Shepherd Crisp
27 October 2025

Data Project 1

I. Project Overview

This project delivers the design and implementation of a data warehouse (DW) using MySQL workbench and the integration of a NoSQL MongoDB data source. The goal of this project was to extract, transform, and load (ETL) data from multiple systems including the Adventureworks OLTP database and simulated MongoDB web event data to form a unified warehouse schema for analytics.

II. Process Summary

A. Data Warehouse Creation (MySQL)

1. In MySQL workbench, I created four SQL scripts that formed the basis of this project after running the Adventureworks database and the Adventureworks queries. The four SQL scripts are listed below:
    a) 1 Create DW.sql – This created the adventure_dw schema and table structures
    b) 2 Load Dim.sql – This loaded and populated dimension tables
    c) 3 Create Date Dim.sql – Built dim_date table
    d) 4 Integrate Dim.sql – Linked fact and dimension tables with foreign keys
2. These established the base DW schema to hold transformed Adventureworks data

B. ETL Pipelines (Python / Jupyter)

1. Two python jupyter notebooks were used to automate and extend the ETL process:
    a) Crisp P1 Adventureworks DW ETL.ipynb
        (1) Extracted data from OLTP database using sqalchemy and pandas
        (2) Transformed date to mate dim model
        (3) Loaded data into MySQL DW tables
    b) Crisp P1 Adventureworks Mongo ETL.ipynb
        (1) Created a simulated web activity dataset (web_events.json) and stored it in MongoDB Atlas
        (2) Extracted the MongoDB data using pymongo and certifi
        (3) Joined it with the DW's customer and date dimensions to for fact_web_events
        (4) Validated some loads with SQL queries

C. Validation & Demo Queries

1. Some SQL queries were executed to validate and demonstrate functionality

III. Deployment Strategy

A. Environments: Jupyter Notebook with local MySQL instance and MongoDB Atlas

B. Execution order:

1. Run SQL scripts in MySQL workbench to initialize DW schema
2. Run DW ETL notebook to load Adventureworks data
3. Run Mongo ETL notebook to load and integrate web events
4. Validate results using SQL queries in Jupyter

IV. Results

A. Fully functional data warehouse with 5 dim tables and 3 fact tables

B. Integrated MongoDB web event data with Adventureworks customer and date dimensions
C. Verified data integrity and product reports through SQL