Name: Ishaan Khan
Class: D15C Roll no. 57

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

1. Denormalized structure: Data is stored in fewer tables to improve query performance.
2. Two main components:
    a. Fact Tables – Store numeric measurements (facts) of the business.
    b. Dimension Tables – Store descriptive attributes that describe the facts.
3. 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.

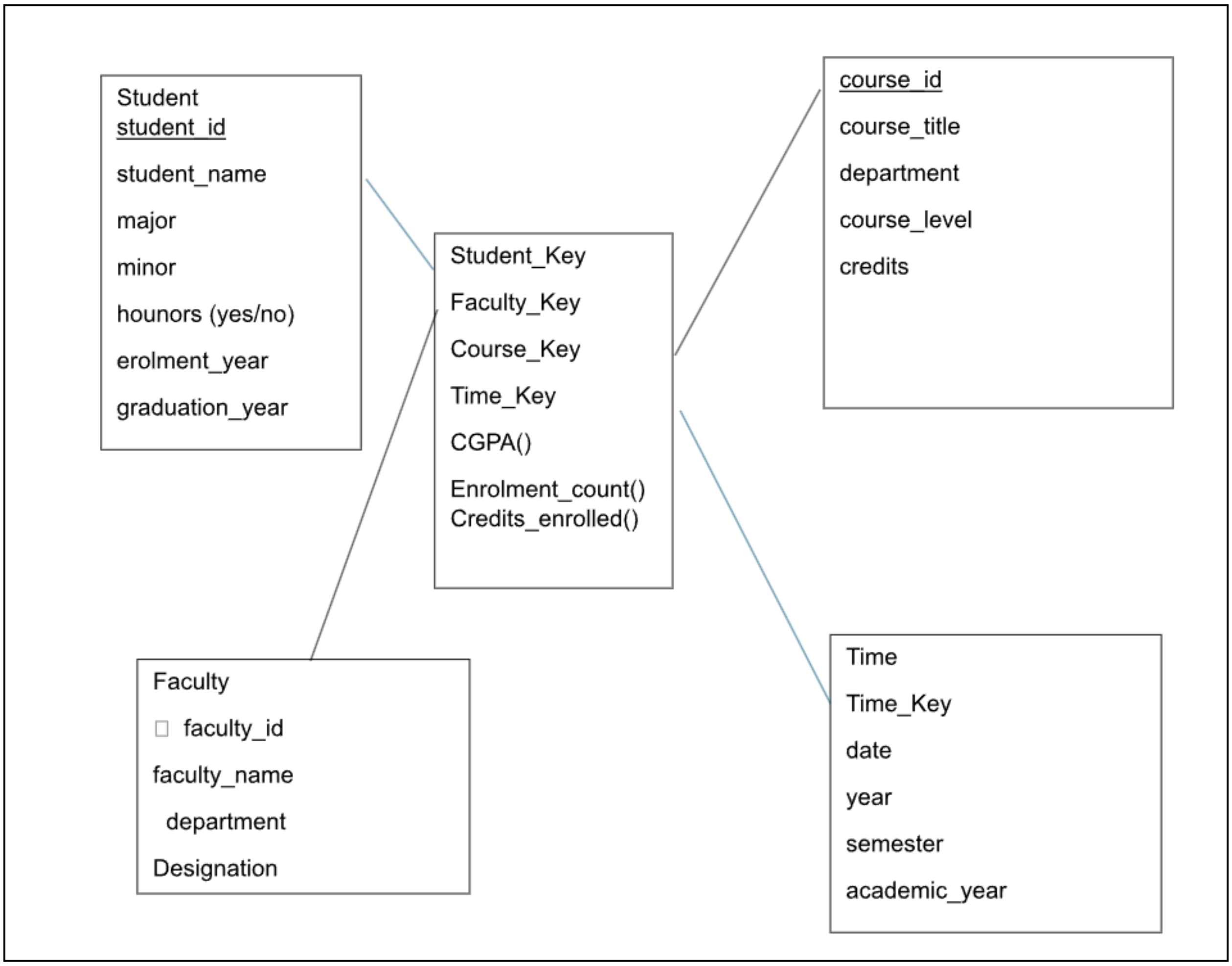
Advantages of Dimensional Modeling

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

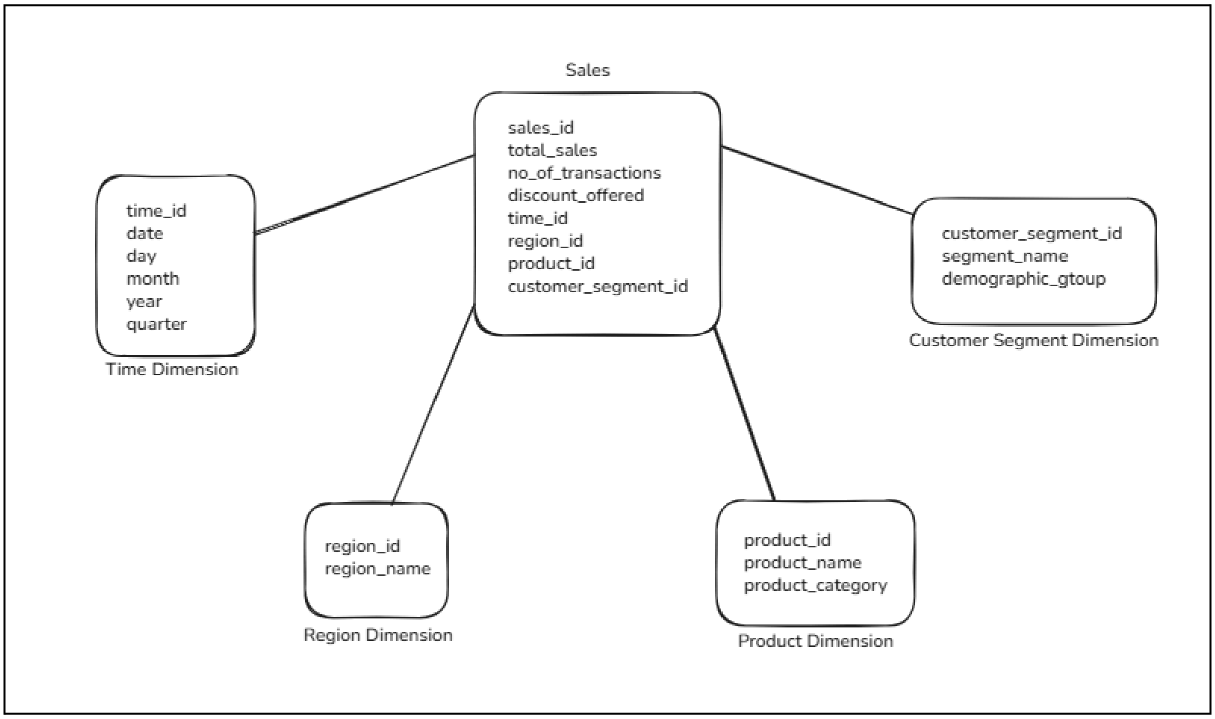
Disadvantages

1. Not suitable for transactional operations.
2. Denormalization can lead to data redundancy.
3. 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.



Sales

sales_id
total_sales
no_of_transactions
discount_offered
time_id
region_id
product_id
customer_segment_id

time_id
date
day
month
year
quarter

Time Dimension

customer_segment_id
segment_name
demographic_gtoup

Customer Segment Dimension

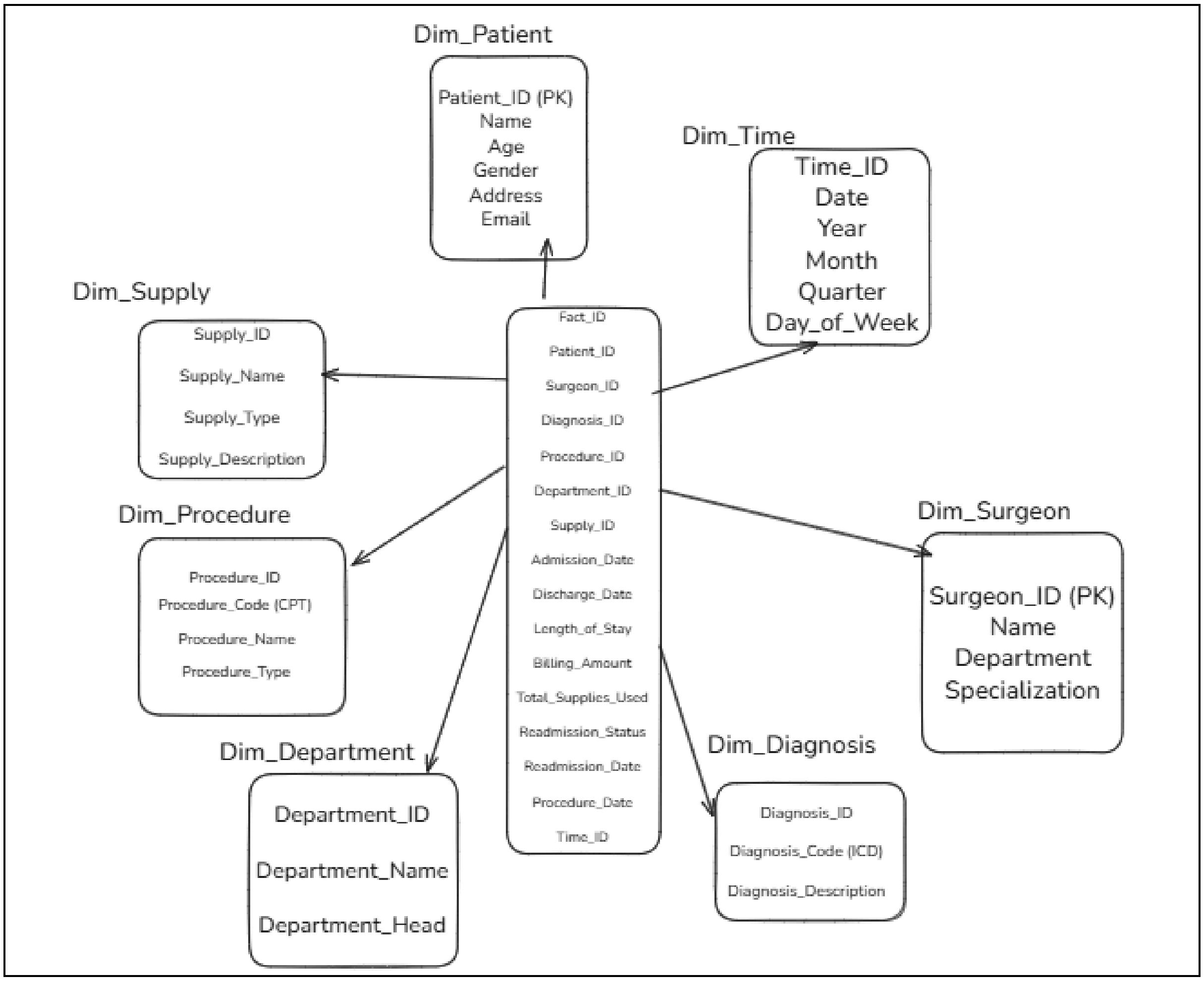
region_id
region_name

Region Dimension

product_id
product_name
product_category

Product Dimension

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:


Dim_Patient
Patient_ID (PK)
Name
Age
Gender
Address
Email
Dim_Time
Time_ID
Date
Year
Month
Quarter
Day_of_Week
Dim_Supply
Supply_ID
Supply_Name
Supply_Type
Supply_Description
Fact_ID
Patient_ID
Surgeon_ID
Diagnosis_ID
Procedure_ID
Department_ID
Supply_ID
Admission_Date
Discharge_Date
Length_of_Stay
Billing_Amount
Total_Supplies_Used
Readmission_Status
Readmission_Date
Procedure_Date
Time_ID
Dim_Procedure
Procedure_ID
Procedure_Code (CPT)
Procedure_Name
Procedure_Type
Dim_Surgeon
Surgeon_ID (PK)
Name
Department
Specialization
Dim_Department
Department_ID
Department_Name
Department_Head
Dim_Diagnosis
Diagnosis_ID
Diagnosis_Code (ICD)
Diagnosis_Description