ACADEMIC SESSION 2022/2023 : SEMESTER I

WQD 7005 : Data Mining

6 May 2023 Time: 1 hours

INSTRUCTIONS TO CANDIDATES:

Answer ALL questions (5 marks)

Part B

Designing a data warehouse is an essential part of building an effective data management system. To design a star schema data warehouse, you need to choose a dataset that can provide you with valuable insights. In this context, your task is to design a star schema data warehouse based on a chosen dataset (group assignment dataset).

Your data warehouse should have at least four dimensions, including a time dimension, and two measures. If the chosen dataset does not meet these requirements, you may recommend adding additional attributes or columns to the dataset to fulfill the design specifications.

The first step in designing the data warehouse is to identify the fact table. The fact table is the table that contains the quantitative data that you want to analyze. You need to identify the fact table and ensure that it meets the design specifications.

Once the fact table has been identified, you need to determine the dimensions. Each dimension will provide a different perspective on the data and enable users to explore the relationships between the different factors. For each dimension, it is important to carefully select the relevant attributes that will be included. The selected attributes should provide useful information to the users of the data warehouse.

If applicable, you also need to include a time dimension. The time dimension enables users to analyze data across different time periods and obtain insights into trends and patterns.

After selecting the dimensions and measures, you can create a star schema diagram for the data warehouse. The star schema diagram will enable users to

easily navigate the dimensions and measures, allowing them to obtain insights into the chosen dataset. It is important to ensure that all attributes included in the star schema are relevant and useful for analysis and reporting. Any irrelevant attributes may detract from the usability and effectiveness of the data warehouse.

In summary, your task is to design a star schema data warehouse based on a chosen dataset. You need to identify the fact table, select relevant attributes for each dimension, including the time dimension, and create a star schema diagram that enables users to explore relationships between factors over time and obtain insights into the chosen dataset. Finally, you need to discuss the importance of ensuring that all attributes included in the star schema are relevant and useful for analysis and reporting, and how irrelevant attributes may detract from the usability and effectiveness of the data warehouse.

[5 marks]