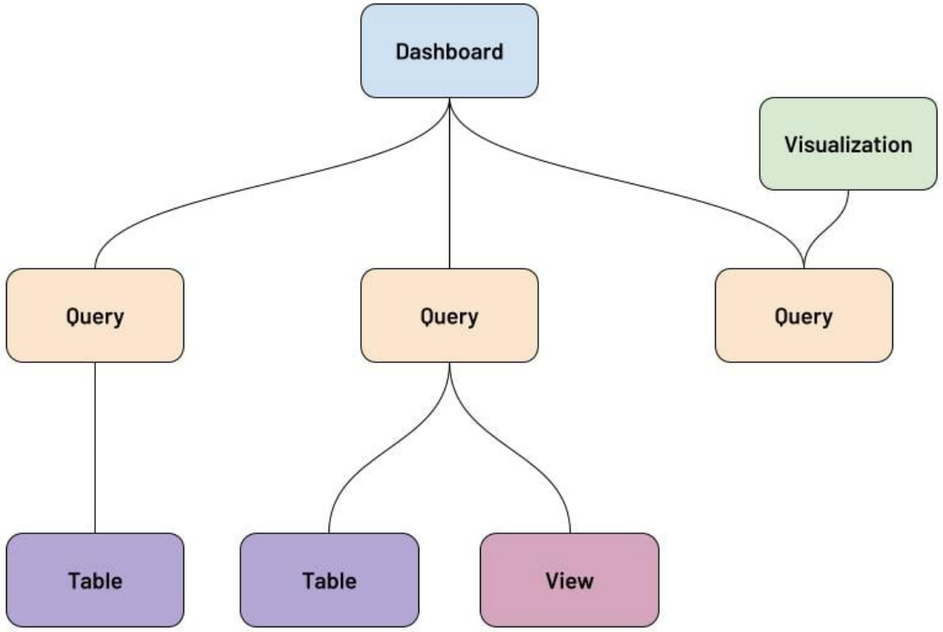
INTRODUCTION TO DATABRICKS SQL

Kevin Barlow
Data Manager

Helpful analogy

A tree consists of many different components, all of which make up the entire entity

In Databricks SQL, different components combine into a data warehouse solution



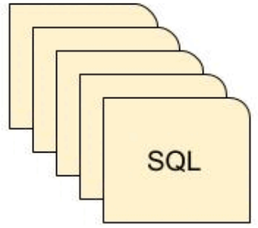
Query

- The base "unit" of analysis in Databricks SQL
- Runs SQL code against compute
- Uses ANSI SQL standard
- Process data from:
  - Unity Catalog
  - Delta tables
  - Data lake files
  - Data streams

```
SELECT
    orderdate AS Date,
    orderpriority AS Priority
    sum(totalprice) AS TotalPrice
FROM sfdc.sales.orders
GROUP BY
    1, 2
ORDER BY
    1, 2
```

SQL Warehouse

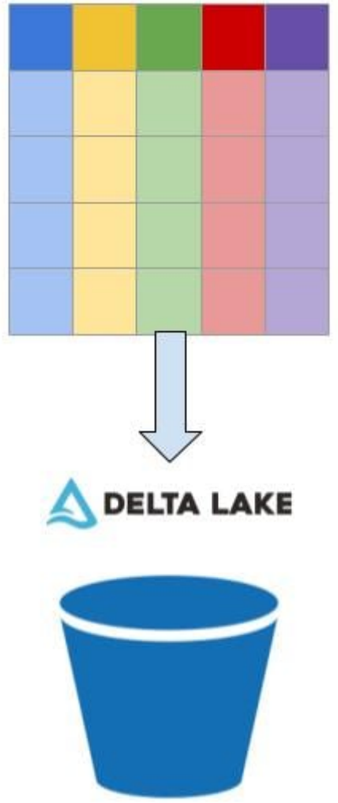
- Compute cluster dedicated for SQL
- Optimizations (e.g. Photon)
- Simpler administration
- Easy scaling
- Queries and BI tools



Tables versus views

Tables

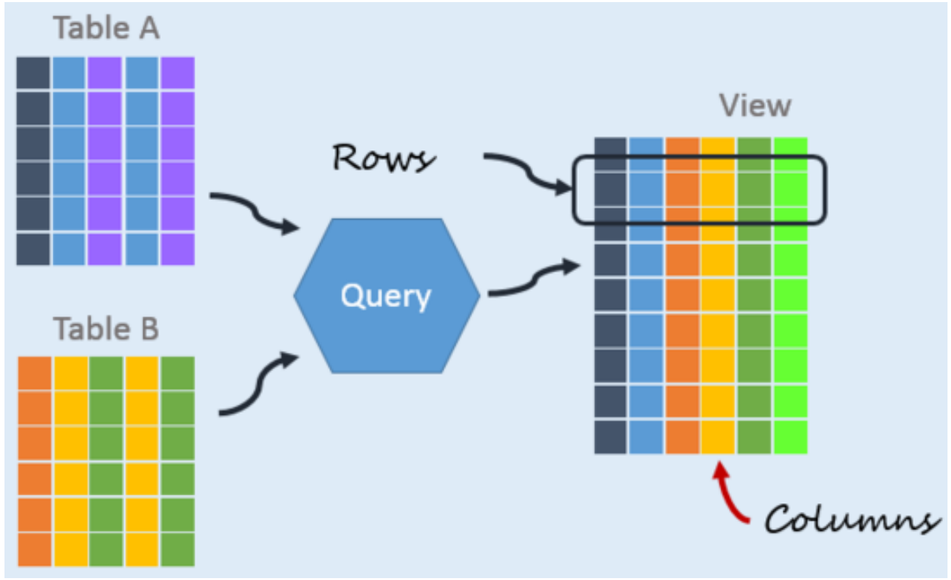
- Physical manifestations of datasets
- Written in Delta format
- Readable and accessible outside of the data pipeline
- Can optimize data layout (partitioning, etc.)



Tables versus views

Views

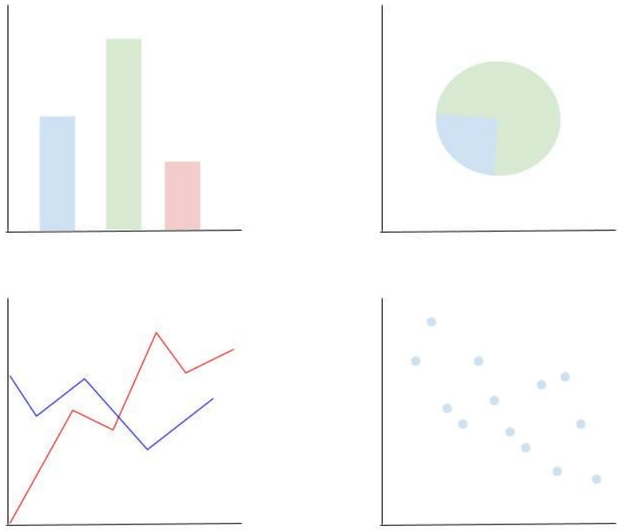
- Virtual representations of query results in Unity Catalog
- Fast performance for reading data
- Great for simplifying downstream queries
  - Source query has many joins, filters, etc.
- Incremental data processing available



Visualizations and dashboards

Visualizations

- Visual representations of a query result
- Created relative to a single query



Dashboards

- Collection of several visualizations
- Across multiple datasets / query results

Let's practice!

INTRODUCTION TO DATABRICKS SQL