BYTEWISE LIMITED

Data Engineering Track

Task: Week – 2 (First Month)

Task No: 3

Task Date: 21-03-2023

Internee Name: Umer Farooq

Mentor Name: Muhammad Ahtisham

Task Details:

This task includes the following:

1. What is SQL?
2. What is DDL?
3. What is DML?
4. What us DQL?

1.What is SQL:

SQL is a Querying language which is **designed for accessing and manipulating** data/information in data repositories (databases). Querying language is an essential part of learning Data engineering and it is one of the most crucial skill that a Data Enthusiast must have.

- SQL is a Structured Query Language.
- SQL is used **to access and manipulate** information from (but not explicitly) databases.
- Using SQL, **we can write a set of instructions**, such as
    - Insert, Update, & Delete,
    - Create new schemas,
    - Write a Stored Procedure **(which means you can write your own set of instructions & call them for later use)**.

We need SQL **because relational databases are an essential part of modern applications and data management**. They provide a **structured way to store and organize data**, making it easier to access, manage, and analyze data. SQL is the standard language used to interact with these databases and retrieve the data we need.

In Data Engineering, **SQL plays a crucial role in the design, implementation, and maintenance of data pipelines**. Data Engineers use SQL to write queries that extract data from different sources, transform and manipulate the data, and load it into a target database or data warehouse. SQL is also used to create and manage the schema of the database, define relationships between tables, and optimize queries for performance. In short, SQL is a fundamental tool for Data Engineers to work with data, and it is an essential skill for anyone working with databases or data pipelines.

Advantages:

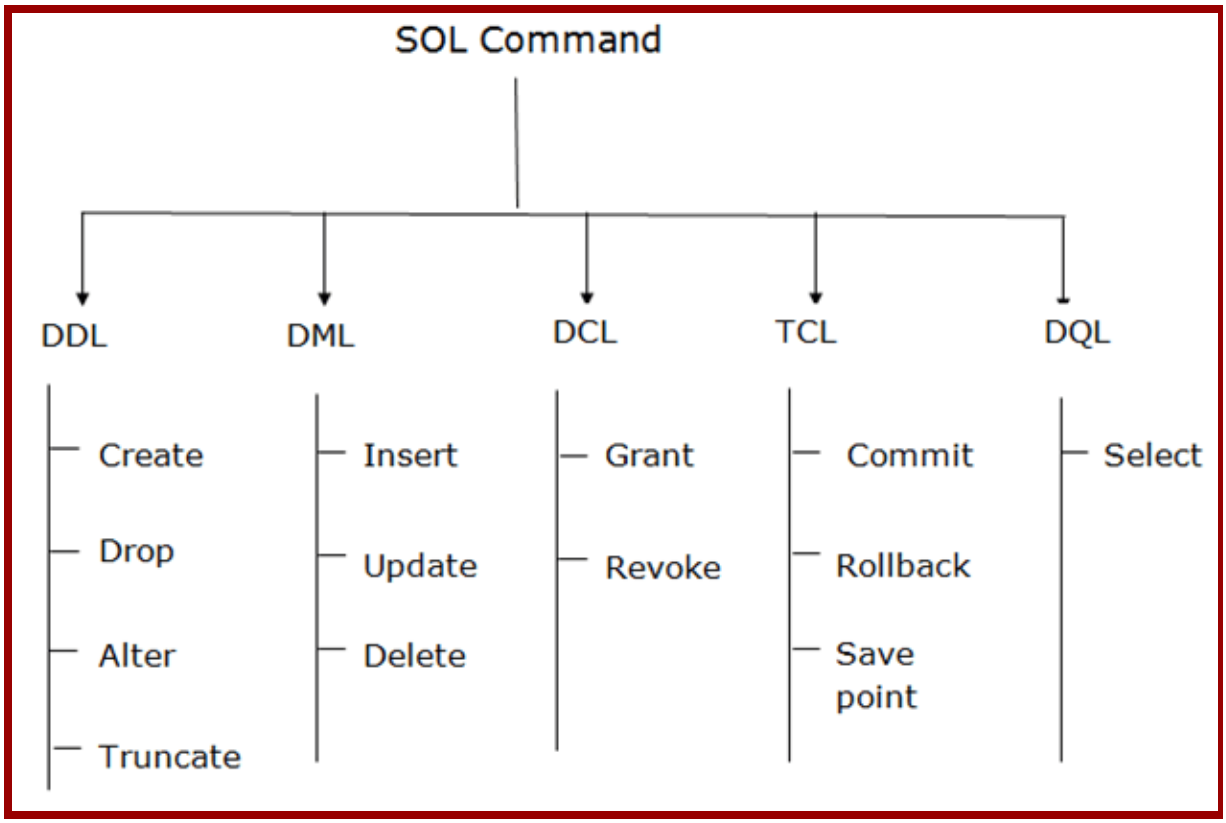
- SQL is portable & platform-independent.
- SQL can be used for querying data in a wide variety of databases & data repositories. **(Although each vendor may have variations & special extensions)**
- SQL has a simple syntax that is similar to the English language.
- Its syntax allows developers to write programs with fewer lines of code using basic keywords such as select, insert, into, & update.
- SQL can retrieve large amounts of **data quickly & effectively**.

- **Runs on an interpreter system** (which means code can be executed as soon as it is written, making prototyping quick & easy).
    - Interpreter system: Directly executes instructions written in a programming/querying/scripting language without previously converting to an object code or machine code.

SQL Commands:

- SQL **commands are instructions**. It is **used to communicate** with the database. It is also used to perform specific tasks, functions, and data queries.
- SQL can perform various tasks like creating a table, adding data to tables, dropping the table, modifying the table, set permission for users.

Types of SQL Commands:



There are five types of SQL commands:

1. DDL,
2. DML,
3. DCL,
4. TCL, and
5. DQL.

2.DDL:

DDL (Data Definition Language) is a set of SQL commands used to define and manage the structure of a database:

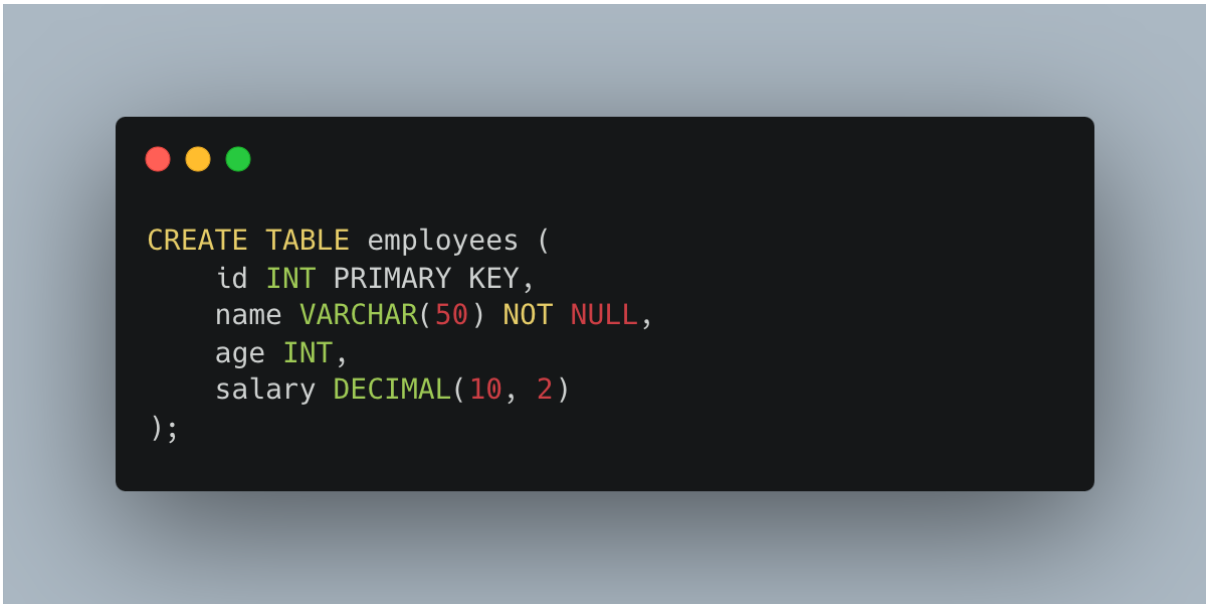
- including tables,
- indexes,
- constraints, and
- other database objects.

All the commands of **DDL are auto-committed** which means it **permanently saves all the changes** in the database.

Some examples of DDL commands are:

a. **CREATE:** Used to create a new database object, such as a table, index, or view.

   Example:



```
CREATE TABLE employees (
    id INT PRIMARY KEY,
    name VARCHAR(50) NOT NULL,
    age INT,
    salary DECIMAL(10, 2)
);
```

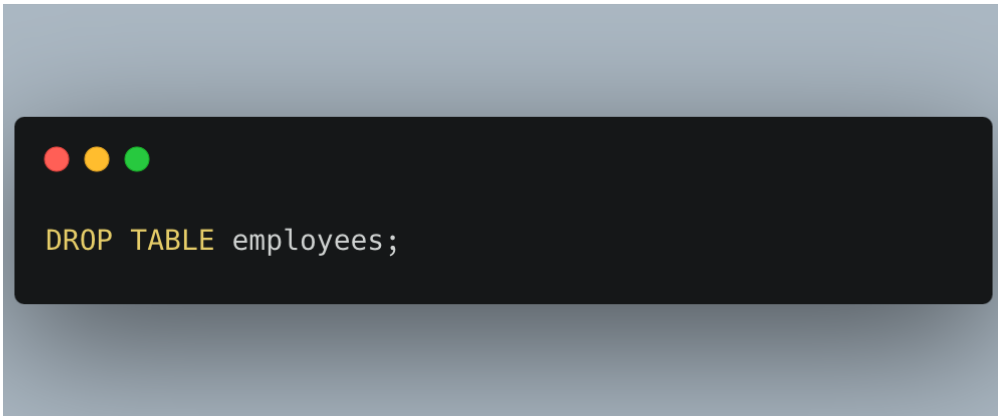
b. **ALTER:** Used to modify the structure of an existing database object, such as a table or column.

Example:

```
ALTER TABLE employees ADD department VARCHAR(50);
```

c. **DROP:** Used to remove an existing database object, such as a table, index, or view.

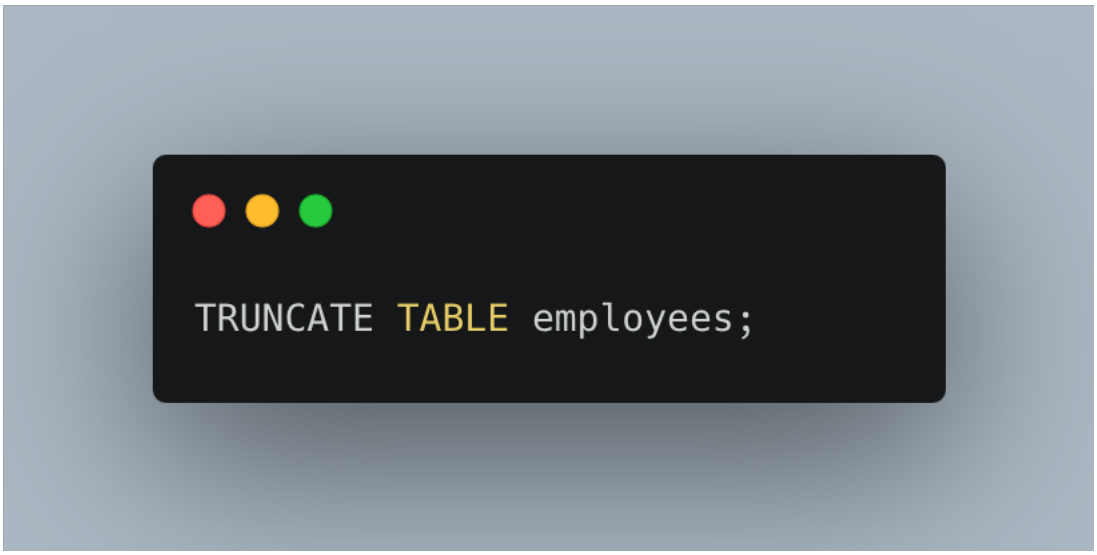
Example:



```
DROP TABLE employees;
```

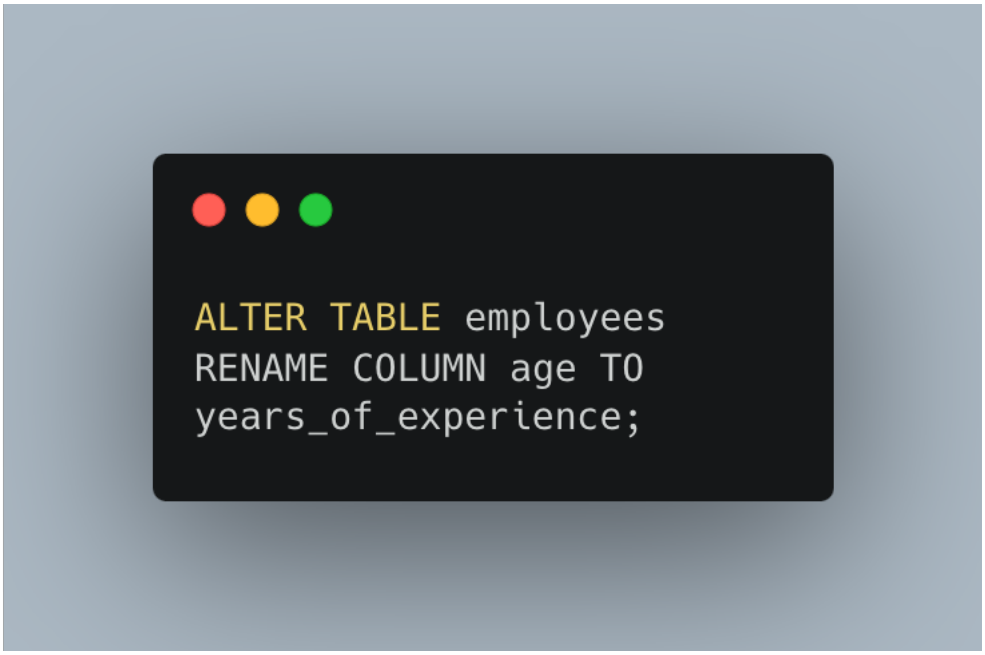
d. **Truncate:** Used to delete all rows from a table, effectively resetting the table.

Example:



e. **Rename:** Used to rename an existing database object, such as a table or column.

Example:



3.DML:

DML (Data Manipulation Language) is a set of SQL commands used to **manipulate the data stored in a database**. DML commands are used to insert, update, and delete data from tables.

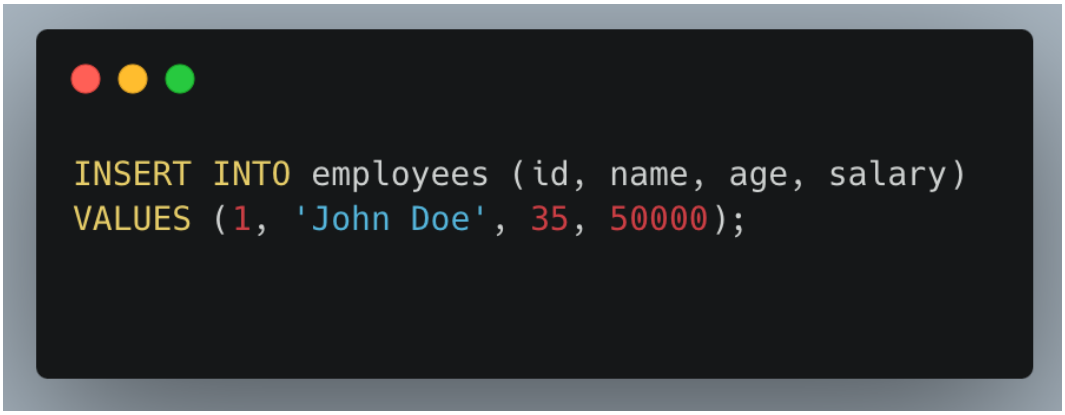
It is responsible for all forms of changes in the database.

The command of **DML is not auto-committed** which means it can't permanently save all the changes in the database. **They can be rollback**.

Some examples of DML commands are:

a. **INSERT:** Used to add new rows of data to a table.

   Example:



```
INSERT INTO employees (id, name, age, salary)
VALUES (1, 'John Doe', 35, 50000);
```

b. **UPDATE:** Used to modify the values of one or more columns in existing rows.

   Example:

```
UPDATE employees
SET salary = 55000
WHERE id = 1;
```

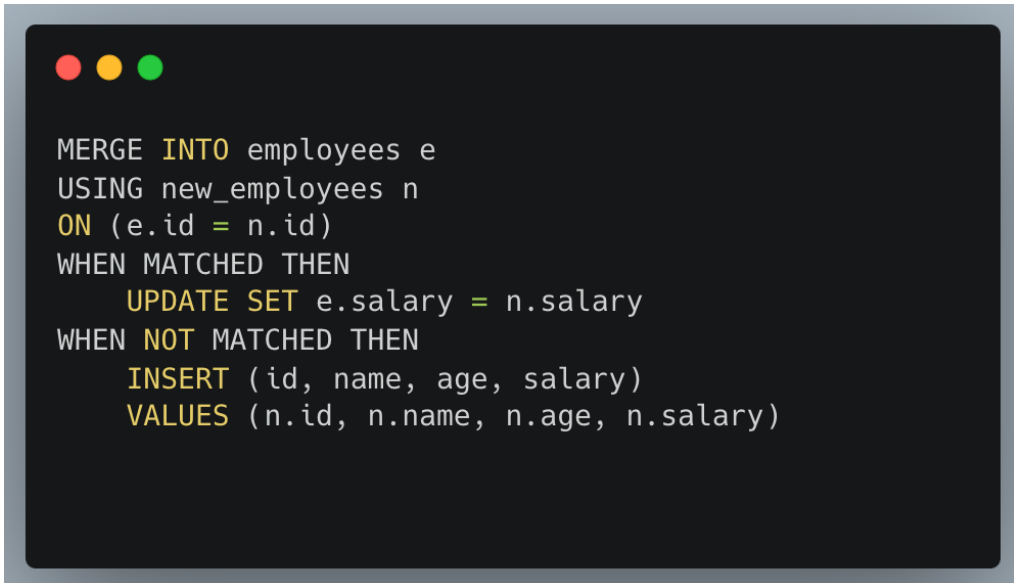
c. **DELETE:** Used to remove one or more rows of data from a table.

Example:

```
DELETE FROM employees
WHERE id = 1;
```

d. **MERGE:** Used to combine INSERT, UPDATE, and DELETE operations into a single command.

Example:



```
MERGE INTO employees e
USING new_employees n
ON (e.id = n.id)
WHEN MATCHED THEN
    UPDATE SET e.salary = n.salary
WHEN NOT MATCHED THEN
    INSERT (id, name, age, salary)
    VALUES (n.id, n.name, n.age, n.salary)
```

This command performs a "upsert" operation, where rows are inserted or updated in the "employees" table based on data from the "new_employees" table. If a matching row already exists in the "employees" table, the "salary" column is updated with the value from the "new_employees" table. If a matching row does not exist, a new row is inserted with values from the "new_employees" table.

These are some examples of DML commands used in SQL to manipulate data stored in a database.

4.DQL:

DQL (Data Query Language) is a subset of SQL **used to retrieve data from a database**. DQL commands do not modify data or the structure of a database; they only retrieve data from existing tables.

Some examples of DQL commands are:

a. **SELECT:** Used to retrieve data from one or more tables.

Example:

b. **DISTINCT:** Used to retrieve unique values of a column.

Example:

c. **WHERE:** Used to filter data based on a condition.

Example:

```
SELECT *
FROM employees
WHERE salary > 40000;
```

d. **GROUP BY:** Used to group data based on one or more columns.

Example:

```
SELECT department, AVG(salary) FROM employees GROUP
BY department;
```

These are some examples of DQL commands used in SQL to retrieve data from a database.