# Roles and Responsibilities

As a Data Engineer for the Retail Sales Data Pipeline project, I was responsible for implementing the data cleanup and automation logic across the pipeline. I handled the ingestion of data from multiple sources, including RESTful APIs, CSV files, and on-prem SQL Server, by developing and deploying Cloud Functions triggered via Pub/Sub and Cloud Scheduler to load raw data into the Bronze Layer of GCP Cloud Storage. I ensured the integrity of the pipeline by performing data validations such as row count checks and schema validations. I developed a Beam pipeline using GCP Dataflow to clean and transform the raw data—removing duplicates, handling nulls, standardizing formats—and then stored the processed output into the Silver Layer and finally into GCP Cloud SQL for further analysis. My role was key in automating the pipeline stages to ensure each stage only executes after the successful completion of the previous one.