Data Modelling - Star vs Snowflake Schema

Q1: Today, we'll dive into data modelling concepts, specifically focusing on star and snowflake schemas. Are you familiar with these concepts?

Ans: They're commonly used in data warehousing to organize and structure data for analytical purposes.

Q2: Could you explain what a star schema is and how it's structured?

Ans: In a star schema, we have a central fact table surrounded by dimension tables. The fact table contains quantitative data, usually numerical metrics or measures, while the dimension tables contain descriptive attributes that provide context to the measures. The fact table is connected to the dimension tables through foreign key relationships, forming a star-like shape."

Q3: Difference between Star and Snowflake Schema

Ans: "In a snowflake schema, the dimension tables are normalized, meaning that they are further broken down into multiple related tables. This results in a more complex network of relationships, resembling the branches of a snowflake. While this normalization can save storage space and reduce data redundancy, it can also lead to increased query complexity due to the need for additional joins."

Q4: In what scenarios would you prefer using a snowflake schema over a star schema, and vice versa?"

Ans: "Choosing between a star and snowflake schema depends on various factors such as the nature of the data, query patterns, and performance requirements. A star schema is simpler and easier to understand, making it suitable for scenarios where performance and simplicity are prioritized. On the other hand, a snowflake schema may be preferred in scenarios where data integrity and storage optimization are critical, and the additional complexity introduced by normalization is acceptable."

Q5: let's consider a hypothetical scenario where you're tasked with designing a data warehouse for an e-commerce company. Would you opt for a star or snowflake schema, and why?"

Ans: "In the case of an e-commerce company, where performance and ease of querying are paramount, I would lean towards a star schema. The simplicity and denormalization of the star schema would facilitate efficient querying of sales data and analytics. However, I would consider normalizing certain dimension tables in a snowflake-like fashion if there are large, frequently updated attributes that could benefit from reduced redundancy and improved data integrity."