

# Practice Questions for Dimensional Modelling

## Exercise-I

Suppose that a data warehouse contains three dimensions *date*, *doctor* and *patient*. There is only measure – *charge* where charge is the fee that a doctor charges to a patient for a visit. Design a star schema for the data warehouse, assuming some concept hierarchy for each dimension. Starting with the base cuboid [date, doctor, patient], write an SQL query to list the total fee collected by each doctor in the year 2019?

## Exercise-II

A consortium of banks wants to develop a data warehouse for effective decision-making about their loan schemes. The banks provide loans to customers for various purposes, like, House Building Loan, Car Loan, Educational Loan, Personal Loan, etc. The whole country is categorized into a number of regions, namely, North, South, East and West. Each region consists of a set of districts. Loan is disbursed to customers at interest rates that change from time to time. Also, at any given point of time, the different types of loans have different rates. The data warehouse should record an entry for each disbursement of loan to customer.

With respect to the above business scenario, answer the following questions. Clearly state any reasonable assumptions you make.

- a. Design a star schema for the data warehouse clearly identifying the fact table(s), dimensional table(s), their attributes and measures along with the primary key and foreign key relationships.
- b. Write an SQL query by which you can display region-wise, bank-wise, year-wise total amount of loans disbursed from your schema.
- c. Starting with the base cuboid, if we want to see the amount of loan disbursed during the year 2018 for the District of Kampala, which sequence of OLAP operations would you need to perform?

## Exercise-III

Consider the following business scenario. A telecom company plans to maintain a CRM data warehouse. There are 10 million customers of the company. Besides the usual attributes, the company wants to maintain additional demographic information like literacy percentage, male/female ratio, average life expectancy and average income of the people belonging to the district to which each customer belongs. The company also wants to maintain information about the age group, income level and marital status of its customers. They also need to run queries like the number of married and unmarried customers they have at any point in time.

- a. Design an efficient data warehouse schema that satisfies the above business scenario. Clearly identify the fact table(s), dimension table(s), primary key(s) and foreign key(s).
- b. Write an SQL statement that generates the number of married and unmarried customers that the company has today.

#### **Exercise-iv**

A university plans to build a data warehouse that would help them in analyzing the performance of the students in various courses in different academic semesters. They want to analyze if there is any relation between the average grade of a course and the number of students attending it. They would also like to know if there were some courses offered but did not have any students registered for them. Relative performance among boys and girls and average grades of students from various districts of the country for each course must be analyzed for each course and also overall CGPA.

- (a) Design a star schema for such a data warehouse. clearly identifying the fact table(s) and dimension table(s), their primary key(s), foreign key(s) and measures. Your schema should at least be able to satisfy the above-mentioned analysis requirements. You may consider other suitable attributes for the dimension table(s).
- (b) Transform the star schema into a snowflake schema.
- (c) Write an SQL query that runs on your schema and returns the average CGPA of boys from Wakiso district for the 1st semester the year 2005.