

Sony Pictures Entertainment - Data Engineering Demo

Snowflake Medallion Architecture Overview

Medallion Architecture

Bronze (Raw) → Silver (Refined) → Gold (Analytics)

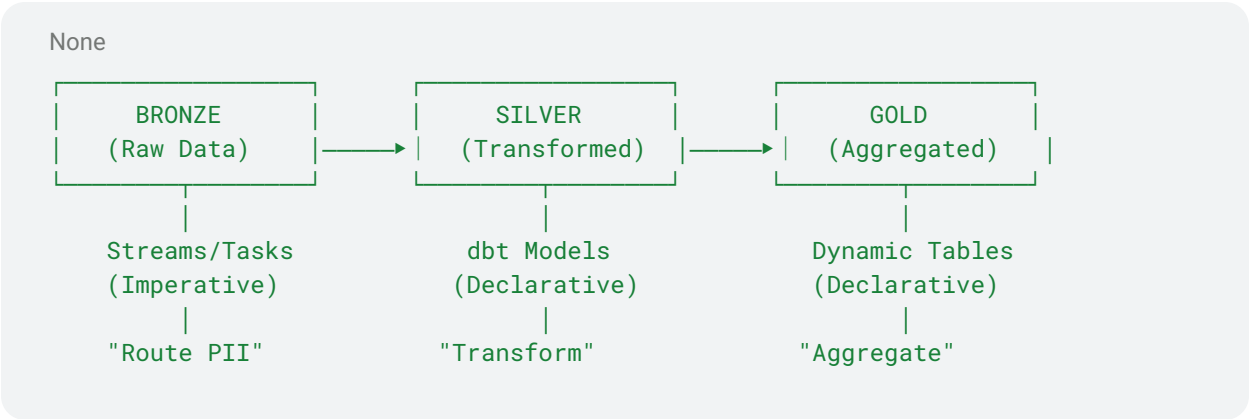
Layer	Purpose	Tables
Bronze	Immutable raw data landing zone	RAW_FAN_INTERACTIONS, RAW_TITLE_METADATA, RAW_BOX_OFFICE
Silver	Cleansed, conformed, deduplicated	STG_* tables, DIM_FANS, DIM_TITLES, FACT_DAILY_PERFORMANCE
Gold	Business-ready aggregations	AGG_FRANCHISE_PERFORMANCE, AGG_FAN_LIFETIME_VALUE

Demo Assets

Component	Technology	Purpose
CDC & Routing	Streams + Tasks	Fan data split: Verified → PII Vault, Guest → Public
Transformations	dbt (9 models)	Bronze→Silver→Gold with version control & testing
Real-time Aggregates	Dynamic Tables	Auto-refreshing franchise & LTV metrics

Component	Technology	Purpose
SCD Type 2	dbt Snapshots	Track fan region/preference changes over time
Data Quality	6 DMFs	<code>valid_imdb_score</code> , <code>valid_email_hash</code> , <code>positive_revenue</code> , etc.
Testing	37 dbt tests	Schema + custom data tests for pipeline integrity
Monitoring	Streamlit Dashboard	Pipeline health, DQ metrics, lineage visualization

Hybrid Architecture Pattern



When to use each:

- **Streams/Tasks:** Complex routing, conditional logic, PII handling
- **dbt:** Transformations, testing, documentation, version control
- **Dynamic Tables:** Continuous aggregations with `TARGET_LAG`

Key Stats (Current Pipeline)

Metric	Value
Bronze Layer Rows	205,015

Metric	Value
Silver Layer Rows	105,000
Gold Layer Rows	11,547
DMF Checks	6/6 Passing
dbt Tests	37 Configured
SCD2 Snapshot Records	27,000+
