Name:Khushi Singh
Class: D15C Roll no. 71

Practical No. 1

AIM: Design a Star Schema for the given system.

THEORY:

Introduction to Dimensional Modeling

Dimensional Modeling is a design technique used in data warehouse and business intelligence systems to structure data in a way that is easy to understand and fast to query for analytical purposes.
Unlike OLTP databases, which are optimized for transaction processing (inserts, updates, deletes), dimensional models are optimized for OLAP (Online Analytical Processing) — queries that aggregate, slice, dice, and drill down through data.

Key Features

- Denormalized structure: Data is stored in fewer tables to improve query performance.
- Two main components:
    1. Fact Tables – Store numeric measurements (facts) of the business.
    2. Dimension Tables – Store descriptive attributes that describe the facts.
- Schema designs such as Star Schema and Snowflake Schema.

Fact Table

- Central table in a dimensional model.
- Stores measurable, quantitative data (e.g., sales amount, quantity sold, length of stay).
- Contains foreign keys to link to dimension tables.
- Can store aggregated or transaction-level data.

Dimension Table

- Contains descriptive attributes that describe the facts.
- Usually textual and denormalized.
- Used for filtering, grouping, and labeling in queries.

Types of Dimensional Models

Dimensional modeling can be implemented using different schema types.

Star Schema

- Structure: One central fact table surrounded by dimension tables.
- Simplicity: Easy for users to understand.

- Performance: Fewer joins make queries faster.
- Usage: Most common schema in data warehouses.

Snowflake Schema

- Structure: Similar to Star Schema, but dimension tables are normalized into multiple related tables.
- Advantage: Saves storage space by reducing data redundancy.
- Disadvantage: More joins are needed, which can slow down queries.

Example:
Dim_Product split into Product → Product_Category → Product_Department.

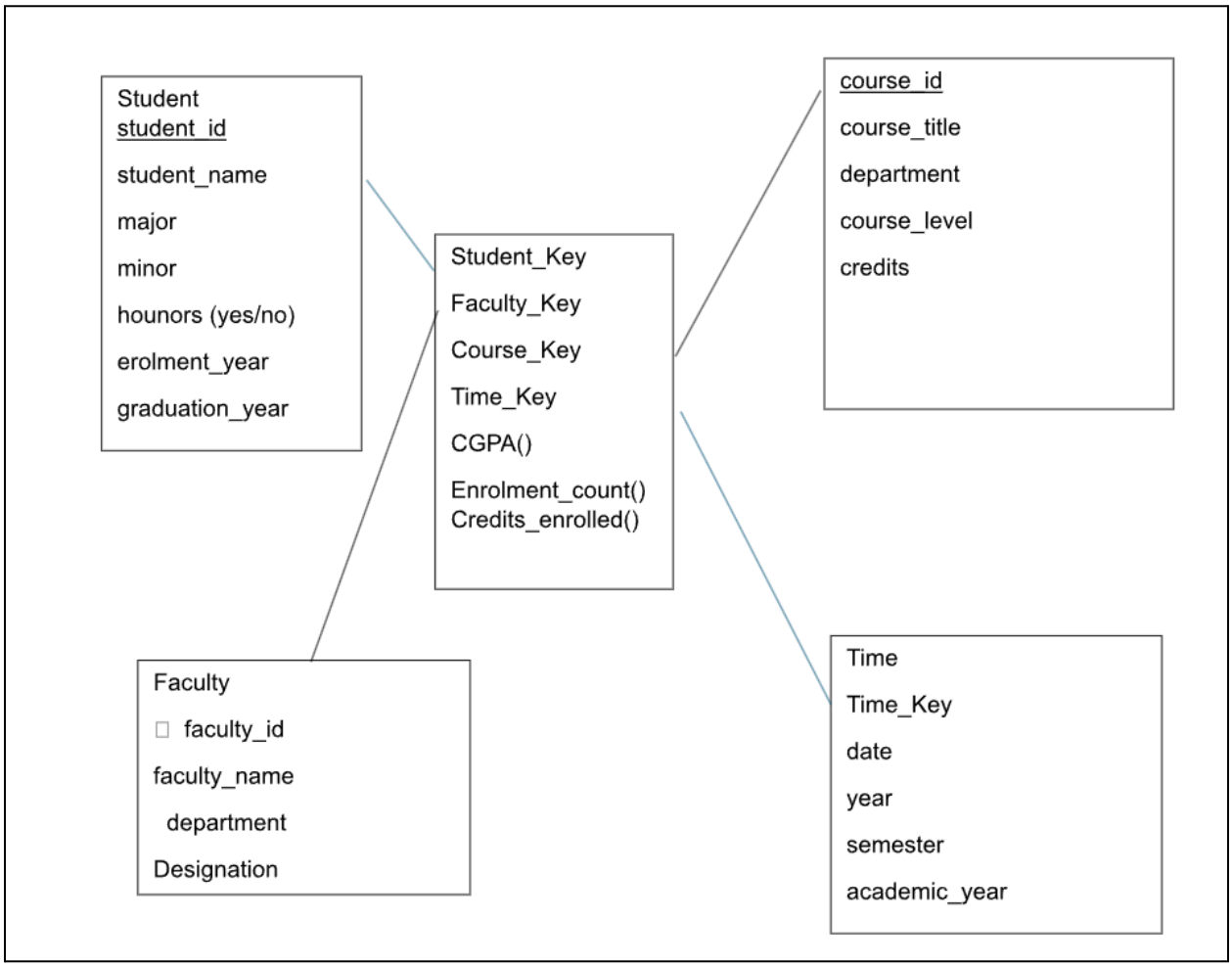
Advantages of Dimensional Modeling

- User-friendly for non-technical analysts.
- High performance for read-heavy queries.
- Supports historical data analysis.
- Scalable for growing business data.

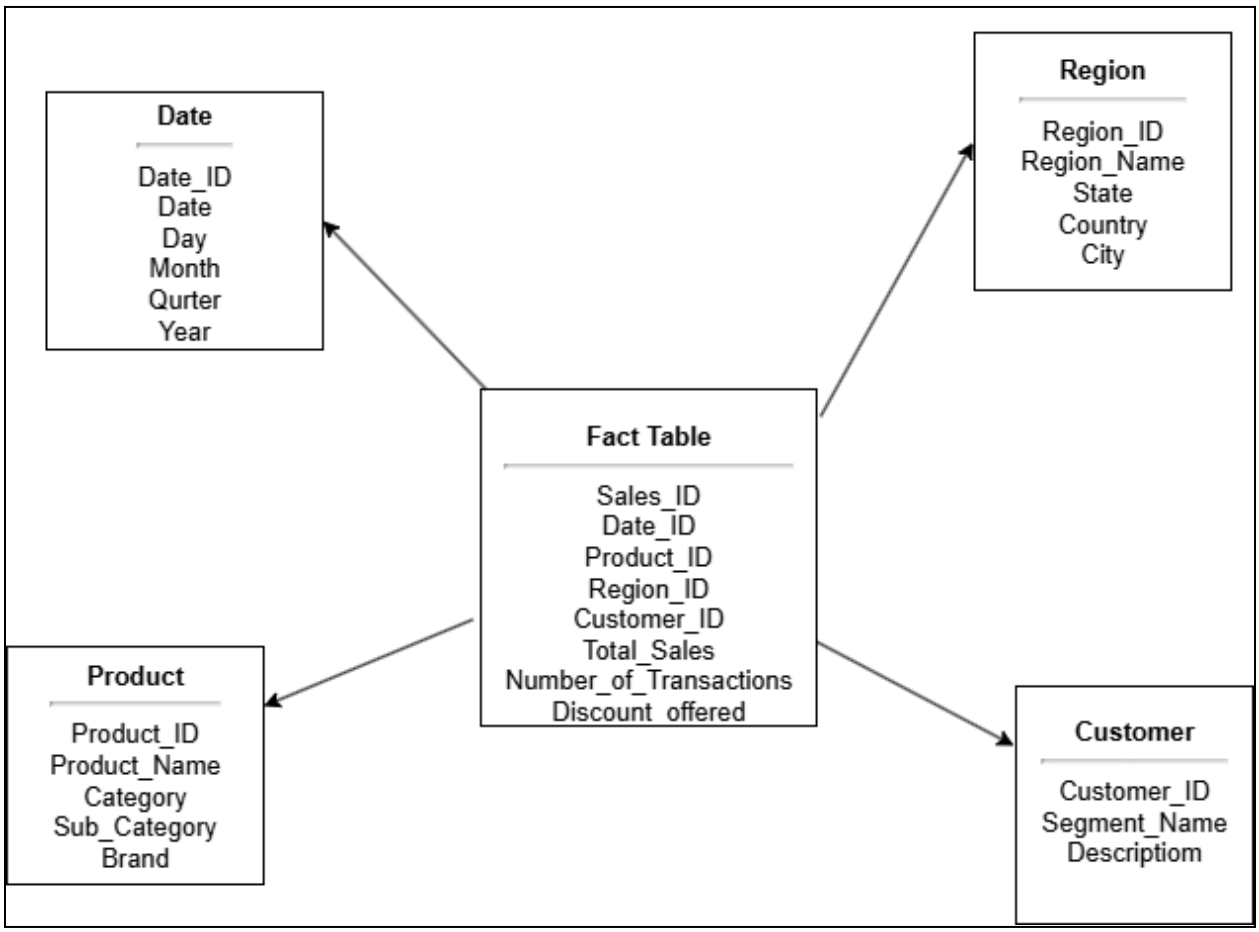
Disadvantages

- Not suitable for transactional operations.
- Denormalization can lead to data redundancy.
- Updates to dimension data can be tricky (slowly changing dimensions).

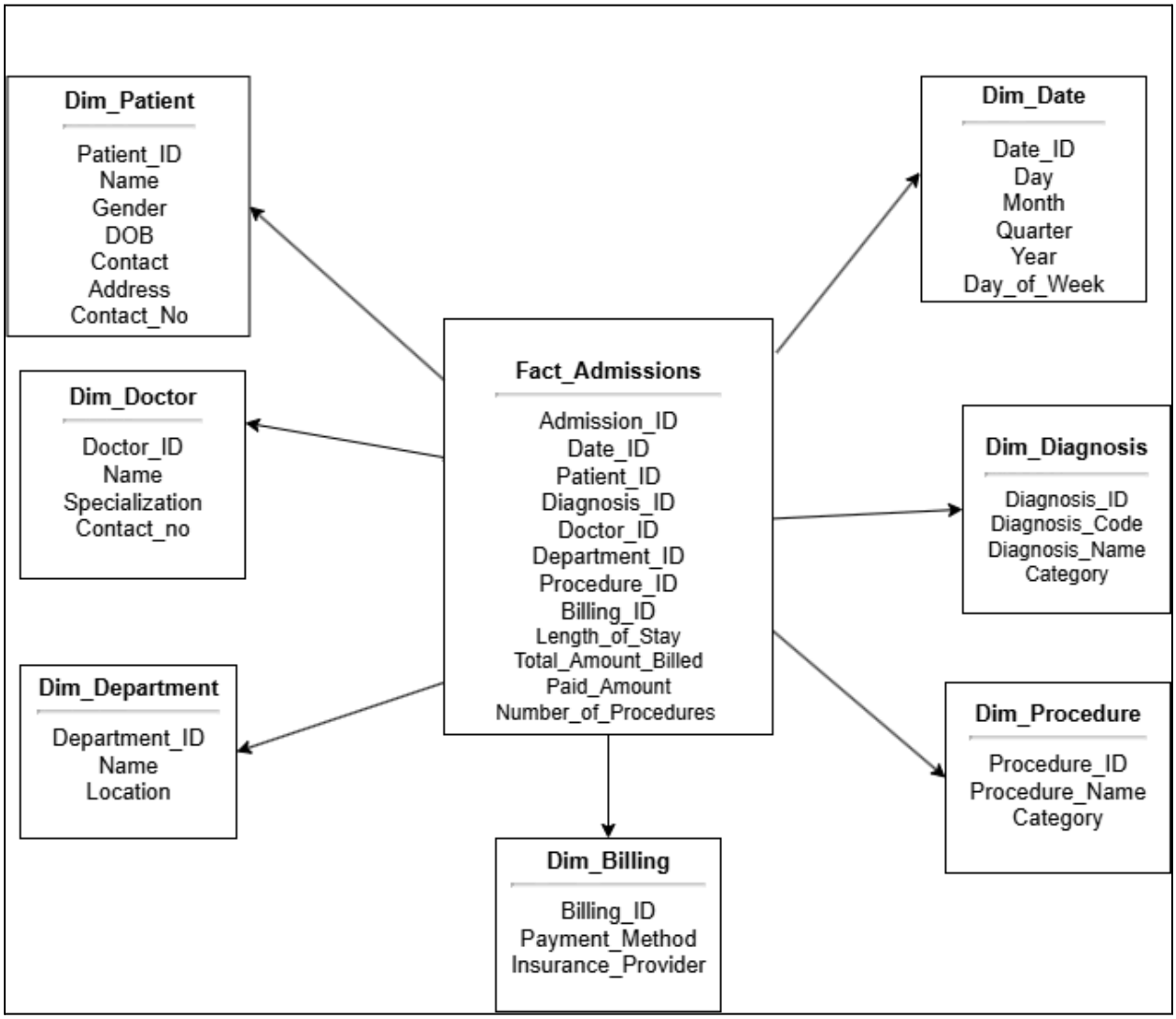
Q1.A university wants to design a data warehouse to analyze student performance, course enrollments, and faculty workload. The university's operational database is highly normalized, making it difficult to perform analytical queries.



Q2.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.



Q3.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:



Conclusion:

In this experiment, I explored Dimensional Modeling as a powerful technique for designing data warehouses, focusing on the Star Schema. Unlike traditional OLTP systems that are highly normalized, dimensional models simplify data into fact and dimension tables, making it easier to perform analytical queries such as aggregation, slicing, and drilling down. Through the three case studies—university performance analysis, retail sales tracking, and hospital management optimization—I designed star schemas with central fact tables (storing measurable data like grades, sales, and billing amounts) surrounded by descriptive dimension tables (such as student, course, faculty, time, product, region, patient, and procedure). This structure demonstrated how the star schema enhances query performance, usability, and decision-making by providing a denormalized yet intuitive view of the business processes. Overall, the exercise showed that dimensional modeling using star schemas is highly effective for real-world analytical systems, enabling organizations to gain insights, improve efficiency, and make data-driven decisions.