Experiment 4

Shashwat Shah
60004220126
TY BTECH COMPS B

Aim: Execute Hive commands to load, insert, retrieve, update or delete data in tables.

Theory: Hive is a data warehouse infrastructure tool to process the structured data in hadoop. It resides on top of hadoop to summarise Big Data, and makes querying and analysing easy.
Initially Hive was developed by facebook. Later the Apache software foundation took it up and developed it further as an open source under the name Apache Hive.

Commands:

1) Show databases
   Show all the existing databases.

2) create database student
   creates a database named student.

3) Use student
   Use the database student for executing queries.

4) create table student
   (id int, name varchar(20));
   Partitioned by (load-date date)
   clustered by (id) into 3 buckets
   Stored as ORC TBLPROPERTIES ('transactional' = 'true');

Shashwat Shah
60004220126

5) Insert into student partition (load_date = '2023-03-09') values (101, 'xyz', 123456);

6) Select * from student
To display contents of table.

7) update student set name = 'xyz' where id = 101
update the values in the table

8) delete from studen where name = 'xyz';
Delete the entry from that table

9) alter table student rename to stud;
- rename the table

10) Relational operators
select * from stud where CGPA < 9.0;

11) Joins in hive
select * from customers c join orders o on (c.id = o.customer_id)

12) Views in hive
create view if not exists customers_vw as
Select * from customers where address = 'Mumba

13) Drop the view
drop view if exists customer_vw;

Conclusion: We have successfully executed HIVE queries using HQL in cmdor.