

## 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