Asmita Porwal
Batch-1
Day-2
18/1/2023
Data engineering

1.RDBMS

- A Relational Database Management System (RDBMS) is a type of database management system that organizes data into tables with rows and columns.

- It is based on the relational model of data, which was introduced by E.F. Codd in 1970.

- In an RDBMS, data is stored in a structured manner, and relationships between different pieces of data are maintained.

Here are some key concepts and components of an RDBMS:
1.Tables
2.Row and records
3.Columns
4.Keys
5.Relationships
6.Normalization
7.sql

2.DATA WAREHOUSING

- Data Warehouse is like a special storage space for data that helps managers make decisions.

- It's organized around specific topics, includes information from different sources, keeps track of changes over time, and doesn't easily lose data.

- The main goal is to give managers a clear and complete view of how the business is doing at any given moment, so they can make informed decisions.

Features of Data Warehouse

1. Subject-oriented
2. Integrated
3. Time-variant
4. Nonvolatile

Subject-oriented:

Think of a Data Warehouse like a giant filing cabinet for a business. In a regular office, you might have different drawers for different purposes, like one for customer information, another for sales records, and so on. In a subject-oriented Data Warehouse, it's organized based on what the business cares about (subjects), not just how the information is used day-to-day. For example, instead of having separate folders for customer details in sales and finance, all customer information is neatly organized in one place for decision-makers to easily analyze.

Integrated:

Imagine the business has information stored in different formats and places, like files on a computer, databases, and records from online transactions. Integrated in a Data Warehouse means bringing all this information together in a way that makes sense. It's like translating everything into a common language . This helps avoid confusion and makes it easier for decision-makers to trust and understand the information because it's all in one unified system.

Time-variant:

Think of the Data Warehouse as a big diary for the business. Unlike day-to-day logs, this diary covers a much longer time imagine looking back at the last 5-10 years. It's like flipping through the pages of a company's history book, showing how things have changed over a significant period. Decision-makers can use this historical perspective to understand trends, patterns, and how the business has evolved.

Non-volatile:

Once information goes into the Data Warehouse, it stays there like a permanent record. It's like adding chapters to the business's storybook, but you can't go back and erase or

change what's already written. The warehouse keeps a complete history, and every new piece of information adds to it. This ensures that decision-makers have a reliable and unchanging reference point for understanding the company's past, helping them make decisions based on a solid historical foundation.

3.OLTP

- OLTP databases are all about managing day-to-day transactions, like adding, removing, or updating data.
- They're like the quick and efficient assistants of the data world, providing users with speedy access to large amounts of information.
- These systems work swiftly and intuitively, helping users make sense of data for decision-making.
- The main job of OLTP is to handle transaction-oriented tasks, especially for things like entering and retrieving data.
- It's the go-to method when you need to deal with lots of transactions quickly.

Benefits of OLTP

- It includes making things simpler and more efficient for businesses.
- It cuts down on paperwork and allows for faster and more accurate predictions of revenues and expenses.
- OLTP systems are great at maintaining data integrity and quickly processing queries, even when lots of people are accessing the data.

OLTP Challenges

- It demands instant updates, and the data it provides isn't always suitable for in-depth analysis.
- Even with a well-organized structure, performing a simple transaction might require querying multiple tables through joins, which can be a bit complex.

OLAP (Online Analytical Processing)

- OLAP is an approach for handling complex analytical queries, including relational reporting and data mining.

- An OLAP cube is a data array understood in terms of dimensions, allowing users to analyze data quickly and interactively.

- Dimensions are the categories like product, time, city, and scenario.
- For instance, financial data can be summarized by product, time, and city.
- The cube cell holds values for a specific combination, like Maruthi sales in Mumbai for May 2006.
- Data in the Data Warehouse is arranged hierarchically into dimensions and facts tables.
- The Star Schema, commonly used in OLAP, combines dimensions and fact tables, with the fact table containing business process measurements.
- Fact table consists of the measurements metrics or facts of a business process.
- It is located at the center of a star schema surrounded by dimension tables.
- In the above example, it holding sales by item, by time, by branch and by location.
- Dimension tables contain descriptive attributes (or fields) which are typically textual fields or discrete numbers that behave like text.
- These attributes are designed to serve two critical purposes:
  -query constraining/filtering
  -query result set labeling.

OLAP Server

- The OLAP Server receives data from the data warehouse and represents it in a user-friendly format, providing analytical functionality for Decision Support Systems (DSS).

- OLAP Server generally performs data analysis in two forms.
  -ROLAP(Relational OLAP)
  -MOLAP(Multi-dimensional OLAP)

ROLAP (Relational OLAP)

- ROLAP performs dynamic multi-dimensional analysis using data stored in a relational database.
- Data processing may occur within the database system, a mid-tier server, or the client.

MOLAP (Multi-dimensional OLAP)

- MOLAP facilitates "slice and dice" analysis, organizing data in a cube structure.

- MOLAP often uses a multi-cube approach, where small, pre-calculated cubes contribute to a larger hypercube.
- Slicing involves viewing a specific "slice" of the cube, like all sales of a particular product.
- Dicing involves focusing on a subset, like all sales in a specific city.

4.SQL (Structured Query Language)

SQL (Structured Query Language) is a domain-specific language designed for managing and manipulating relational databases**.**

Key Features:

Data Query Language (DQL):
SQL allows users to query databases using the SELECT statement to retrieve specific data from one or more tables**.**

Data Definition Language (DDL):
DDL statements like CREATE, ALTER, and DROP are used to define, modify, and remove database structures such as tables, indexes, and views.

Data Manipulation Language (DML):
DML statements like INSERT, UPDATE, and DELETE are employed to add, modify, or remove data in the database.

Data Control Language (DCL):
DCL statements like GRANT and REVOKE manage access rights and permissions to users, controlling who can perform specific actions on the database**.**

Transaction Control Language (TCL):
TCL statements like COMMIT and ROLLBACK ensure the integrity of transactions, allowing changes to be permanently saved or rolled back in case of errors.

Relational Database Management System (RDBMS) Support:
SQL is used with RDBMS systems, such as MySQL, PostgreSQL, Oracle, SQL Server, and SQLite, providing a standardized way to interact with these databases.

Data Integrity:
SQL includes features like constraints (e.g., PRIMARY KEY, FOREIGN KEY) to enforce data integrity, ensuring that data adheres to predefined rules.

Normalization:
SQL supports normalization techniques to organize data efficiently, minimizing redundancy and improving database structure.

Joins:
SQL allows users to combine data from multiple tables using JOIN operations, facilitating the retrieval of related information.

Indexing:
Indexing in SQL improves query performance by creating data structures that enable faster data retrieval.

Views:
SQL enables the creation of virtual tables known as views, allowing users to present data in a specific way without modifying the underlying tables.

Stored Procedures and Triggers:
SQL supports the creation of stored procedures and triggers, which are precompiled sets of SQL statements that can be executed on demand or in response to certain events.

Security:
SQL provides security mechanisms to control access to databases, including user authentication and authorization.

Concurrency Control:

SQL includes features for managing concurrent access to data, preventing conflicts when multiple users attempt to modify the same data simultaneously.