



# Week 3

Foundations of Data Warehousing

September 9, 2026

[illegible]

# The Medallion Architecture



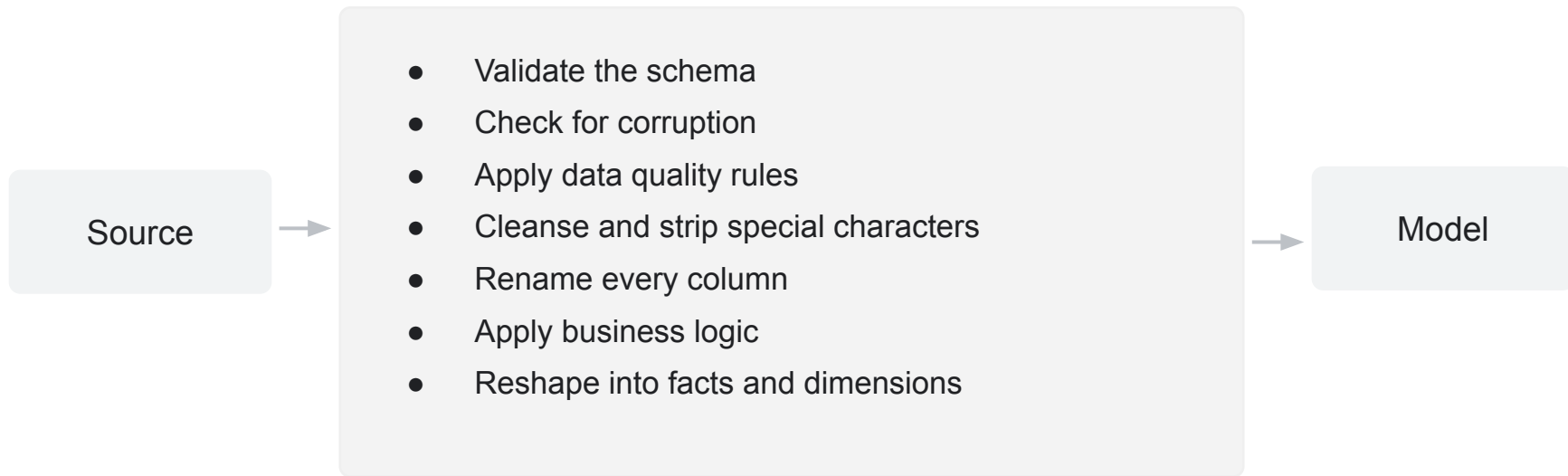
*Databricks coined this term, but the same concepts have been around since the '90s.*

# Why can't we go from source straight to the data model?



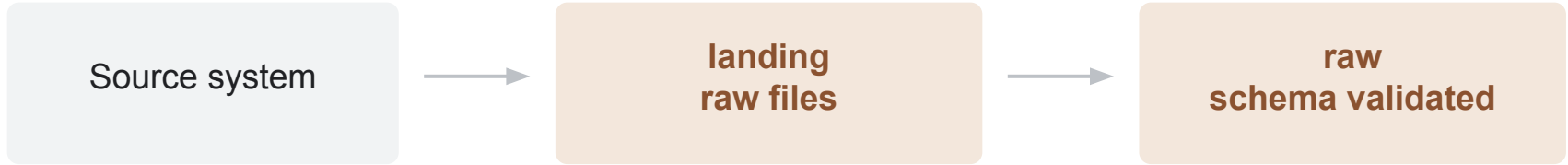
Revenue is \$10M or \$9.3M or \$7.8M.

# Why can't you have one pipeline that does everything?



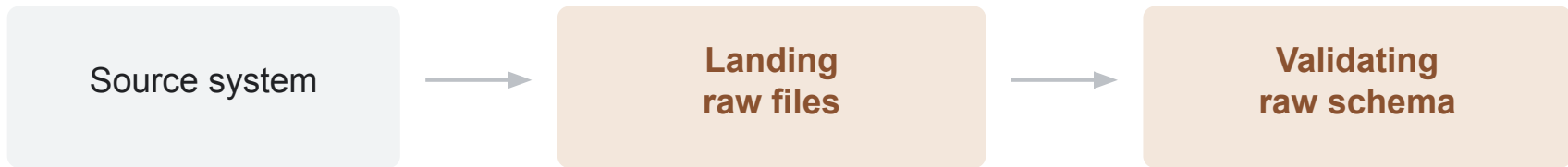
*What happens when I need to change a join condition?*

## Bronze has two steps



*Get the data into a queryable state. It may still be wrong.*

## Bronze has two steps



*Get the data into a queryable state. It may still be wrong.*

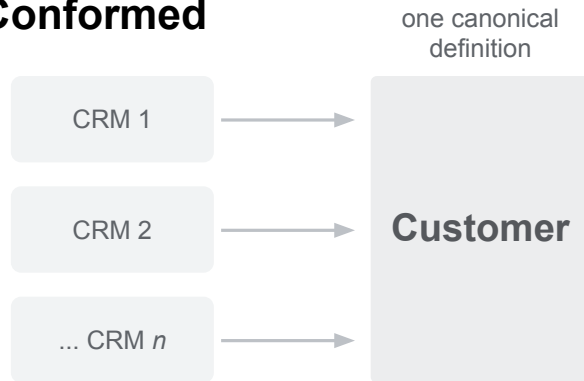
- How do we connect to all the different sources?
- How do we know the data we received is what we expected?
- What happens when things goes wrong?

# Silver has two parts

## Cleaned and validated

- Convert raw into clean tables
- Apply primary keys
- Reject rows that aren't real
- Merge new rows with existing ones
- Audit tables and logging

## Conformed



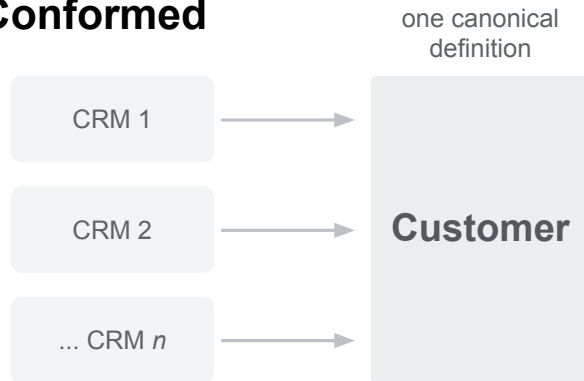
*The data in Silver still looks like the source system, just a cleaner, validated version of it.*

# Silver has two parts

## Cleaned and validated

- Convert raw into clean tables
- Apply primary keys
- Reject rows that aren't real
- Merge new rows with existing ones
- Audit tables and logging

## Conformed



*The data in Silver looks similar to the source system, just a cleaner, validated version of it.*

- What is the primary key when the source doesn't have one?
- When CRM 1 and CRM 3 disagree, which one wins?
- Can we prove *what* we changed and *why*, six months from now?

## Gold also has two parts

### Analytical models

Transform conformed data into facts and dimensions, ready for consumption.

### Semantic models

Business definitions, reporting tables, the knowledge graph.

*The data in Gold is shaped around the business questions.*

- Who decides what "revenue" or "active customer" means, and where does that definition physically live?
- How do we handle slowly changing dimensions?
- What happens when we have late-arriving facts?

## Gold also has two parts

### **Analytical models**

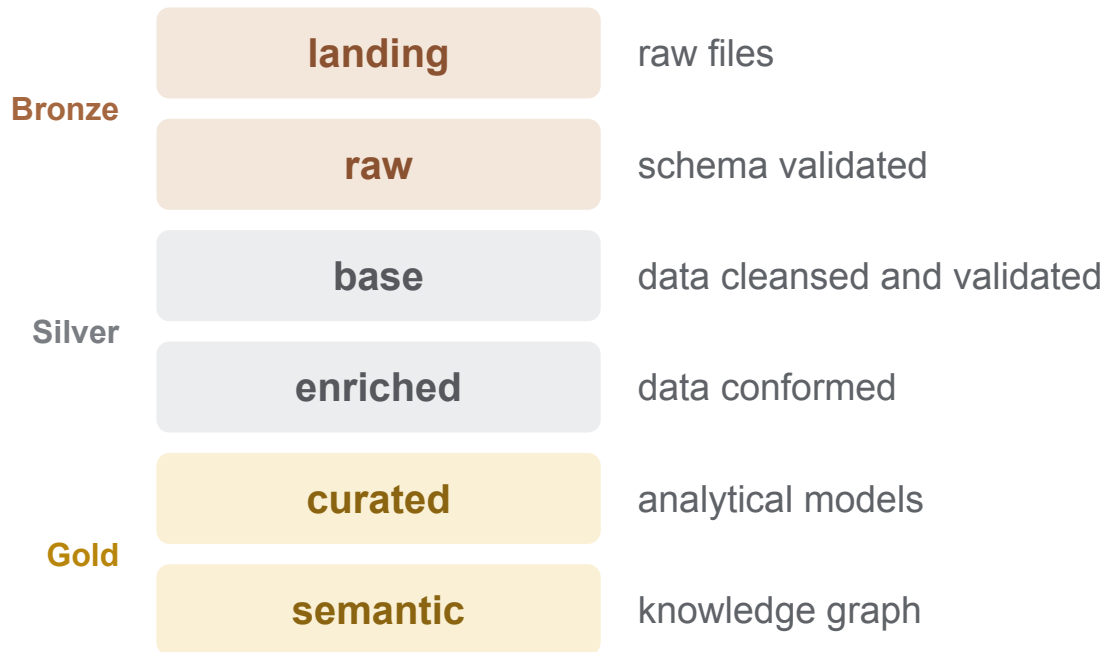
Transform conformed data into facts and dimensions, ready for consumption.

### **Semantic models**

Business definitions, reporting tables, the knowledge graph.

*The data in Gold is shaped around the business questions.*

# The 6 layers we're building



*Not every project needs all six layers. It depends on what your data looks like.*