

# Open Table Formats

## The Foundation of the Data Lakehouse

# Obligatory Intro Slide

- Long time data person holding various roles over the years
- Consultant specializing in Data Engineering and Architecture
- Working with AWS for 7 years
- I'll post the Github link in the Meetup and LinkedIn groups
- Please jump in with questions along the way or at the end



# Outline

- Some history
- Open Table Formats Intro
- Apache Iceberg Intro
- Athena and Iceberg Demo

# Some History

- First there was the Data Warehouse
- Then the Data Lake
- And now a hybrid approach called Data Lakehouse



# Open Table Formats Intro

- Its the foundation of the Data Lakehouse
- A layer that sits on top of object stores that brings database-like features to Data Lakes
- Delivering on the Data Lake hype
- Cheaper than Data Warehouses



# The Database-Like features

- ACID transactions
- Deletes, updates and inserts
- Time Travel queries
- Schema Evolution



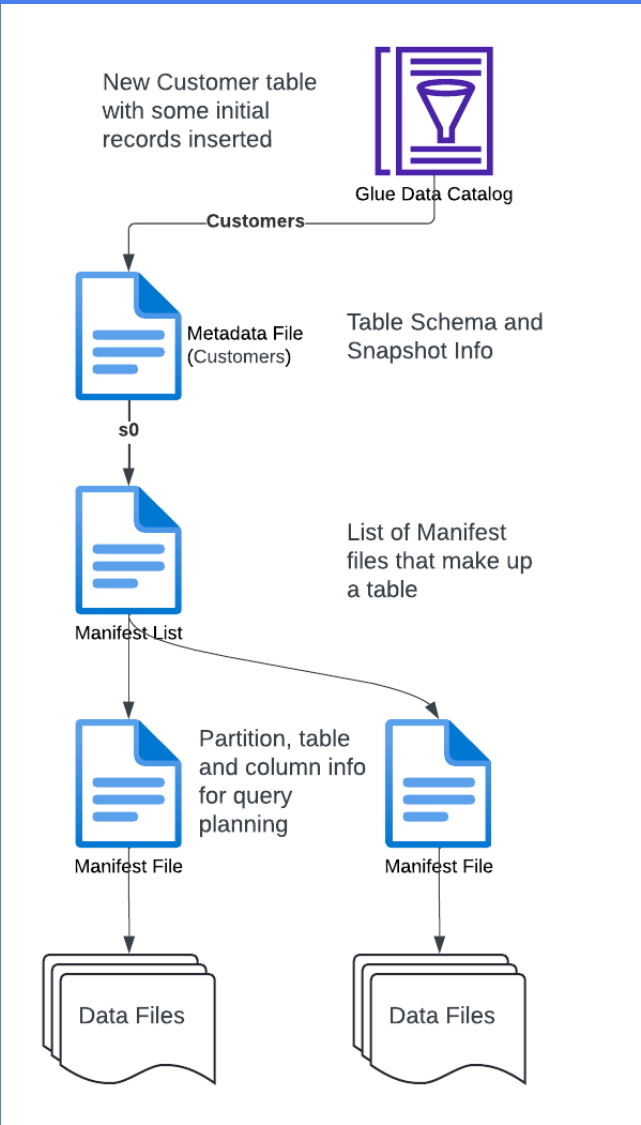
# The Open Table Format Players

- Delta Lake
- Apache Hudi
- Apache Iceberg
- Also, Apache XTable formerly OneTable

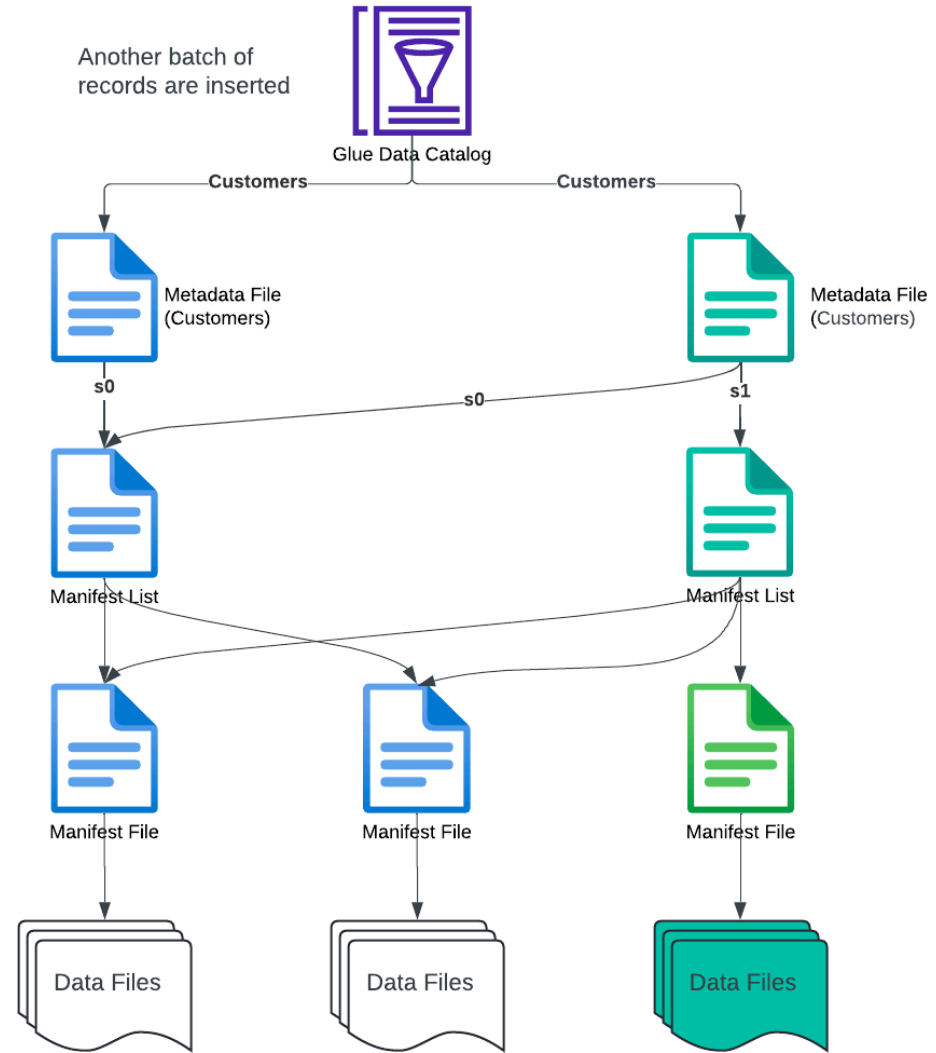


# Apache Iceberg Intro

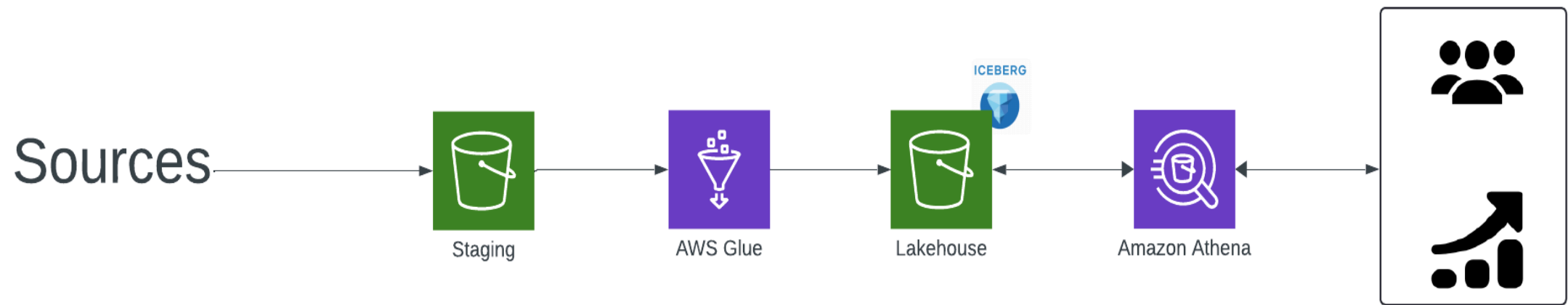




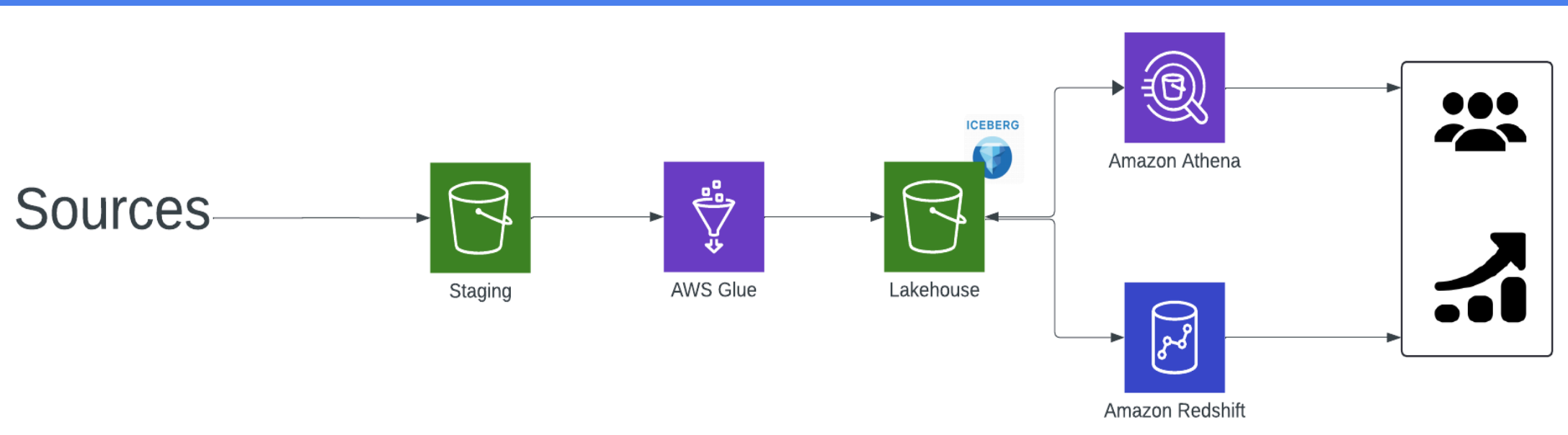
Another batch of records are inserted



# How does this all fit in a Data Platform?



# How does this all fit in a Data Platform?



# Athena and Iceberg Demo

