**NAME: PRAKHAR MISHRA**

**DATE: 17/01/2024**

**QUESTIONS**

1. **What do you mean by Data engineering?**
   * In simple terms, it involves storage and management of large amount of data as well as processing the data.
   * It helps the organization in creating proper infrastructure for storage and management of the large amount of data.
   * It ensures reliability and scalability of the data systems.
   * It gives data analyst to analyze the data and give meaningful insights from the data.
2. **What do you mean by Data Warehousing?**

* In simple terms, data warehousing is like a specialized storage system for organizing and managing large amounts of business data.
* It's a centralized and structured repository that brings together data from various sources within a company.
* This organized data is then used for analysis, reporting, and making informed business decisions.

1. **What are the features of Data Warehousing? Explain them**.

There are four basic features of Data Warehousing:

1. **Subject-Oriented**: Data are organized according to the subject instead of application.

**Example**: In a retail data warehouse, data is organized around subjects like sales, inventory, and customer demographics rather than being separated by individual applications. This makes it easier for analysts to focus on specific business areas when extracting insights.

1. **Time-Invariant**: The time duration for data warehouse is longer than that of operational systems.

**Example**: A financial institution tracks daily account balances, and this historical data is stored in the data warehouse. This enables analysts to assess trends, monitor account fluctuations over time, and identify patterns in customer transactions.

1. **Integrated**: Constructed by integrating multiple, heterogeneous data sources like relational databases, flat files, on-line transaction records.

Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources.

**Example:** A manufacturing company integrates data from production, supply chain, and sales systems into a data warehouse. This integration ensures that information about inventory levels, production schedules, and sales performance is consistent and can be analyzed together.

1. **Non-Volatile**: No updates are allowed. Once the data entered into the data warehouse, they are never removed.

The data in warehouse represent company’s history, the operational data representing near term history are always added to it.

**Example:** A healthcare organization stores patient records in a data warehouse. Once a patient's data is entered, it remains unchanged in the data warehouse, providing a reliable historical record for medical research and compliance purposes.