

DATA WAREHOUSING ASSIGNMENT

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

Option A:

Strengths: Simple to implement and understand.

Weaknesses: It may not accurately represent courses with multiple instructors

Option B:

Strengths: Maintains granularity on individual instructor level. Accurately represents courses with multiple instructors.

Weaknesses: More complex to implement and understand.

Option C:

Strengths: Maintains granularity on individual instructor level. Allows for flexibility in querying.

Weaknesses: More complex to implement and understand. It required creation and maintenance of two separate fact tables.

Question 2: Which option would you choose and why?

If the majority of classes had multiple instructors, I would choose Option C. This option allows for the most granular data on individual instructors, and the complexity of maintaining two fact tables it allow to use frequently courses with multiple instructors. If only one or two classes had multiple instructors, I would choose Option A. The simplicity of this option and easy understanding.

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.)

If the majority of classes had multiple instructors, my answer to Question 2 would not change. Option C would still be the best choice as it allows for granular data on individual instructors and is justified by the high frequency of courses with multiple instructors. If only one or two classes had multiple instructors, my answer to Question 2 would change. In this case, the simplicity of Option A and the low impact on overall data accuracy outweigh the benefits of maintaining granularity on the individual instructor level, so I would choose Option A.

Scenario II:

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

Option A:

Strengths: Simple to implement. No need to maintain multiple copies of customer data.

Weaknesses: Loses historical data on customer scores. Cannot track changes in customer scores over time.

Option B:

Strengths: Allows for tracking of changes in customer scores over time. We can analysis of how and why customer scores change.

Weaknesses: Requires maintenance of multiple copies of customer data. It require more storage space.

Option C:

Strengths: Allows for tracking of changes in customer scores over time. Allows for analysis of how and why customer scores change.

Weaknesses: Requires maintenance of a separate CustomerScores dimension. May require additional storage space.

Option D:

Strengths: Allows for tracking of changes in customer scores over time. Allows for analysis of how and why customer scores change.

Weaknesses: Requires maintenance of a separate CustomerScores outrigger table. May require additional storage space. May be more complex to implement and to understand .

Question 6: Which option would you choose and why?

My Option would be B. This option allows for tracking of changes in customer scores over time, which is important for the company's data analysts to understand how and why customer scores change.

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.)

If number of customers was much larger, my answer to Question 6 would not change. Option B would still be the best choice as it allows for tracking of changes in customer scores over time. If the number of customers was much smaller, my answer to Question 6 may change. In this case, the added complexity and storage requirements of Option B may not be justified. If the number of customers is small, there may be fewer changes in customer scores over time. This means that there will be fewer new rows added to the Customer dimension to represent changes in scores, and the added complexity of maintaining multiple copies of customer data may not be justified. Option A (Slowly Changing Dimension) or Option C (separate CustomerScores dimension) may be more good choices for this senario.