**Batch: T2**

**Assignment No: 7**

**Title of Assignment: Data Warehouse for X-mart**

**Student Name: Shreeyash Shripad Dongarkar**

**Student PRN: 22510025**

**1. Introduction**

Objective:

The objective of this project is to design and implement a **data warehouse** for **X-Mart**, a retail company operating multiple stores in different locations. The management faces challenges in analyzing sales data due to the absence of an integrated system. The goal of this project is to build a data warehouse that consolidates sales, customer, and product data to enable **faster decision-making** and **better business insights**.

Scope:

This data warehouse will facilitate:

* **Data integration** from multiple stores into a single database.
* **Historical data storage** for long-term analysis.
* **Fast retrieval of analytical reports** for sales performance, product demand, and customer behaviour.
* **Online Analytical Processing (OLAP)** capabilities using data cubes to analyse sales trends over different time dimensions.

**2. Business Requirement**

**Need for Data Warehousing**

Currently, X-Mart's sales data is stored in a **transactional database**, making it difficult to perform large-scale trend analysis and business forecasting. The management requires a system that allows them to:

* Analyse **daily, weekly, monthly, and yearly sales trends**.
* Compare **store-wise and product-wise performance**.
* Study **customer purchasing behaviour**.
* Identify **peak sales periods and high-demand products**.

**Intended Users and Their Requirements**

* **Business Executives & Managers** require summarized reports on revenue trends, store performance, and overall profitability.
* **Marketing Teams** need insights into customer preferences and product demand to plan promotions and stock management.
* **Inventory Managers** require real-time and historical stock-level tracking to optimize supply chain operations.

**3. Functional Specification**

**Input Data (Extracted from Operational Database)**

The operational database consists of transactional sales data, customer information, store details, and product information. The following tables provide the necessary inputs:

1. Sales Data: Invoice details, sales date, sales time, store ID, salesperson ID, total amount.
2. Customer Data: Customer name, contact details, location, purchase history.
3. Product Data: Product ID, product name, category, price, and stock availability.
4. Store Data: Store ID, store name, location, and sales performance metrics.

**Output Data (Reports and Analytics)**

1. Sales Performance Reports: Comparison of sales across different time periods.
2. Product Demand Reports: Analysis of which products perform best in each location.
3. Customer Behaviour Analysis: Identification of frequent buyers and high-value customers.
4. Store Performance Reports: Ranking of stores based on revenue generation.

**4. Data Warehousing Design**

**Step 1: Requirement Gathering**

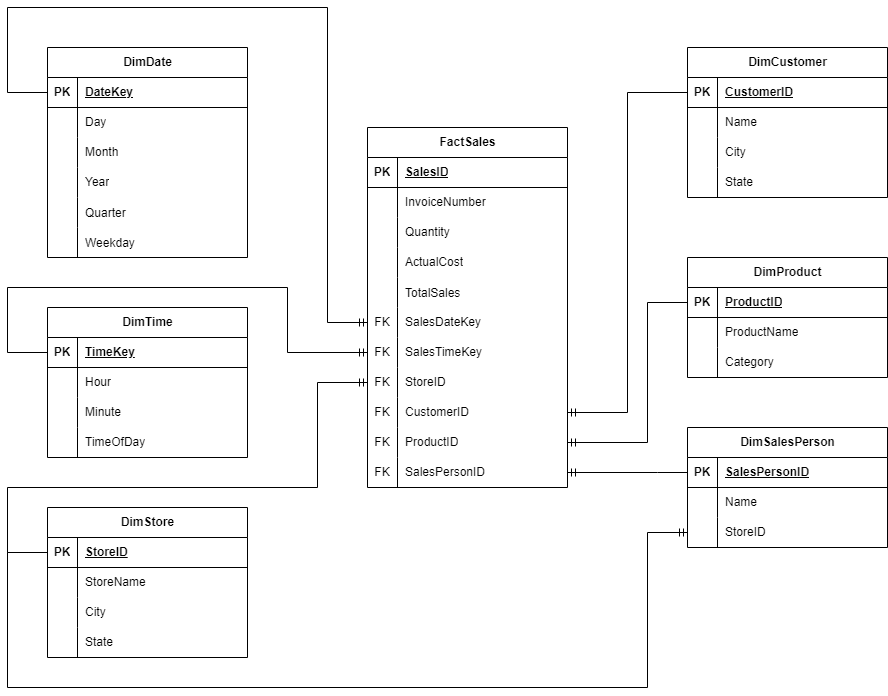
The first step involved identifying the key performance indicators (KPIs) that the business wants to track. These include **monthly revenue, product demand trends, customer purchasing behaviour, and store-wise performance metrics**.

**Step 2: Designing the Schema**

For efficient query performance and analytical capabilities, a **Star Schema** was chosen. This schema consists of:

* **A central fact table (FactSales)** that contains transactional data.
* **Multiple dimension tables (DimDate, DimStore, DimProduct, DimCustomer, DimSalesPerson)** to store descriptive information.

**Step 3: Schema Design Diagram**

****

**Step 4: Implementing the Schema in SQL**

The schema was implemented using SQL by creating fact and dimension tables, ensuring that primary and foreign keys were properly defined for maintaining referential integrity.

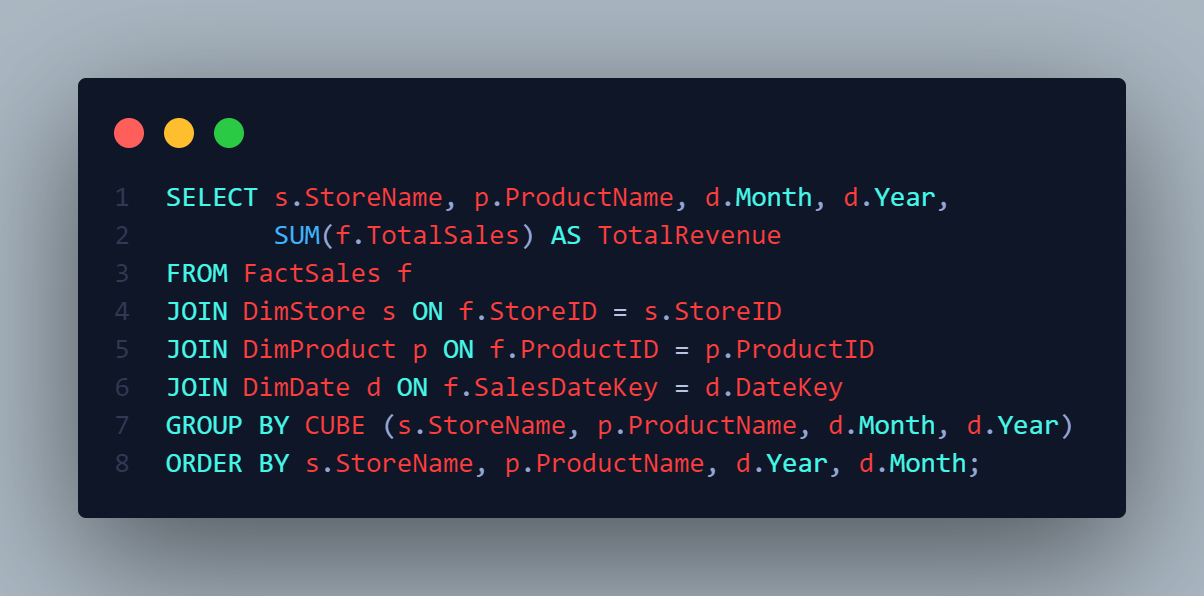
**5. Data Cube Implementation**

**Automating Data Warehousing & ETL Process**

The **ETL (Extract, Transform, Load) process** was implemented to:

1. **Extract data** from the operational database.
2. **Transform data** to fit into dimension and fact tables.
3. **Load the transformed data** into the data warehouse.

**Creating a Data Cube for Multi-Dimensional Analysis**



**6. Conclusion**

The implementation of the **X-Mart data warehouse** successfully enables the management to **analyse sales performance, customer behaviour, and store efficiency**. The **data cube** allows for **multi-dimensional querying**, making it easier to generate reports dynamically.

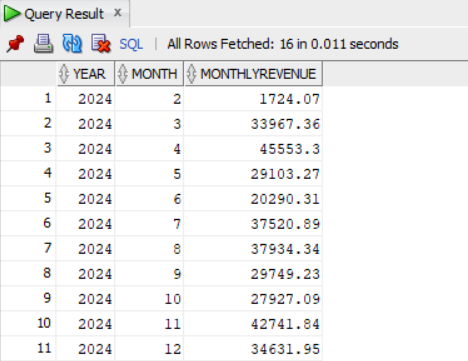
**Key Achievements:**

* **Consolidated data storage** for all sales transactions.
* **Fast OLAP query execution** using **CUBE () and ROLLUP ()** operations.
* **Business insights through automated reports** on revenue, product trends, and customer spending habits.

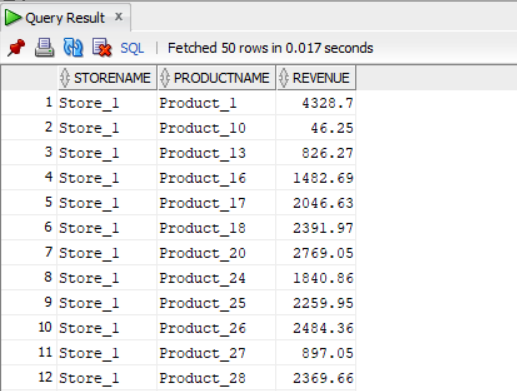
**Future Enhancements:**

1. **Automate ETL Jobs:** Schedule ETL processes for **daily updates** to the data warehouse.
2. **Predictive Analysis:** Implement **machine learning models** to forecast sales trends.
3. **Visualization & Dashboards:** Use **BI tools (Power BI, Tableau)** for better data representation.

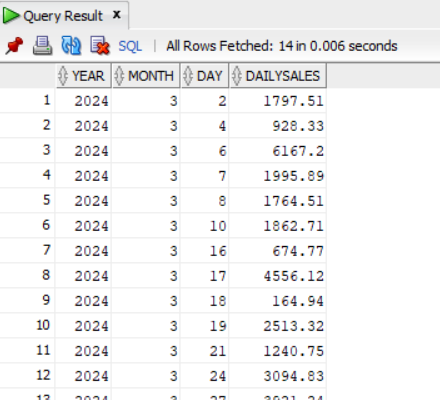
Summarizing Sales by Year



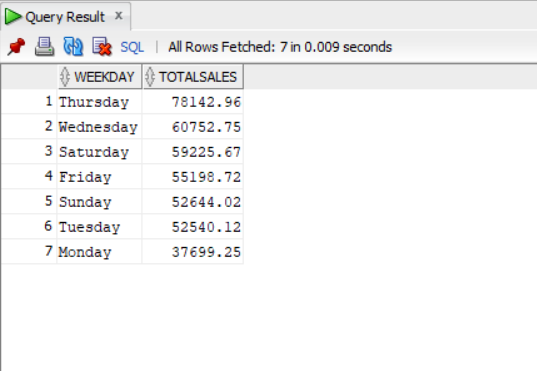
Store wise product information



Daily sales for given month



Comparison Weekday vs Weekends



Sales duration by time of day

