SQL ETL Pipeline Simulation Project

Prepared by: Nandana S

Date: 12-07-2025

Introduction & Abstract

This project simulates an ETL (Extract, Transform, Load) pipeline using SQL with SQLite and DB Browser. The objective is to demonstrate data ingestion, cleaning, transformation, and tracking of data load processes through automation.

Tools Used

- SQLite (via DB Browser for SQLite)
- CSV files as raw input

Steps Involved in Building the Project

1. Extract:
   - Raw sales data was imported from a CSV file into a staging table (`staging_sales`).

2. Transform:
   - Removed duplicates and rows with NULL values.
   - Inserted cleaned data into a new table: `production_sales`.

3. Load:
   - Created `etl_audit_log` to track insert actions.
   - Added a trigger on `production_sales` to automate logging into `etl_audit_log`.

4. Export:
   - Exported cleaned data (`production_sales.csv`), audit logs (`etl_audit_log.csv`), full SQL backup (`etl_pipeline_backup.sql`), and the database file (`etl_pipeline.db`).

Conclusion

This project showcases the fundamentals of ETL using SQL alone. It successfully demonstrates how to simulate an enterprise-grade pipeline with minimal tools and manual data handling using DB Browser for SQLite. The audit and automation features enhance the robustness and traceability of the pipeline.

Deliverables

1. production_sales.csv – Cleaned production table
2. etl_audit_log.csv – Audit tracking insert logs
3. etl_pipeline.db – SQLite database
4. etl_pipeline_backup.sql – Full SQL dump of the project