Data Engineering Batch 1 **Day 2: 18/01/2024**

Name: Pradip Bochare

Statement: RDBMS, OLAP, OLTP, SQL

DSS architectural styles:

- OLTP (Online Transaction Processing)
  used by traditional operational systems (RDBMS).
- OLAP (Online Analytical Processing)
  used by Data Warehouse.

OLTP (Online Transaction Processing):

- **OLTP** is a methodology to provide end users with access to large amounts of data
- It works in an intuitive and rapid manner to assist with deductions based on investigative reasoning.
- **OLTP** refers to a class of systems that facilitate and manage transaction-oriented applications, typically for data entry and retrieval transaction processing.

❖ Benefits of OLTP:

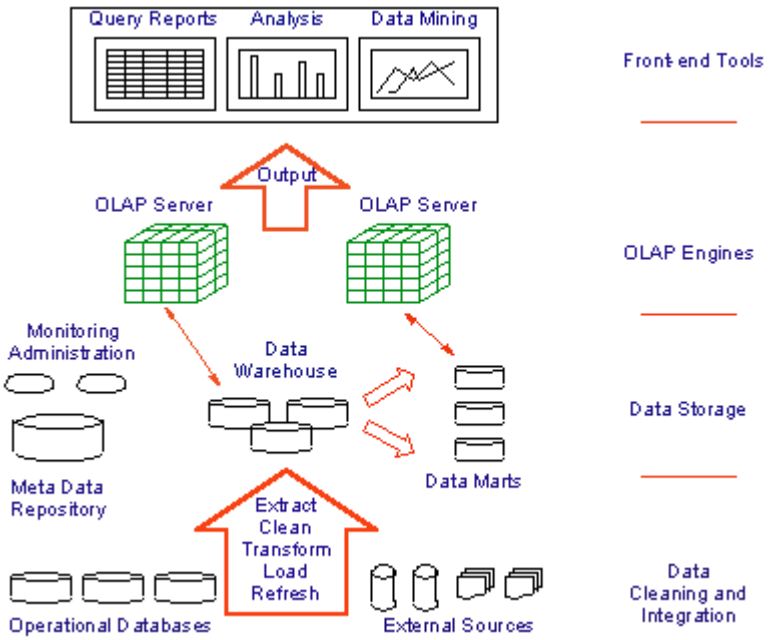
1. Simplicity & Efficiency:
   Reduced paper trails and the faster and more accurate forecasts for revenues and expenses are both examples of how OLTP makes things simpler for businesses.
2. OLTP systems maintain data integrity and they also provide fast query processing in environments having multiple access.

❖ Pitfalls of OLTP:

1. OLTP requires instant update.
2. The data what we get from OLTP is not suitable for data analysis.
3. To perform one simple transaction even with the normalized structure, we need to query multiple tables by using joins.

OLAP (Online Analytical Processing) :

- OLAP is an approach to answer multi-dimensional analytical queries which also encompasses relational reporting and data mining.
- An **OLAP cube** is an array of data that is understood in terms of its 0 or more dimensions which enables the users to gain insight into their data in a fast, interactive, easy-to-use manner.
- The data in Data Warehouse is arranged in the form of hierarchical groups often called dimensions and into facts tables and aggregate facts**.**
- OLAP data is typically stored in a Star Schema**.** Which is a combination of dimensions and fact tables.



OLAP Architecture

OLAP Server:

- OLAP Server receives the data from data warehouse by which it represents the data in a user understandable format which actually supply analytical functionality for the DSS system.
- OLAP Server generally performs data analysis in two forms.

  ROLAP (Relational OLAP)

  MOLAP (Multi-dimensional OLAP)

ROLAP (Relational OLAP):

- It is a form of OLAP that performs dynamic multi- dimensional analysis of data stored in a **relational database** rather than in a multi-dimensional database (which is usually considered the OLAP standard).
- Data processing may take place within the database system, a mid-tier server, or the client.
- In two-tier architecture, the user submits a Structured **Query Language (SQL)** query to the database and receives back the requested data.

MOLAP (Multi-dimensional OLAP):

- It is a form of OLAP that helps the user to **"slice and dice"** information, providing multi-dimensional analysis of data by putting data in a cube structure.
- Most MOLAP products use a multi-cube approach in which a series of small, dense, pre-calculated cubes make a **hypercube.**

warehouse Applications:

- Information processing
    - supports querying, basic statistical analysis, and reporting using crosstabs, tables, charts and graphs.
- Analytical processing
    - multidimensional analysis of data warehouse data.
    - supports basic OLAP operations, slice-dice, drilling, pivoting.
- Data mining
    - knowledge discovery from hidden patterns.
    - supports associations, constructing analytical models, performing classification and prediction, and presenting the mining results using visualization tools.

RDBMS:

- A database management system (DBMS) defines, creates, and maintains a database.
- RDBMS data is structured in database tables, fields and records.

Different Types of Databases:

- **Relational Databases**
- **Operational Databases**
- **Database Warehouses**
- **Distributed Databases**
- **End-User Databases**

MySQL:

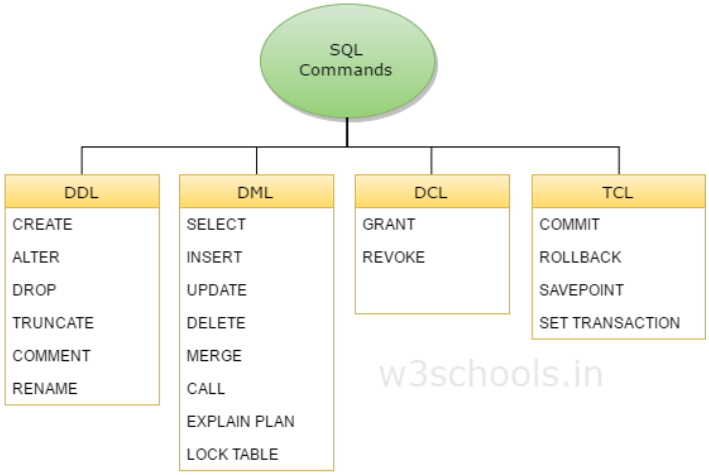
- MySQL is the world's most popular open-source database software, with over 100 million copies of its software downloaded or distributed throughout its history.
- MySQL is RDBMS which runs a server, providing multi-user access to a number of databases.
- With its superior speed, reliability, and ease of use, MySQL has become the preferred choice for IT in all sectors or domains.

Features of MySQL:

- MySQL is written in C and C++ and its SQL parser is written in yacc (**Yet Another Compiler Compiler).**
- MySQL uses only just under 1 MB of RAM on your laptop while Oracle 9i installation uses 128 MB
- MySQL is great for database enabled websites while Oracle is made for enterprises.
- MySQL is portable.
- MySQL default port number is 3306.

SQL Language statements:

- Data Definition Language (DDL) for defining the database structure and controlling access to the data.
- Data Manipulation Language (DML) for retrieving and updating data.
- Data Control Language (DCL) concerns with rights, permissions and other controls of the database system.



SQL data types:

- Numeric data types
  - TINYINT, SMALLINT, MEDIUMINT,
  - INT, BIGINT
  - FLOAT (display_length, decimals)
  - DOUBLE (display_length, decimals)
  - DECIMAL (display_length, decimals)
- Date and time types
  - DATE
    - format is YYYY-MM-DD
  - DATETIME
    - format YYYY-MM-DD HH:MM:SS
  - TIMESTAMP
    - format YYYYMMDDHHMMSS
  - TIME
    - format HH:MM:SS
  - YEAR
    - default length is 4

- String types
  - CHAR
    - fixed length string, e.g., CHAR (20)
  - VARCHAR
    - variable length string, e.g., VARCHAR (20)
  - BLOB, TINYBLOB, MEDIUMBLOB, LONGBLOB
    - same as TEXT, TINYTEXT ...
  - ENUM
    - list of items from which value is selected

SQL commands SHOW, USE:

- SHOW
  - Display databases or tables in current database;
  - show databases;
  - show tables;
- USE
  - Specify which database to use
  - use bookstore;

CREATE Command:

```
CREATE  TABLE   table_name
(
    Column_name1  column_type1,
    Column_name2  column_type2,
    …..
    Column_nameN  column_typeN,
    PRIMARY KEY (column_name1)
) ;
```

DROP & INSERT Commands:

- To delete databases and tables use the DROP command
- Examples

```
DROP DATABASE db_name;

DROP TABLE table_name;
```

- Inserting rows into a table using INSERT command

```
INSERT INTO table_name
    ( col_1, col_2, ..., col_N)
VALUES
    ( val_1, val_2, ..., val_N);
```

SELECT Command:

- Simplest form: select all columns

```
SELECT * FROM table_name;
```

- Select specified columns

```
SELECT column_list FROM table_name;
```

- Conditional selection of rows

```
SELECT  column_list  FROM  table_name  WHERE condition;
```

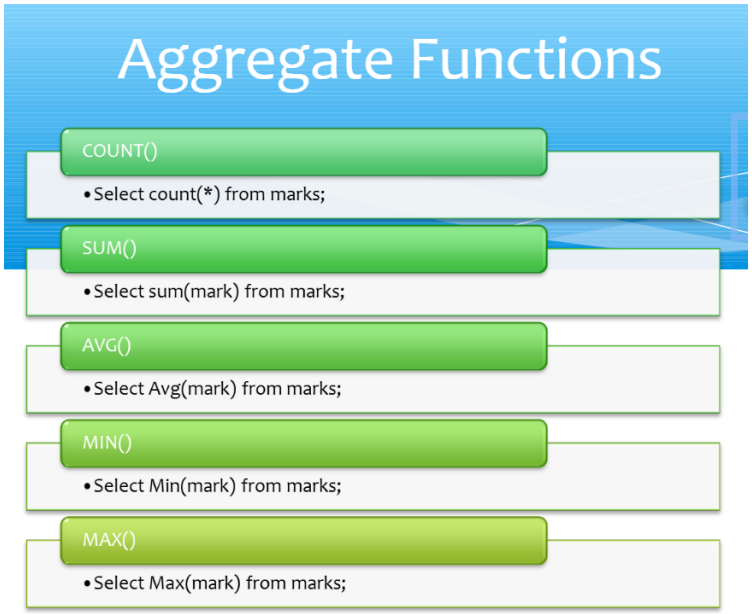
UPDATE Command:

```
UPDATE table_name

SET col_1 = 'new_value1',

..., col_n = 'new_value2'

WHERE condition;
```



Cloud SQL:

- A fully managed relational mySQL databases on cloud hosted by Google platform.
- Runs on Google infrastructure
- Google + MySQL
- Cloud SQL provides a database infrastructure for applications running anywhere
- WordPress sites, e-commerce applications, CRM tools, or any other application that is compatible with MySQL.
- It doesn't require any software installation.
- Security: Cloud SQL customer's data is encrypted i.e. Every Cloud SQL instance includes a network firewall.