

1. What is the difference between a data warehouse and a traditional database system?

Answer:

A database is a structured collection of connected data that is stored in a tabular style, whereas a data warehouse is a single place that houses combined data from several databases. This is the major distinction between the two. More differences are given below in a table.

Database	Data Warehouse
1. Contains detailed data.	1. Contains summarized data.
2. It uses Online Transactional Processing (OLTP).	2. It uses Online Analytical Processing (OLAP).
3. Helps to perform the fundamental operations of a business.	3. Helps to analyze the business.
4. It is slower and less accurate. Moreover application oriented.	4. It is faster, accurate and subject oriented.
5. Tables and joins of a database are complex as they are normalized.	5. Table and joins are simple in a data warehouse because they are denormalized.
6. ER modeling techniques are used for designing.	6. Data modeling techniques are used for designing.
7. Flat Relational Approach method is used for data storage.	7. Data Ware House uses dimensional and normalized approach for the data structure. Example: Star and snowflake schema.

2. Explain the concept of dimensional modeling and how it differs from normalized data modeling.

Answer:

Dimensional modeling is a database modeling technique used primarily in data warehousing. Instead of using an intricate relational structure, it includes arranging data into a collection of connected and simple dimensions and measures. A fact table containing business metrics is often accompanied by one or more dimension tables that explain the context of those metrics in dimensional models.

In dimensional modeling, the data is organized into a star schema or snowflake schema. A star schema consists of a central fact table connected to dimension tables via foreign keys. A snowflake schema is a variation of the star schema, in which the dimensions are normalized into multiple related tables.

In contrast, normalized data modeling is a database design strategy that divides data into a number of related tables to reduce data redundancy. Each table in a normalized data model represents a distinct entity, and the data is divided into a number of linked tables according to their logical relationships. A relational schema is often used to describe the normalized data model, with each table having its own set of columns. The table given below shows the significant differences in tabular form.

Dimensional Modeling	Normalized Data Modeling
1. Emphasizes the ease of use and query performance for business intelligence purposes.	1. Focuses on minimizing data redundancy and improving data consistency.
2. Optimized for query performance and analysis.	2. Optimized for data integrity and consistency.
3. Dimensional models, on the other hand, are simpler and easier to understand, as they use a star	3. Normalized data models tend to be more complex than dimensional models, as they involve

schema or snowflake schema that can be easily visualized.	multiple tables and complex relationships between them.
4. Dimensional modeling is an effective technique for building data warehouses and supporting business intelligence reporting and analysis.	4. Normalized data modeling is more appropriate for transactional databases and systems where data integrity and consistency are critical.

3. What is ETL? Describe the key steps involved in an ETL process.

Answer:

Extract, Transform, and Load is referred to as ETL. Data is extracted from different sources, transformed into the necessary format, and then loaded into a destination system, like a data warehouse or a data lake. ETL is used to combine data from disparate systems, databases, and applications into a single, consolidated view and is a crucial component of data integration. The key steps of the ETL process are listed below.

- a. Extraction: In the extraction step, data is extracted from various sources such as databases, flat files, web services, and APIs.
- b. Transformation: In the transformation step, the extracted data is transformed into a desired format that can be loaded into the target system. This involves cleaning the data, removing duplicates, merging data from multiple sources, and applying business rules and calculations to the data.
- c. Loading: In the loading step, the transformed data is loaded into the target system such as a data warehouse, data lake, or a reporting tool. This involves identifying the target schema, mapping the transformed data to the target schema, and loading the data into the target system.

4. What are the benefits of data warehousing for an organization?

Answer:

Data warehousing is the process of gathering, managing, and storing data from multiple sources for the benefit of corporate information. Data from various sources must be integrated into a single, centralized repository in order to provide a consolidated picture of the data to enable analytics and decision-making. The benefits of data warehousing for an organization are listed below

- a. Data warehousing helps to improve data quality by ensuring that data is accurate, consistent, and complete.
- b. Data warehousing provides a single, consolidated view of the data, making it easier for decision-makers to access and analyze data.
- c. Helps to provide insights into customer behavior, market trends, and business performance, enabling organizations to make informed decisions.
- d. Enables faster access to data by providing a centralized repository that is optimized for query performance

e. Enables organizations to scale their data processing and storage capabilities as their data volumes and analytics requirements grow.

f. Data warehousing provides enhanced data security by centralizing data storage and providing controlled access to data.

g. Data warehousing allows organizations to implement security measures such as access controls, encryption, and data backup and recovery.

5. What are the different types of data that can be stored in a data warehouse?

Answer:

Depending on the objectives and requirements of a company, data warehousing can store a variety of data kinds. The different types of data that can be stored in a data warehouse are

Serial	Type	Description	Example
1.	Structured Data	Data is organized and formatted in a specific way and can be easily processed using traditional database management systems.	Customer transactions, Sales data, Financial data, Inventory data.
2.	Unstructured Data	Unstructured data refers to data that is not organized in a specific way and does not fit into a traditional database management system.	Emails, Social Media posts, Customer reviews, Photos, Videos.
3.	Semi-structured Data	Semi-structured data is not organized in a specific way but has some organizational elements that make it easier to process.	Web logs, XML files, Social media data.
4.	Time-series Data	Time-series data is a type of structured data that is organized based on time.	Stock prices, Weather data, Sensor data.
5.	Metadata	It is actually the data about data.	Data definitions, Data lineage, Data relationships.