Name: Abhishek Kanoujia

DATA ENGINEERING BATCH 1

DAY 2 ASSIGNMENT

Data Warehouse:

A database used for data reporting and analysis.

Data uploaded from operational systems.

Contains business data and business data models.

Business data extracted from operational databases and external sources.

Uses ETL (Extract Transform Load) process, including staging, integration, and access layers (Data Marts).

DSS (Decision Support System) Data vs Operational Data:

DSS data differs from operational data in time span, granularity, and dimensionality.

Business data in DSS represents a snapshot of the company's situation.

Components of DSS:

Data store, extraction, filtering components.

End-user query and presentation tools.

Data Marts:

Subsets of the data warehouse oriented to specific business lines or teams.

OLAP (Online Analytical Processing):

OLAP is an approach to answer multi-dimensional analytical queries.

OLAP cube represents data in terms of dimensions for fast, interactive analysis.

OLAP data typically stored in a Star Schema with dimensions and fact tables.

Star Schema:

Fact tables hold measurements or facts of a business process.

Dimension tables contain descriptive attributes for query constraining and result set labeling.

OLAP Architecture:

OLAP Server receives data from the data warehouse and represents it in a user-understandable format.

ROLAP (Relational OLAP) and MOLAP (Multi-dimensional OLAP) are two forms of OLAP.

ROLAP (Relational OLAP):

Performs dynamic multi-dimensional analysis on data stored in a relational database.

Data processing may occur within the database system, mid-tier server, or the client.

MOLAP (Multi-dimensional OLAP):

Provides multi-dimensional analysis using cube structures.

Slicing and dicing enable users to analyze data along different dimensions.

Three Kinds of Data Warehouse Applications:

Information Processing:

Supports querying, basic statistical analysis, and reporting.

Analytical Processing:

Involves multidimensional analysis with basic OLAP operations.

Data Mining:

Focuses on knowledge discovery from hidden patterns.

Supports associations, constructing analytical models, classification, prediction, and visualization.

RDBMS (Relational Database Management System):

A Relational Database Management System (RDBMS) is a type of database management system that stores data in the form of tables and relationships between those tables. Key features of RDBMS include:

Tables: Data is organized into tables with rows and columns. Each table represents an entity, and each row represents a record.

Relationships: Tables can be related to each other using keys, creating relationships between entities. Common types of relationships include one-to-one, one-to-many, and many-to-many.

Integrity Constraints: RDBMS enforces integrity constraints to maintain the accuracy and consistency of data. These include primary keys, foreign keys, unique constraints, and check constraints.

Normalization: The process of organizing data to reduce redundancy and improve data integrity. Normalization helps avoid data anomalies and ensures efficient storage.

ACID Properties: RDBMS follows ACID (Atomicity, Consistency, Isolation, Durability) properties to ensure reliable and secure transactions.

SQL (Structured Query Language):

SQL is a domain-specific language used to manage and manipulate relational databases. It is used for tasks such as querying data, updating data, inserting data, and managing database structures. Key features of SQL include:

Data Query Language (DQL): Allows users to retrieve data from the database using SELECT statements.

Data Definition Language (DDL): Involves the creation, alteration, and deletion of database structures. Statements include CREATE, ALTER, and DROP.

Data Manipulation Language (DML): Involves the manipulation of data stored in the database. Statements include INSERT, UPDATE, and DELETE.

Data Control Language (DCL): Manages access to data within the database. Statements include GRANT and REVOKE.

Transaction Control Language (TCL): Manages transactions within the database. Statements include COMMIT, ROLLBACK, and SAVEPOINT.

Key Features of SQL:

Declarative Language: SQL is a declarative language, meaning users specify the desired outcome, and the database management system determines the best way to achieve it.

Data Integrity: SQL enforces constraints to maintain data accuracy and consistency, ensuring that data meets certain criteria.

Portability: SQL is a standardized language, allowing users to interact with different database systems using a common set of commands.

Scalability: SQL databases can handle large amounts of data and are scalable to accommodate growth.

Security: SQL provides security features, including user authentication, access control, and encryption, to protect data from unauthorized access.

Transactions: SQL supports transactions, allowing users to group multiple SQL statements into a single, atomic operation.

SQL and RDBMS together provide a powerful and widely used framework for managing and querying structured data in various applications and industries.