Name: Shrey Pendurkar
Class: D15C
Batch: C
Roll No: 64

DMBI - Experiment 1

Aim: Design a Star Schema for the given system

Dimensional Modeling and Its Types

Dimensional Modeling is a widely used design technique in data warehousing and business intelligence to structure data for efficient querying and reporting. It organizes data into a format that is intuitive for business users and optimized for read access. The main goal of dimensional modeling is to make complex data easier to understand and analyze by categorizing it into measurable events and their descriptive contexts.

Core Concepts

At the heart of dimensional modeling are two key components:

- Fact Tables: These tables store quantitative data or metrics, called facts, which represent business events such as sales transactions, orders, or inventory levels. Fact tables typically contain numeric values like revenue, quantity sold, or cost, along with foreign keys that reference related dimensions.

- Dimension Tables: These provide descriptive information or attributes about the facts. Dimensions give context to the facts, such as time (date, month, year), product details (name, category, brand), customer demographics, or geographical location. Dimension tables are designed to be easily readable and understandable by end users.

Importance of Dimensional Modeling

Dimensional modeling simplifies complex data structures into easy-to-navigate schemas that speed up query performance and improve the usability of reports and dashboards. It is particularly beneficial for online analytical processing (OLAP) systems, where fast and efficient aggregation and slicing/dicing of data are required.

Types of Dimensional Modeling Schemas

1. **Star Schema:**

   The star schema is the simplest and most common dimensional model. It features a single central fact table connected directly to multiple denormalized dimension tables. The dimension tables contain all the attributes needed for filtering and grouping data.

   - Advantages: Simplified queries, fast query performance, and easy to understand.
   - Disadvantages: Dimension tables can have redundant data, which may increase storage requirements.

2. **Snowflake Schema:**

   The snowflake schema is a more complex form of dimensional modeling where dimension tables are normalized into multiple related tables. For example, a product dimension might be broken down into product, product category, and product brand tables.

   - Advantages: Reduces data redundancy and saves storage space.
   - Disadvantages: More complex joins and slower query performance compared to the star schema.

Conclusion:

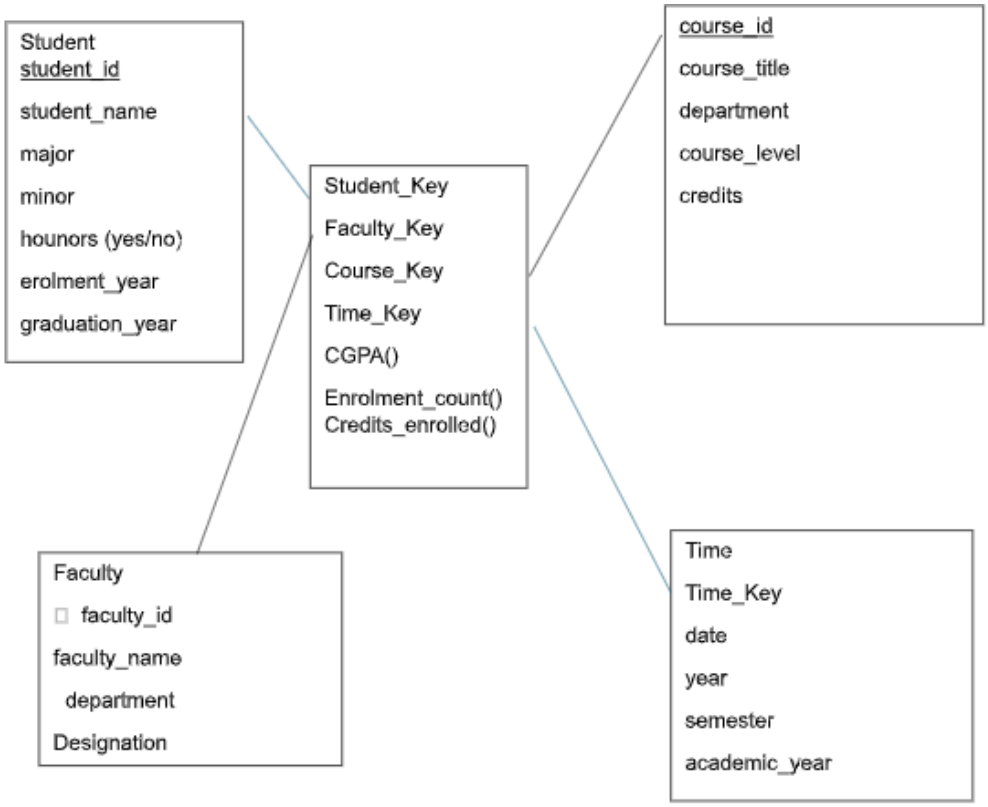
Through this experiment, we demonstrated that star schemas are an effective approach for structuring analytical data warehouses across different industries. They provide:

- Ease of use for business reporting,
- Improved query performance, and
- An intuitive “business view” of data that translates operational complexity into actionable insights.

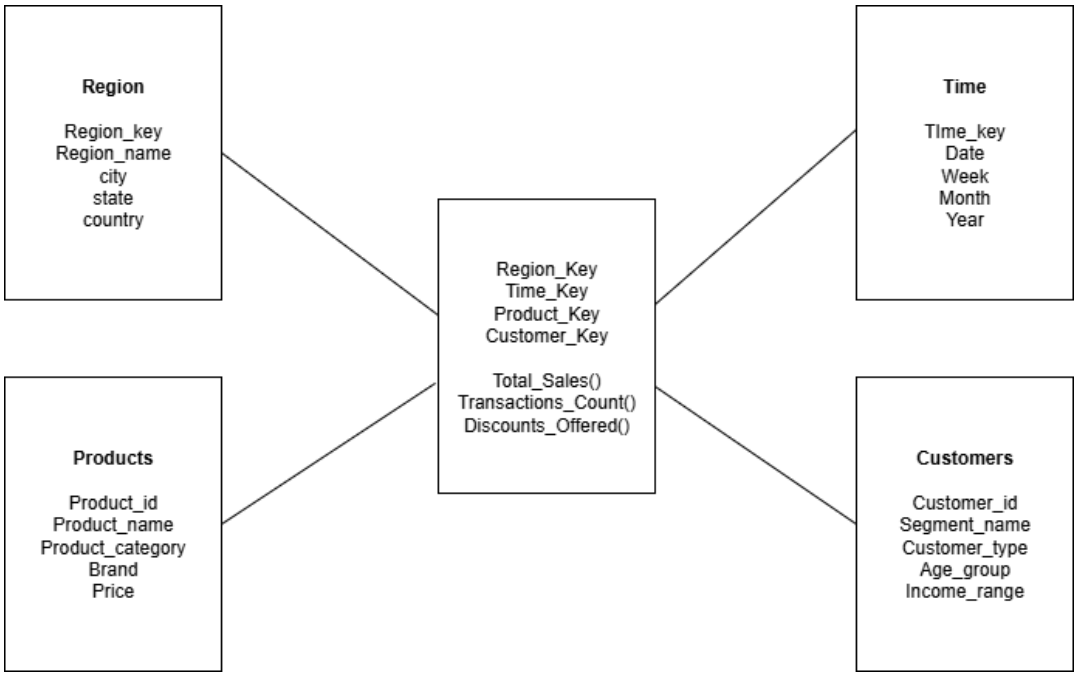
However, the main challenges lie in design decisions (grain, fact selection, SCD handling) and trade-offs between simplicity and redundancy. In real-world data warehouse projects, addressing these challenges requires close collaboration between technical teams and business users to ensure the schema supports both performance needs and business goals.

Questions:

1. 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. The management needs to answer questions like:



2. 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.



3. 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: