**Data Warehouse**: A data warehouse stores structured data that is cleaned, processed, and organized into tables for analytics. It is optimized for querying and reporting.

Data mining is the process of collecting information in order to find patterns, trends, and usable data that will help a company to make data-driven decisions from large amounts of data

Transactional Fact Table: This is a very basic and fundamental view of corporate processes. It can be used to depict the occurrence of an event at any given time

Snapshot Fact Table: The snapshot depicts the condition of things at a specific point in time, sometimes known as a "picture of the moment.

Accumulating Fact Table: These are used to depict the activity of any process with a well-defined beginning and end.

A table in a data warehouse's star schema is referred to as a dimension table. Dimensional data models, which are made up of fact and dimension tables, are used to create data warehouses. Dimension tables contain dimension keys, values, and attributes and are used to describe dimensions. It is usually of a tiny size. The number of rows might range from a few to thousands. It is a description of the objects in the fact table. The term "dimension table" refers to a collection or group of data pertaining to any quantifiable occurrence

In Online Analytical Processing (OLAP), a **data cube** is a way to organize and structure data for fast, multidimensional analysis. It allows users to explore and analyze data across various dimensions, making it easy to slice, dice, and drill down into data for insights.

In simple terms, a **data cube** in OLAP is a structured way to organize data that makes it easy to analyze from different angles. Imagine data organized in a 3D cube, where each side (or dimension) represents a category, like **Time**, **Product**, or **Location**. You can quickly look at the data by filtering one of these categories or drilling down into details

A data mart is a Data Warehouse's single subtype. It is created to fulfill the requirements of a certain user group.

A data cube is a multidimensional data model that stores optimized, summarized, or aggregated data for quick and easy analysis using OLAP technologies. The precomputed data is stored in a data cube, which makes online analytical processing easier. We all think of a cube as a three-dimensional structure, however in data warehousing, an n-dimensional data cube can be implemented. A data cube stores information in terms of dimensions and facts.

 **Data Warehouse**: A data warehouse stores structured data that is cleaned, processed, and organized into tables for analytics. It is optimized for querying and reporting.

 **Data Lake**: A data lake stores raw, unstructured, semi-structured, and structured data. It can hold data in its native format and is more flexible, supporting advanced analytics and machine learning

 **Star Schema**: A simple schema with a central fact table connected to dimension tables. It is easier to understand and query, offering faster performance for complex queries due to fewer joins.

 **Snowflake Schema**: A more normalized version of the star schema, where dimension tables are split into sub-dimensions. It reduces data redundancy but can make querying more complex and slower due to more joins

**4. How would you ensure data quality in a data warehouse environment?**

* **Data Validation**: Ensuring data adheres to expected formats and ranges before being loaded into the warehouse.
* **Cleansing**: Removing duplicates, correcting errors, and handling missing values.
* **Consistency**: Ensuring data is consistent across different sources and historical datasets.
* **Monitoring**: Continuously monitoring data loads and transformation processes for errors.

**5. What is data partitioning, and why is it important in data warehouse design?**

* **Data Partitioning**: Splitting large datasets into smaller, more manageable pieces based on certain attributes (e.g., date, region). This improves query performance and load times, as only relevant partitions are scanned during queries.

**12. What are the common challenges when integrating data from multiple sources into a data warehouse, and how would you address them?**

* **Data Consistency**: Different source systems may use different formats or definitions. Address this by using a data governance framework to standardize data.
* **Data Quality**: Ensuring data from various sources is clean and accurate. Implement data validation and cleansing processes.
* **Latency**: Data from various sources may be updated at different times, creating synchronization issues. Use real-time or near-real-time integration tools like CDC.
* **Scalability**: Handling large volumes of data from diverse sources can be challenging. Ensure that the data pipeline is scalable and optimized for performance.