Assignment 8.3

1. Setting HIVE properties 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>
```

2. Created table that supports HIVE transactions

```
hive> CREATE TABLE employee(
    > emp_id int,
    > emp_name string,
    > emp_salary int,
    > emp_department string,
    > unit int
    > )
    > clustered by (emp_id) into 5 buckets stored as orc TBLPROPERTIES('transactional'='true');
OK
Time taken: 0.376 seconds
hive>
```

3. Insert data into HIVE table:

```
hive> INSERT INTO table employee values (101,'Amitabh',256,'Finance',1),(102,'Shahrukh',78,'IT_Dept',2),(103,'Akshay',110,'HR',3),
    > (104,'Anubhav',50,'Network_Team',4),(105,'Pawan',250,'Admin',5),(106,'Aamir',25,'Finance',1),(107,'Salman',175,'IT_Dept',2),(108,'Ranbir',142,'HR',3),(109,'
Katrina',100,'Network_Team',4),(110,'Priyanka',222,'Admin',5),(111,'Tushar',500,'Finance',1),(112,'Ajay',58,'IT_Dept',2),(113,'Jubeen',100,'Finance',1),(114,'Madh
uri',200,'IT_Dept',2);
Query ID = acadgild_20171120210101_cc63d169-5ecc-4de4-b64c-673a01a0c0a5
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1511187342194_0001, Tracking URL = http://localhost:8088/proxy/application_1511187342194_0001/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1511187342194_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2017-11-20 21:02:23,998 Stage-1 map = 0%,  reduce = 0%
2017-11-20 21:02:37,420 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.96 sec
MapReduce Total cumulative CPU time: 2 seconds 960 msec
Ended Job = job_1511187342194_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://localhost:9000/tmp/hive/acadgild/59ebbe3f-29b8-4435-b816-7df81cab3cdd/hive_2017-11-20_21-01-56_712_7984791112795110294-1/-ext-10000
Loading data to table sumona.employee
Table sumona.employee stats: [numFiles=1, numRows=14, totalSize=729, rawDataSize=2688]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 2.96 sec   HDFS Read: 622 HDFS Write: 802 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 960 msec
OK
Time taken: 44.174 seconds
hive>
```

```
hive> SELECT * FROM employee;
OK
101     Amitabh 256         Finance 1
102     Shahrukh            78      IT_Dept 2
103     Akshay  110         HR      3
104     Anubhav 50          Network_Team    4
105     Pawan   250         Admin   5
106     Aamir   25          Finance 1
107     Salman  175         IT_Dept 2
108     Ranbir  142         HR      3
109     Katrina 100         Network_Team    4
110     Priyanka            222     Admin   5
111     Tushar  500         Finance 1
112     Ajay    58          IT_Dept 2
113     Jubeen  100         Finance 1
114     Madhuri 200         IT_Dept 2
Time taken: 0.221 seconds, Fetched: 14 row(s)
hive>
```

4. Updating the values in the table:

UPDATE employee set emp_id = 108 where emp_id = 112;

The above command is used to update a row in HIVE table.

```
hive> UPDATE employee set emp_id = 108 where emp_id = 112;
FAILED: SemanticException [Error 10302]: Updating values of bucketing columns is not supported.  Column emp_id.
```

UPDATE college set clg_name = 'IIT' where clg_id = 6;

The above command is used to update on NON bucketed column

```
hive> UPDATE employee set emp_name = 'Anushka' where emp_id = 105;
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 usin
g Hive 1.X releases.
Query ID = acadgild_20171121122113_90f19e37-b1ef-4f1a-b889-97c30ff4fe07
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_1511246161557_0002, Tracking URL = http://localhost:8088/proxy/application_1511246161557_0002/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1511246161557_0002
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-11-21 12:21:22,561 Stage-1 map = 0%,   reduce = 0%
2017-11-21 12:21:47,201 Stage-1 map = 20%,  reduce = 0%, Cumulative CPU 1.88 sec
2017-11-21 12:21:49,763 Stage-1 map = 40%,  reduce = 0%, Cumulative CPU 3.68 sec
2017-11-21 12:21:52,519 Stage-1 map = 60%,  reduce = 0%, Cumulative CPU 7.1 sec
2017-11-21 12:21:53,805 Stage-1 map = 80%,  reduce = 0%, Cumulative CPU 8.89 sec
2017-11-21 12:21:55,055 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.09 sec
2017-11-21 12:22:12,743 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 10.16 sec
2017-11-21 12:22:14,006 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 11.21 sec
2017-11-21 12:22:15,463 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 12.54 sec
2017-11-21 12:22:16,643 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 14.46 sec
2017-11-21 12:22:18,071 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 16.29 sec
MapReduce Total cumulative CPU time: 16 seconds 290 msec
Ended Job = job_1511246161557_0002
Loading data to table sumona.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 16.29 sec   HDFS Read: 62182 HDFS Write: 1113 SUCCESS
Total MapReduce CPU Time Spent: 16 seconds 290 msec
OK
Time taken: 66.574 seconds
hive>
```

Now, if we compare the initial table and the table after updating the row, you will notice that the row 105 has been updated with the name Anushka.

INITIAL TABLE

```
hive> SELECT * FROM employee;
OK
101     Amitabh 256     Finance 1
102     Shahrukh        78      IT_Dept 2
103     Akshay  110     HR      3
104     Anubhav 50      Network_Team    4
105     Pawan   250     Admin   5
106     Aamir   25      Finance 1
107     Salman  175     IT_Dept 2
108     Ranbir  142     HR      3
109     Katrina 100     Network_Team    4
110     Priyanka        222     Admin   5
111     Tushar  500     Finance 1
112     Ajay    58      IT_Dept 2
113     Jubeen  100     Finance 1
114     Madhuri 200     IT_Dept 2
Time taken: 0.221 seconds, Fetched: 14 row(s)
hive>
```

TABLE AFTER UPDATING THE ROW

```
hive> SELECT * FROM employee;
OK
105     Anushka 250     Admin   5
110     Priyanka        222     Admin   5
106     Aamir   25      Finance 1
101     Amitabh 256     Finance 1
111     Tushar  500     Finance 1
107     Salman  175     IT_Dept 2
112     Ajay    58      IT_Dept 2
102     Shahrukh        78      IT_Dept 2
113     Jubeen  100     Finance 1
103     Akshay  110     HR      3
108     Ranbir  142     HR      3
109     Katrina 100     Network_Team    4
104     Anubhav 50      Network_Team    4
114     Madhuri 200     IT_Dept 2
Time taken: 0.184 seconds, Fetched: 14 row(s)
hive>
```

5. Deleting a Row from HIVE table:

```
hive> delete from employee where emp_id= 107;
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 usin
g Hive 1.X releases.
Query ID = acadgild_20171121123313_50eaf3bc-cb02-4c9a-929a-c83e6bed4881
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_1511246161557_0003, Tracking URL = http://localhost:8088/proxy/application_1511246161557_0003/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1511246161557_0003
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-11-21 12:33:21,782 Stage-1 map = 0%,   reduce = 0%
2017-11-21 12:33:41,829 Stage-1 map = 20%,   reduce = 0%, Cumulative CPU 1.76 sec
2017-11-21 12:33:45,857 Stage-1 map = 40%,   reduce = 0%, Cumulative CPU 3.57 sec
2017-11-21 12:33:50,104 Stage-1 map = 60%,   reduce = 0%, Cumulative CPU 6.95 sec
2017-11-21 12:33:51,439 Stage-1 map = 80%,   reduce = 0%, Cumulative CPU 8.72 sec
2017-11-21 12:33:54,018 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU 8.99 sec
2017-11-21 12:34:06,472 Stage-1 map = 100%,   reduce = 13%, Cumulative CPU 10.08 sec
2017-11-21 12:34:07,901 Stage-1 map = 100%,   reduce = 27%, Cumulative CPU 11.15 sec
2017-11-21 12:34:09,165 Stage-1 map = 100%,   reduce = 40%, Cumulative CPU 11.78 sec
2017-11-21 12:34:10,320 Stage-1 map = 100%,   reduce = 53%, Cumulative CPU 12.9 sec
2017-11-21 12:34:11,447 Stage-1 map = 100%,   reduce = 60%, Cumulative CPU 13.33 sec
2017-11-21 12:34:12,631 Stage-1 map = 100%,   reduce = 100%, Cumulative CPU 15.94 sec
MapReduce Total cumulative CPU time: 15 seconds 940 msec
Ended Job = job_1511246161557_0003
Loading data to table sumona.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5   Cumulative CPU: 15.94 sec   HDFS Read: 57207 HDFS Write: 757 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 940 msec
OK
Time taken: 61.455 seconds
hive>
```

Now, if we compare the initial table and the table after deleting the row, you will notice that the row 107 has been deleted.

INITIAL TABLE   **TABLE AFTER DLEELTING A ROW**

```
hive> SELECT * FROM employee;
OK
101     Amitabh 256     Finance 1
102     Shahrukh        78      IT_Dept 2
103     Akshay  110     HR      3
104     Anubhav 50      Network_Team    4
105     Pawan   250     Admin   5
106     Aamir   25      Finance 1
107     Salman  175     IT_Dept 2
108     Ranbir  142     HR      3
109     Katrina 100     Network_Team    4
110     Priyanka        222     Admin   5
111     Tushar  500     Finance 1
112     Ajay    58      IT_Dept 2
113     Jubeen  100     Finance 1
114     Madhuri 200     IT_Dept 2
Time taken: 0.221 seconds, Fetched: 14 row(s)
hive>
```

```
hive> SELECT * FROM employee;
OK
105     Anushka 250     Admin   5
110     Priyanka        222     Admin   5
106     Aamir   25      Finance 1
101     Amitabh 256     Finance 1
111     Tushar  500     Finance 1
112     Ajay    58      IT_Dept 2
102     Shahrukh        78      IT_Dept 2
113     Jubeen  100     Finance 1
103     Akshay  110     HR      3
108     Ranbir  142     HR      3
109     Katrina 100     Network_Team    4
104     Anubhav 50      Network_Team    4
114     Madhuri 200     IT_Dept 2
Time taken: 0.142 seconds, Fetched: 13 row(s)
hive>
```