



Week 7

Foundations of Data Warehousing

October 5, 2026

Review Milestone 4 (data ingestion pipelines)

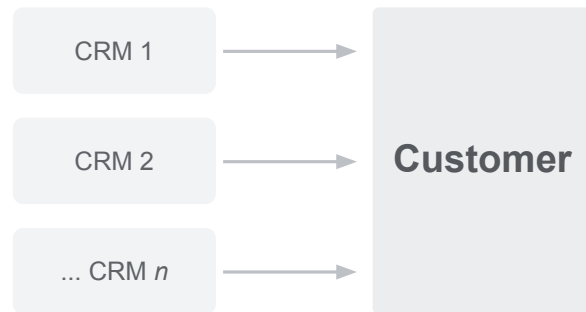


Recall: Silver has two main parts

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

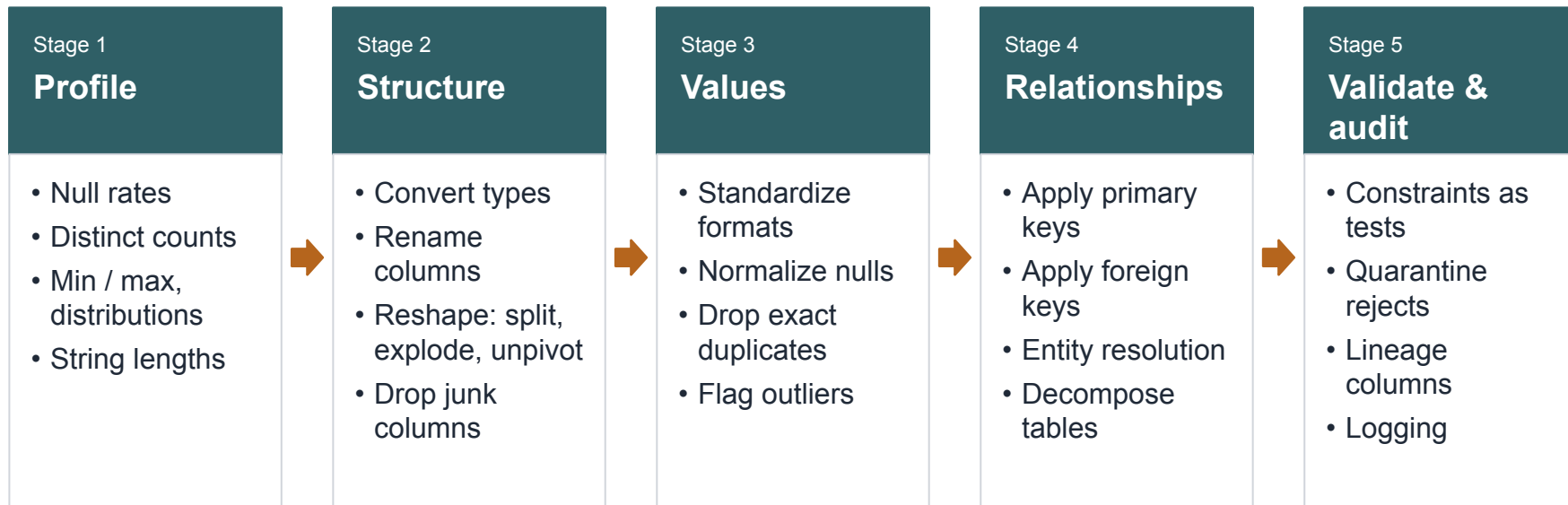
Conformed



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

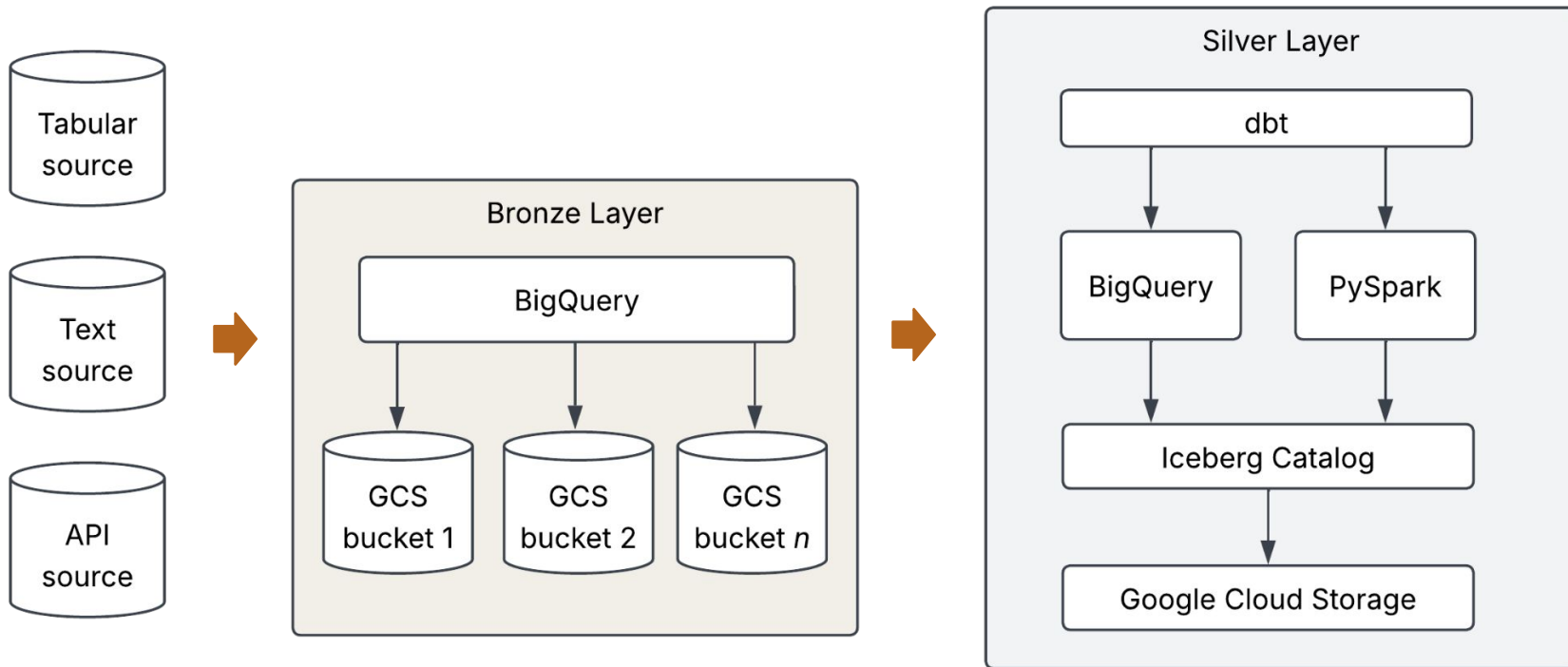
- What is the primary key when the source doesn't have one?
- When CRM 1 and CRM 3 disagree, which one wins?
- Can we prove *what* we changed and *why*, six months from now?

Silver cleaning pipeline

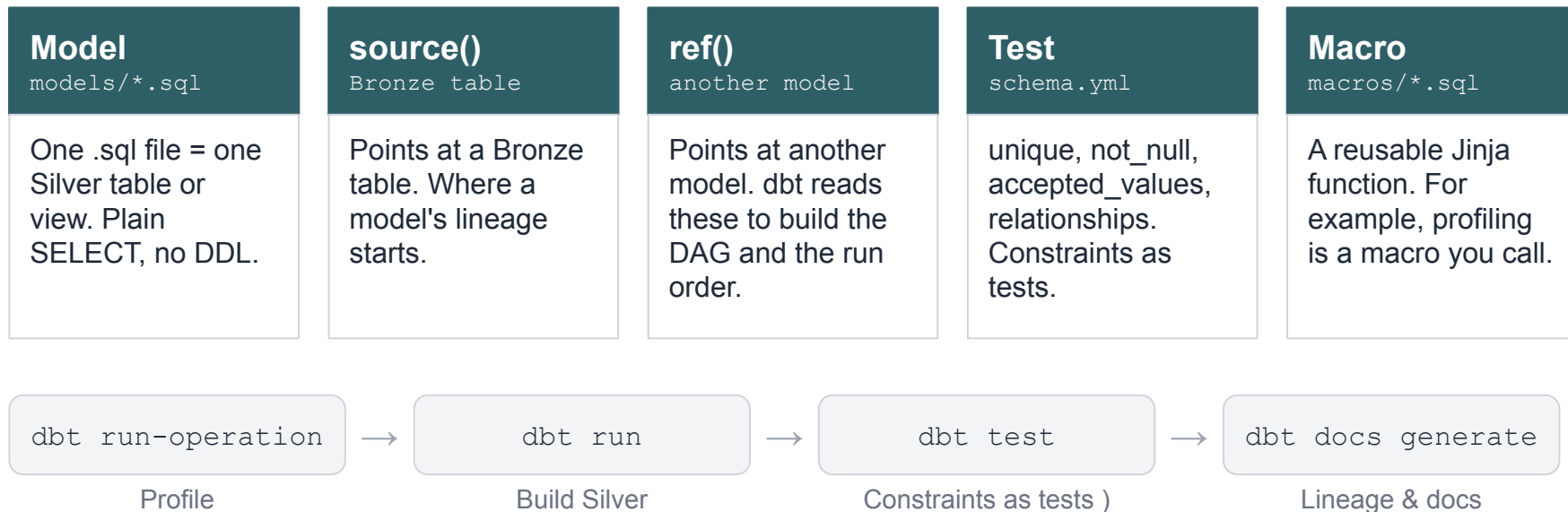


Each task is version-controlled code, and every one is a deterministic rule except one. Which one is probabilistic?

Where Silver lives



dbt overview



You write SELECT statements with Jinja snippets; dbt compiles them, orders them by dependency, and runs them in BigQuery.

A dbt model with Jinja

Model (SQL + Jinja)

```
-- models/silver/customers.sql
{{ config(materialized='table') }}

{% set text_cols = ['email', 'phone', 'city'] %}

select
  customer_id,
  {% for col in text_cols %}
  nullif(trim({{ col }}), '') as {{ col }},
  {% endfor %}
  cast(signup_date as date) as signup_date
from {{ source('bronze', 'customers') }}
```

Compiled SQL (what BigQuery runs)



```
select
  customer_id,
  nullif(trim(email), '') as email,
  nullif(trim(phone), '') as phone,
  nullif(trim(city), '') as city,
  cast(signup_date as date) as
  signup_date
from `my-project.bronze.customers`
```

Jinja runs first and produces plain SQL. Only the compiled SQL reaches BigQuery.

Iceberg catalog setup

Flavor	Written by	Read by	How it is set up
BigQuery-managed Iceberg tables	dbt / BigQuery	BigQuery Python via <i>catalog federation</i> (read-only)	A GCS bucket + a BigQuery Cloud resource connection. No catalog resource.
Iceberg REST catalog tables	Python (PyIceberg)	Python BigQuery (read-only)	A catalog + namespace in the Lakehouse runtime catalog.

Rule of thumb: whoever writes the table chooses the catalog flavor.