Data Warehousing Assignment-2

Scenario I

Question 1. What are the strengths and weaknesses of each option?

Answer 1:

Option A: The advantage of this option is that it preserves the Instructor dimension in the fact table and allows for the tracking of instructor-related attributes. However, it may lead to increased complexity in the data model and querying, as well as potential confusion among users when dealing with instructor teams.

Option B: This option also preserves the Instructor dimension in the fact table and allows for tracking of instructor-related attributes, but it requires modifying the grain of the fact table, which may cause confusion among users and require additional explanation in documentation.

Option C: This option simplifies the data model by separating the Instructor dimension into a separate fact table, but it requires users to know which fact table to use for different queries, which may be confusing.

Question 2. Which option would you choose and why?

Answer 2:

I would choose Option B, as it preserves the Instructor dimension in the fact table while still maintaining a clear grain for the fact table. Modifying the Instructor dimension or separating it into a separate fact table may add unnecessary complexity to the data model and querying, and Option B allows for easy aggregation and reporting of enrollment counts by course, instructor, and other dimensions.

Question 3. Would your answer to Question 2 be different if the majority of classes had multiple instructors? How

about if only one or two classes had multiple instructors? (Explain your answer.)

Answer 3:

If the majority of classes had multiple instructors, I would still choose Option B, as it allows for the tracking of instructor-related attributes while maintaining a clear grain for the fact table. However, if only one or two classes had multiple instructors, it may be simpler to use Option A and modify the Instructor dimension by adding special rows representing instructor teams.

Question 4. [OPTIONAL] Can you think of another reasonable alternative design besides Options A, B, and C? If

so, what are the advantages and disadvantages of your alternative design?

Answer 4:

Another reasonable alternative design could be to create a separate fact table for courses with multiple instructors, rather than modifying the grain of the fact table or separating the Instructor dimension into a separate fact table. This would allow for the tracking of instructor-related attributes while still maintaining a clear grain for the fact table, and would avoid confusion among users about which fact table to use for different queries. However, it may require additional maintenance and management of multiple fact tables.

Scenario II

Question 5. What are the strengths and weaknesses of each option?

Answer 5:

Option A:

---------------------------Strengths-----------------------------------

Simple and easy to implement.

Minimal storage requirement as there is no need for additional tables.

---------------------------Weaknesses-------------------------------------

Historical data cannot be preserved. As a result, there is no way to see how a customer’s activity or profitability score has changed over time.

It is not suitable for reporting or analysis purposes, as the previous scores will be overwritten with the new scores.

Option B:

---------------------------Strengths-----------------------------------

Historical data can be preserved by creating new rows with updated scores.

Suitable for analysis and reporting purposes as it can track changes in scores over time.

---------------------------Weaknesses-------------------------------------

Requires more storage than Option A as there is a need for additional rows.

More complex to implement than Option A.

Option C:

---------------------------Strengths-----------------------------------

The number of rows in the CustomerScores table is fixed, which reduces storage requirements.

Suitable for analysis and reporting purposes.

---------------------------Weaknesses-------------------------------------

The combination of activity and profitability scores may not be sufficient to represent customer behavior accurately.

Does not track changes in scores over time.

Option D:

---------------------------Strengths-----------------------------------

Allows for historical tracking of scores and reduces storage requirements by using an outrigger table.

---------------------------Weaknesses-------------------------------------

Complex to implement.

May not be suitable for analysis and reporting purposes as the fact table does not contain foreign keys to the outrigger table.

Question 6. Which option would you choose and why?

Answer 6:

Option B is the best choice because it allows for historical tracking of scores while still being suitable for analysis and reporting purposes. Although it requires more storage than Option A, it is more flexible and useful in the long run.

Question 7. Would your answer to Question 6 be different if the number of customers and/or the time interval

between score recalculations was much larger or much smaller? (Explain your answer.)

Answer 7:

If the number of customers or the time interval between score recalculations were much larger, then Option C may become more attractive due to its reduced storage requirements. However, if the time interval between score recalculations were much smaller, then Option D might be more suitable as it would minimize the number of updates to the Customer dimension.

Question 8. [OPTIONAL] Can you think of another reasonable alternative design besides Options A, B, C, and D?

If so, what are the advantages and disadvantages of your alternative design?

Answer 8:

Another possible design would be to use a hybrid approach between Option A and Option B, where the current score is stored in the Customer dimension as a type 1 slowly changing dimension, while historical scores are stored in a separate CustomerScoreHistory table as a type 2 slowly changing dimension. This approach would minimize storage requirements while still allowing for historical tracking of scores. However, it would be more complex to implement and maintain than Option B.