Understanding the Hive Metastore

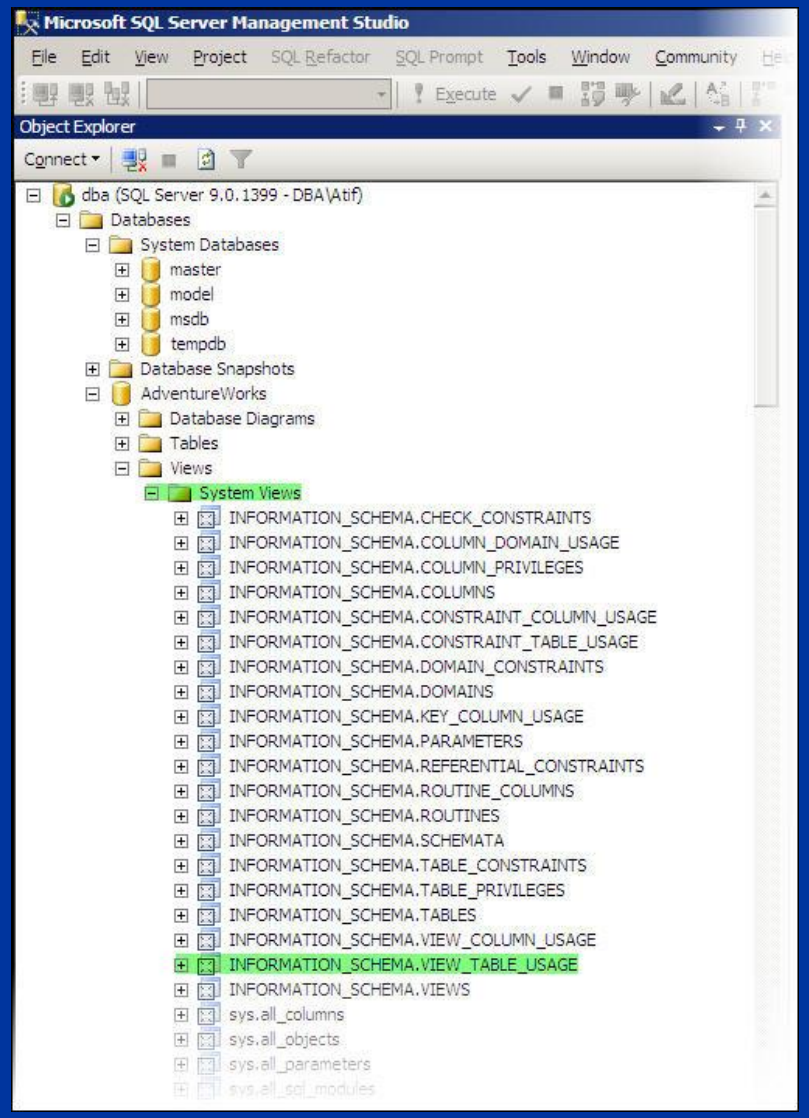
What is the Hive Metastore?

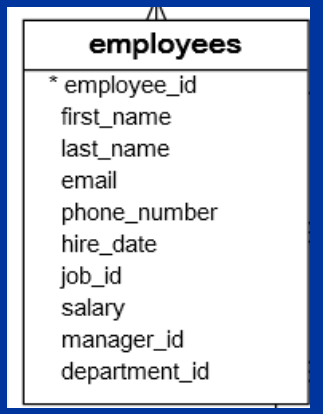
- A Catalog that Stores Data Describing Data (Metadata)
- Used to Support SQL
- How Spark Finds the Data
- Makes Spark Simulate DBMS Functionality
- It was Carried Over from Hadoop Map Reduce

A Metadata Store is called a Catalog

SELECT FirstName,
LastName, Age
From dbo.Employees

FirstName	LastName	Age
Mary	Jones	28
Benny	Fits	32



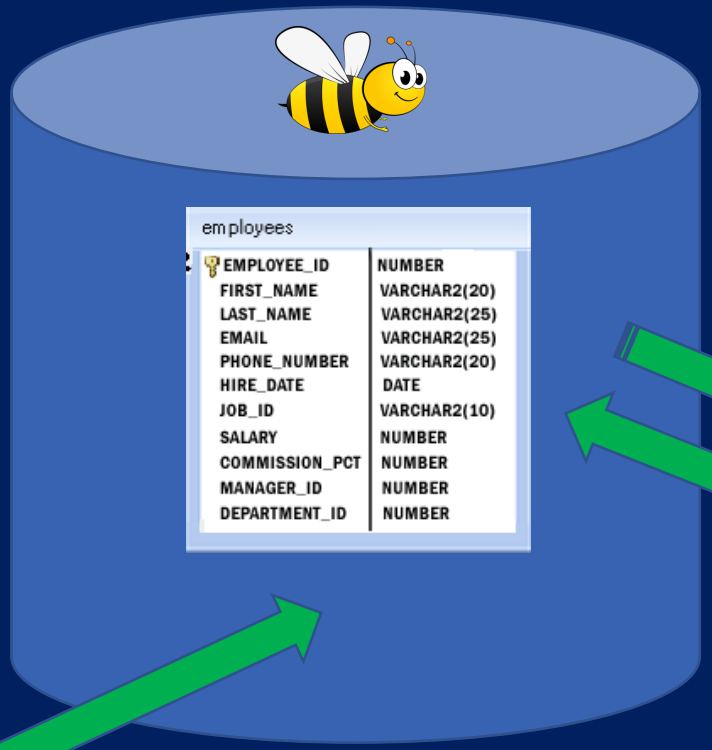


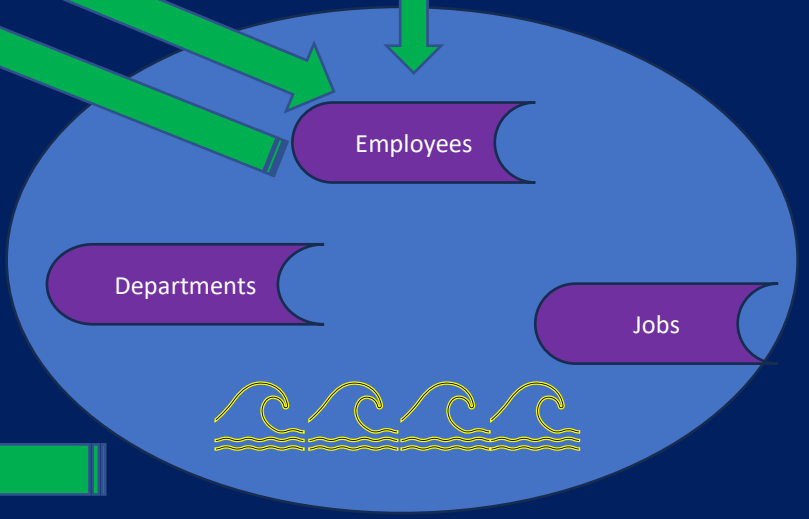
SQL Server
Database Data
File

Databricks Workspace

The Hive Metastore – Allows SQL Support

SELECT FirstName,
LastName, Age
From Employees



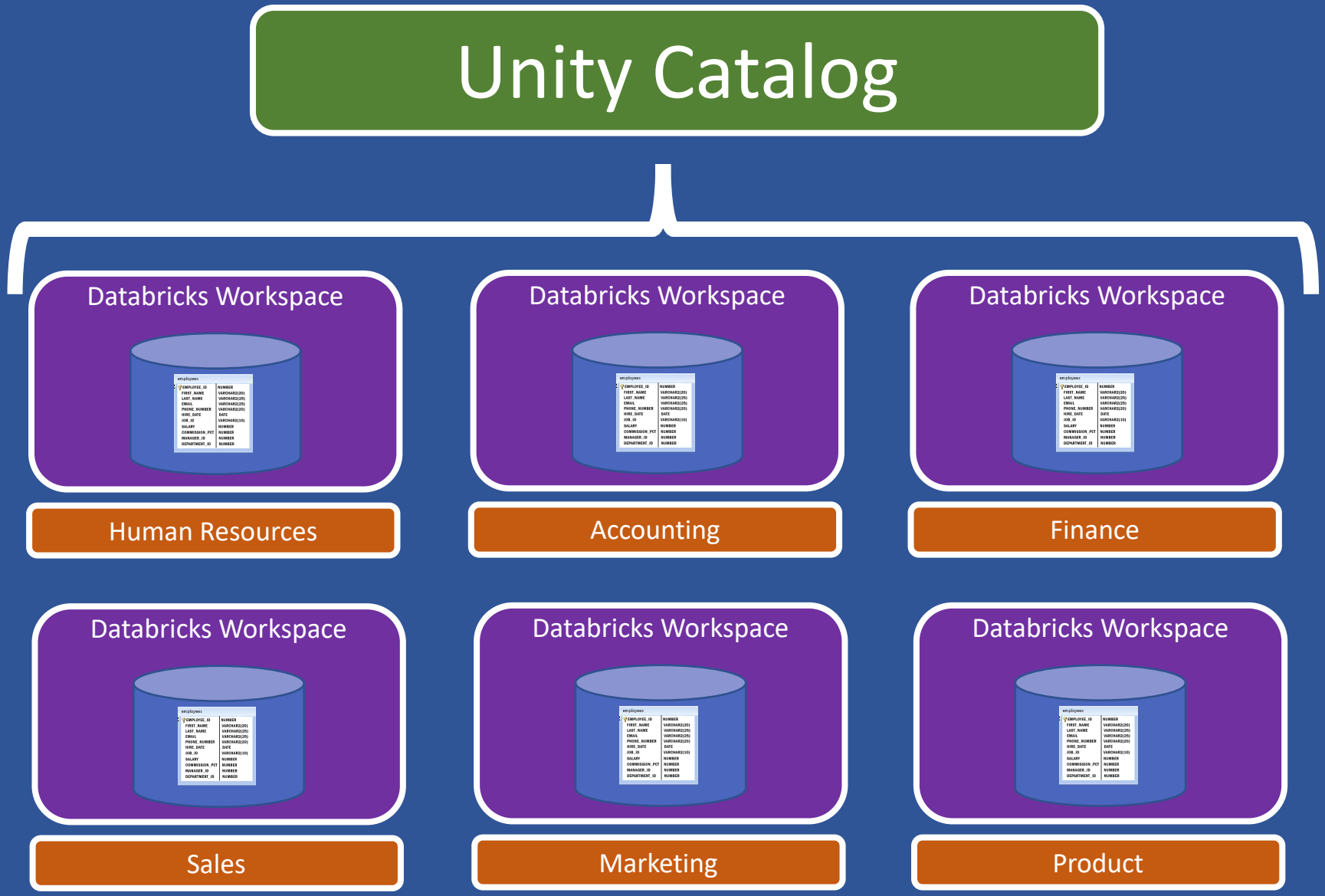


FirstName	LastName	Age
Mary	Jones	28
Benny	Fits	32

Power BI

CSV
JSON
TXT
Delta

Databricks Workspace Hive Catalog



Each Workspace Has Its Own Hive Catalog