

Created a full DLT Medallion pipeline with two branches:

Branch 1: Bronze → Silver → Gold (SCD1)

Branch 2: Bronze → Silver_v2 → Stage → Gold (SCD2)

Each table was implemented using:

- **Bronze**: Streaming ingestion
- **Silver**: Transformations
- **Silver_v2**: CDF-enabled transformations
- **Stage SCD2**: DLT apply_changes with stored_as_scd_type = 2
- **Gold SCD1**: DLT apply_changes with stored_as_scd_type = 1

Raw_2_bronze.ipynb

- **Performs minimal cleaning**, such as standardizing column names.
- **Creates Bronze streaming tables** using DLT:
 - bronze_patients_cdf ◦ bronze_conditions_cdf ◦ bronze_encounters_cdf
 - bronze_providers_cdf

bronze_2_silver_2_gold_SCD1.ipynb

1. Reads Bronze tables (not CDF).
2. Creates Silver views:
3. These Silver views allow:
 - type fixes ◦ trimming ◦ upper casing
 - timestamp normalization
4. Creates Gold SCD1 tables using DLT
5. SCD1 overwrites data (latest value only)

bronze_2_silver_2_stage_SCD2.ipynb

Transform Bronze → Silver (with streaming CDF) → Stage SCD2 tables.

This notebook is responsible for SCD Type 2 logic.

1. Reads Bronze tables using CDF
2. **Builds Silver_v2 views** for each domain:
3. **Creates SCD2 stage tables** using DLT:
4. SCD2 stage tables automatically generate: start_at, end_at, is_active, historical versions

Stage_2_gold_scd2.ipynb

1. This notebook takes the SCD2 stage tables and pushes them into final Gold SCD1 tables.
2. This is the final branch, producing your slowly changing dimension Type-2 gold layer.
3. Read Stage SCD2 Tables
4. Used Silver_v2 → stage tables (patients_type2_stage, conditions_type2_stage.)
5. Create Gold SCD2 Tables from Stage

Pipeline fails because:

- The SILVER table is being read as a streaming source
- Performed an update, because to test SCD
- But CDF was not enabled on that table
- So Spark cannot track the updated rows
- And the pipeline stops to avoid corrupting your SCD logic

Fix:

- Turned on CDF for ALL Bronze tables
- Updated Silver_v2 to use streaming CDF
- Removed unsupported window functions from streaming views