Aim: Implementation of HIVE Commands

Commands:

1. To enter hive terminal
   Command: hive

```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive>
```

2. To check the databases
   Command: show databases;

```
hive> show databases;
OK
dbs
default
Time taken: 2.864 seconds, Fetched: 2 row(s)
hive>
```

3. To check the tables
   Command: show tables;

```
hive> show tables;
OK
Time taken: 0.228 seconds
hive>
```

4. To use a particular database
   Command: use dbname;

```
hive> use dbs;
OK
Time taken: 0.047 seconds
hive>
```

5. To create database
   Command: create database retail;

```
hive> create database retail;
OK
Time taken: 4.362 seconds
hive> use retail;
OK
Time taken: 0.151 seconds
hive>
```

6. To create table emp in retail database
   Command: create table <tablename>;
   Output:
   hive> use retail;

   hive> create table emp(id INT,name STRING,sal DOUBLE) row format delimited fields terminated by ',' stored as textfile;

```
hive> use retail;
OK
Time taken: 0.151 seconds
hive> create table emp(id INT,name STRING,sal DOUBLE) row format delimited fields terminated by ',' stored as
textfile;
OK
Time taken: 0.717 seconds
hive>
```

7. Schema information of table
   Command: describe <tablename>;
   Output:
   hive> use retail;
   hive> describe emp;

```
hive> describe emp;
OK
id                      int
name                    string
sal                     double
Time taken: 0.36 seconds, Fetched: 3 row(s)
hive>
```

8. To create file in training folder and save as demo.txt
   1,abc,2000
   2,pqr,4500

   To view contents of demo.txt file
   [training@localhost ~]$ cat /home/training/demo.txt
   1,abc,2000
   2,pqr,4500

```
[cloudera@quickstart ~]$ mkdir trainingPracs
[cloudera@quickstart ~]$ cd trainingPracs/
[cloudera@quickstart trainingPracs]$ cat > demoP.txt
1,abc,2000
2,pqr,4500
^Z
[3]+  Stopped                 cat > demoP.txt
[cloudera@quickstart trainingPracs]$ cat demoP.txt
1,abc,2000
2,pqr,4500
[cloudera@quickstart trainingPracs]$
```

To load data from local path

hive> load data local inpath '/home/training/demo.txt' into table emp;

```
hive> load data local inpath '/home/cloudera/trainingPracs/demoP.txt' into table emp;
Loading data to table retail.emp
Table retail.emp stats: [numFiles=1, totalSize=22]
OK
Time taken: 2.405 seconds
hive>
```

9. To view contents of table
   Command: select * from emp;
   Output:

```
hive> select * from emp;
OK
1       abc     2000.0
2       pqr     4500.0
Time taken: 1.192 seconds, Fetched: 2 row(s)
hive>
```

10. To rename table name
    Command: ALTER TABLE old_table_name RENAME TO new_table_name;
    Output:

    hive> use retail;
    hive> alter table emp rename to emp_sal;

```
hive> use retail;
OK
Time taken: 0.037 seconds
hive> alter table emp rename to emp_sal;
OK
Time taken: 0.261 seconds
hive> describe emp_sal;
OK
id                      int
name                    string
sal                     double
Time taken: 0.191 seconds, Fetched: 3 row(s)
hive> show tables;
OK
emp_sal
Time taken: 0.059 seconds, Fetched: 1 row(s)
hive>
```

11. Selecting data

hive> select * from emp_sal where id=1;

```
hive> select * from emp_sal where id=1;
OK
1       abc     2000.0
Time taken: 0.672 seconds, Fetched: 1 row(s)
hive>
```

12. To count number of records in table

hive> select count(*) from emp_sal;

```
hive> select count(*) from emp_sal;
Query ID = cloudera_20201125032121_ce22c049-0bab-4780-a604-2ac39bd06d28
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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>
Starting Job = job_1606301819383_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1606301819
383_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1606301819383_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-25 03:22:28,710 Stage-1 map = 0%,  reduce = 0%
2020-11-25 03:22:54,306 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.45 sec
2020-11-25 03:23:20,091 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.99 sec
MapReduce Total cumulative CPU time: 7 seconds 990 msec
Ended Job = job_1606301819383_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.99 sec   HDFS Read: 6886 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 990 msec
OK
2
Time taken: 95.97 seconds, Fetched: 1 row(s)
hive>
```

13. Try using aggregate commands using HQL(Try creating tables with group by fields and execute the aggregate commands)

Firstly, creating table and adding data into the table

```
hive> create table emp_agrr(id INT,name STRING,sal DOUBLE,dept STRING) row format delimited fields terminated by ',' stored
as textfile;
OK
Time taken: 0.486 seconds
hive> load data local inpath '/home/cloudera/trainingPracs/demoP.txt' into table emp_agrr;
Loading data to table retail.emp_agrr
Table retail.emp_agrr stats: [numFiles=1, totalSize=80]
OK
Time taken: 2.283 seconds
hive> select * from emp_agrr;
OK
1       abc     2000.0  Admin
2       def     4500.0  Tech
3       ghi     5600.0  PR
4       jkl     8000.0  Tech
5       mno     2300.0  Admin
Time taken: 0.941 seconds, Fetched: 5 row(s)
hive>
```

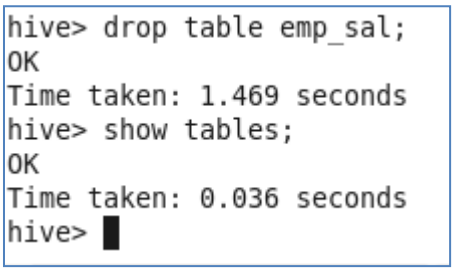
Now applying group by and aggregate commands.

```
hive> select dept,count(*) from emp_agrr group by dept;
Query ID = cloudera_20201125033636_9ecc5a2d-fa4d-4f6b-a633-4f38a7fd7e4d
Total jobs = 1
Launching Job 1 out of 1
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>
Starting Job = job_1606301819383_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1606301819383_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1606301819383_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-25 03:37:15,307 Stage-1 map = 0%,  reduce = 0%
2020-11-25 03:37:41,868 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.55 sec
2020-11-25 03:38:07,637 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.51 sec
MapReduce Total cumulative CPU time: 8 seconds 510 msec
Ended Job = job_1606301819383_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.51 sec   HDFS Read: 7252 HDFS Write: 20 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 510 msec
OK
Admin   2
PR      1
Tech    2
Time taken: 96.815 seconds, Fetched: 3 row(s)
hive>
```

```
hive> select max(sal),dept from emp_agrr group by dept;
Query ID = cloudera_20201125034141_616fcca2-267c-4a06-b064-2c6906a83303
Total jobs = 1
Launching Job 1 out of 1
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>
Starting Job = job_1606301819383_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1606301819383_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1606301819383_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-25 03:41:50,417 Stage-1 map = 0%,  reduce = 0%
2020-11-25 03:42:12,445 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.32 sec
2020-11-25 03:42:36,176 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.8 sec
MapReduce Total cumulative CPU time: 7 seconds 800 msec
Ended Job = job_1606301819383_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.8 sec   HDFS Read: 7512 HDFS Write: 35 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 800 msec
OK
2300.0  Admin
5600.0  PR
8000.0  Tech
Time taken: 72.617 seconds, Fetched: 3 row(s)
hive>
```

```
hive> select min(sal),dept from emp_agrr group by dept;
Query ID = cloudera_20201125035252_7aef7a8a-3e50-42d7-808d-3a2db024804c
Total jobs = 1
Launching Job 1 out of 1
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>
Starting Job = job_1606301819383_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1606301819383_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1606301819383_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-25 03:52:32,710 Stage-1 map = 0%,  reduce = 0%
2020-11-25 03:53:04,845 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.55 sec
2020-11-25 03:53:32,434 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.9 sec
MapReduce Total cumulative CPU time: 8 seconds 900 msec
Ended Job = job_1606301819383_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.9 sec   HDFS Read: 7512 HDFS Write: 35 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 900 msec
OK
2000.0  Admin
5600.0  PR
4500.0  Tech
Time taken: 89.464 seconds, Fetched: 3 row(s)
hive>
```

14. To drop table
    hive> drop table emp_sal;



15. To exit from Hive terminal
    hive> exit;