

## LAB- 8

**AIM:** To implement HIVE commands.

### **COMMANDS/QUERIES:**

**Command 1:** CREATE database mydb;

CREATE TABLE employee(ID int, name string, salary float, age int)

Row format delimited

Fields terminated by ',';

```
hive> CREATE database mydb;
OK
Time taken: 19.722 seconds
hive> show databases;
OK
default
mydb
Time taken: 0.62 seconds, Fetched: 2 row(s)
hive> CREATE TABLE employee(ID int, name string, salary float, age int)
> row format delimited
> fields terminated by ',';
OK
Time taken: 0.875 seconds
```

**Command 2:** DESCRIBE employee;

```
OK
id                int
name              string
salary            float
age               int
Time taken: 0.863 seconds, Fetched: 4 row(s)
hive> describe formatted employee;
# Detailed Table Information
Database:         default
Owner:            cloudera
CreateTime:       Tue May 02 11:03:25 PDT 2023
LastAccessTime:   UNKNOWN
Protect Mode:     None
Retention:        0
Location:         hdfs://quickstart.cloudera:8020/user/hive/warehouse/emp
oyee
Table Type:       MANAGED_TABLE
Table Parameters:
    transient_lastDdlTime    1683050605
```

**Command 3:** ALTER table employee2 RENAME TO emptable;

```
OK
Time taken: 0.559 seconds
hive> DESCRIBE emptable;
OK
id                int
name              string
salary            float
age               int
Time taken: 0.522 seconds, Fetched: 4 row(s)
```

**Command 4:** set hive.exec.dynamic.partition = true;  
set hive.exec.dynamic.partition.mode = nonstrict;

```
OK
Loading partition {course=23}
Loading partition {course=28}
Time taken for adding to write entity : 6
Partition student2.student_part{course=22} stats: [numFiles=1, numRows=1, totalSize=14, rawDataSize=13]
Partition student2.student_part{course=23} stats: [numFiles=1, numRows=2, totalSize=27, rawDataSize=25]
Partition student2.student_part{course=24} stats: [numFiles=1, numRows=1, totalSize=14, rawDataSize=13]
Partition student2.student_part{course=26} stats: [numFiles=1, numRows=1, totalSize=13, rawDataSize=12]
Partition student2.student_part{course=28} stats: [numFiles=1, numRows=2, totalSize=27, rawDataSize=25]
Partition student2.student_part{course=30} stats: [numFiles=1, numRows=1, totalSize=15, rawDataSize=14]
Partition student2.student_part{course=_HIVE_DEFAULT_PARTITION_} stats: [numFiles=1, numRows=1, totalSize=11, rawDataSize=10]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 3.31 sec HDFS Read: 4598 HDFS Write: 532 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 310 msec
OK
Time taken: 88.157 seconds
```

**Command 5:** set hive.enforce.bucketing = true;

```
OK
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 3
2023-05-02 12:01:41,690 Stage-1 map = 0%, reduce = 0%
2023-05-02 12:02:12,141 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.33 sec
2023-05-02 12:02:47,361 Stage-1 map = 100%, reduce = 22%, Cumulative CPU 5.68 sec
2023-05-02 12:02:49,682 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 10.1 sec
2023-05-02 12:02:51,262 Stage-1 map = 100%, reduce = 78%, Cumulative CPU 11.1 sec
2023-05-02 12:02:54,418 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 12.86 sec
MapReduce Total cumulative CPU time: 12 seconds 860 msec
Ended Job = job_1682366041039_0005
Loading data to table edubucket.studbucketx
Table edubucket.studbucketx stats: [numFiles=3, numRows=9, totalSize=148, rawDataSize=139]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 3 Cumulative CPU: 12.86 sec HDFS Read: 15459 HDFS Write: 379 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 860 msec
OK
Time taken: 116.103 seconds
```

## COMMAND 6: SELECT COUNT(\*) FROM customers;

```
Query ID = cloudera_20230521011919_6f75c52c-7625-43e7-9a20-f6ad565061dc
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_1683016188545_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1683016188545_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1683016188545_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-21 01:19:55,845 Stage-1 map = 0%, reduce = 0%
2023-05-21 01:20:24,662 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.74 sec
2023-05-21 01:20:54,589 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.74 sec
MapReduce Total cumulative CPU time: 8 seconds 960 msec
Ended Job = job_1683016188545_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.96 sec HDFS Read: 7989 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 960 msec
OK
12
Time taken: 109.609 seconds, Fetched: 1 row(s)
```

## COMMAND 7: SELECT \*FROM customers LIMIT 4;

```
OK
7      Ramasundar      Bangalore      15      077-25814763
3      Alex      London      13      075-12458969
8      Alford      New York      12      044-25874365
11     Ravi      Bangalore      15      077-45625874
Time taken: 0.147 seconds, Fetched: 4 row(s)
hive> █
```

## COMMAND 8: SELECT \*FROM customers ORDER BY cust\_id;

```
Query ID = cloudera_20230521012424_c1b16064-f71f-4921-a7cf-244ca5fcec68
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_1683016188545_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1683016188545_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1683016188545_0019
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-21 01:24:32,760 Stage-1 map = 0%, reduce = 0%
2023-05-21 01:24:55,338 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.18 sec
2023-05-21 01:25:19,711 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.48 sec
MapReduce Total cumulative CPU time: 7 seconds 480 msec
Ended Job = job_1683016188545_0019
MapReduce Jobs Launched:
```

**COMMAND 9:** SELECT \*FROM customers JOIN purchase ON purchase.id = cust.order\_id;

```
hive> SELECT * FROM cust JOIN purchase ON purchase.id=cust.order_id;
Query ID = cloudera_20230521042525_69d2cb70-1bcb-441f-8a48-82f0c28e7373
Total jobs = 1
Total MapReduce CPU Time Spent: 5 seconds 630 msec
OK
7      Ramasundar      Bangalore      15      077-25814763      15      SSD card      30
00
3      Alex      London      13      075-12458969      13      hard disk      43000
8      Alford      New York      12      044-25874365      12      motherboard      60000
11     Ravi      Bangalore      15      077-45625874      15      SSD card      3000
10     Santakumar      Chennai      14      007-22388644      14      cooling pad      1500
12     Lucida      San Jose      12      044-52981425      12      motherboard      60000
5      Anderson      Brisban      13      045-21447739      13      hard disk      43000
1      Subbarao      Bangalore      14      077-12346674      14      cooling pad      15
00
2      Mukesh      Mumbai      11      029-12358964      11      gaming PC      89000
6      McDen      London      15      078-22255588      15      SSD card      3000
4      Ivan      Toronto      15      008-22544166      15      SSD card      3000
9      Benjamin      Hampshair      11      008-22536178      11      gaming PC      89
000
```

**COMMAND 10:** SELECT MAX(price) FROM purchase;

```
Query ID = cloudera_20230521043232_2c7632bd-f36e-491b-9248-2faf26338884
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_1683016188545_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1683016188545_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1683016188545_0021
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-21 04:33:06,507 Stage-1 map = 0%, reduce = 0%
2023-05-21 04:33:27,841 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.15 sec
2023-05-21 04:33:51,842 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.58 sec
MapReduce Total cumulative CPU time: 7 seconds 580 msec
Ended Job = job_1683016188545_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.58 sec HDFS Read: 7664 HDFS Write: 6 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 580 msec
OK
89000
Time taken: 77.166 seconds, Fetched: 1 row(s)
```