



DSA 2040A US 2025 LAB 1

Lab Manual: Building a Mini Data Warehouse for a Retail Chain

Overview

Objective: Students will design and build a star schema-based data warehouse, load data from CSV files, and run analytical queries to gain business insights.

Tools:

- PostgreSQL / MySQL / SQL Server / Snowflake (pick one DBMS)
- SQL client (pgAdmin, DBeaver, SSMS, etc.)
- Provided CSV files

Part 1: Introduction to Data Warehousing

Briefing:

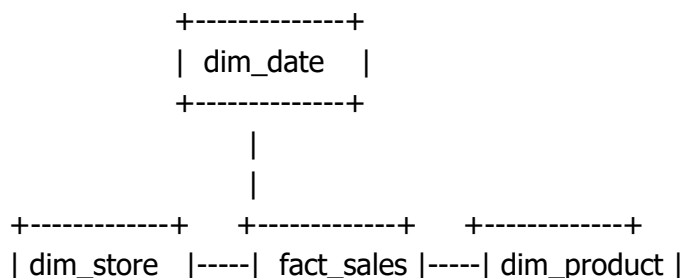
Imagine you're working as a junior data engineer at a retail company. The company wants to analyze sales across different products, stores, and time periods. Your task is to design a simple data warehouse and answer business questions using SQL.

Key Concepts:

- Data Warehouse
- Star Schema
- Fact and Dimension Tables
- ETL (Extract, Transform, Load)

Part 2: Schema Design and Table Creation

Star Schema:



+-----+ +-----+ +-----+

SQL Tables:
(see schema.sql)

Part 3: Load Data

Load the CSV files into the database using either the SQL client's import tool or the SQL COPY command.

See CSV files provided:

- dim_date.csv
- dim_product.csv
- dim_store.csv
- fact_sales.csv

SQL Loading Example (PostgreSQL):

```
COPY dim_date FROM '/path/to/dim_date.csv' DELIMITER ',' CSV HEADER;
```

Part 4: Querying the Warehouse

Run the following analytical queries:

1. Total Revenue by Product Category
2. Monthly Sales Trends
3. Revenue by Region
4. Top Products by Quantity Sold

(See queries.sql)

Part 5: Reflection & Discussion

Discussion Questions:

1. Why use a star schema instead of a normalized schema?
2. What are the benefits of separating facts from dimensions?
3. What types of business decisions could this warehouse support?

Part 6: Push to GitHub

Steps:

1. Organize project files
2. Create a GitHub repo(PUBLIC)
3. Initialize git and push code
4. Add collaborators

5. Create a README.md with project details

(See README.md template in zip)