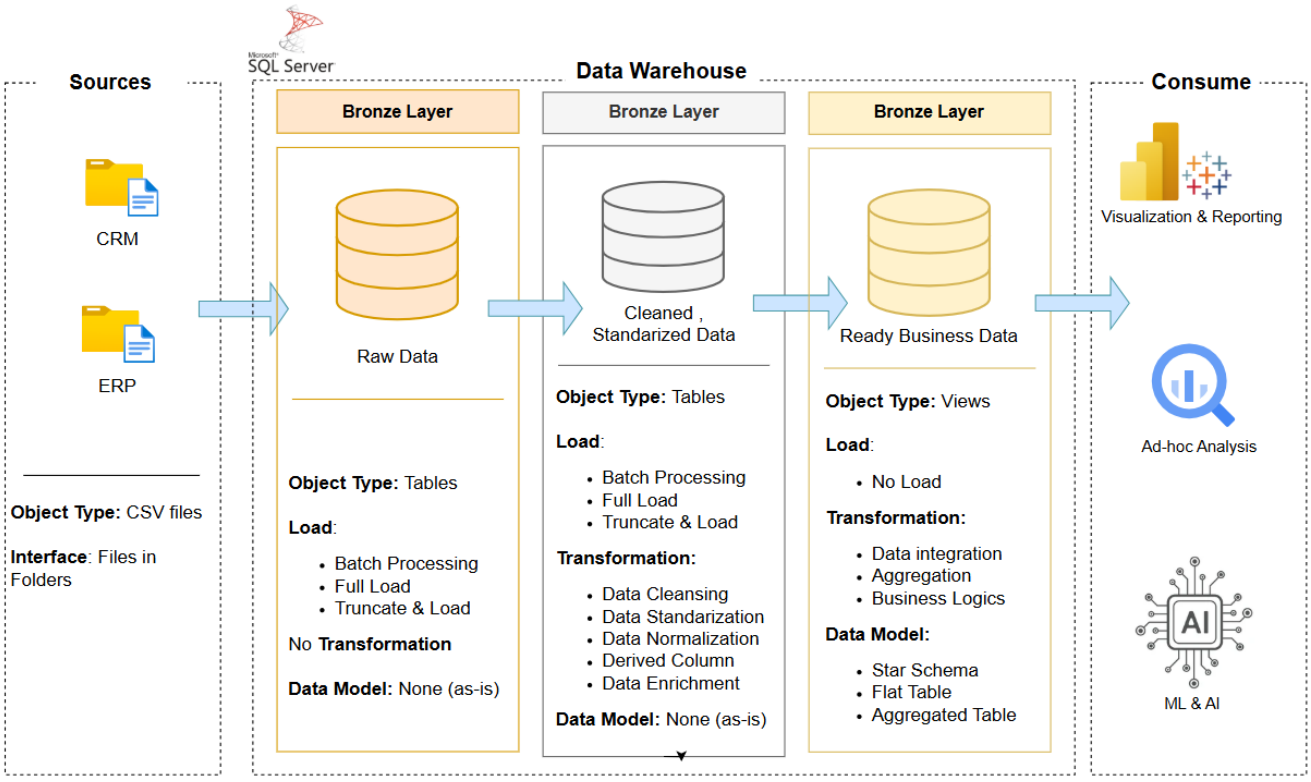
# Data Warehouse and Analytics Project

Welcome to the **Data Warehouse and Analytics Project** repository! 🚀
This project demonstrates a comprehensive data warehousing and analytics solution, from building a data warehouse to generating actionable insights. Designed as a portfolio project, it highlights industry best practices in data engineering and analytics.

## 🏗️ Data Architecture

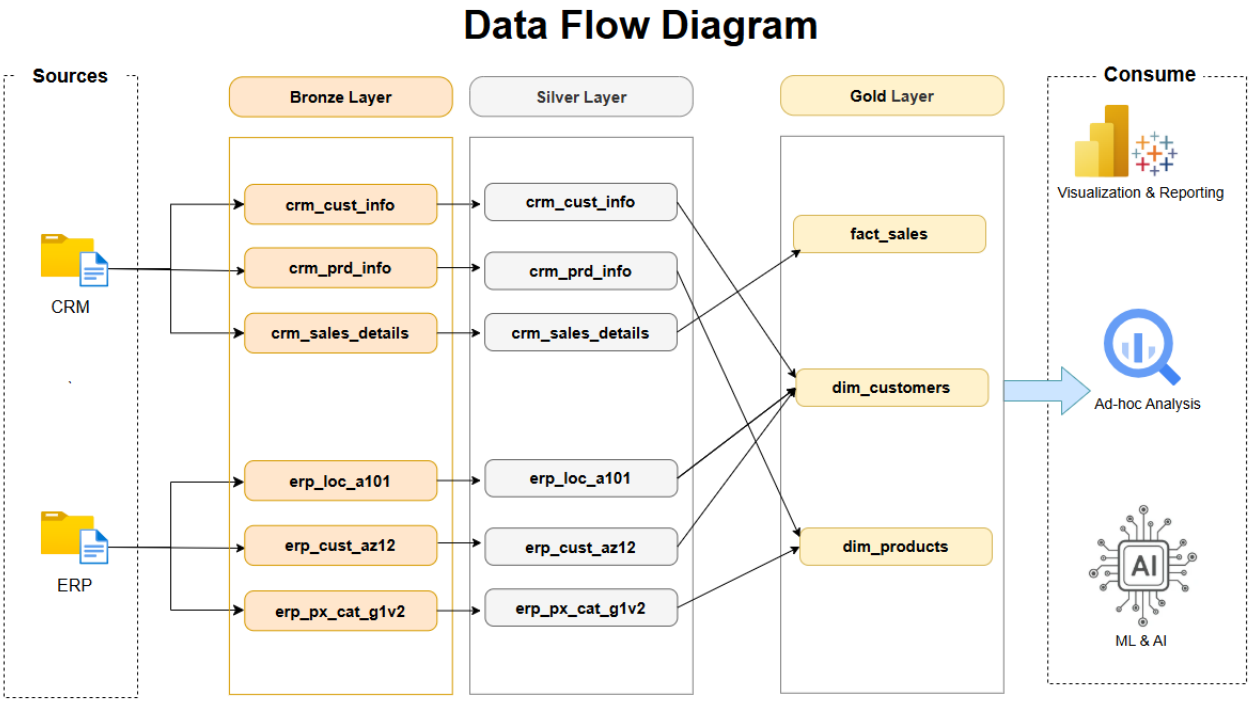
The data architecture for this project follows Medallion Architecture **Bronze**, **Silver**, and **Gold** layers:




1. **Bronze Layer**: Stores raw data as-is from the source systems. Data is ingested from CSV Files into SQL Server Database.
2. **Silver Layer**: This layer includes data cleansing, standardization, and normalization processes to prepare data for analysis.
3. **Gold Layer**: Houses business-ready data modeled into a star schema required for reporting and analytics.

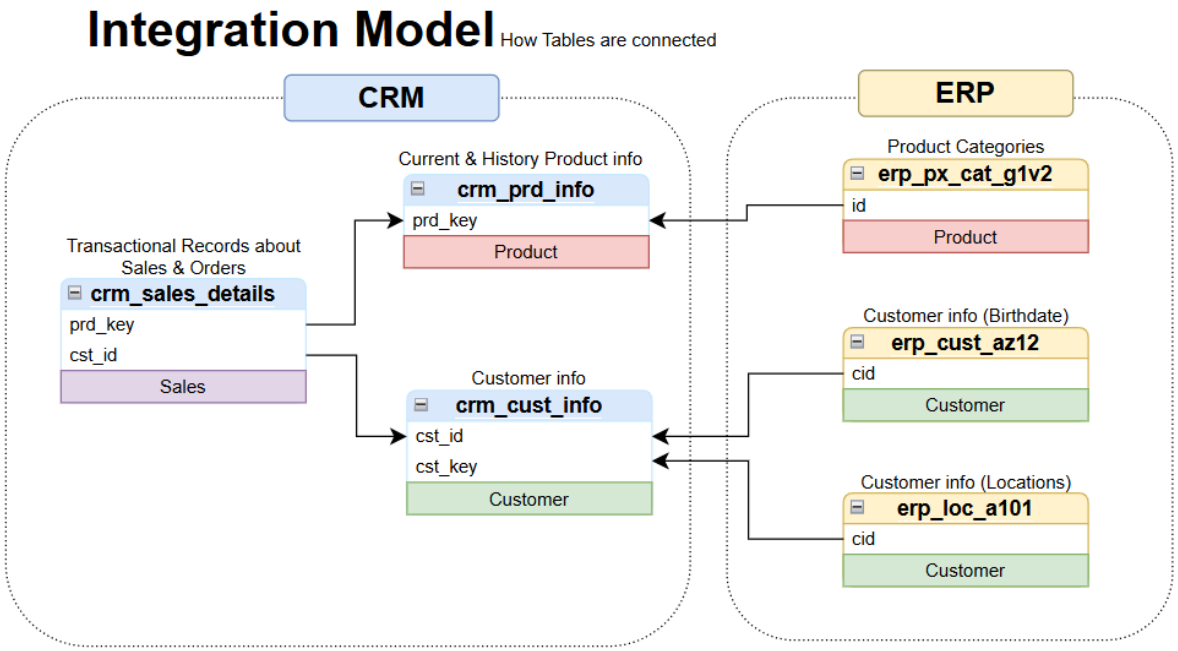
## 🏗️ Data Flow Diagram

Data flow from Bronze to Silver to Gold layers showing ETL transformations and integration.




## Integration Model

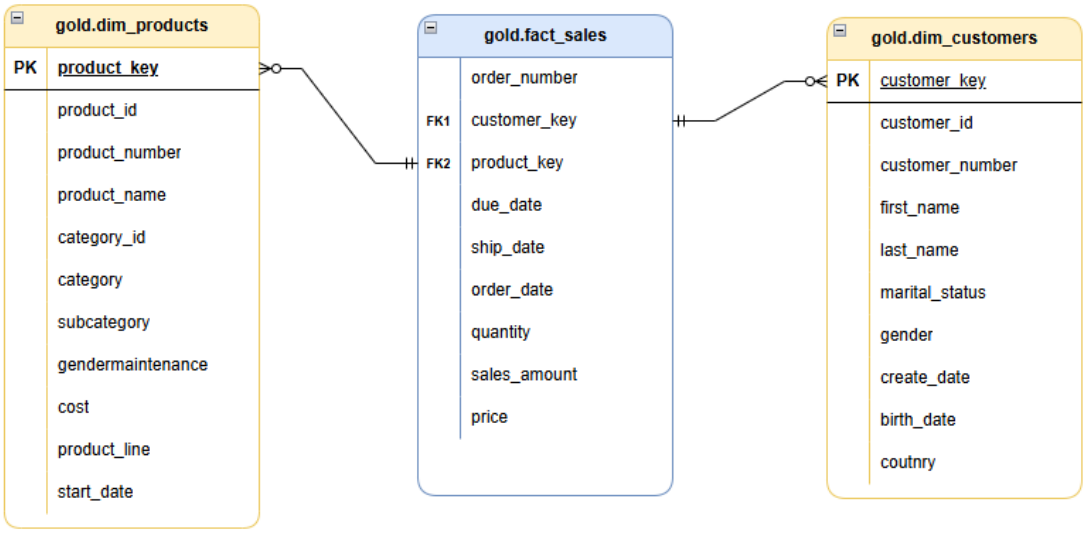
The following diagram shows the **Integration Model**, illustrating how CRM and ERP tables are connected:




## Sales Data Mart

The following diagram represents the **Sales Data Mart** in a star schema format:

**Sales Data Mart** (Star Schema)




## 📖 Project Overview

This project involves:

1. **Data Architecture**: Designing a Modern Data Warehouse Using Medallion Architecture **Bronze**, **Silver**, and **Gold** layers.
2. **ETL Pipelines**: Extracting, transforming, and loading data from source systems into the warehouse.
3. **Data Modeling**: Developing fact and dimension tables optimized for analytical queries.
4. **Analytics & Reporting**: Creating SQL-based reports and dashboards for actionable insights.

🎯 This repository is an excellent resource for professionals and students looking to showcase expertise in:

- SQL Development
- Data Architect
- Data Engineering
- ETL Pipeline Developer
- Data Modeling
- Data Analytics

---

## 🚀 Project Requirements

### Building the Data Warehouse (Data Engineering)

**Objective**

Develop a modern data warehouse using SQL Server to consolidate sales data, enabling analytical reporting and informed decision-making.

**Specifications**

- **Data Sources**: Import data from two source systems (ERP and CRM) provided as CSV files.
- **Data Quality**: Cleanse and resolve data quality issues prior to analysis.
- **Integration**: Combine both sources into a single, user-friendly data model designed for analytical queries.
- **Scope**: Focus on the latest dataset only; historization of data is not required.
- **Documentation**: Provide clear documentation of the data model to support both business stakeholders and analytics teams.

---

### BI: Analytics & Reporting (Data Analysis)

**Objective**

Develop SQL-based analytics to deliver detailed insights into:

- **Customer Behavior**
- **Product Performance**
- **Sales Trends**

These insights empower stakeholders with key business metrics, enabling strategic decision-making.

For more details, refer to docs/requirements.md.

## 🗁 Repository Structure

```
data-warehouse-project/
│
├── datasets/                           # Raw datasets used for the project (ERP
and CRM data)
│
├── docs/                               # Project documentation and architecture
details
│   ├── etl.drawio                      # Draw.io file shows all different
techniquies and methods of ETL
│   ├── data_architecture.drawio        # Draw.io file shows the project's
architecture
│   ├── data_catalog.md                 # Catalog of datasets, including field
descriptions and metadata
│   ├── data_flow.drawio                # Draw.io file for the data flow diagram
│   ├── data_models.drawio              # Draw.io file for data models (star
schema)
│   ├── naming-conventions.md           # Consistent naming guidelines for tables,
columns, and files
│
├── scripts/                            # SQL scripts for ETL and transformations
│   ├── bronze/                         # Scripts for extracting and loading raw
data
│   ├── silver/                         # Scripts for cleaning and transforming
data
```

```
│   ├── gold/                          # Scripts for creating analytical models
│
├── tests/                             # Test scripts and quality files
│
├── README.md                          # Project overview and instructions
├── LICENSE                            # License information for the repository
├── .gitignore                         # Files and directories to be ignored by
Git
└── requirements.txt                   # Dependencies and requirements for the
project
```

---## About Me

Hi, I'm **Mohamed Hanafy** – a passionate **Data Analyst & BI Developer** with experience in designing data warehouses, building ETL workflows, and creating interactive dashboards and reports. I have a strong background in **Power BI**, SQL, and data integration, and I enjoy turning complex datasets into actionable insights.

- 💼 Experienced in **Tableau, Alteryx, SingleStore, and SQL**
- 📊 Specialize in **business intelligence, data visualization, and dashboard optimization**
- 🌐 Strong understanding of **Medallion Architecture (Bronze, Silver, Gold layers)**
- 🎯 Driven to deliver **clean, reliable, and analytics-ready data**

I love learning new tools and techniques to make data analysis more effective and impactful.

## Stay Connected

Let's stay in touch! Feel free to connect with me on the following platforms:

YOUTUBE LINKEDIN WEBSITE