DATA MODEL DOCUMENTATION

Bronze Layer – Raw Data

Purpose

The Bronze layer acts as a **landing zone** for raw data. It preserves the original structure and content of the source files.

Tables

- bronze_sales_transactions
- bronze_product_master
- bronze_store_region

Characteristics

- Append-only ingestion
- No data transformations
- Schema preserved as-is
- Metadata added:
  - ingestion_timestamp
  - ingestion_date
  - source_system

Why Bronze Exists

- Supports replay and reprocessing
- Enables audits and debugging
- Acts as a data backup layer

Silver Layer – Cleaned & Conformed Data

Purpose

The Silver layer applies **data quality rules, cleansing logic, and calibration** to prepare trustworthy data.

Silver Tables

silver_sales_transactions

- Cleaned transactional fact table
- Deduplicated using transaction_id
- Corrected monetary values

- Currency normalized to USD

silver_product_master

- Cleaned product dimension

- Standardized text fields

- Invalid records filtered

silver_store_region

- Cleaned store dimension

- Standardized country, region, city fields

silver_sales_quarantine

- Stores invalid records

- Captures rejection reason

- Supports data quality analysis

Silver Layer Characteristics

- Schema enforced using Delta

- Incremental MERGE logic

- Idempotent pipeline runs

- Invalid data isolated

Gold Layer – Analytics Ready

Purpose

The Gold layer provides **business-ready aggregated datasets** optimized for reporting and dashboards.

Gold Datasets

gold_daily_sales_summary

- Daily revenue metrics

- Used for trend analysis

gold_monthly_revenue_by_region

- Monthly revenue grouped by region and city

- Used for regional performance dashboards

gold_product_performance

- Product-level metrics

- Used for product performance analysis

Gold Layer Characteristics

- Stable schema

- Derived only from Silver

- Minimal transformations in Power BI

- Optimized for analytics