

Aim: Design a Star Schema for the given systems.

Theory:

Dimensional Modeling (DM) is a **design technique for data warehouses** that organizes data into **facts** (measurable events) and **dimensions** (descriptive context) to make analytical queries fast and easy. It's **optimized for querying and reporting**, not for transaction processing.

- It separates data into **facts** (measurable, numeric events) and **dimensions** (descriptive attributes providing context).
- Optimized for **OLAP (Online Analytical Processing)**, reporting, and decision-making, not for transaction processing.
- Designed to be **simple, intuitive, and query-friendly** for business users.

Types of Dimensional Models

1. Star Schema

- One **fact table** in the center linked to multiple **dimension tables**.
- Dimensions are **denormalized**.
- **Simple** and easy to query.
- Example: Sales fact table with dimensions for Product, Customer, Time, Region.
- **Structure:** A central **fact table** stores quantitative data (e.g., sales, revenue, transactions) and connects to multiple **dimension tables** (e.g., product, customer, time, region).
- **Denormalization:** Dimension tables are denormalized, meaning attributes are stored in a single table for quick access (e.g., Product table includes product name, brand, category).
- **Performance:** Optimized for **fast querying** and OLAP (Online Analytical Processing) since fewer joins are required.

- **Use Case:** Best suited for **business intelligence and reporting systems** where simplicity and query speed are prioritized over storage efficiency.

2. Snowflake Schema

- Like a star schema, but dimensions are **normalized** into multiple related tables.
- Reduces redundancy but adds complexity (more joins).
- Example: Product dimension split into Product → Category → Department.

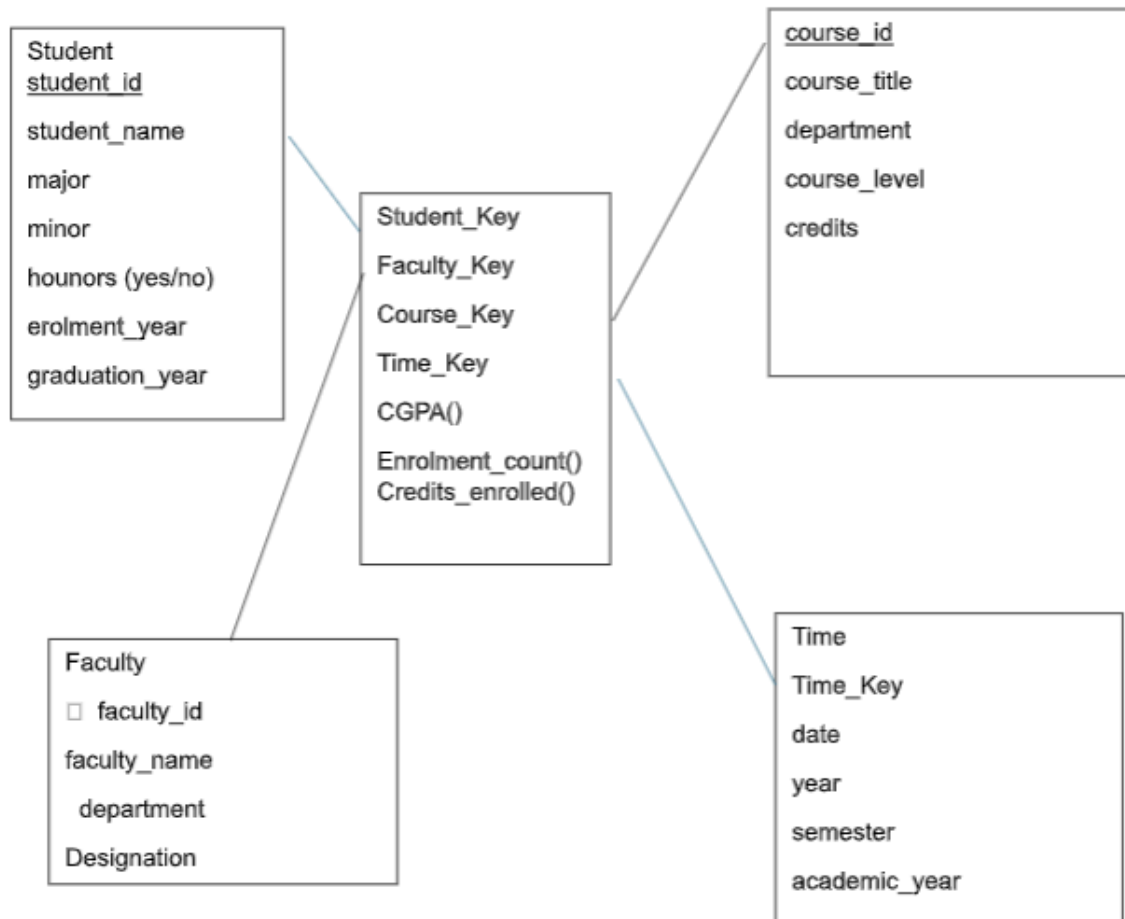
3. Galaxy Schema (or Fact Constellation)

- Multiple fact tables share dimension tables.
- Useful for complex systems with related subject areas.
- Example: Sales Fact and Returns Fact sharing Product, Customer, and Time dimensions.

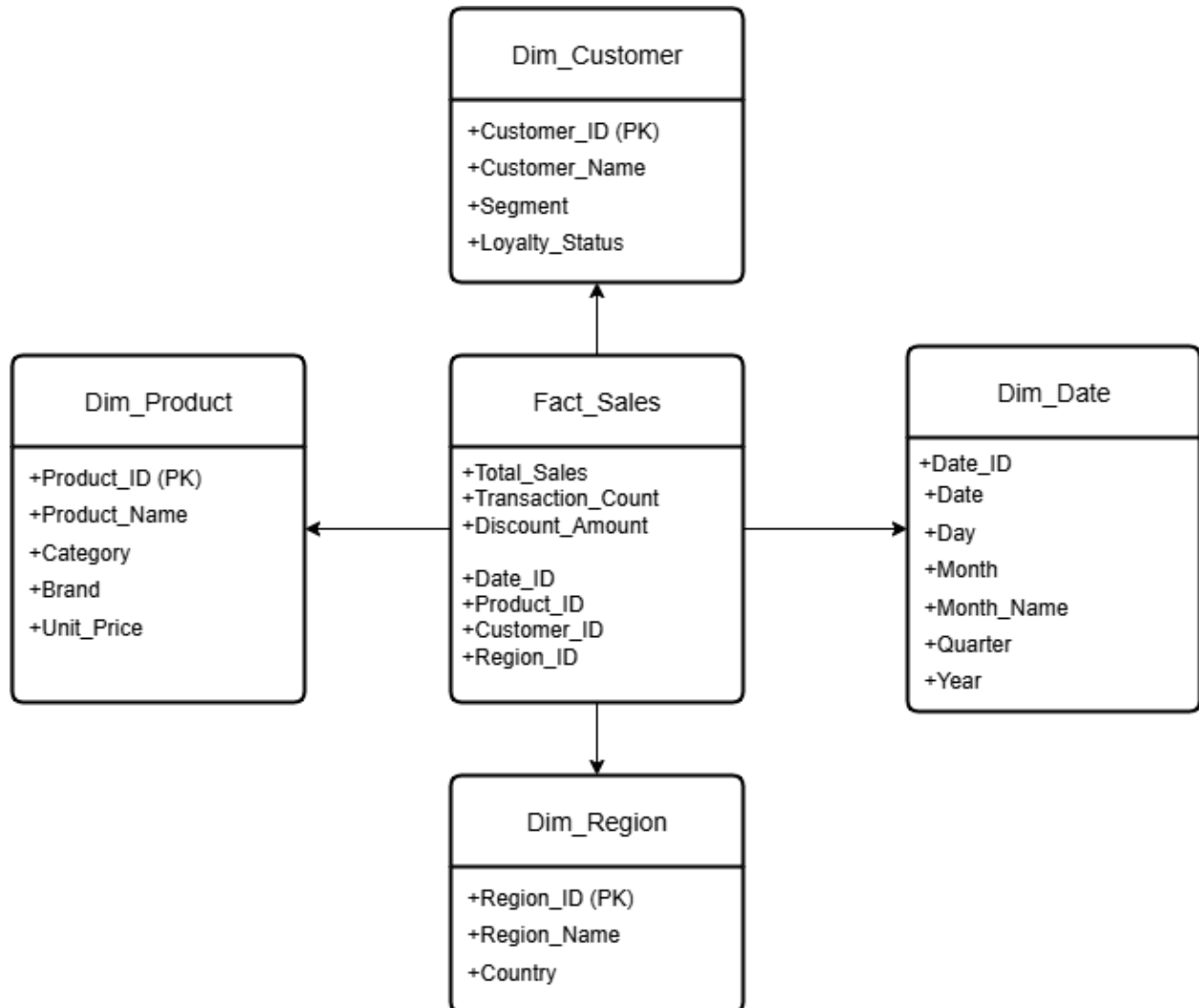
Conclusion:

Dimensional Modeling makes analytical databases **simple, efficient, and business-friendly**. The Star Schema is the most common due to its **simplicity and speed**. Snowflake improves **data integrity**, while Galaxy handles **complex business needs**. Overall, DM enables **better decision-making** through clear, optimized data structures.

Q.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.



Q.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.



Q.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.

