Name: **Gargi Devendra Shintre**
Div: **D15C** Roll No. **69**

Practical No. 1

Aim: Design Star Schema for given problem statements.

Theory:

The Star Schema is the simplest and most widely used data modeling technique in data warehouse design. It is called a "star" because the structure visually resembles a star: a large central fact table at the center, surrounded by multiple dimension tables, with each dimension table connected directly to the fact table.

Key Components:

1. Fact Table:

- Contains the measurable, quantitative data (also called "facts") used for analysis.
- Stores numerical values such as sales amount, revenue, quantity, duration, etc.
- Includes foreign keys linking to the dimension tables.
- Facts are usually additive (can be summed up across dimensions).

2. Dimensions Table:

- Contain descriptive attributes (also called "dimensions") that provide context to the facts.
- Examples: Date, Product, Customer, Region, Department.
- Each dimension has a primary key that connects to the fact table's foreign key.
- Dimensions are denormalized to simplify querying and improve performance.

Characteristics of a Star Schema:

- **Single Fact Table**: All measures are stored in one central table.
- **Direct Relationships**: Each dimension connects directly to the fact table (no intermediate joins).
- **Denormalized Dimensions**: Dimension tables often store redundant data to reduce the number of joins and speed up queries.
- **Optimized for OLAP**: Supports fast aggregation and reporting in Online Analytical Processing systems.

Advantages:

- **Simplicity**: Easy to understand for business users and analysts.
- **High Query Performance**: Fewer joins compared to normalized schemas.

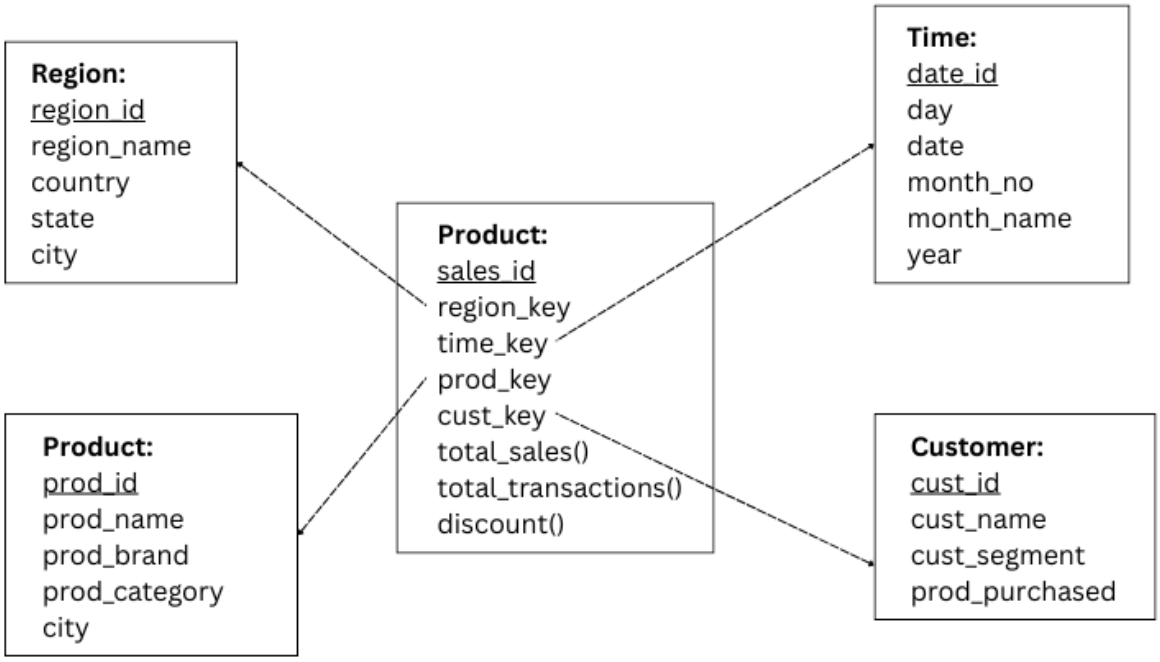
- **Clear Separation**: Facts and dimensions are clearly separated, improving manageability.
- **Supports Drill-Down Analysis**: Enables users to navigate from summary-level data to detailed data.

Disadvantages:

- **Redundancy in Dimensions**: Denormalized dimension tables may lead to data duplication.
- **Limited Flexibility for Complex Relationships**: Works best for simpler analysis; complex many-to-many relationships may require alternative designs.
- **Update Overhead**: Maintaining denormalized dimensions can require more effort when updating data.

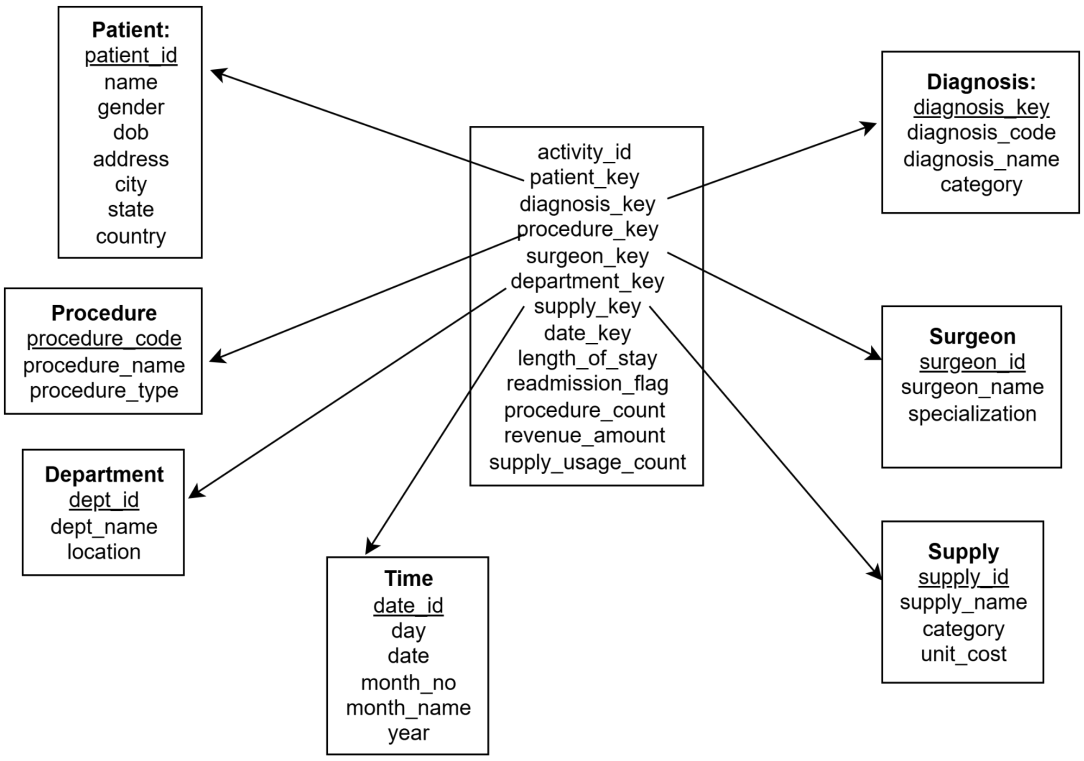
Questions:

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 discounts offered.



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:

- What is the average length of stay for patients with a specific diagnosis?
- How many surgical procedures were performed by each surgeon last month?
- What is the total revenue generated by a particular department (e.g., Cardiology, Orthopedics) per quarter?
- Which medical supplies are most frequently used in the emergency department?
- What is the readmission rate for patients who had a certain procedure?



Conclusion:

The Star Schema is a highly efficient and user-friendly data modeling approach for organizing data in a data warehouse. Its structure, with a single fact table linked to multiple dimension tables, simplifies query execution and enhances analytical performance. By enabling fast aggregations, drill-down analysis, and clear separation of facts and descriptive attributes, it supports effective decision-making. Although it may introduce some redundancy in dimension tables, the performance benefits and ease of understanding make the Star Schema one of the most preferred designs for business intelligence and analytical applications.