Lab 04 - HIVE Commands

Full name: Chloe Tee Rouyi

Student ID: 0354731

Tasks:

1. ls

```
[cloudera@quickstart ~]$ ls
bdtsec03          Desktop    enterprise-deployment.json  parcels    testdir
chloetee          Documents  kerberos                    Pictures   testDir2
cloudera-manager  Downloads  lib                         Public     Videos
cm_api.py         eclipse    Music                       Templates  workspace
```

2. pwd

```
[cloudera@quickstart ~]$ pwd
/home/cloudera
```

3. sudo hive

```
[cloudera@quickstart ~]$ sudo 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> show database;
NoViableAltException(74@[697:1: ddlStatement : ( createDatabaseStatement | switc
hDatabaseStatement | dropDatabaseStatement | createTableStatement | dropTableSta
tement | truncateTableStatement | alterStatement | descStatement | showStatement
 | metastoreCheck | createViewStatement | dropViewStatement | createFunctionStat
ement | createMacroStatement | createIndexStatement | dropIndexStatement | dropF
unctionStatement | reloadFunctionStatement | dropMacroStatement | analyzeStateme
nt | lockStatement | unlockStatement | lockDatabase | unlockDatabase | createRol
eStatement | dropRoleStatement | grantPrivileges | revokePrivileges | showGrants
 | showRoleGrants | showRolePrincipals | showRoles | grantRole | revokeRole | se
tRole | showCurrentRole );])
        at org.antlr.runtime.DFA.noViableAlt(DFA.java:158)
        at org.antlr.runtime.DFA.predict(DFA.java:116)
        at org.apache.hadoop.hive.ql.parse.HiveParser.ddlStatement(HiveParser.ja
va:2503)
        at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.j
ava:1589)
```

4. show databases;

```
hive> show databases;
OK
default
Time taken: 0.442 seconds, Fetched: 1 row(s)
```

5. create database riyazdb;

```
hive> create database chloe;
OK
Time taken: 2.133 seconds
```

Note: instead of the name “riyaz”, use your own shortest name.

6. create database riyazbackupstudy comment ‘holds study backup’;

```
hive> create database chloebackupstudy comment 'this DB holds study BKP of chloe
';
OK
Time taken: 0.151 seconds
```

7. create database riyazbackup with DBPROPERTIES ('creator'='abc','dateyy'='15-05-2020');

```
hive> create database chloebackup with DBPROPERTIES ('creator'='abc','dateyy'='2
3-10-2024');
OK
Time taken: 0.081 seconds
```

8. describe database extended riyazbackup;

```
hive> describe database extended chloebackup;
OK
chloebackup             hdfs://quickstart.cloudera:8020/user/hive/warehouse/chlo
ebackup.db      root    USER    {dateyy=23-10-2024, creator=abc}
Time taken: 0.077 seconds, Fetched: 1 row(s)
```

9. create table in particular database

10. use riyazdb;

```
hive> use chloe;
OK
Time taken: 0.022 seconds
```

11. create table researchpaper (paperid int, papername string, pic string) row format delimited fields terminated by ',';

```
hive> create table researchpaper (paperid int, papername string, pic string)row format delimited fields terminated by ',';
OK
Time taken: 0.471 seconds
```

12. show tables;

```
hive> show tables;
OK
researchpaper
Time taken: 0.089 seconds, Fetched: 1 row(s)
```

13. select * from researchpaper;

```
hive> select*from researchpaper;
OK
Time taken: 0.669 seconds
```

Open terminal in another window and perform the following tasks:

14. gedit researchpaper.input

Note: After the text editor is open, add the content below and save and close the editor.

001,bigdata,riyaz

002,deeplearning,viky

003,machinelearning,Ong

```
[cloudera@quickstart ~]$ gedit researchpaper.input
[cloudera@quickstart ~]$ cat researchpaper.input
001,'Big Data Overview','Chloe'
002,'Apache Hive Overview','Obama'
003,'bigdata','riyaz'
004,'deeplearning','viky'
005,'machinelearning','Ong'
```

15. pwd

16. gedit researchpaper.input

17. Go back to hive CLI and preform the following tasks:

Syntax: LOAD DATA LOCAL INPATH ‘ path’ INTO TABLE tablename;

18. LOAD DATA LOCAL INPATH '/home/cloudera/researchpaper.input' INTO TABLE researchpaper;

```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/researchpaper.i
nput' INTO TABLE researchpaper;
Loading data to table chloe.researchpaper
Table chloe.researchpaper stats: [numFiles=1, totalSize=148]
OK
Time taken: 1.167 seconds
```

```
hive> select*from researchpaper;
OK
1       'Big Data Overview'     'Chloe'
2       'Apache Hive Overview'  'Obama'
3       'bigdata'       'riyaz'
4       'deeplearning'  'viky'
5       'machinelearning'       'Ong'
NULL    NULL    NULL
NULL    NULL    NULL
Time taken: 0.12 seconds, Fetched: 7 row(s)
```

Syntax: describe tablename

19. describe extended tablename

```
hive> describe extended reseachpaper;
FAILED: SemanticException [Error 10001]: Table not found reseachpaper
hive> describe extended researchpaper;
OK
paperid                 int
papername               string
pic                     string

Detailed Table Information      Table(tableName:researchpaper, dbName:chloe, owner:root, createTime:1729650668, lastAccessTime:0, rete
ntion:0, sd:StorageDescriptor(cols:[FieldSchema(name:paperid, type:int, comment:null), FieldSchema(name:papername, type:string, commen
t:null), FieldSchema(name:pic, type:string, comment:null)], location:hdfs://quickstart.cloudera:8020/user/hive/warehouse/chloe.db/rese
archpaper, inputFormat:org.apache.hadoop.mapred.TextInputFormat, outputFormat:org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputForm
at, compressed:false, numBuckets:-1, serdeInfo:SerDeInfo(name:null, serializationLib:org.apache.hadoop.hive.serde2.lazy.LazySimpleSerD
e, parameters:{serialization.format=,, field.delim=,}), bucketCols:[], sortCols:[], parameters:{}, skewedInfo:SkewedInfo(skewedColName
s:[], skewedColValues:[], skewedColValueLocationMaps:{}), storedAsSubDirectories:false), partitionKeys:[], parameters:{numFiles=1, tra
nsient_lastDdlTime=1729651388, COLUMN_STATS_ACCURATE=true, totalSize=148}, viewOriginalText:null, viewExpandedText:null, tableType:MAN
AGED_TABLE)
Time taken: 0.108 seconds, Fetched: 5 row(s)
```

Syntax: drop table tablename;

20. create another table and then, show the table in the database. after that drop the table. Then show the table is not in the database anymore.

```
hive> create table chloetable (id int, name string, pic string)row format delimited fields terminated by ',';
OK
Time taken: 0.19 seconds
hive> show tables in chloe;
OK
chloetable
researchpaper
Time taken: 0.02 seconds, Fetched: 2 row(s)


hive> drop table chloetable;
OK
Time taken: 0.23 seconds
```

21. show tables in riyazdb;

```
hive> show tables in chloe;
OK
researchpaper
Time taken: 0.041 seconds, Fetched: 1 row(s)
```

22. describe extended researchpaper;

```
hive> describe extended researchpaper;
OK
paperid                 int
papername               string
pic                     string

Detailed Table Information      Table(tableName:researchpaper, dbName:chloe, owner:root, createTime:1729650668, lastAccessTime:0, rete
ntion:0, sd:StorageDescriptor(cols:[FieldSchema(name:paperid, type:int, comment:null), FieldSchema(name:papername, type:string, commen
t:null), FieldSchema(name:pic, type:string, comment:null)], location:hdfs://quickstart.cloudera:8020/user/hive/warehouse/chloe.db/rese
archpaper, inputFormat:org.apache.hadoop.mapred.TextInputFormat, outputFormat:org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputForm
at, compressed:false, numBuckets:-1, serdeInfo:SerDeInfo(name:null, serializationLib:org.apache.hadoop.hive.serde2.lazy.LazySimpleSerD
e, parameters:{serialization.format=,, field.delim=,}), bucketCols:[], sortCols:[], parameters:{}, skewedInfo:SkewedInfo(skewedColName
s:[], skewedColValues:[], skewedColValueLocationMaps:{}), storedAsSubDirectories:false), partitionKeys:[], parameters:{numFiles=1, tra
nsient_lastDdlTime=1729651388, COLUMN_STATS_ACCURATE=true, totalSize=148}, viewOriginalText:null, viewExpandedText:null, tableType:MAN
AGED_TABLE)
Time taken: 0.155 seconds, Fetched: 5 row(s)
```

23. select count(*) from researchpaper;

```
hive> select count(*) from researchpaper;
Query ID = root_20241022233737_620d61f2-b7ad-49dc-8113-a9230e37d1f1
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_1727834699991_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1727834699991_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1727834699991_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-10-22 23:37:57,784 Stage-1 map = 0%,  reduce = 0%
2024-10-22 23:38:08,041 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.13 sec
2024-10-22 23:38:19,381 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.77 sec
MapReduce Total cumulative CPU time: 2 seconds 770 msec
Ended Job = job_1727834699991_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.77 sec   HDFS Read: 6843 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 770 msec
OK
7
Time taken: 43.125 seconds, Fetched: 1 row(s)
```

24. Go back to terminal and perform the following tasks:

25. hdfs dfs -ls /user/hive/warehouse/

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/warehouse/
Found 4 items
drwxrwxrwx   - root supergroup          0 2024-10-22 23:32 /user/hive/warehouse/chloe.db
drwxrwxrwx   - root supergroup          0 2024-10-22 19:20 /user/hive/warehouse/chloebackup.db
drwxrwxrwx   - root supergroup          0 2024-10-22 19:19 /user/hive/warehouse/chloebackupstudy.db
drwxrwxrwx   - root supergroup          0 2024-10-22 23:27 /user/hive/warehouse/chloedb2.db
```

26. hadoop dfs -ls /user/hive/warehouse/riyazbackup.db

```
[cloudera@quickstart ~]$ hadoop dfs -ls /user/hive/warehouse/chloebackup.db/
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
```

Error: wrong syntax

Error troubleshooting: use *hdfs dfs* instead of hadoop dfs for the correct syntax.

27. hadoop fs -ls /user/hive/warehouse/riyazbackup.db

```
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/chloebackup.db/
```

```
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/chloe.db/
Found 1 items
drwxrwxrwx   - root supergroup          0 2024-10-22 19:43 /user/hive/warehouse/chloe.db/researchpaper
[cloudera@quickstart ~]$
```

Explanation: Eventhough it doesn't use the conventional hdfs dfs syntax, hadoop fs works.

28. hadoop fs -lsr /user/hive/warehouse/riyazbackup.db

```
[cloudera@quickstart ~]$ hadoop fs -lsr /user/hive/warehouse/chloe.db/
lsr: DEPRECATED: Please use 'ls -R' instead.
drwxrwxrwx   - root supergroup          0 2024-10-22 19:43 /user/hive/warehouse/chloe.db/researchpaper
-rwxrwxrwx   1 root supergroup        148 2024-10-22 19:43 /user/hive/warehouse/chloe.db/researchpaper/researchpaper.input
```

Error: Wrong syntax again

Error troubleshooting: Use *hadoop fs -ls -R* to list the directories in reverse order.

29. open Hue and try to perform the same tasks via Hue GUI.