

SQL ETL Pipeline Simulation Project

Internship Completion Report – Elevate Labs

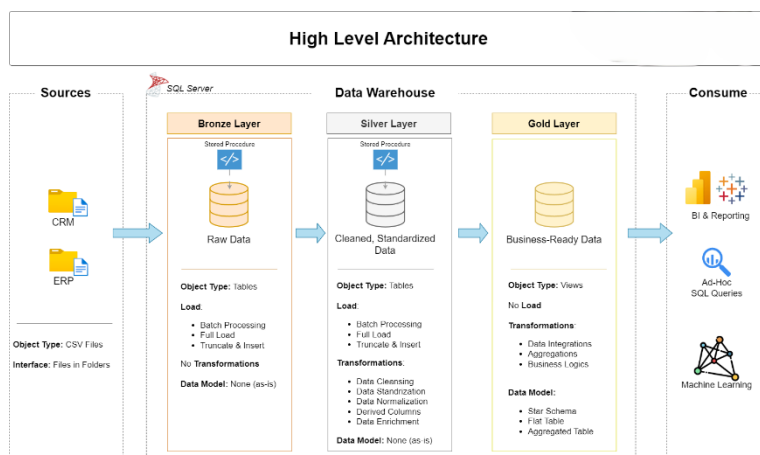
By Manasi Mane

1. Introduction

In real-world systems, raw data collected from different sources cannot be used directly for analysis due to quality issues such as missing values, duplicates, and inconsistent formats. To address this, organizations rely on **ETL (Extract, Transform, Load)** pipelines to prepare data for analytics.

This project was developed as part of the **Elevate Labs Internship Program** to simulate a real-world **SQL-based ETL pipeline**. The project also serves as a **portfolio project**, demonstrating practical skills in SQL, data transformation, and layered data architecture.

2. Abstract



The SQL ETL Pipeline Simulation Project demonstrates an end-to-end ETL workflow using SQL. Raw CSV data is first ingested into a staging layer, then cleaned and standardized in a transformation layer, and finally loaded into production-ready tables.

The project follows a **layered architecture (Staging/Bronze, Transformation/Silver, Production/Gold)** inspired by industry best practices. The final output consists of clean, structured tables suitable for analytical queries and reporting.

3. Tools Used

- **SQL Server / MySQL / PostgreSQL**
- **SQL Server Management Studio (SSMS) / DBeaver**
- **CSV Files** (raw data sources)
- **Draw.io** (ETL architecture diagrams)
- **Git & GitHub** (version control and documentation)

4. Steps Involved in Building the Project

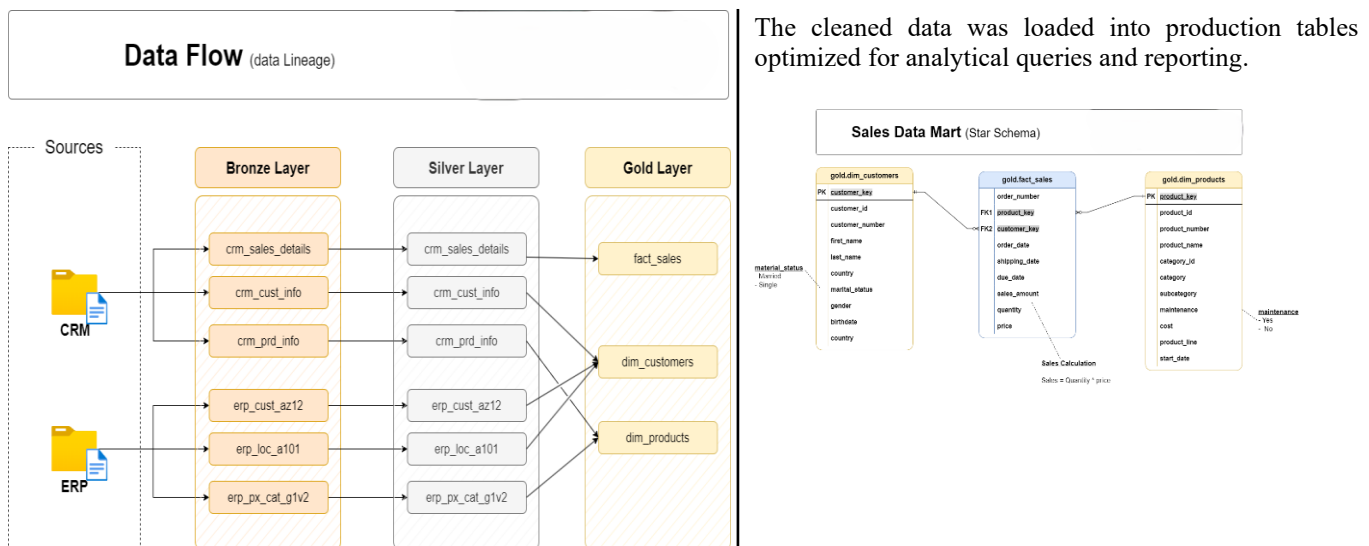
Step 1: Data Extraction (Bronze Layer)

Raw data was imported from CSV files into staging tables without modification. This layer acts as a landing zone and preserves the original source data.

Step 2: Data Transformation (Silver Layer)

SQL transformations were applied to clean the data by removing duplicates, handling null values, and standardizing formats to ensure data consistency and reliability.

Step 3: Data Loading (Gold Layer)



Step 4: Validation and Reporting

Basic analytical queries were executed on the production tables to validate the ETL process and simulate real-world data consumption.

5. Conclusion

This project successfully demonstrates a complete **SQL-based ETL pipeline** using a layered architecture. It provided hands-on experience in data extraction, transformation, and loading using SQL. Completing this project as part of the **Elevate Labs Internship Program** strengthened my understanding of ETL workflows and enhanced my confidence in explaining SQL-based data processing during interviews.