**Data Engineering Assignment-2**

Question: RDMS, Data Warehouse, OLTP, OLAP, SQL and its features

Answer:

**RDBMS:**

RDBMS stands for Relational Database Management System. It is a type of database management system that stores data in a structured format, using tables to organize information. The data in these tables is related to each other based on common fields, creating a relational model.

The components of RDBMs includes tables, rows, columns, primary key, foreign key and many more. Relational Database Management Systems (RDBMS) organize data in tables, each row representing a unique record and each column a specific attribute. Tables utilize primary keys for unique identification and foreign keys to establish relationships between them. RDBMS ensures data integrity through ACID properties (Atomicity, Consistency, Isolation, Durability). One such example of RDBMS is SQL. SQL is the standard language for interacting with RDBMS, allowing users to query, insert, update, and delete data. Normalization is applied to reduce redundancy and enhance data organization. Popular examples include MySQL, PostgreSQL, Oracle, SQL Server, and SQLite.

**Data Warehouse:**

Data warehouse is a collection of data designed to support management decision making by presenting a coherent picture of business conditions at a single point of time. In simple terms we can say that, a data warehouse is a centralized repository that is used for storing, integrating, and managing large volumes of data from various sources within an organization.

It is designed to support decision-making processes by providing a unified, historical, and consistent view of data across the organization. The main purpose of data warehouse is to provide a platform for efficient data analysis and reporting. The data stored in data warehouse are uploaded from operational systems.

The data store contains mainly two types of data.

* Business data
* Business model data

Features of data warehouse:

* Subject-oriented: Instead of organizing the data according to application, the data is organized based on subject. Its main focus is to model and analyze the data for decision makers.
* Integrated: Data in data warehouse is collected from various heterogenous sources like relational databases, flat files, online transaction records etc.
* Time-variant: The time variant feature of a data warehouse refers to its ability to manage and store historical data over time. Unlike operational databases that often focus on the current state of data, a data warehouse is designed to capture, organize data over different points in time.
* Non-volatile: Non-volatile feature states that no updates are allowed. Once the data entered into the warehouse, it cannot be changed.

Application of Data Warehouse:

* Information processing:
* Analytical processing:
* Data mining:

**OLTP:**

OLTP is a methodology where end users are provided access to large amounts of data. It refers to a class of systems and processes that handle the real-time, day-to-day transactional operations of an organization. These transactions typically involve interactions with customers, such as sales, order processing, and other routine business operations. OLTP systems are designed to ensure the consistency and accuracy of data in the operational databases that support an organization's core business functions.

Example of OLTP is Atm.

**OLAP:**

Unlike OLTP (Online Transaction Processing), which focuses on transactional data and day-to-day operations, OLAP is designed for complex queries and data analysis, providing a multidimensional view of historical and aggregated data. In other words, we can say that OLAP is an approach to answer multi-dimensional analytical queries which also encompasses relational reporting and data mining. All the OLAP data is typically stored in a star schema, which is a combination of dimensions and fact tables.

There are mainly two types of OLAP systems.

* ROLAP (Relational OLAP): Data is stored in relational databases, and aggregation is performed on-the-fly during query execution. Examples include Oracle OLAP, SAP BW.
* MOLAP (Multidimensional OLAP): Data is pre-aggregated and stored in a multidimensional cube format. Examples include Microsoft Analysis Services, IBM Cognos TM1.

**SQL:**

SQL (Structured Query Language) is a standard programming language designed for managing and manipulating relational databases. It provides a set of commands for defining, querying, updating, and managing relational database systems. It includes Data Definition Language (DDL) commands for creating, altering, and deleting database structures, as well as Data Manipulation Language (DML) commands for querying, inserting, updating, and deleting data. SQL supports features such as transaction control, data integrity through constraints, and the ability to define views for simplified data access. With its powerful capabilities, SQL is essential for tasks ranging from defining database schemas to retrieving and analyzing data efficiently.

**Features of SQL:**

* Data Definition Language (DDL): SQL provides commands for defining and managing the structure of a database. DDL commands include CREATE, ALTER and DROP.
* Data Manipulation Language (DML): SQL supports commands for interacting with data in a database. DML commands include INSERT, and DELETE.
* Transaction Control: SQL supports transactions, allowing users to group multiple SQL statements into a single, atomic unit. Transaction control commands include COMMIT and ROLLBACK.
* Data Integrity: SQL enforces data integrity through constraints such as PRIMARY KEY, FOREIGN KEY, UNIQUE, and CHECK.
* Normalization: SQL supports normalization, a process of organizing data to reduce redundancy and improve data integrity.
* Security: SQL includes features for managing database security, including user authentication, authorization, and permissions.
* Aggregation Functions: SQL provides aggregate functions such as SUM, AVG, COUNT, MIN, and MAX for performing calculations on sets of data. These functions are useful for obtaining summary information from the database.