Program 6 : Steps for Execution of Hive Queries

HIVE INSTALLATION

1. download the file “apache-hive-3.1.2-bin.tar.gz” from following link:
https://downloads.apache.org/hive/

2. Run the following command from terminal

Note: This command installs hive and should be run only once

hive_shell.sh

CREATING TABLE and EXECUTING HIVE QUERIES

1. Run following commands from terminal

stop-all.sh
hadoop namenode -format
start-all.sh

2. get hive prompt : run following command at terminal

hive

3. Create database
show databases;

create database dbl;

use dbl ;

4. Create Table and alter table

create table flight (fno int , year int , dest varchar(10) , delay float);

5. Alter table

alter table flight rename to air_flight;

alter table air_flight add columns (new_col1 varchar(10) , new_col2 varchar(10));

6. Modigfy column
Note we can modify only one column at a time

alter table air_flight change new_col1 new_source1 varchar(15);

7. Drop columns
Note : use alter table: list the columns which you want to keep, no direct drop column command

alter table air_flight replace columns (fno int, year int , dest varchar(20), delay float) ;

8. drop table

```
drop table air_flight;
```

9. Create new database and create table

```
create database mydb;

use mydb;

create table flight (fno int , year int , dest varchar(10) , delay float)  row format delimited  fields terminated by ','  lines terminated by '\n'  stored as textfile;

desc flight;
```

10. insert into table

```
# copy this block of insert commands and paste at hive prompt,
# It runs as a shell script

insert into flight values ( 110 , 2008 , "Banglore" , 30.0);
insert into flight values ( 111 , 2008 , "Pune" , 40.0);
insert into flight values ( 112 , 2008 , "Mumbai" , 50.0);
insert into flight values ( 113 , 2008 , "Banglore" , 60.0);
insert into flight values ( 114 , 2009 , "Pune" , 70.0);
insert into flight values ( 115 , 2009 , "Mumbai" , 80.0);

select * from flight ;
```

11. Load data from text file into hive table

```
load data local inpath "/home/mangal/flight_data.txt" overwrite into table flight;

# 1. The LOAD DATA  statement is used to load data into the hive table

#2 . Syntax:

# LOAD DATA [LOCAL] INPATH '<The table data location>' [OVERWRITE] INTO TABLE
# <table_name>;

# 3. If the LOCAL switch is not used, the hive will consider the location as an HDFS
# 4. The OVERWRITE switch allows us to overwrite the table data.
# 5. each line of the data text file consists of 4 values seperated by comma ( for four columns of our
# flight table)
# 6. Note : There should not be any space any space before or after the seperating commas in the
# text file
# 7. integers or floats need to be specified without double quotes, whereas strings must be specified
# within  double quotes
```

12. create directory in HDFS and copy data into it

```
hadoop fs -mkdir /my_dir_in_hadoop
hadoop fs -ls /
hadoop fs -put  /home/mangal/flight_data.txt /my_dir_in_hadoop;
```

==========

13 in case of error like ...NO NAME NODE FOUND run following block commands from terminal

```
stop-all.sh

sudo rm -R /tmp/*
sudo rm -r /app/hadoop/tmp  # this doesnt run in my case

hdfs namenode -format

start-all.sh

hadoop fs -mkdir /my_dir_in_hadoop

hadoop fs -put  /home/mangal/flight_data.txt /my_dir_in_hadoop

load data inpath "/my_dir_in_hadoop/flight_data.txt" into table flight;

select * from flight;
```

14. Table join

```
# creating new table
create table nflight ( fno , year int ,new_source varchar (10 ) )
row format delimited
fields terminated by ','
lines terminated by '\n'
stored as textfile;

insert into nflight values ( 110 , 2021 , 'Pune');
insert into nflight values ( 111 , 2022 , 'Pune');
insert into nflight values ( 112 , 2023 , 'Pune');

# joining table
select a.fno , a.year , a.dest , a.delay , b.source
from flight a join nflight b
on ( a.fno = b.fno );
```

15. create index

```
create index flight_index on table flight ( fno)
as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
WITH DEFERRED REBUILD;

show tables;
```

16 Query to find the average departure delay per day in year 2008

```
select avg(delay) from flight where year = 2008;
```

17 Exit hive prompt: run following command from terminal

```
exit
```