Nupur Ghangarekar
D15C
Rollno:33
Date of submission: 8 August 2025

Practical 1

Aim: To design star schema for given problem statements

Theory:

Dimensional Modelling

Dimensional modeling is a specialized data design technique widely used in data warehousing and business intelligence to organize data for efficient retrieval, reporting, and analysis. It simplifies complex data by structuring it into two key components: **facts** and **dimensions**.

Facts are typically numeric measurements that capture business transactions or events. For example, sales revenue, quantities sold, or profit margins. These facts provide the quantitative metrics needed for analysis.

Dimensions, on the other hand, are descriptive attributes that give context to the facts. They include information such as dates, product details, customer demographics, or geographical locations, enabling analysts to slice and dice the data from multiple perspectives.

The foundation of dimensional modeling is the **fact table**, which stores the measurable data and includes foreign keys that link to the **dimension tables**. Dimension tables contain detailed attributes and are connected to the fact table, providing rich context for analysis.

There are two primary schema designs used in dimensional modeling:

- The **star schema**, where a central fact table is directly connected to multiple dimension tables, offering simplicity and fast query performance.

- The **snowflake schema**, where dimensions are normalized into related tables to reduce redundancy, at the cost of increased complexity.

Dimensional modeling optimizes query speed, making it easier for business users to understand and analyze data. It supports multidimensional data analysis, which is essential for effective data mining, trend detection, and decision-making processes in OLAP (Online Analytical Processing) environments

Questions:

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 discount offered.
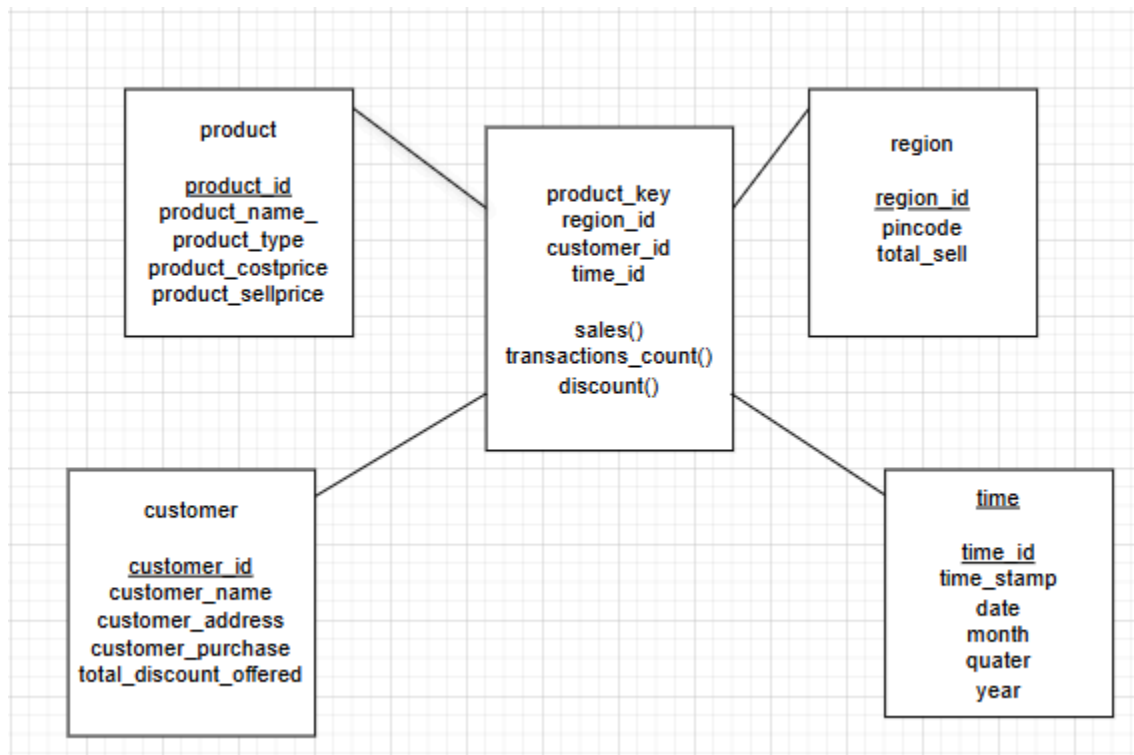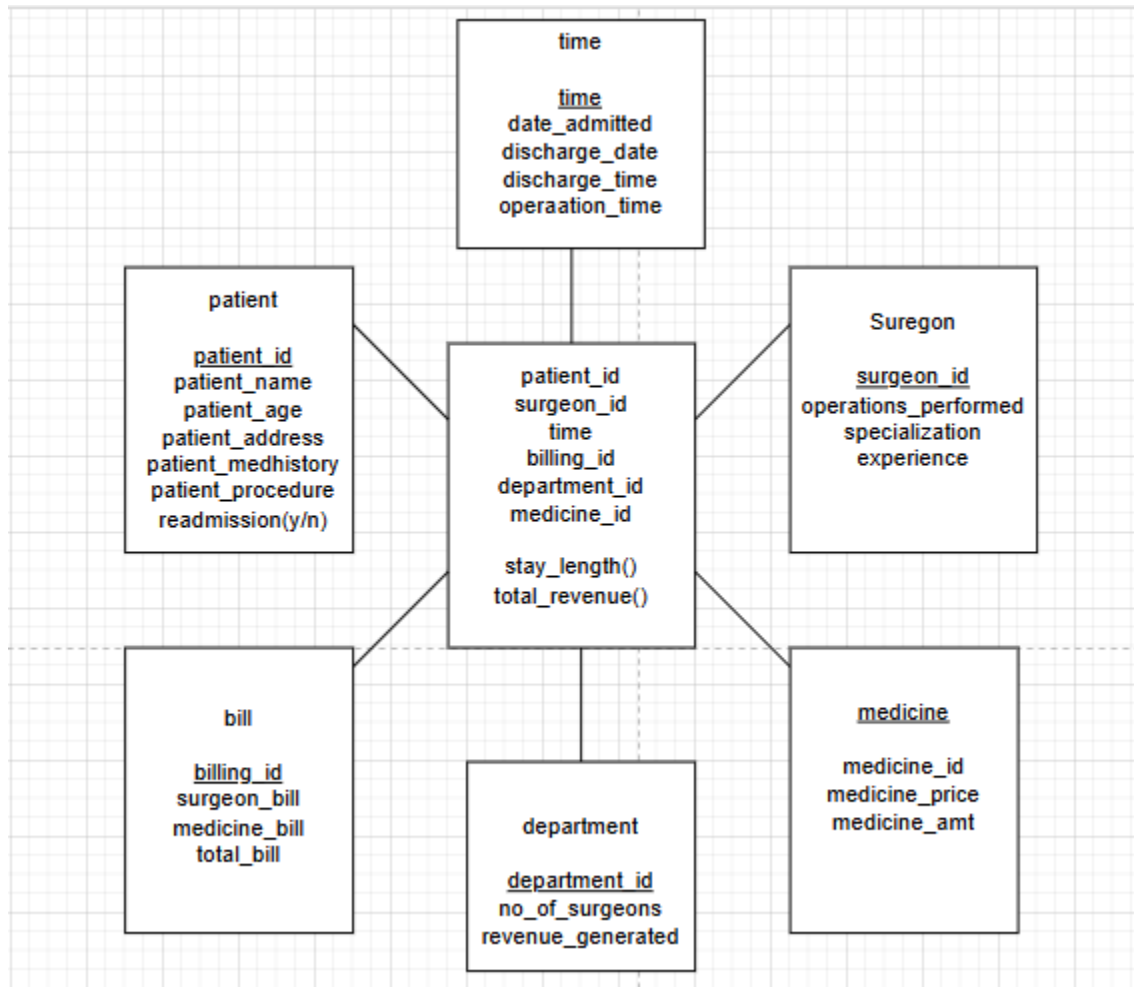

Q3.

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:
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?

time

time
date_admitted
discharge_date
discharge_time
operaation_time

patient

patient_id
patient_name
patient_age
patient_address
patient_medhistory
patient_procedure
readmission(y/n)

patient_id
surgeon_id
time
billing_id
department_id
medicine_id

stay_length()
total_revenue()

Suregon

surgeon_id
operations_performed
specialization
experience

bill

billing_id
surgeon_bill
medicine_bill
total_bill

department

department_id
no_of_surgeons
revenue_generated

medicine

medicine_id
medicine_price
medicine_amt

Conclusion:

Dimensional modeling is a data design technique that organizes data into facts and dimensions for easy retrieval and analysis. We created a **star schema** for the question set (QS), with a central fact table linked directly to dimension tables. This structure simplifies queries and improves performance by clearly connecting measurable data with descriptive context. In summary, dimensional modeling especially using star schemas enables fast, intuitive, and multidimensional data analysis. It is essential for data warehousing, business intelligence, and data mining, helping organizations make informed decisions from well-structured and accessible data.