

Data Warehouse Assignment

Objectives of the project

1. Understand Different Types of SQL Projects

A **data warehouse** provides a **centralized, automated, reliable** system for analytics and reporting.

SQL projects vary depending on **business needs**, data complexity, and reporting requirements. The main types include:

- **ETL Projects**
- **Data Integration**
- BI/Reporting Projects

2. Learn What a Data Warehouse Is and Its Role

A **Data Warehouse House** is a centralized system that stores large volumes of historical data for **analysis and reporting**.

Key Characteristics:

- Stores data from multiple sources.
- Data is cleaned, transformed, and structured using **ETL**.
- Supports historical analysis, business intelligence, and decision-making.
- Helps in tracking performance.
- Enables complex queries and dashboards.
- Provides a **"single source of truth"** for data.

3. Understand ETL and Its Process

ETL (Extract, Transform, Load) is a key process for moving and shaping data before it enters the data warehouse.

Steps:

1. **Extract:** Collect data from sources

2. **Transform:** Clean, convert, rename, join, and filter data
3. **Load:** Insert the data into the Data Warehouse

4. Explore Provided Project Materials

- **CSV files** from CRM and ERP (your source data)
- **SQL Server Express + SSMS** (your local database)
- **Draw.io** for ER diagram design
- **Notion** for planning tasks (epics, subtasks)
- **GitHub repo** (for code, versioning)
- Pull-based extraction from files
- Staging, transformation, and loading into final tables
- Full batch loading process with update logic (SCD1)

Summary of ETL & Data Warehouse (DWH) Concepts

ETL (Extract, Transform, Load)

ETL is the **core process** that moves data from source systems into a **Data Warehouse**. It includes:

1. Extract

Pull raw data from sources (e.g., databases, CSV files, APIs)

Example: Read CRM and ERP data from CSV files

2. Transform

Clean, filter, enrich, and structure data

Convert formats, handle nulls, rename columns, remove duplicates

3. Load

Insert transformed data into the target (data warehouse)

Use techniques like `INSERT`, `MERGE`.

Data Warehouse (DWH)

A **Data Warehouse** is a centralized repository for storing **historical, structured, cleaned data** used for business intelligence and analytics.

Key Features:

- Stores data from **multiple sources**
- Uses **denormalized schema**
- Enables reporting, dashboards, KPIs

Types of Data Sources

Data sources can be **structured, semi-structured, or unstructured**, and come from internal or external systems:

- Flat Files
- Databases
- **APIs**
- **Web Scraping**
- **Cloud Storage**
- **Logs**

Materials Checklist

- **Data Files**
- **SQL Server Express**
- **SQL Server Management Studio**
- **Draw.io**
- **Notion**
- **Python or SQL**
- **GitHub / Git**

Short Report of Key concepts

Data Architecture

Convention Name

Data Warehouse Project

<https://github.com/Abdul-wahab456/SQL-Data-Warehouse-Project>

Data Base & Schema

bronze Layer DLD

bronze SQL Load Script

Silver Explore and understand data

Silver DLD tables

Silver Load and Clean Table

Gold Business Logics