Data Mining and Business Intelligence

Star Schema

Name: **Abhishek Vishwakarma**
Class: **D15C**
Roll No: **73**

Aim: To Design Star Schema's

Theory:

Introduction

The **Star Schema** is one of the most commonly used data modeling techniques in data warehousing and business intelligence. It is designed to organize data into a structure that is simple, fast for querying, and optimized for analytical processing (OLAP). Its name comes from the diagram's appearance — a central fact table surrounded by multiple dimension tables, resembling a star.

Structure of Star Schema

A star schema consists of two main components:

1. **Fact Table**
    - Contains the **quantitative data** (measures) that an organization wants to analyze.
    - Examples: sales amount, quantity sold, revenue, profit.
    - Typically has **foreign keys** linking to each dimension table.
    - Usually very large in size and contains aggregated or transactional data.
2. **Dimension Tables**
    - Contain **descriptive attributes** (context) related to the facts.
    - Examples: product details, customer information, time periods, geographic locations.
    - Dimension tables are denormalized to simplify queries and improve performance.
    - Each dimension table has a **primary key** that uniquely identifies a record.

Key Characteristics

- **Central Fact Table**: Stores measurable business metrics.
- **Surrounding Dimension Tables**: Provide descriptive context for analysis.
- **Denormalized Design**: Reduces complexity and improves query speed.
- **Simple Joins**: Queries typically join the fact table to dimension tables via primary–foreign key relationships.

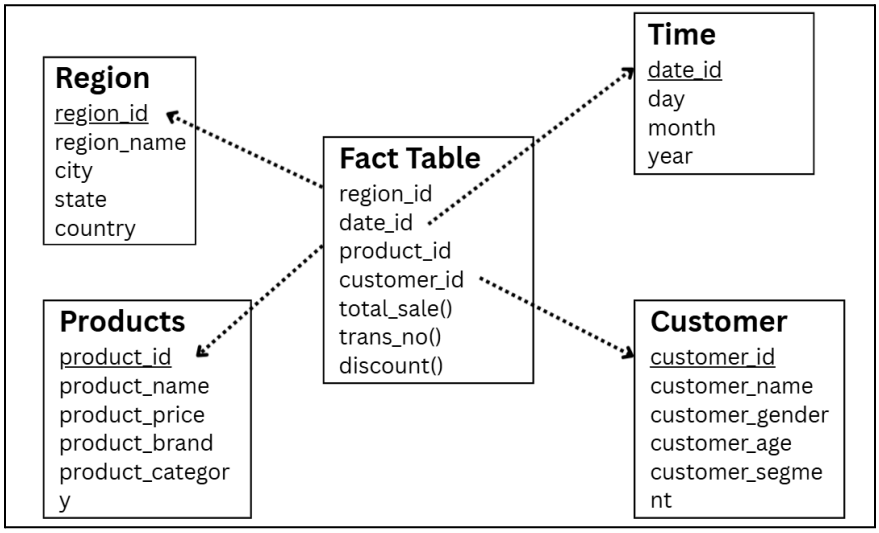
Advantages

1. **Simplicity** – Easy to understand and navigate for business users.
2. **High Query Performance** – Fewer joins and denormalized dimensions speed up retrieval.
3. **Intuitive Design** – Closely matches the way business reports and dashboards are structured.
4. **Supports OLAP** – Well-suited for slicing, dicing, drilling down, and aggregations.

Questions:

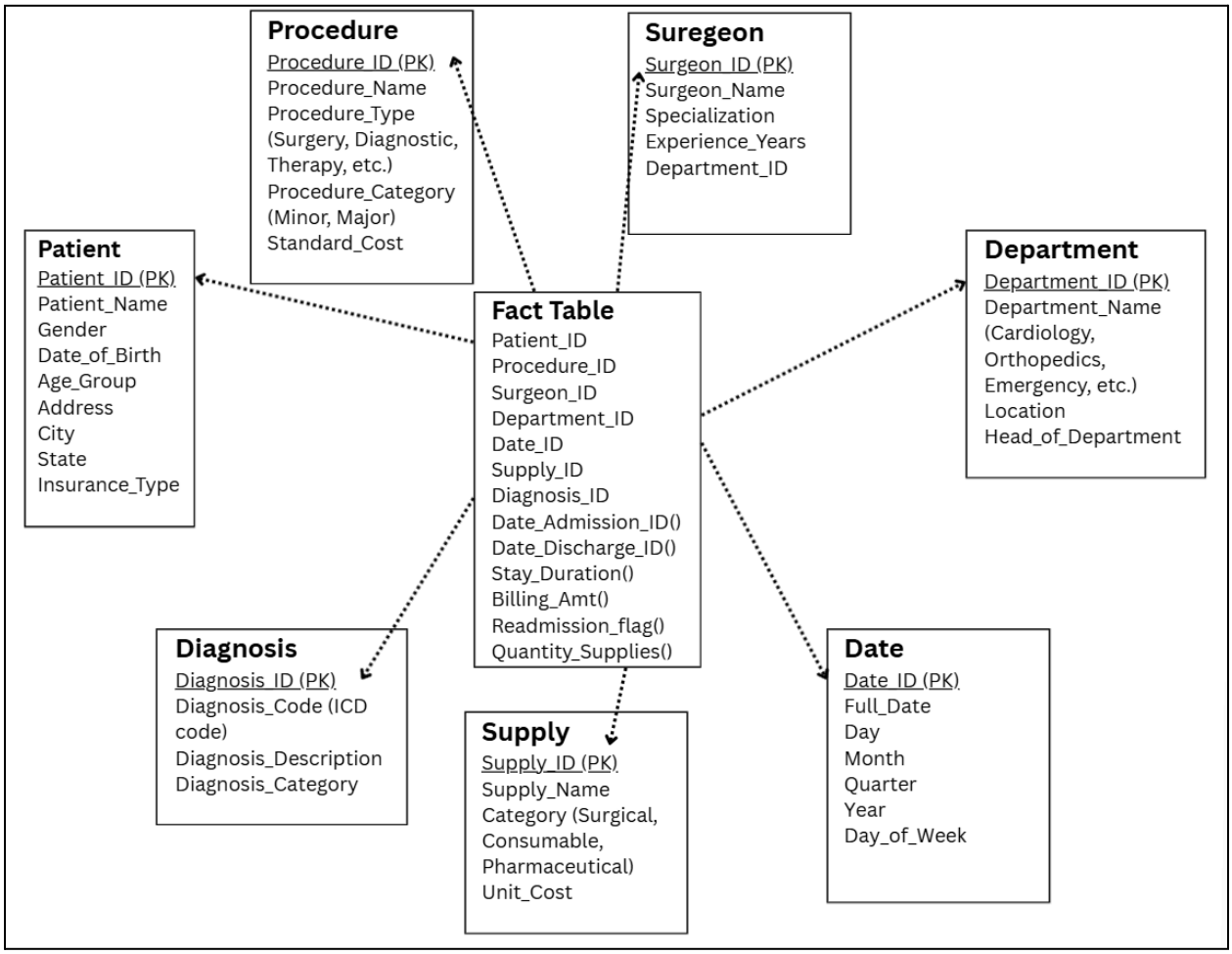
1) A retail company wants to analyze its sales performance across different regions, time periods, products, and customer segments. The company wants to track total sales, number of transactions, and discount offered.

Sol'n



2)A hospital management wants to create a data warehouse to analyze patient admissions, procedures, and billing information. The goal is to improve operational efficiency and patient care by answering questions such as:
i) What is the average length of stay for patients with a specific diagnosis?
ii) How many surgical procedures were performed by each surgeon last month?
iii)What is the total revenue generated by a particular department (e.g., Cardiology, Orthopedics) per quarter?
iv)Which medical supplies are most frequently used in the emergency department?
What is the readmission rate for patients who had a certain procedure?

Sol'n



Conclusion: Through this experiment, we learned how to transform normalized operational data into Star Schema models suitable for analytical processing. By applying Star Schema design to varied domains—education, retail, and healthcare—we saw how it simplifies queries, improves performance, and supports decision-making in different industries.