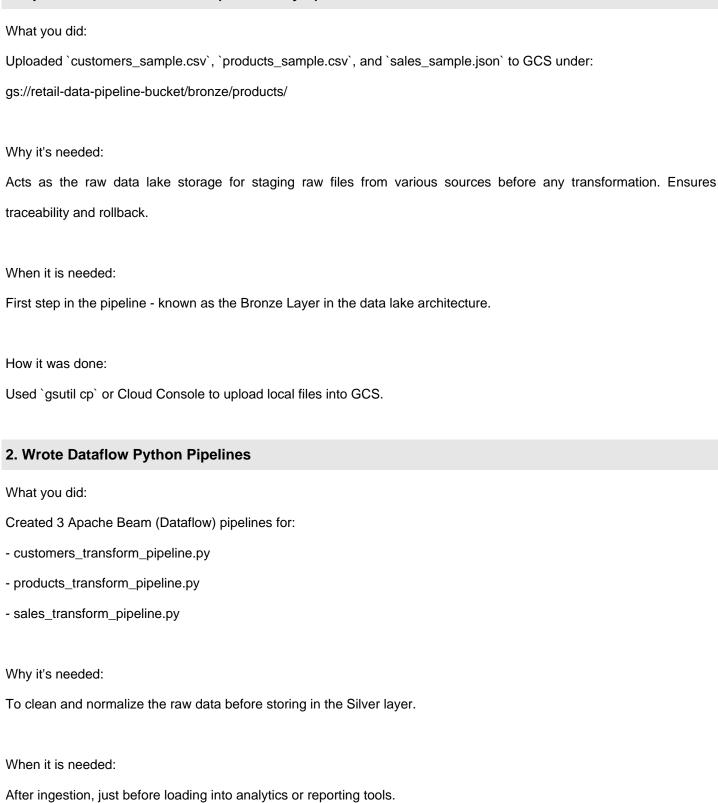
Retail Sales Data Pipeline - Task Summary

1. Uploaded Raw Data to GCS (Bronze Layer)



How it was done:

Used Apache Beam with Python, defined custom DoFn classes to transform each row, saved `.py` files using nano

Retail Sales Data Pipeline - Task Summary

editor in Cloud Shell. 3. Created Metadata Files What you did: Created `metadata.json` files for each pipeline (e.g., sales_metadata.json). Why it's needed: Defines runtime parameters like input/output paths for Flex Template execution. When it is needed: At the time of building Flex Templates and launching Dataflow jobs. How it was done: Used nano to create JSON files with name, description, and parameter keys. 4. Built Dataflow Flex Templates What you did: Built templates using gcloud CLI for: - customers_template.json - products_template.json - sales_template.json Why it's needed: So that the pipeline can be triggered on-demand or on-schedule without having to rebuild each time. When it is needed: Before setting up automation using Pub/Sub or Scheduler.

Used `gcloud dataflow flex-template build` commands with proper flags, linking metadata and Python files.

How it was done: