

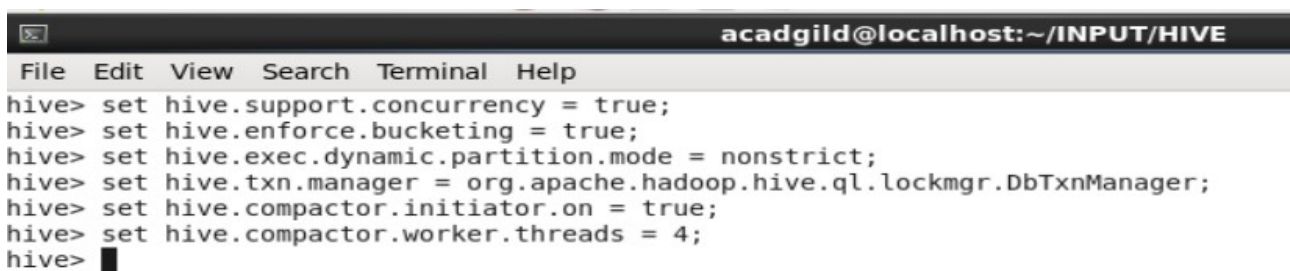
Bigdata Assignment 3.6

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 -

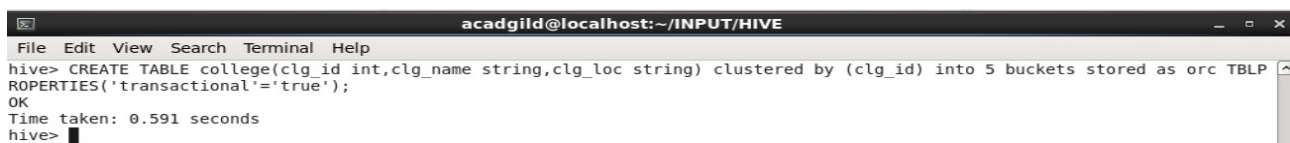
- Entered into hive shell by -
start-all.sh
hive
- The properties of the hive is set so that Insert , Update and Delete operation can be performed.

A terminal window titled 'acadgild@localhost:~/INPUT/HIVE' showing the Hive shell interface. The prompt is 'hive>'. Several properties are being set: 'hive.support.concurrency = true', 'hive.enforce.bucketing = true', 'hive.exec.dynamic.partition.mode = nonstrict', 'hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager', 'hive.compactor.initiator.on = true', and 'hive.compactor.worker.threads = 4'. The prompt returns to 'hive>' after the last command.

```
acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
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 = 4;
hive>
```

- A table is created with name college and having columns like college id , name and location. The table format is set to 'orc' and the transactions are enabled in the table by specifying it inside the *TBLPROPERTIES* as 'transactional'='true'

**CREATE TABLE college(clg_id int, clg_name String, clg_loc String) into 5 buckets stored as orc
TBLPROPERTIES('transactional'='true') ;**

A terminal window titled 'acadgild@localhost:~/INPUT/HIVE' showing the Hive shell interface. The prompt is 'hive>'. The command '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');' is entered. The prompt returns to 'hive>' after the command is executed. Below the command, it says 'Time taken: 0.591 seconds' and 'hive>'.

```
acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
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: 0.591 seconds
hive>
```

- Table is created as shown in the below screenshot
show tables ;

```

File Edit View Search Terminal Help
hive> show tables;
OK
college
employee
olympic_data
temp_table
temperature_data
temperature_data_vw
Time taken: 0.138 seconds, Fetched: 6 row(s)
hive> █

```

- Data is inserted into the table

INSERT INTO table college values(1,'cet','bbsr') , (2,'vit','vlr') , (3,'srm','chen') , (4,'lpu','del') , (5,'stanford','uk') , (6,'bits','pilani') , (7,'cambridge','us') ;

```

acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
hive> INSERT INTO table college values(1,'cet','bbsr'),(2,'vit','vlr'),(3,'srm','chen'),(4,'lpu','del'),(5,'stanford','uk'),(
6,'bits','pilani'),(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 execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180606190941_89ec4b30-110a-4b44-b040-7fb9682f6143
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_1528279835710_0005, Tracking URL = http://localhost:8088/proxy/application_1528279835710_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528279835710_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-06-06 19:09:50,671 Stage-1 map = 0%, reduce = 0%
2018-06-06 19:09:58,384 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.98 sec
2018-06-06 19:10:16,385 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 5.0 sec
2018-06-06 19:10:23,770 Stage-1 map = 100%, reduce = 27%, Cumulative CPU 6.82 sec
2018-06-06 19:10:24,942 Stage-1 map = 100%, reduce = 47%, Cumulative CPU 11.73 sec
2018-06-06 19:10:28,616 Stage-1 map = 100%, reduce = 60%, Cumulative CPU 13.93 sec
2018-06-06 19:10:29,911 Stage-1 map = 100%, reduce = 73%, Cumulative CPU 16.0 sec
2018-06-06 19:10:31,155 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 21.86 sec
2018-06-06 19:10:34,566 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 27.13 sec
MapReduce Total cumulative CPU time: 27 seconds 130 msec
Ended Job = job_1528279835710_0005
Loading data to table custom.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 27.13 sec HDFS Read: 27059 HDFS Write: 4004 SUCCESS
Total MapReduce CPU Time Spent: 27 seconds 130 msec
OK
Time taken: 56.36 seconds

```

- The content of the table -

```

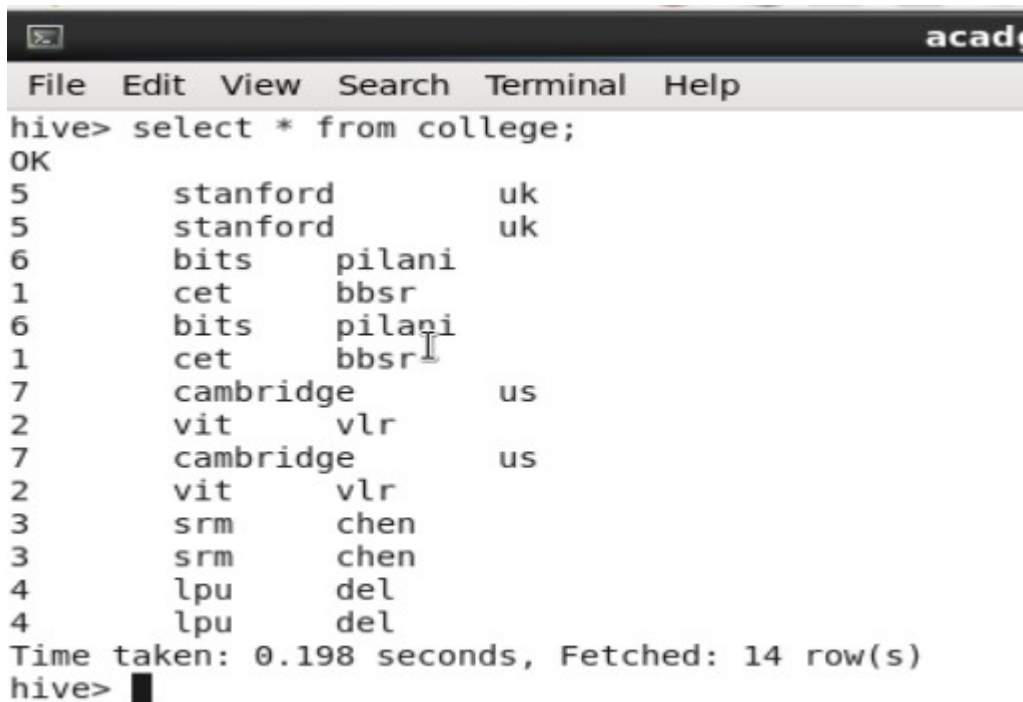
File Edit View Search Terminal Help
hive> select * from college;
OK
5      stanford      uk
6      bits          pilani
1      cet           bbsr
7      cambridge     us
2      vit           vlr
3      srm           chen
4      lpu           del
Time taken: 0.298 seconds, Fetched: 7 row(s)
hive> █

```

**select * from
college ;**

- If it is inserted again , it gets appended.

```
INSERT INTO table college values(1,'cet','bbsr') , (2,'vit','vlr') ,  
(3,'srm','chen') , (4,'lpu','del') , (5,'stanford','uk') ,  
(6,'bits','pilani') , (7,'cambridge','us') ;  
select * from college ;
```



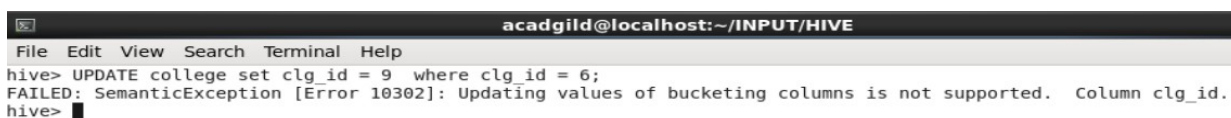
The screenshot shows a Hive terminal window with the following output for the query `select * from college;`:

clg_id	clg_name	clg_city
5	stanford	uk
5	stanford	uk
6	bits	pilani
1	cet	bbsr
6	bits	pilani
1	cet	bbsr
7	cambridge	us
2	vit	vlr
7	cambridge	us
2	vit	vlr
3	srm	chen
3	srm	chen
4	lpu	del
4	lpu	del

Time taken: 0.198 seconds, Fetched: 14 row(s)

- If we try to update a bucketed column then it will give a semantic exception.

```
UPDATE college set clg_id = 9 where clg_id = 6;
```



The screenshot shows a Hive terminal window where the command `UPDATE college set clg_id = 9 where clg_id = 6;` was executed. The output is an error message:

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

- We can update only non-bucketed columns.

```
UPDATE college set clg_name = 'NIT' where clg_id = 5;
```

```
acagdild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
hive> UPDATE college set clg_name = 'NIT' 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 = acagdild_20180606193325_929f7605-4d2e-4c1d-a093-9d1e5fc798db
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_1528279835710_0010, Tracking URL = http://localhost:8088/proxy/application_1528279835710_0010/
Kill Command = /home/acagdild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528279835710_0010
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-06-06 19:33:34,483 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 3.92 sec
2018-06-06 19:33:56,989 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 7.53 sec
2018-06-06 19:33:59,358 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 14.81 sec
2018-06-06 19:34:04,225 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 18.57 sec
2018-06-06 19:34:06,584 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 18.57 sec
2018-06-06 19:34:20,862 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 20.5 sec
2018-06-06 19:34:22,075 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 24.43 sec
2018-06-06 19:34:24,493 Stage-1 map = 100%, reduce = 60%, Cumulative CPU 27.45 sec
2018-06-06 19:34:26,660 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 33.1 sec
MapReduce Total cumulative CPU time: 33 seconds 100 msec
Ended Job = job_1528279835710_0010
Loading data to table custom.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 33.1 sec HDFS Read: 56427 HDFS Write: 971 SUCCESS
Total MapReduce CPU Time Spent: 33 seconds 100 msec
OK
Time taken: 63.139 seconds
hive>
```

- Content of the table is displayed to confirm that it is updated.

```
aca
File Edit View Search Terminal Help
hive> select * from college;
OK
5      NIT      uk
5      NIT      uk
6      bits     pilani
1      cet      bbsr
6      bits     pilani
1      cet      bbsr
7      cambridge      us
2      vit      vlr
7      cambridge      us
2      vit      vlr
3      srm       chen
3      srm       chen
4      lpu       del
4      lpu       del
Time taken: 0.191 seconds, Fetched: 14 row(s)
hive>
```

- Deleted a row
delete from college where clg_id=3 ;

```

acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
hive> delete from college where clg_id=3;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180606193710_92136c58-5e3e-4146-88e5-5c46b4005360
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_1528279835710_0011, Tracking URL = http://localhost:8088/proxy/application_1528279835710_0011/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528279835710_0011
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-06-06 19:37:18,725 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 7.22 sec
2018-06-06 19:37:43,398 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 7.95 sec
2018-06-06 19:37:45,632 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 11.63 sec
2018-06-06 19:37:46,842 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 18.43 sec
2018-06-06 19:37:50,445 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 18.95 sec
2018-06-06 19:37:52,845 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.57 sec
2018-06-06 19:38:08,320 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 23.22 sec
2018-06-06 19:38:09,512 Stage-1 map = 100%, reduce = 27%, Cumulative CPU 27.42 sec
2018-06-06 19:38:10,756 Stage-1 map = 100%, reduce = 60%, Cumulative CPU 31.92 sec
2018-06-06 19:38:13,173 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 33.44 sec
2018-06-06 19:38:14,303 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 34.91 sec
2018-06-06 19:38:15,349 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 34.91 sec
MapReduce Total cumulative CPU time: 34 seconds 910 msec
Ended Job = job_1528279835710_0011
Loading data to table custom.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 34.91 sec HDFS Read: 54645 HDFS Write: 762 SUCCESS
Total MapReduce CPU Time Spent: 34 seconds 910 msec
OK
Time taken: 68.049 seconds
hive>

```

- The content of the table is displayed to confirm it is deleted or not .
As 12 rows are there so it is deleted.

```

acad
File Edit View Search Terminal Help
hive> select * from college;
OK
5      NIT      uk
5      NIT      uk
6      bits    pilani
1      cet     bbsr
6      bits    pilani
1      cet     bbsr
7      cambridge    us
2      vit     vlr
7      cambridge    us
2      vit     vlr
4      lpu     del
4      lpu     del
Time taken: 0.166 seconds, Fetched: 12 row(s)
hive>

```