ASSIGNMENT-2

QUESTION:

Explain RBMS, Data Warehouse, OLTP, OLAP, MySQL and its features.

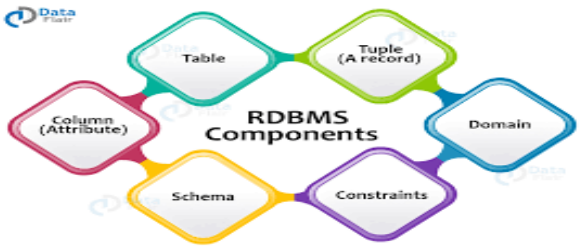
SOLUTION:

1) RDBMS:

- RDBMS stands for Relational Database Management System.
- RDBMS is a software system that enables us to define, create and maintain relational databases.

Terminology:

- It stores the data in the form of tables and the table is known as Entity or Relation
- Named column of a relation is Attribute
- Set of allowable values for one or more attributes is Domain. Ex: Domain(marks)→int
- Row of a relation is known as Tuple.
- Number of attributes is known as Degree.
- Number of tuples is known as Cardinality.
- Number of records at a point of time is known as Instance. Instance is always variable.



Attributes:

Attributes are the properties used to describe an entity.
There are different types of attributes. They are : 1) Single attribute
2) Multi-valued attribute
3) Derived attribute
4) Composite attribute

Entity:

Entities are specific objects or things that are represented in a database.
Ex: Person, book
There are two types of Entities. They are: 1) Strong/ Regular Entity (It has a key attribute).
2) Weak Entity (dependent on strong entity cannot exist on own.

Keys:

There are different types of keys. They are:

1) Primary key
2) Candidate key
3) Foreign Key
4) Alternate Key
5) Composite Key
6) Super Key

Primary Key: It is a column that uniquely identifies rows of table. It is always not null and unique.
Candidate Key: It is minimal subset of super key that can become Primary Key. It is set of one or more columns that can identify record uniquely in table.
Foreign Key: It creates relationship between two or more tables, a foreign key of one table is referred as foreign key in another table.
Alternate Key: A key that can be work as Primary Key if required but right now it is not a Primary Key.
Composite Key: It is a combination of more than one columns of table. It can be candidate key and Primary Key.
Super Key: Primary Key, Unique Key, Alternate Key are subset of Super Keys.

Popular examples of RDBMS include MySQL, PostgreSQL, Oracle Database, Microsoft SQL Server, and SQLite. Each RDBMS may have its own specific features and extensions, but they all adhere to the relational model principles.

Constraints
Constraints are rules and conditions applied to the data stored in tables to ensure data integrity and maintain consistency.
Types of Constraints:

1) **Primary Key Constraint:**
   Ensures that each row in a table is uniquely identified by a specific column or combination of columns
2) **Foreign Key Constraint:**
   Establishes a link between two tables by enforcing referential integrity. The foreign key in one table refers to the primary key in another table.
3) **Unique Constraint:**
   Ensures that all values in a specified column or combination of columns are unique, similar to a primary key constraint, but without the requirement for null values.
4) **Check Constraint:**
   Imposes a condition on the values allowed in a column. It restricts the range of values that can be entered.

5) **Not null Constraint:**
   Ensures that a column cannot contain null values, indicating that every row must have a value in that column.
6) **Default Constraint:**
   Specifies a default value for a column. If a value is not provided during an INSERT operation, the default value is used.

2) Data Warehouse:

Data warehouse is a database that allows organizations to store, manipulate and analyze large amounts of data which is collected from various sources . It mainly helps in making effective decisions by the organizations.

Elements of Data Warehouse:

1) Relational database (to store and manage data).
2) ETL (Extraction, Transformation and Load)
3) Statistical analysis, reporting, and data mining capabilities
4) Client analysis tools for visualizing and presenting data to business users.

Characteristics of Data Warehouse:

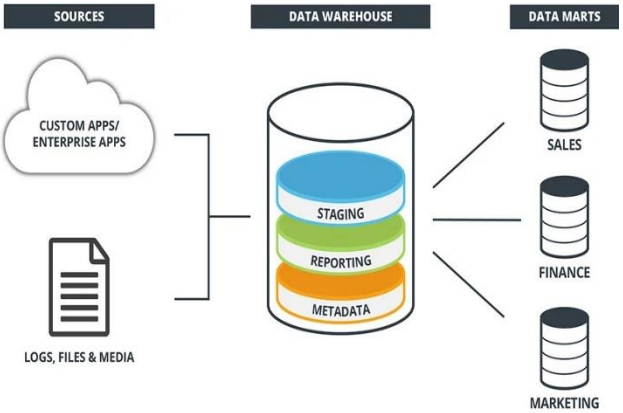
Characteristics of data warehouse include subject-oriented, Integrated, Time Variant and non-volatile.

1) **Subject-Oriented:**
   A data warehouse is also subject-oriented, which means that the data is organized around specific subjects, such as customers, products, or sales. This allows for easy access to the data relevant to a specific subject, as well as the ability to track the data over time.
2) **Integrated:**
   Data from different sources is integrated into the data warehouse in a standardized format. Integration ensures consistency and accuracy, allowing users to analyze data across the organization without worrying about issues that occur.

3) **Time Variant:**
   The data is stored with a time dimension. This allows users for easy access to data for specific time periods, such as last quarter or last year. This makes it possible to track trends and patterns over time. Data warehouses include historical data and support time-based analysis.

4) **Non-Volatile:**

Data in the warehouse is never updated or deleted. This is important because it allows for the preservation of historical data, making it possible to track trends and patterns over time. This ensures data consistency for reporting and analysis.



Components of Data Warehouse:

1) **Data Warehouse Database:**
   This is the central repository that stores integrated, historical, and subject-oriented data from various sources within an organization. The data warehouse database is designed to support complex queries and analysis.
2) **ETL:**
   ETL (Extract, Transform, Load) or ELT (Extract, Load, Transform) tools extract data from source systems, transform it into a suitable format for analysis, and load it into the data warehouse.
3) **Meta Data:**
   Metadata is information about the data stored in the data warehouse. It provides details about the data. Metadata helps users understand the structure of the data warehouse, the origin of data, and how data elements are related.
4) **Access Tools:**
   These tools enable end-users to interact with and retrieve information from the data warehouse. The goal is to provide a user-friendly interface for accessing and exploring the data stored in the data warehouse.

3) OLTP:

1. OLTP refers to Online Transaction Processing.
2. It allows users to access large amount of data.
3. OLTP systems are designed for processing day-to-day, operational transactions that occur in real-time and used by traditional operational systems i.e., RDBMS.

4. These transactions typically involve adding, updating, or retrieving small amounts of data, and they are critical for the daily operations of an organization.
5. OLTP systems ensures to maintain ACID properties i.e., Atomicity, Consistency, Isolation and Durability.
6. OLTP handles the daily transactions and interactions with customers.
7. They ensure that data is accurately and promptly recorded.
8. OLTP systems are mainly used e-commerce, banking and many other applications where real time processing is required.

Example: **Online Book Store**
Consider an online bookstore where customers can browse, search, and purchase books.

1) **Inserts:** When a new customer creates an account or places an order, new data is added to the system.
2) **Updates:** If a customer updates their shipping address or modifies their order, the system needs to reflect those changes immediately.
3) **Retrieves:** When a customer searches for a book or checks the status of their order, the system retrieves and displays the relevant data.

Advantages of OLTP

1) Data Manipulation is easy.
2) The tasks include insertion, updation, or deletion of data which allows users to use efficiently.
3) It supports bigger databases.
4) It ensures to maintain ACID properties.
5) Maintain Data Integrity.
6) Faster query processing.

Disadvantages of OLTP

1) OLTP requires instant update.
2) Data obtained from OLTP is not suitable for analysis.
3) To perform single transaction we have to query multiple tables using joins.

4) OLAP:

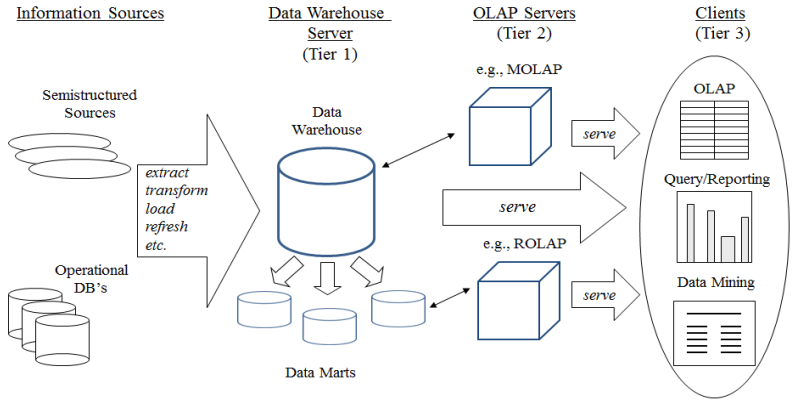
- OLAP refers to Online Analytical Processing .
- It is a category of software tools and technologies used for multidimensional analysis of data in a data warehouse.
- OLAP systems are designed to provide users with a fast and interactive way to analyze and explore large volumes of data from different perspectives.
- OLAP systems organize data into a multidimensional structure, typically in the form of data cubes.

- Dimensions are the categorical attributes along which data is analyzed.
- For example, in a sales data cube, dimensions might include time (e.g., year, quarter, month), geography (e.g., region, country), and product (e.g., category, brand).
- There are two types of OLAP:
  1) **ROLAP(Relational OLAP):** Data is stored in relational databases, and OLAP operations are translated into SQL queries at runtime.
     Example: AP BW and Oracle OLAP.
  2) **MOLAP (Multidimensional OLAP):** Data is stored in a multidimensional cube, and operations are performed directly on this cube.
     Example: Microsoft Analysis Services and IBM Cognos TM1.

OLAP Server:

OLAP servers are components that manage and process OLAP queries. They are responsible for efficiently retrieving and aggregating data from the underlying data warehouse to respond to user queries.

OLAP Architecture:



5) SQL:

Structured Query Language, is designed for managing and manipulating relational database systems. SQL is used for tasks such as querying data, updating data, inserting data, and managing database structures

MySQL:

- MySQL is a popular open-source relational database management system (RDBMS) known for its speed, reliability, and ease of use.
- It is developed by Oracle Corporation.
- MySQL offers a range of features that make it a popular choice for web applications and various other types of software.

Features of MySQL:

1) MySQL is open source software and used by large community of people.

2) MySQL is designed to run on various operating systems, including Windows, Linux, macOS, and others.

3) It is scalable.

4) It offers high performance by using various methods like indexing, caching.

5) MySQL follows ACID properties (Atomicity, Consistency, Isolation, Durability) to ensure data reliability and integrity.

6) It offers good security to data stored in it.

7) It supports multiple data engines.

8) MySQL supports replication, allowing data to be duplicated across multiple database servers.

9) MySQL provides tools for backup and restore operations.

Uses of MySQL:

MySQL is used in following areas:

1) Web applications
2) E commerce platforms
3) Data Warehousing
4) Business Intelligence
5) Telecommunications
6) Health care
7) Financial applications
8) Educational Institutions
9) Mobile applications

Vuthur Sriganga
Hexaware Data Engineering Batch-1