

Task 3

Link: <https://acadgild.com/blog/transactions-in-hive/>

Refer the above given link for transactions in Hive and implement the operations given in the blog using your own sample data set and send us the screenshot.

SOLUTION:

```
hive> set hive.support.concurrency = true;
hive> set hive.enforce.bucketing= true;
hive> set hive.txn.manager =
org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;
hive> set hive.compactor.initiator.on = true;
hive> set hive.compactor.worker.threads = a positive number on at least
one instance of the Thrift metastore service;

hive> CREATE TABLE college(clg_id int,clg_name string,clg_loc string)
clustered by (clg_id) into 5 buckets stored as orc
TBLPROPERTIES('transactional'='true');
OK
Time taken: 2.233 seconds
hive> show tables
> ;
OK
college
olympics
temperature_data
temperature_data_vw
Time taken: 0.35 seconds, Fetched: 4 row(s)
hive> INSERT INTO table college
values(1,'nec','nlr'),(2,'vit','vlr'),(3,'srm','chen'),(4,'lpu','del'),(5
,'stanford','uk'),(6,'JNTUA','atp'),(7,'cambridge','us');
WARNING: 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.
Query ID = acadgild_20180524021447_23c03468-ad54-4882-9f62-e1ecbf74a2a7
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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_1527096493792_0012, Tracking URL =
http://localhost:8088/proxy/application_1527096493792_0012/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill
job_1527096493792_0012
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 5
2018-05-24 02:15:17,265 Stage-1 map = 0%, reduce = 0%
```

2018-05-24 02:15:46,157 Stage-1 map = 100%, reduce = 0%, Cumulative CPU
4.33 sec
2018-05-24 02:16:46,288 Stage-1 map = 100%, reduce = 0%, Cumulative CPU
4.33 sec
2018-05-24 02:17:02,315 Stage-1 map = 100%, reduce = 13%, Cumulative CPU
5.17 sec
2018-05-24 02:17:06,724 Stage-1 map = 100%, reduce = 67%, Cumulative CPU
10.43 sec
2018-05-24 02:17:37,155 Stage-1 map = 100%, reduce = 92%, Cumulative CPU
30.58 sec
2018-05-24 02:17:44,281 Stage-1 map = 100%, reduce = 100%, Cumulative
CPU 34.65 sec

MapReduce Total cumulative CPU time: 34 seconds 650 msec

Ended Job = job_1527096493792_0012

Loading data to table custom.college

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 34.65 sec HDFS Read:

26601 HDFS Write: 3993 SUCCESS

Total MapReduce CPU Time Spent: 34 seconds 650 msec

OK

Time taken: 181.629 seconds

hive> *select * from college;*

OK

```
5      stanford      uk
6      JNTUA atp
1      nec      nlr
7      cambridge      us
2      vit      vlr
3      srm      chen
4      lpu      del
```

Time taken: 0.986 seconds, Fetched: 7 row(s)

hive> *UPDATE college set clg_id = 4 where clg_id = 8;*

*FAILED: SemanticException [Error 10302]: Updating values of bucketing
columns is not supported. Column clg_id.*

hive> *UPDATE college set clg_name = 'AIEEE' where clg_id =4;*

WARNING: 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.

Query ID = acadgild_20180524023116_58f2039c-80e3-4d1c-a4e7-c1ecf1ee0377

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 5

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_1527096493792_0013, Tracking URL =

http://localhost:8088/proxy/application_1527096493792_0013/

Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill

job_1527096493792_0013

```

Hadoop job information for Stage-1: number of mappers: 5; number of
reducers: 5
2018-05-24 02:31:43,726 Stage-1 map = 0%,   reduce = 0%
2018-05-24 02:32:46,221 Stage-1 map = 0%,   reduce = 0%
2018-05-24 02:33:57,490 Stage-1 map = 0%,   reduce = 0%, Cumulative CPU
17.08 sec
2018-05-24 02:34:15,758 Stage-1 map = 60%,   reduce = 0%, Cumulative CPU
27.34 sec
2018-05-24 02:34:18,250 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU
34.72 sec
2018-05-24 02:35:19,728 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU
34.72 sec
2018-05-24 02:35:45,097 Stage-1 map = 100%,   reduce = 67%, Cumulative CPU
39.35 sec
2018-05-24 02:36:06,616 Stage-1 map = 100%,   reduce = 93%, Cumulative CPU
52.5 sec
2018-05-24 02:36:08,041 Stage-1 map = 100%,   reduce = 99%, Cumulative CPU
55.39 sec
2018-05-24 02:36:09,689 Stage-1 map = 100%,   reduce = 100%, Cumulative
CPU 56.05 sec
MapReduce Total cumulative CPU time: 56 seconds 50 msec
Ended Job = job_1527096493792_0013
Loading data to table custom.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5   Reduce: 5   Cumulative CPU: 56.05 sec   HDFS Read:
56156 HDFS Write: 962 SUCCESS
Total MapReduce CPU Time Spent: 56 seconds 50 msec
OK
Time taken: 298.278 seconds
hive> select * from college;
OK
5      stanford      uk
6      JNTUA atp
1      nec      nlr
7      cambridge      us
2      vit      vlr
3      srm      chen
4      AIEEE del
Time taken: 0.887 seconds, Fetched: 7 row(s)

```

```

hive> delete from college where clg_id=5;
WARNING: 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.
Query ID = acadgild_20180524023756_6f8543ce-0dd5-4a2f-91af-af03bbcfcc272
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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_1527096493792_0014, Tracking URL =
http://localhost:8088/proxy/application_1527096493792_0014/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill
job_1527096493792_0014
Hadoop job information for Stage-1: number of mappers: 5; number of
reducers: 5
2018-05-24 02:38:29,246 Stage-1 map = 0%,   reduce = 0%
2018-05-24 02:39:30,096 Stage-1 map = 0%,   reduce = 0%
2018-05-24 02:40:13,548 Stage-1 map = 40%,   reduce = 0%, Cumulative CPU
19.45 sec
2018-05-24 02:40:15,929 Stage-1 map = 80%,   reduce = 0%, Cumulative CPU
26.17 sec
2018-05-24 02:40:17,115 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU
28.17 sec
2018-05-24 02:41:17,269 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU
28.17 sec
2018-05-24 02:41:28,223 Stage-1 map = 100%,   reduce = 53%, Cumulative CPU
31.89 sec
2018-05-24 02:41:32,347 Stage-1 map = 100%,   reduce = 67%, Cumulative CPU
35.46 sec
2018-05-24 02:41:48,111 Stage-1 map = 100%,   reduce = 99%, Cumulative CPU
47.63 sec
2018-05-24 02:41:51,287 Stage-1 map = 100%,   reduce = 100%, Cumulative
CPU 48.69 sec
MapReduce Total cumulative CPU time: 48 seconds 690 msec
Ended Job = job_1527096493792_0014
Loading data to table custom.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5   Reduce: 5   Cumulative CPU: 48.69 sec   HDFS Read:
54253 HDFS Write: 751 SUCCESS
Total MapReduce CPU Time Spent: 48 seconds 690 msec
OK
Time taken: 238.864 seconds
hive> select * from employee;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found
'employee'
hive> select * from college;
OK
6      JNTUA atp
1      nec   nlr
7      cambridge   us
2      vit    vlr
3      srm    chen
4      AIEEE del
Time taken: 0.988 seconds, Fetched: 6 row(s)--->in output there is No
record with clg_id =5;

```