DWM VIVA QUESTIONS

1. Data Warehouse:
   - It is a repository of integrated information which can be available for queries and analysis.
   - It is a system for the storage of data from various sources in an orderly manner.

2. Data Warehousing:
   - It is the process of constructing and using a data warehouse.
   - It is constructed by integrating data from multiple different sources that support analytical reporting, structured and/or ad hoc queries, and decision making.
   - It involves data cleaning, data integration, and data consolidations.

3. Stages of Data Warehousing:
   a. Offline Operational Databases
   b. Offline Data Warehouse
   c. Real-Time Data Warehouse
   d. Integrated Data Warehouse

4. Data Warehouse Architecture:
   - A data warehouse architecture is a method of defining the overall architecture of data communication processing and presentation that exists for end-clients computing within the enterprise.
   - Each data warehouse is different, but all are characterized by standard vital components.
   - There are 3 approaches for constructing Data Warehouse layers: Single Tier, Two-tier and Three tier.

   a. Single-tier architecture:
      - The objective of a single layer is to minimize the amount of data stored.
      - This goal is to remove data redundancy.
      - This architecture is not frequently used in practice.
   b. Two-tier architecture:
      - Two-layer architecture is one of the Data Warehouse layers which separates physically available sources and data warehouses.
      - This architecture is not expandable and also not supporting a large number of end-users.
      - It also has connectivity problems because of network limitations.
   c. Three-Tier Data Warehouse Architecture:
      - This is the most widely used Architecture of a Data Warehouse.
      - It consists of the Top, Middle and Bottom Tier.
      i. Bottom Tier:
         - The database of the Datawarehouse serves as the bottom tier.
         - It is usually a relational database system.
         - Data is cleansed, transformed, and loaded into this layer using back-end tools.
      ii. Middle Tier:
         - The middle tier in the Data warehouse is an OLAP server which is implemented using either ROLAP or MOLAP model.
         - For a user, this application tier presents an abstracted view of the database.
         - This layer also acts as a mediator between the end-user and the database.
      iii. Top-Tier:

- The top tier is a front-end client layer. The top tier is the tools and API that you connect and get data out from the data warehouse.
- It could be Query tools, reporting tools, managed query tools, Analysis tools and Data mining tools.

5. Steps to build Data Warehouse:
   a. Gathering business requirements
   b. Identifying the necessary sources
   c. Identifying the facts
   d. Defining the dimensions
   e. Defining the attributes
   f. Redefine the dimensions and attributes if required
   g. Organize the Attribute hierarchy
   h. Define Relationships
   i. Assign unique Identifiers

6. Types of Data Warehousing:
   - Enterprise Data Warehousing
   - Operational Data Store
   - Data Mart

7. VLDB (Very Large Database):
   - It is a database that contains a very large amount of data so much that it can require specialized architectural, management, processing and maintenance methodologies.
   - It is a decision support system that is used to serve a large number of users.

8. ER Diagram (Entity-Relationship diagram):
   - It illustrates the interrelationships between the entities in the database.
   - This diagram shows the structure of each table and the links between the tables.

9. ER Modelling:
   - It is a graphical approach to database design.
   - It is a high-level data model that defines data elements and their relationship for a specified software system.
   - ER model is used to represent real-world objects.

10. Dimension Modelling:
    - It is a data structure technique optimized for data storage in a Data warehouse.
    - The purpose of dimensional modelling is to optimize the database for faster retrieval of data.

11. Difference between ER Modeling and Dimensional Modeling:
    - ER modelling will have a logical and physical model but Dimensional modelling will have only a Physical model.
    - ER Modeling is used for normalizing the OLTP database design whereas Dimensional Modeling is used for de-normalizing the ROLAP and MOLAP design.

12. Data Aggregation:
    - It is the process where data is collected and presented in a summarized format for statistical analysis and to effectively achieve business objectives.
    - It is vital to data warehousing as it helps to make decisions based on vast amounts of raw data.

13. Dimension Table:
    - A dimension table is a table that contains attributes of measurements stored in fact tables.
    - This table consists of hierarchies, categories and logic that can be used to traverse in nodes.

14. Fact Table:
    - Fact table contains the measurement of business processes, and it contains foreign keys for the dimension tables.
    - Example – If the business process is the manufacturing of bricks
    - The average number of bricks produced by one person/machine – a measure of the business process

15. Factless Fact Tables:
    - These are the fact table which doesn't contain a numeric fact column in the fact table.

16. Aggregate Tables:
    - These are the tables that contain the existing warehouse data which has been grouped to a certain level of dimensions.
    - It is easier to retrieve data from the aggregated tables than the original table which has more records.

17. Star Schema:
    - It is a type of organizing the tables in such a way that results can be retrieved from the database quickly in the data warehouse environment.

18. Snowflake Schema:
    - Snowflake schema which has a primary dimension table to which one or more dimensions can be joined.
    - The primary dimension table is the only table that can be joined with the fact table.

19. Datamart:
    - It is a specialized version of Data Warehousing.
    - It contains a snapshot of operational data that helps the business people to decide with the analysis of past trends and experiences.
    - This helps to emphasize easy access to relevant information.

20. ODS(Operational Data Store):
    - It is a repository of real-time operational data rather than long term trend data.

21. Metadata:
    - Metadata is defined as data about the data.
    - It contains information like the number of columns used, fixed-width and limited width, ordering of fields and data types of the fields.
    - Examples of metadata for a document might include a collection of information like the author, file size, the date the document was created, and keywords to describe the document.

22. Data Dictionary:
    - It is a centralized repository of metadata.
    - It is a collection of the names, definitions, and attributes for data elements and models. The data in a data dictionary is the metadata about the database.

23. Difference between metadata and data dictionary:
    - Metadata is defined as data about the data.
    - Data dictionary contains information about the project information, graphs, commands and server information.

24. Data Mining:
    - Data Mining is a process of analyzing the data in different dimensions or perspectives and summarizing it into useful information.
    - It can be queried and retrieved the data from the database in its own format.

25. Data Cleaning:
    - The name itself implies that it is a self-explanatory term.
    - Cleaning of Orphan records, Data breaching business rules, Inconsistent data and missing information in a database.

26. OLAP (Online Analytical Processing):
    - It is a system that collects, manages, processes multi-dimensional data for analysis and management purposes.

27. OLTP (On-Line Transaction Processing):
    - It is an application that modifies the data whenever it is received and has a large number of simultaneous users.

28. Difference between OLAP and OLTP:

OLTP	OLAP
Data is from the original data source	Data is from various data sources
Simple queries by users	Complex queries by system
Normalized small database	De-normalized Large Database
Fundamental business tasks	Multi-dimensional business tasks

29. ETL (Extract, Transform and Load):
    - It is a process that extracts the data from different source systems.
    - Then transforms the data (like applying calculations, concatenations, etc.).
    - Finally loads the data into the Data Warehouse system.

30. Difference between Business Intelligence vs Data Warehouse:

Business Intelligence	Data Warehouse
It is a set of tools and methods to analyze data and discover, extract and formulate actionable information that would be useful for business decisions	It is a system for the storage of data from various sources in an orderly manner.
It is a Decision Support System (DSS)	It is a data storage system
Serves at the front end	Serves at the back end

Collects data from the data warehouse for analysis	Collects data from various different sources and organises it for efficient BI analysis
Comprises business reports, charts, graphs, etc.	Comprises of data held in "fact tables" and "dimensions" with business meaning incorporated into them