Bigdata Assignment 3.3

Explain the below concepts with an example in brief.

- **Hive Data Definitions** - Hive DDL are used to define the structure of the data structure by CREATE, DROP and ALTER of databases, tables, views, partitions and buckets.

1. CREATE DATABASE/SCHEMA, TABLE, VIEW, FUNCTION, INDEX – For creating databases , views , tables , etc ;

2. DROP DATABASE/SCHEMA, TABLE, VIEW, INDEX – For dropping table , database , etc.

3. TRUNCATE TABLE – To truncate – a table

4. ALTER DATABASE/SCHEMA, TABLE, VIEW – To modify the database , schema , table

5. MSCK REPAIR TABLE (or ALTER TABLE RECOVER PARTITIONS)

6. SHOW DATABASES/SCHEMAS, TABLES, TBLPROPERTIES, VIEWS, PARTITIONS, FUNCTIONS, INDEX[ES], COLUMNS- Shows existing tables/views/databases

7. DESCRIBE DATABASE/SCHEMA, table_name, view_name – Describes the structure of database/schema.

Example -
CREATE DATABASE acadgild_db;
This will create a database with name acadgild_db. If the DB already exists it will throw error. To avid this error IF NOT EXISTS can be used with CREATE statement.

DROP TABLE student_db;
This will drop the table named student_db.

- **Hive Data Manipulations**

Hive DML provides means to manipulate data by inserting data, sorting, merging tables , updating data etc. It is used to manipulate and transform the data.

1. LOAD DATA [LOCAL] INPATH 'filepath' [OVERWRITE] INTO TABLE tablename [PARTITION (partcol1=val1, partcol2=val2 ...)] - Load data into the hive table from the file system

2. INSERT OVERWRITE TABLE tablename1 [PARTITION (partcol1=val1, partcol2=val2 ...) [IF NOT EXISTS]] select_statement1 FROM from_statement; - insert thed ata into an existing table

3. UPDATE tablename SET column = value [, column = value ...] [WHERE expression]

4. DELETE FROM tablename [WHERE expression] – Delete data which satisfies certain condition

5. MERGE INTO <target table> AS T USING <source expression/table> AS S ON <boolean expression1>WHEN MATCHED [AND <boolean expression2>] THEN UPDATE SET <set clause list>

Examples:
UPDATE my_table SET sport='cricket' WHERE city='BBSR';
This will change the value of the existing data based on condition.

INSERT INTO TABLE my_table VALUES ('Adam','32'),   ('John', 28);
This will insert data uinto table named my_table.

- **HiveQL Manipulations**

    HQL is a SQL like language used in Hive. It reduces the complexity of MapReduce by making it similar to SQL. It reuses familiar concepts of RDBMS.Hive supports different file formats like TEXTFILE, SEQUENCEFILE, ORC, RC etc.Along with SQL like queries HQL introduces concepts of optimizing data storage, manipulation of data based on table definition etc.

Examples:

CREATE TABLE IF NOT EXISTS my_tbl (name String, age int)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS ORC;

Here it creates table like in SQL but additionally it also tells the type of data. So that when the data is stored it the table's location, it can read the data based on this criteria. It also defines the storing format (ORC) which makes it optimized.