Data Science & AI

ENGINEERING

warehousing

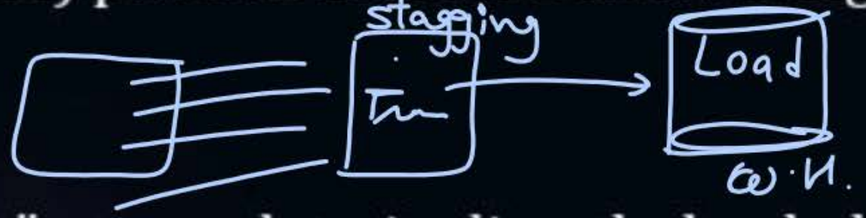
Part- 01

By- Vishvadeep Gothi sir

Topic : Datawarehousing

B, D

#Q. Which of the following statements are true regarding the ETL (Extract, Transform. Load) process in data warehousing?



A In the "Extract" stage, data is directly loaded into the data warehouse to avoid corruption.

B The "Transform" stage involves applying rules or functions to convert data into a standard format.

C ETL is a one-time process and does not need to be repeated as new data is added.

D The "Load" stage involves creating physical data structures for the transformed data.

Topic : Datawarehousing

#Q. Which of the following are advantages of the ETL (Extract, Transform.. Load) process in data warehousing?

A ✓ Improved data quality

B ✓ Increased data security

C ✗ Limited flexibility

D ✓ Improved scalability

Topic : Datawarehousing

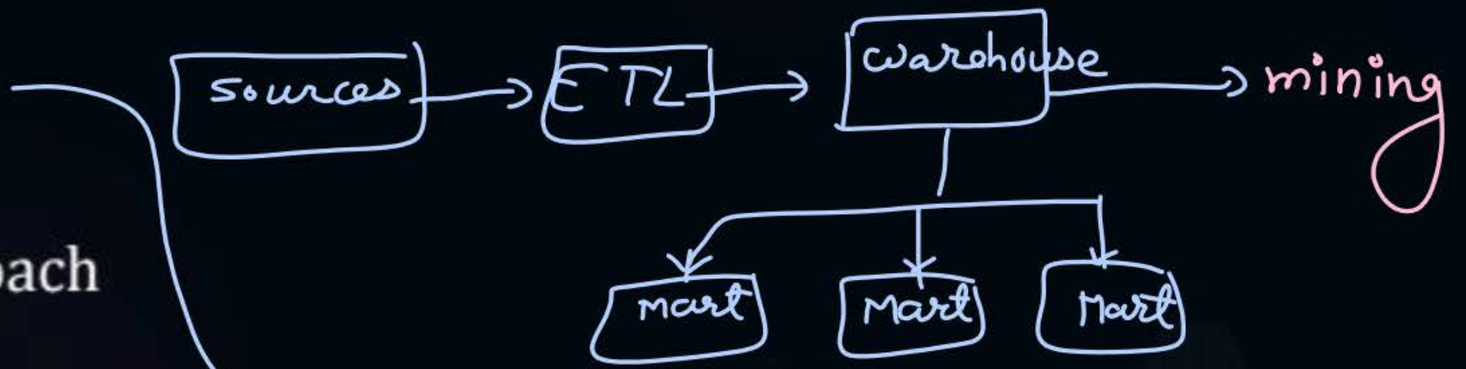
#Q. Which approaches can be used in the design process of a data warehouse?

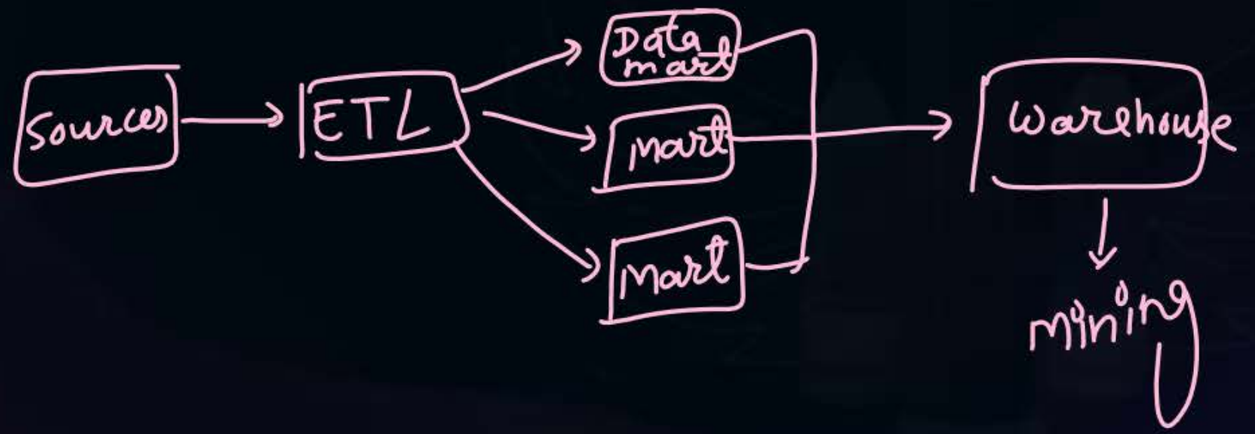
A Top-Down approach

B Bottom-Up approach

C Middle-Out approach

D Inside-Out approach





Topic : Data warehousing

#Q. Which characteristics describe a data warehouse?

A ✗ Volatility

B ✗ Subject-Oriented → data mart

C ✗ Real-time updates → in database

D ✓ Integrated

Topic : Datawarehousing

#Q. Which characteristics describe a data warehouse?

A ✗ Real-time updates

B ✓ Time-Variant → with time, warehouse changes
→ not on real time

C ✗ Volatility

D ✓ Non-Volatile

Topic : Datawarehousing

#Q. Which characteristic does NOT describe a data warehouse?

A Volatility

B Subject-Oriented

C Integrated

D Time-Variant

Topic : Datawarehousing

#Q. What are the stages involved in the ETL (Extract. Transform.. Load) process?

A Extract

B Adapt

C Transform

D Load

Topic : Datawarehousing

#Q. What is the importance of the ETL (Extract. Transform, Load) process in data warehousing?

A ✓ Ensures data in the warehouse is accurate, complete, and up-to-date.

B ✗ Provides a way to manage and analyze large amounts of data.

C ✗ Reduces the need for data integration.

D ✓ Simplifies data mining and reporting.

Topic : Datawarehousing

#Q. Why is the ETL (Extract. Transform, Load) process considered an iterative process?

A It is a one-time process with no need for repetition.

B It is repeated as new data is added to the warehouse.

C It does not require data validation.

D It is solely dependent on the technology used.

Topic : Datawarehousing

#Q. What is an advantage of using ETL tools in the data warehousing process?

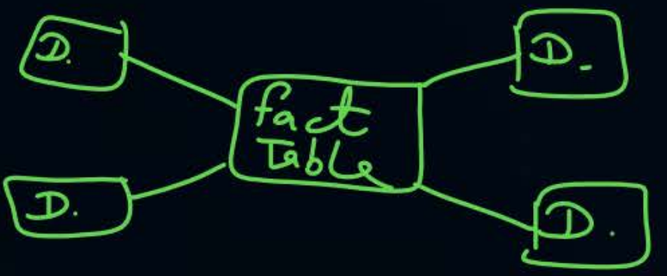
A. Increased cost of implementation

B. Improved scalability

C. Limited automation

D. Simplification of the ETL process

Topic : Datawarehousing



#Q. In a star schema for a data warehouse:

- What is the role of the central table (fact table)?
- What characterizes the attendant tables (dimension tables)?

A ✗ The central table contains redundant data. → duplicate

B ✓ The central table is large and contains the bulk of the data.

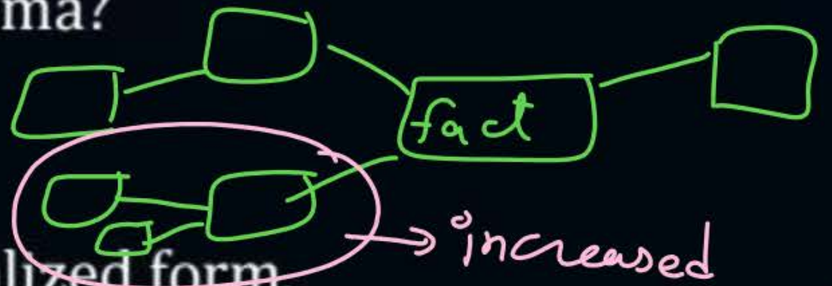
C Each dimension is represented by multiple tables.

D ✓ Each dimension is represented by only one table.

Topic : Datawarehousing

#Q. What characterizes the snowflake schema in a data warehouse?

- How does it differ from the star schema?



A Dimension tables are kept in the normalized form.

B It forms a shape similar to a snowflake.

C It increases system performance by reducing joins.

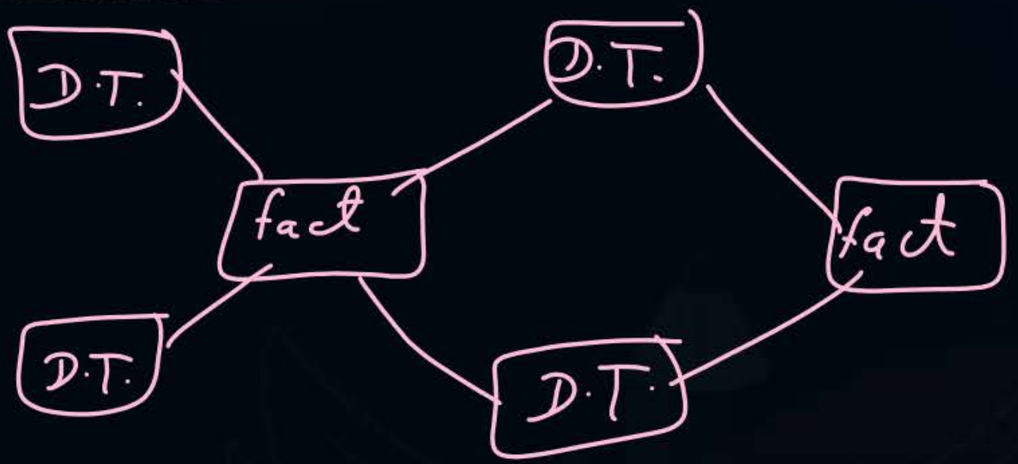
D The fact table is smaller in magnitude compared to the star schema.

↳ similar

Topic : Datawarehousing

#Q. What characterizes a fact constellation schema in a data warehouse?

- How does it differ from the star schema?



A ~~It~~ includes multiple fact tables.

B ✗ It reduces redundancy.

C ~~It~~ can be viewed as a collection of stars.

D ✗ Each dimension is represented by multiple tables.

Topic : Datawarehousing

#Q. Which schema model(s) may be used in a data warehouse design?

A Star Schema

B Snowflake Schema

C Constellation Schema

D Zigzag Schema

Topic : Datawarehousing

#Q. What is an impact of the snowflake schema on system performance in comparison to the star schema?

A. It increases system performance.

B. It reduces the number of joins needed for queries.

C. It can adversely impact system performance.

D. It improves data browsing effectiveness.

Topic : Datawarehousing

#Q. How are dimension tables treated in a snowflake schema compared to a star schema?

A Dimension tables are kept in the normalized form.

B Dimension tables are denormalized.

C Dimension tables are duplicated.

D Dimension tables are eliminated.

Topic : Datawarehousing

#Q. In a fact constellation schema, what is a characteristic related to dimension tables?

A. Each dimension is represented by only one table.

B. Dimension tables are duplicated for each fact table.

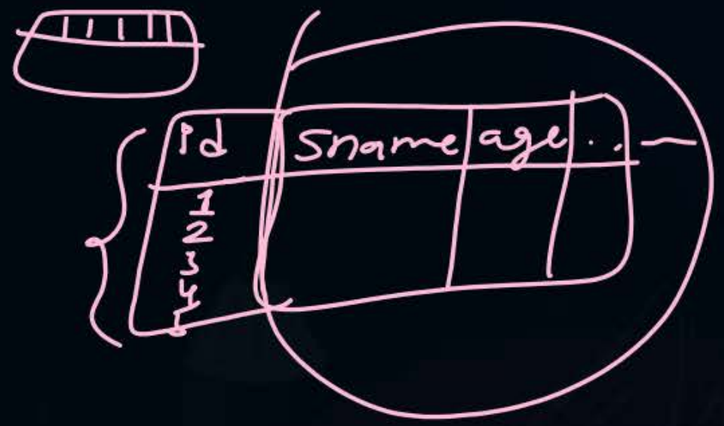
C. Dimension tables are shared between fact tables.

D. Dimension tables have no impact on fact tables.

Topic : Datawarehousing

#Q. What characterizes the attributes of dimension tables in a star schema?

→ columns



A Each dimension is represented by multiple tables.

B Each table contains a set of attributes.

C Dimension identifiers are system-generated.

→ surrogate keys

D ✗ Each table contains redundant data.

Topic : Datawarehousing

#Q. Consider a concept hierarchy for the dimension "location." Given the values Vancouver. Toronto, New York, and Chicago as cities:

- What do these cities map to in the concept hierarchy?
- What does the concept hierarchy illustrate?

A Vancouver maps to British Columbia.

B Toronto maps to Ontario.

C New York maps to New York State.

D The concept hierarchy maps cities to countries.

Topic : Datawarehousing

#Q. In the context of a data cube, how are measures categorized based on their computation characteristics? - Identify the correct categories.

A Distributive

B Algebraic

C Holistic

D Accumulative

Topic : Datawarehousing

#Q. What characterizes an aggregate function in a data cube as distributive?

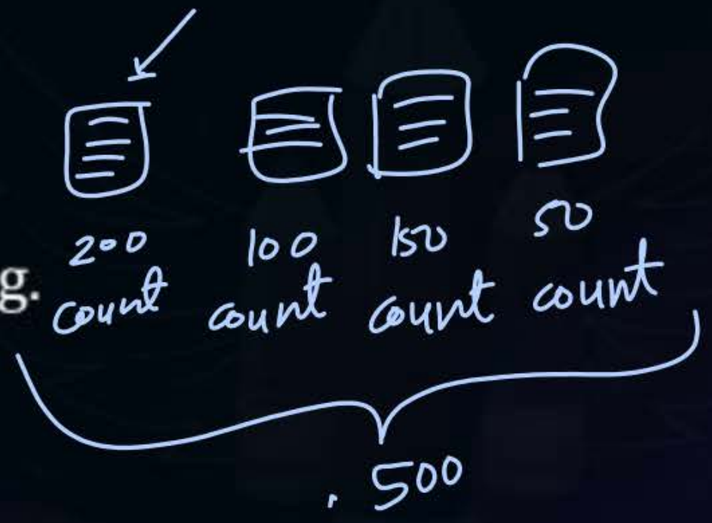
- Identify the correct characteristics.

A ~~It~~ can be computed in a distributed manner.

B ✗ It involves partitioning the cube.

C The result remains the same after partitioning.

D It requires centralized computation.



Topic : Datawarehousing

#Q. What characterizes an aggregate function in a data cube as algebraic?

- Identify the correct characteristics.

A. It can be computed by an algebraic function.

B. It involves partitioning the cube.

C. It requires a holistic approach.

D. It is not bounded by a constant storage size.

Topic : Datawarehousing

#Q. What characterizes an aggregate function in a data cube as holistic?

- Identify the correct characteristics.

A ✗ It involves partitioning the cube.

B It requires a constant bound on storage size.

C It can be computed in a distributed manner.

D ✓ It includes median(), mode(), and rank()-

Happy Learning

THANK - YOU