

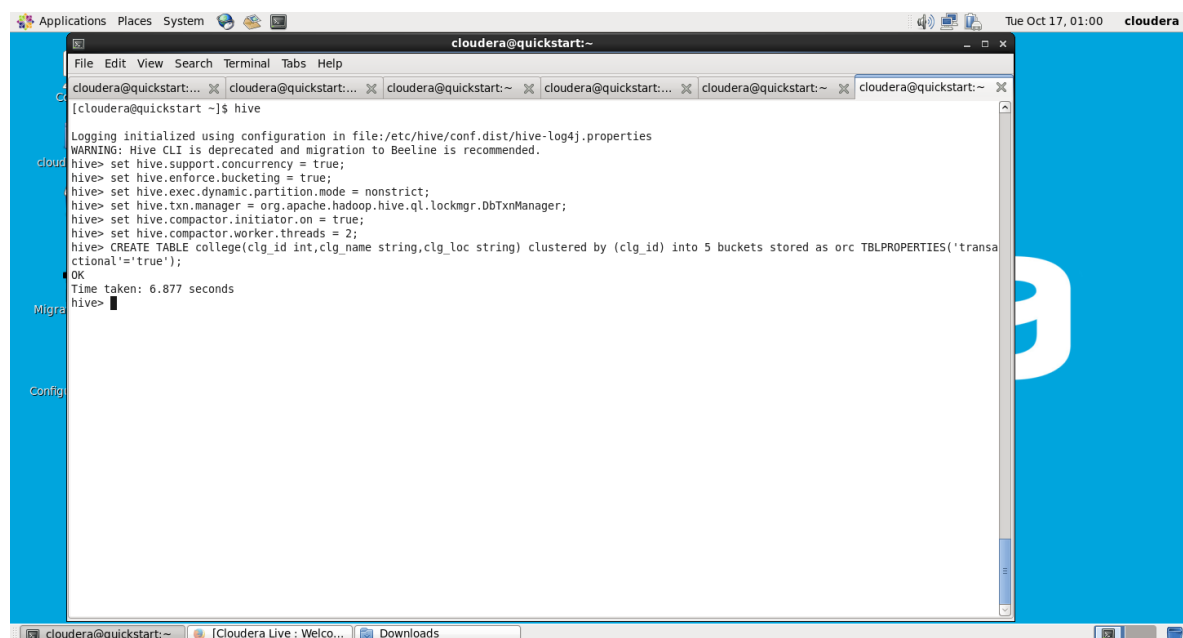
## Transactions in Hive:

The different row-level transactions available in Hive 0.14 are as follows:

- 1 Insert
- 2 Delete
- 3 Update

The below properties need to be set appropriately in *hive shell*, order-wise to work with transactions in Hive. By default they are turned off in Hive:

### Creating a Table That Supports Hive Transactions:



```
cloudera@quickstart:~$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
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 = 2;
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: 6.877 seconds
hive>
```

The create table can be checked using the command *show tables*.

### Inserting Data into a Hive Table:

```
Applications Places System cloudera@quickstart:~ Tue Oct 17, 01:05 cloudera
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~
hive> show tables;
OK
college
hive oozie
cloud Time taken: 1.019 seconds, Fetched: 2 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');
Query ID = cloudera_20171017010202_08cc8ef2-bcd6-4b93-b5e1-ed0aa53ebfa5
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>
Migra In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1507287275802_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1507287275802_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1507287275802_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2017-10-17 01:02:59,887 Stage-1 map = 0%, reduce = 0%
2017-10-17 01:03:24,285 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.34 sec
2017-10-17 01:04:20,704 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 5.87 sec
2017-10-17 01:04:23,662 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 7.47 sec
2017-10-17 01:04:26,519 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 9.18 sec
2017-10-17 01:04:33,877 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 15.91 sec
2017-10-17 01:04:36,247 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 17.4 sec
MapReduce Total cumulative CPU time: 17 seconds 400 msec
Ended Job = job_1507287275802_0016
Loading data to table default.college
Table default.college stats: [numFiles=5, numRows=7, totalSize=3626, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 17.4 sec HDFS Read: 21991 HDFS Write: 4001 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 400 msec
OK
Time taken: 138.666 seconds
hive>
```

Contents of the table can be viewed using :

```
Applications Places System cloudera@quickstart:~ Tue Oct 17, 01:35 cloudera
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~
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>
Migra In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1507287275802_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1507287275802_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1507287275802_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2017-10-17 01:02:59,887 Stage-1 map = 0%, reduce = 0%
2017-10-17 01:03:24,285 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.34 sec
2017-10-17 01:04:20,704 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 5.87 sec
2017-10-17 01:04:23,662 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 7.47 sec
2017-10-17 01:04:26,519 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 9.18 sec
2017-10-17 01:04:33,877 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 15.91 sec
2017-10-17 01:04:36,247 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 17.4 sec
MapReduce Total cumulative CPU time: 17 seconds 400 msec
Ended Job = job_1507287275802_0016
Loading data to table default.college
Table default.college stats: [numFiles=5, numRows=7, totalSize=3626, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 17.4 sec HDFS Read: 21991 HDFS Write: 4001 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 400 msec
OK
Time taken: 138.666 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: 0.251 seconds, Fetched: 7 row(s)
```

Updating the Data in Hive Table:

The image displays two screenshots of a Cloudera Quickstart terminal window. The top screenshot shows a Hive query being executed: `hive> UPDATE college set clg_id = 8 where clg_id = 7;`. The output indicates that 7 rows were fetched and 1 job was launched. The job progress shows various stages of map and reduce tasks, with a final status of 'SUCCESS'. The bottom screenshot shows a similar Hive query: `hive> UPDATE college set clg_name = 'IIT' where clg_id = 6;`. The output shows 7 rows fetched and 1 job launched. The job progress is also shown, with a final status of 'SUCCESS'. Both screenshots include a table of results for the 'college' table, showing columns like 'stanford', 'JNTUA', 'nec', 'cambridge', 'vit', 'vlr', 'uk', 'atp', 'nlr', and 'us'.

```
cloudera@quickstart:~  
File Edit View Search Terminal Tabs Help  
cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x  
Time taken: 0.251 seconds, Fetched: 7 row(s)  
hive> UPDATE college set clg_id = 8 where clg_id = 7;  
Query ID = cloudera_20171017010606_1b7d941d-ee1b-4301-ad4b-492714faf57d  
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_1507287275802_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1507287275802_0017/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1507287275802_0017  
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5  
2017-10-17 01:06:57,832 Stage-1 map = 0%, reduce = 0%  
2017-10-17 01:07:51,514 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 7.58 sec  
2017-10-17 01:07:52,800 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 10.01 sec  
2017-10-17 01:07:54,004 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.76 sec  
2017-10-17 01:08:41,064 Stage-1 map = 100%, reduce = 20%, Cumulative CPU 14.72 sec  
2017-10-17 01:08:42,498 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 20.93 sec  
2017-10-17 01:08:43,939 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 23.21 sec  
MapReduce Total cumulative CPU time: 23 seconds 410 msec  
Ended Job = job_1507287275802_0017  
Loading data to table default.college  
Table default.college stats: [numFiles=6, numRows=7, totalSize=4362, rawDataSize=0]  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 23.41 sec HDFS Read: 50173 HDFS Write: 975 SUCCESS  
Total MapReduce CPU Time Spent: 23 seconds 410 msec  
OK  
Time taken: 129.685 seconds  
hive> select * from college;  
OK  
5      stanford      uk  
6      JNTUA      atp  
1      nec      nlr  
8      cambridge      us  
-----  
cloudera@quickstart:~  
File Edit View Search Terminal Tabs Help  
cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x cloudera@quickstart:~ x  
4      lpu      del  
Time taken: 0.262 seconds, Fetched: 7 row(s)  
hive> UPDATE college set clg_name = 'IIT' where clg_id = 6;  
Query ID = cloudera_20171017011212_90be54b6-5a42-4b9d-9fal-ad488eea8600  
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_1507287275802_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1507287275802_0018/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1507287275802_0018  
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5  
2017-10-17 01:13:19,160 Stage-1 map = 0%, reduce = 0%  
2017-10-17 01:14:05,620 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 9.21 sec  
2017-10-17 01:14:06,869 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 9.65 sec  
2017-10-17 01:14:08,066 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 10.58 sec  
2017-10-17 01:14:51,673 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 19.27 sec  
MapReduce Total cumulative CPU time: 19 seconds 560 msec  
Ended Job = job_1507287275802_0018  
Loading data to table default.college  
Table default.college stats: [numFiles=7, numRows=7, totalSize=5061, rawDataSize=0]  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 19.56 sec HDFS Read: 50145 HDFS Write: 938 SUCCESS  
Total MapReduce CPU Time Spent: 19 seconds 560 msec  
OK  
Time taken: 117.04 seconds  
hive> select * from college;  
OK  
5      stanford      uk  
6      IIT      atp  
1      nec      nlr  
8      cambridge      us  
2      vit      vlr
```

## Deleting a Row from Hive Table:

```
Applications Places System cloudera@quickstart:~ Tue Oct 17, 01:20 cloudera
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~
hive> delete from college where clg_id=5;
Query ID = cloudera_20171017011717_e7de6ead-fc96-4edf-b482-cc546912b536
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_1507287275802_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1507287275802_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1507287275802_0019
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-10-17 01:18:02,026 Stage-1 map = 0%, reduce = 0%
2017-10-17 01:18:45,108 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 1.98 sec
2017-10-17 01:18:46,392 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 5.87 sec
2017-10-17 01:18:47,504 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 9.57 sec
2017-10-17 01:18:48,645 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.77 sec
2017-10-17 01:19:34,571 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 13.16 sec
2017-10-17 01:19:35,828 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 16.72 sec
2017-10-17 01:19:37,006 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 18.76 sec
MapReduce Total cumulative CPU time: 18 seconds 760 msec
Ended Job = job_1507287275802_0019
Loading data to table default.college
Table default.college stats: [numFiles=0, numRows=6, totalSize=5578, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 18.76 sec HDFS Read: 48164 HDFS Write: 756 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 760 msec
OK
Time taken: 113.721 seconds
hive> select * from college;
```

Check deleted row using select \* from college:

```
Applications Places System cloudera@quickstart:~ Tue Oct 17, 01:21 cloudera
File Edit View Se Preferred Mail Reader Send email Help
cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~
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_1507287275802_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1507287275802_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1507287275802_0019
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-10-17 01:18:02,026 Stage-1 map = 0%, reduce = 0%
2017-10-17 01:18:45,108 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 1.98 sec
2017-10-17 01:18:46,392 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 5.87 sec
2017-10-17 01:18:47,504 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 9.57 sec
2017-10-17 01:18:48,645 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.77 sec
2017-10-17 01:19:34,571 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 13.16 sec
2017-10-17 01:19:35,828 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 16.72 sec
2017-10-17 01:19:37,006 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 18.76 sec
MapReduce Total cumulative CPU time: 18 seconds 760 msec
Ended Job = job_1507287275802_0019
Loading data to table default.college
Table default.college stats: [numFiles=0, numRows=6, totalSize=5578, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 18.76 sec HDFS Read: 48164 HDFS Write: 756 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 760 msec
OK
Time taken: 113.721 seconds
hive> select * from college;
OK
6      IIT      atp
1      nec      nlr
8      cambridge  us
2      vit      vlr
3      srm      chen
4      lpu      del
Time taken: 0.122 seconds, Fetched: 6 row(s)
hive>
```