

Ex.no:	EXECUTING HIVE QL QUERIES
Date:	

Aim:

To execute data definition and data manipulation queries in hive using the hive query language.

Procedure:

1. Start Hadoop using the commands **start-dfs.sh** and **start-yarn.sh**.

```
karthikeyan@karthikeyan-VirtualBox:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [karthikeyan-VirtualBox]
karthikeyan@karthikeyan-VirtualBox:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
karthikeyan@karthikeyan-VirtualBox:~$ jps
4738 SecondaryNameNode
5058 NodeManager
4949 ResourceManager
4439 NameNode
4557 DataNode
5215 Jps
karthikeyan@karthikeyan-VirtualBox:~$
```

2. Navigate to the bin directory of hive and start the hive shell using the command **hive**.

```
karthikeyan@karthikeyan-VirtualBox:~$ cd $HIVE_HOME/bin
karthikeyan@karthikeyan-VirtualBox:~/apache-hive-3.1.2-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/karthikeyan/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/karthikeyan/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.30.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = 129cb439-8cdc-4d2a-a737-e4bdda5cf484

Logging initialized using configuration in jar:file:/home/karthikeyan/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e . spark, tez) or using Hive 1.X releases.
Hive Session ID = ce8976e3-54d9-44a8-91ab-3bb59822e8a2
hive>
```

3. Create a database using the create database command and view the list of databases using show databases command.

Queries:

create database college;
show databases;

```
hive> create database college;  
OK  
Time taken: 1.277 seconds  
hive> show databases;  
OK  
college  
default  
Time taken: 0.259 seconds, Fetched: 2 row(s)
```

4. Set the database created in the previous step as the working database using use command.

Queries:

use college;

```
hive> use college;  
OK  
Time taken: 0.116 seconds  
hive> 
```

5. Create a table using create table command and view the table properties using describe command.

Queries:

create table student (name string, regno int, cgpa float, address string) partitioned by (department string);

describe extended student;

```

hive> create table student (name string, regno int, cgpa float, address string) partitioned by(department string);
OK
Time taken: 0.522 seconds
hive> describe extended student;
OK
name                string
regno                int
cgpa                  float
address              string
department            string

# Partition Information
# col_name            data_type            comment
department            string

```

6. Insert a few records into the table.

Queries:

```

insert into student values ('karthi',12,9.10,'chennai','cse');
insert into student values ('praveen',14,9.4,'coimbatore','ece');
insert into student values ('kamal',13,9.5,'tanjore',eee);

```

```

hive> insert into student values ('Karthi',12,9.10,'chennai','cse');
Query ID = karthikeyan_20231110164501_57848df3-159e-4efa-aece-8c5ef163cc1a
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2023-11-10 16:45:09,028 Stage-1 map = 0%,  reduce = 0%
2023-11-10 16:45:11,059 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local900315746_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/college.db/stud
7039223206157301-1/-ext-10000
Loading data to table college.student partition (department=null)

Time taken to load dynamic partitions: 0.55 seconds
Time taken for adding to write entity : 0.006 seconds
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 0 HDFS Write: 216 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 11.095 seconds

```

7. . View the data using select command.

Queries:

```

select * from student;

```

```
hive> select * from student;
OK
Karthi 12      9.1    chennai cse
praveen 14     9.4    coimbatore ece
Kamal 13      9.5    tanjore eee
Time taken: 0.777 seconds, Fetched: 3 row(s)
```

- View the partitions of the table using show partitions command.

Queries:

show partitions student;

```
hive> show partitions student;
OK
department=cse
department=ece
department=eee
Time taken: 0.168 seconds, Fetched: 3 row(s)
```

- Create a new table to load data from a csv file. Create a csv file of data and load it into the table using Load data.

Queries:

create table employee (id int, name string, designation string) partitioned by (dept string) row formatted fields terminated by ',' stored as text file;

load data local inpath '/home/karthikeyan/Documents/data.csv';

select * from employee;

```
hive> create table employee( id int, name string, designation string) partitioned by (dept string) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.249 seconds
hive> load data local inpath '/home/karthikeyan/Documents/data.csv' into table employee partition(dept='cse');
Loading data to table college.employee partition (dept=cse)
OK
Time taken: 0.442 seconds
```

data.csv	
1	1,Nithish,AP,cse
2	2,Vignesh,Professor,cse

```
Time taken: 0.442 seconds
hive> select * from employee;
OK
1      Nithish AP      cse
2      Vignesh Professor      cse
Time taken: 0.394 seconds, Fetched: 2 row(s)
```

10. Drop the database using drop command.

Queries:

drop database collegea cascade;

```
Time taken: 0.21 seconds
hive> drop database collegea cascade;
OK
Time taken: 0.823 seconds
```

Result:

Thus the data manipulation and data definition queries were executed successfully using Hive QL.