Transactions in Hive

NOTE: Before working in hive shell, start all hadoop daemons using start-all.sh command, and start mysqld service using sudo service mysqld start, then launch hive using hive command.

Below steps are followed to work with transactions in Hive:

Step 1: Check hive version, using following command:

```
[acadgild@localhost ~]$ hive --version
Hive 0.14.0
Subversion file:///Users/ghagleitner/Projects/hive-svn/rel-prep/hive-14-rel-prep
 -r Unknown
Compiled by ghagleitner on Sat Nov 8 23:25:06 PST 2014
From source with checksum 49c2182a0856f7917f571802a7594b00
[acadgild@localhost ~]$
```

Hive Version

Step 2: Below properties are set order-wise in order to work with transactions in Hive:

```
hive> set hive.support.concurrency = true;
hive> set hive.enforce.bucketing = true;
hive> set hive.exec.dynamic.partition.mode = nonstrict;
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 = 1;
hive>
```

Step 3: Creation of table that supports Hive transactions

```
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: 3.03 seconds
hive> describe college;
OK
clg_id                  int
clg_name                string
clg_loc                 string
Time taken: 1.1 seconds, Fetched: 3 row(s)
hive> show tables;
OK
college
first
locations
users
Time taken: 0.109 seconds, Fetched: 4 row(s)
hive>
```

Creation of table "college" with 5 fields

Schema of "college" table

"college" table created successfully inside default database under /user/hive/warehouse directory

```
hive> dfs -ls /user/hive/warehouse;
Found 8 items
drwxr-xr-x   - acadgild supergroup          0 2015-11-25 15:25 /user/hive/wareho
use/b1.db
drwxr-xr-x   - acadgild supergroup          0 2017-08-16 22:56 /user/hive/wareho
use/college
drwxr-xr-x   - acadgild supergroup          0 2017-07-06 14:33 /user/hive/wareho
use/custom.db
drwxr-xr-x   - acadgild supergroup          0 2017-08-16 13:55 /user/hive/wareho
use/emp.db
drwxr-xr-x   - acadgild supergroup          0 2015-11-05 13:14 /user/hive/wareho
use/first
drwxr-xr-x   - acadgild supergroup          0 2017-07-09 18:13 /user/hive/wareho
use/locations
drwxr-xr-x   - acadgild supergroup          0 2017-07-08 00:58 /user/hive/wareho
use/olympicsdb.db
drwxr-xr-x   - acadgild supergroup          0 2017-07-09 18:22 /user/hive/wareho
use/users
hive>
```

Using above command besides "show tables" command, we can check "college" table is created successfully inside "/user/hive/warehouse" directory.

Explanation of "create table" command:

The above syntax creates a table with name '*college'* and the columns present in the table are '*clg_id, clg_name, clg_loc'*. Table is being *bucketed* by '*clg_id'* and the table format is '*orc',* also transactions in table are enabled by specifying *'transactional'='true'* inside the *TBLPROPERTIES.*

Step 4: Inserting Data "row-wise" (each row separated by ()) into Hive Table "college"

```
hive> INSERT INTO TABLE college values(1,'Miranda House','New Delhi'),
    > (2,'Loyola College','Chennai'),
    > (3,'Shri Ram College of Commerce','New Delhi'),
    > (4,'Bishop Herber College','Tiruchi'),
    > (5,'Atma Ram Sanatan Dharma College','New Delhi'),
    > (6,'St. Xavier College','Kolkata'),
    > (7,'Lady Shri Ram College of Women','New Delhi'),
    > (8,'Dayal Singh College','New Delhi'),
    > (9,'Deen Dayal Upadhyaya College','New Delhi'),
    > (10,'The Women Christian College','Chennai');
Query ID = acadgild_20170816231111_d3055e81-17f2-4061-b558-a37bde432759
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_1502902094643_0001, Tracking URL = http://localhost:8088/prox
y/application_1502902094643_0001/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_15029020946
43_0001
```

insert command

```
2017-08-16 23:24:18,044 Stage-1 map = 100%,  reduce = 84%, Cumulative CPU 52.03
sec
2017-08-16 23:25:43,252 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 58.41
sec
2017-08-16 23:28:12,161 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 72.76
sec
2017-08-16 23:29:16,740 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 72.76
sec
2017-08-16 23:30:27,221 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 107.5
9 sec
MapReduce Total cumulative CPU time: 1 minutes 47 seconds 590 msec
Ended Job = job_1502902094643_0001
Loading data to table default.college
Table default.college stats: [numFiles=5, numRows=10, totalSize=4428, rawDataSiz
e=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 111.58 sec   HDFS Read: 628 H
DFS Write: 4803 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 51 seconds 580 msec
OK
Time taken: 1179.178 seconds
hive>
```

Data is inserted successfully, to check the data use below command.

```
hive> select * from college;
OK
10      The Women Christian College     Chennai
5       Atma Ram Sanatan Dharma College New Delhi
6       St. Xavier College      Kolkata
1       Miranda House   New Delhi
7       Lady Shri Ram College of Women  New Delhi
2       Loyola College  Chennai
8       Dayal Singh College     New Delhi
3       Shri Ram College of Commerce    New Delhi
9       Deen Dayal Upadhyaya College    New Delhi
4       Bishop Herber College   Tiruchi
Time taken: 7.668 seconds, Fetched: 10 row(s)
```

If above insert command is fired once more, then data will get appended to the table, refer below three screenshots:

```
hive> INSERT INTO TABLE college values(1,'Miranda House','New Delhi'),
    > (2,'Loyola College','Chennai'),
    > (3,'Shri Ram College of Commerce','New Delhi'),
    > (4,'Bishop Herber College','Tiruchi'),
    > (5,'Atma Ram Sanatan Dharma College','New Delhi'),
    > (6,'St. Xavier College','Kolkata'),
    > (7,'Lady Shri Ram College of Women','New Delhi'),
    > (8,'Dayal Singh College','New Delhi'),
    > (9,'Deen Dayal Upadhyaya College','New Delhi'),
    > (10,'The Women Christian College','Chennai');
Query ID = acadgild_20170816233434_ae5cc011-4971-4dbc-99b3-cc8c61e1181e
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_1502902094643_0002, Tracking URL = http://localhost:8088/prox
y/application_1502902094643_0002/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_15029020946
43_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2017-08-16 23:35:59,542 Stage-1 map = 0%,  reduce = 0%
```

```
2017-08-16 23:45:45,810 Stage-1 map = 100%,  reduce = 94%, Cumulative CPU 77.22
sec
2017-08-16 23:46:01,806 Stage-1 map = 100%,  reduce = 97%, Cumulative CPU 89.38
sec
2017-08-16 23:46:10,659 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 102.1
5 sec
MapReduce Total cumulative CPU time: 1 minutes 42 seconds 150 msec
Ended Job = job_1502902094643_0002
Loading data to table default.college
Table default.college stats: [numFiles=10, numRows=20, totalSize=8857, rawDataSi
ze=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 111.67 sec   HDFS Read: 628 H
DFS Write: 4804 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 51 seconds 670 msec
OK
Time taken: 758.361 seconds
hive>
```

```
hive> select * from college;
OK
10      The Women Christian College     Chennai
5       Atma Ram Sanatan Dharma College New Delhi
10      The Women Christian College     Chennai
5       Atma Ram Sanatan Dharma College New Delhi
6       St. Xavier College      Kolkata
1       Miranda House   New Delhi
6       St. Xavier College      Kolkata
1       Miranda House   New Delhi
7       Lady Shri Ram College of Women  New Delhi
2       Loyola College  Chennai
7       Lady Shri Ram College of Women  New Delhi
2       Loyola College  Chennai
8       Dayal Singh College     New Delhi
3       Shri Ram College of Commerce    New Delhi
8       Dayal Singh College     New Delhi
3       Shri Ram College of Commerce    New Delhi
9       Deen Dayal Upadhyaya College    New Delhi
4       Bishop Herber College   Tiruchi
9       Deen Dayal Upadhyaya College    New Delhi
4       Bishop Herber College   Tiruchi
Time taken: 2.101 seconds, Fetched: 20 row(s)
hive>
```

Output of above select command shows that data has got added in the “college” table

Step 5: Updating the data in the Hive table “college”

CASE 1: Updating bucketed column i.e. clg_id

```
hive> UPDATE college
    > set clg_id=11
    > where clg_id=7;
Query ID = acadgild_20170817114949_cc0b3539-cb80-4abb-8040-95b50666f52d
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_1502950180410_0001, Tracking URL = http://localhost:8088/prox
y/application_1502950180410_0001/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_15029501804
10_0001
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-08-17 11:50:31,534 Stage-1 map = 0%,  reduce = 0%
2017-08-17 11:51:31,813 Stage-1 map = 0%,  reduce = 0%
2017-08-17 11:52:31,911 Stage-1 map = 0%,  reduce = 0%
2017-08-17 11:54:24,580 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 17.45 sec
2017-08-17 11:55:45,198 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 17.45 sec
```

```
2017-08-17 11:57:17,655 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 55.7 se
c
2017-08-17 11:58:19,084 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 61.53 s
ec
2017-08-17 11:58:20,814 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 63.88
sec
2017-08-17 11:58:43,636 Stage-1 map = 100%,  reduce = 70%, Cumulative CPU 66.61
sec
2017-08-17 11:58:45,878 Stage-1 map = 100%,  reduce = 97%, Cumulative CPU 79.99
sec
2017-08-17 11:59:04,049 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 88.51
 sec
MapReduce Total cumulative CPU time: 1 minutes 28 seconds 510 msec
Ended Job = job_1502950180410_0001
Loading data to table default.college
Table default.college stats: [numFiles=11, numRows=20, totalSize=9716, rawDataSi
ze=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 88.92 sec   HDFS Read: 12108
HDFS Write: 1098 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 28 seconds 920 msec
OK
Time taken: 614.696 seconds
hive>
```

Records updated successfully

Using select we can see records with clg_id=7 have been updated to clg_id=11

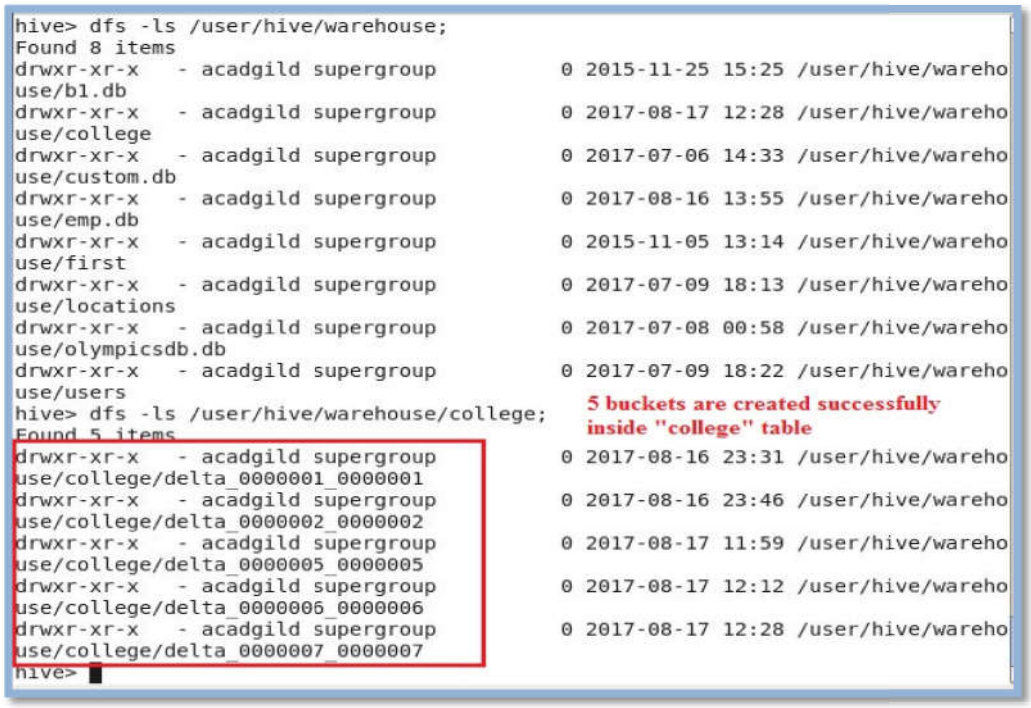
```
hive> select * from college;
OK
10      The Women Christian College     Chennai
5       Atma Ram Sanatan Dharma College New Delhi
10      The Women Christian College     Chennai
5       Atma Ram Sanatan Dharma College New Delhi
6       St. Xavier College      Kolkata
1       Miranda House   New Delhi
6       St. Xavier College      Kolkata
1       Miranda House   New Delhi
11      Lady Shri Ram College of Women  New Delhi
2       Loyola College  Chennai
11      Lady Shri Ram College of Women  New Delhi
2       Loyola College  Chennai
8       Dayal Singh College     New Delhi
3       Shri Ram College of Commerce    New Delhi
8       Dayal Singh College     New Delhi
3       Shri Ram College of Commerce    New Delhi
9       Deen Dayal Upadhyaya College    New Delhi
4       Bishop Herber College   Tiruchi
9       Deen Dayal Upadhyaya College    New Delhi
4       Bishop Herber College   Tiruchi
Time taken: 1.046 seconds, Fetched: 20 row(s)
hive>
```

This college was earlier having clg_id=7, now it's clg_id is updated to 11

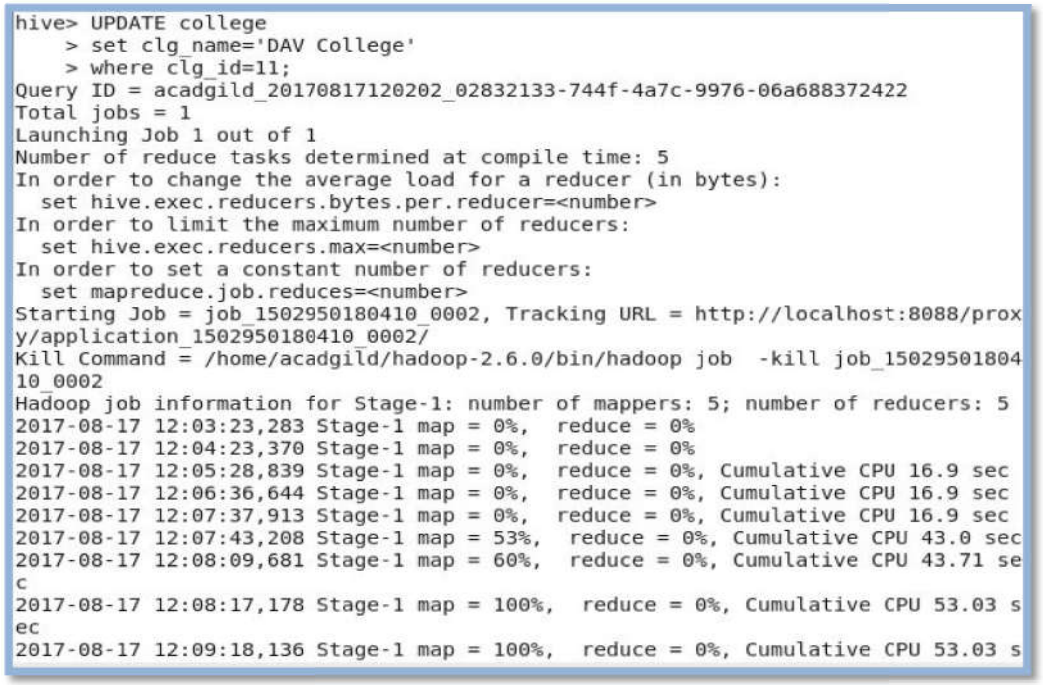
Note: Above UPDATE command, updates all the rows with clg_id=7 to clg_id=11, however, it shouldn't happen, because clg_id is bucketed column and update command is not supported for columns that are bucketed, so below error would have been returned:

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

Below screenshot shows that buckets are created inside "college" table for "clg_id" column, still it allows update operation on this bucketed column



CASE 2: Updating non-bucketed column i.e. clg_name



```
2017-08-17 12:11:19,101 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 66.48
sec
2017-08-17 12:11:22,325 Stage-1 map = 100%,  reduce = 97%, Cumulative CPU 77.76
sec
2017-08-17 12:11:54,176 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 90.65
 sec
MapReduce Total cumulative CPU time: 1 minutes 30 seconds 650 msec
Ended Job = job_1502950180410_0002
Loading data to table default.college
Table default.college stats: [numFiles=12, numRows=20, totalSize=10495, rawDataS
ize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 90.85 sec   HDFS Read: 13241
HDFS Write: 1018 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 30 seconds 850 msec
OK
Time taken: 574.429 seconds
hive>
```

Records with clg_id=11, updated successfully

Using select command, we can see records with clg_id=11 have been updated with clg_name= 'DAV College'

```
hive> select * from college;
OK
10      The Women Christian College      Chennai
5       Atma Ram Sanatan Dharma College New Delhi
10      The Women Christian College      Chennai
5       Atma Ram Sanatan Dharma College New Delhi
6       St. Xavier College       Kolkata
1       Miranda House    New Delhi
6       St. Xavier College       Kolkata
1       Miranda House    New Delhi
11      DAV College      New Delhi
2       Loyola College   Chennai
11      DAV College      New Delhi
2       Loyola College   Chennai
8       Dayal Singh College      New Delhi
3       Shri Ram College of Commerce     New Delhi
8       Dayal Singh College      New Delhi
3       Shri Ram College of Commerce     New Delhi
9       Deen Dayal Upadhyaya College     New Delhi
4       Bishop Herber College    Tiruchi
9       Deen Dayal Upadhyaya College     New Delhi
4       Bishop Herber College    Tiruchi
Time taken: 2.658 seconds, Fetched: 20 row(s)
hive>
```

Step 6: Deleting Row from Hive Table "college"

```
hive> DELETE from college
    > where clg_id=11;
Query ID = acadgild_20170817121616_13603567-9af4-434f-9eb4-5bb78f19a6e3
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_1502950180410_0003, Tracking URL = http://localhost:8088/prox
y/application_1502950180410_0003/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_15029501804
10_0003
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-08-17 12:17:13,594 Stage-1 map = 0%,  reduce = 0%
2017-08-17 12:18:14,598 Stage-1 map = 0%,  reduce = 0%
2017-08-17 12:19:15,770 Stage-1 map = 0%,  reduce = 0%
2017-08-17 12:20:18,004 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 20.1 sec
2017-08-17 12:21:22,503 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 20.1 sec
2017-08-17 12:22:27,953 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 20.1 sec
2017-08-17 12:23:30,174 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 20.1 sec
2017-08-17 12:23:34,726 Stage-1 map = 13%,  reduce = 0%, Cumulative CPU 30.32 se
c
2017-08-17 12:23:37,330 Stage-1 map = 27%,  reduce = 0%, Cumulative CPU 40.95 se
c
2017-08-17 12:23:40,231 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 41.3 sec
```

```
2017-08-17 12:24:16,492 Stage-1 map = 60%,  reduce = 0%, Cumulative CPU 51.82 se
c
2017-08-17 12:24:32,110 Stage-1 map = 80%,  reduce = 0%, Cumulative CPU 62.0 sec
2017-08-17 12:24:33,650 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 72.75 s
ec
2017-08-17 12:25:33,728 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 72.93 s
ec
2017-08-17 12:26:33,690 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 79.67
sec
2017-08-17 12:26:57,237 Stage-1 map = 100%,  reduce = 97%, Cumulative CPU 96.84
sec
2017-08-17 12:27:36,623 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 109.0
7 sec
MapReduce Total cumulative CPU time: 1 minutes 49 seconds 70 msec
Ended Job = job_1502950180410_0003
Loading data to table default.college
Table default.college stats: [numFiles=13, numRows=18, totalSize=10986, rawDataS
ize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 112.82 sec   HDFS Read: 14275
 HDFS Write: 730 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 52 seconds 820 msec
OK
Time taken: 714.614 seconds
hive>
```

Record(s) deleted successfully

Using select command, we can see that records with clg_id=11 deleted successfully

```
hive> select * from college;
OK
10      The Women Christian College     Chennai
5       Atma Ram Sanatan Dharma College New Delhi
10      The Women Christian College     Chennai
5       Atma Ram Sanatan Dharma College New Delhi
6       St. Xavier College      Kolkata
1       Miranda House   New Delhi
6       St. Xavier College      Kolkata
1       Miranda House   New Delhi
2       Loyola College  Chennai
2       Loyola College  Chennai
8       Dayal Singh College     New Delhi
3       Shri Ram College of Commerce    New Delhi
8       Dayal Singh College     New Delhi
3       Shri Ram College of Commerce    New Delhi
9       Deen Dayal Upadhyaya College    New Delhi
4       Bishop Herber College   Tiruchi
9       Deen Dayal Upadhyaya College    New Delhi
4       Bishop Herber College   Tiruchi
Time taken: 1.68 seconds, Fetched: 18 row(s)
hive>
```

Records with clg_id=11 have been deleted