Assignment 8.3

Problem Statement :-

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:-

Transactions in Hive:-

Transactions are provided at the row-level in Hive 0.14. The different row-level transactions available in Hive 0.14 are as follows:

- Insert
- Delete
- Update

There are numerous limitations with the present transactions available in Hive 0.14. ORC is the file format supported by Hive transaction. It is now essential to have ORC file format for performing transactions in Hive. The table needs to be bucketed in order to support transactions.

Row-level Transactions:-

The below properties needs to be set appropriately in hive shell , order-wise to work with transactions in Hive:-

```
hive> set hive.support.concurrency = true;
hive> set hive.enforce.bucketing = true;
Query returned non-zero code: 1, cause: hive configuration hive.enforce.bucketing does not exists.
hive> set hive.exec.dynamic.partitio.mode = nonstrict;
Query returned non-zero code: 1, cause: hive configuration hive.exec.dynamic.partitio.mode does not exists.
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 = 5;
hive> set hive.compactor.worker.threads;
hive.compactor.worker.threads=5
hive> use custom;
OK
Time taken: 2.573 seconds
```

Creating a Table That Supports Hive Transactions:-

```
hive> CREATE TABLE employee_details_8_3
    > (
    > emp_id INT,
    > emp_name STRING,
    > salary INT,
    > department STRING,
    > unit INT
    > )
    > clustered by (emp_id)
    > into 5 buckets
    > stored as orc
    > TBLPROPERTIES('transactional' = 'true');
OK
Time taken: 2.766 seconds
hive> show tables;
OK
employee
employee_details
employee_details_8_3
olympix_data
temperature_data_vw
temperature_table
Time taken: 1.048 seconds, Fetched: 6 row(s)
hive>
```

Inserting Data into a Hive Table

```
hive> INSERT INTO table employee_details_8_3 values (101,'Amitabh',256,'Finance',1),(102,'Shahrukh',78,'IT_Dept',2),(103,'Akshay',110,'HR',3),(104,'Anubha
v',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',1
00,'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,'Madhur
i',200,'IT_Dept',2);
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_20171116000512_41576dc2-f001-49e9-97a7-f49fda296e33
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_1510770641633_0001, Tracking URL = http://localhost:8088/proxy/application_1510770641633_0001/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510770641633_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2017-11-16 00:06:05,146 Stage-1 map = 0%,  reduce = 0%
2017-11-16 00:06:28,343 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.77 sec
2017-11-16 00:07:29,768 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.77 sec
2017-11-16 00:07:50,213 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 7.92 sec
2017-11-16 00:07:52,149 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 9.16 sec
2017-11-16 00:08:20,732 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 23.29 sec
2017-11-16 00:08:22,720 Stage-1 map = 100%,  reduce = 79%, Cumulative CPU 31.32 sec
2017-11-16 00:08:27,637 Stage-1 map = 100%,  reduce = 96%, Cumulative CPU 34.42 sec
2017-11-16 00:08:29,685 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 35.1 sec
MapReduce Total cumulative CPU time: 35 seconds 100 msec
Ended Job = job_1510770641633_0001
Loading data to table custom.employee_details_8_3
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 35.1 sec   HDFS Read: 31028 HDFS Write: 5159 SUCCESS
Total MapReduce CPU Time Spent: 35 seconds 100 msec
OK
Time taken: 201.3 seconds
hive>
```

```
hive> select * from employee_details_8_3;
OK
105     Pawan   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: 0.504 seconds, Fetched: 14 row(s)
hive>
```

```
hive> INSERT INTO table employee_details_8_3 values (101,'Amitabh',256,'Finance',1),(102,'Shahrukh',78,'IT_Dept',2),(103,'Akshay',110,'HR',3),(104,'Anubha
v',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',1
00,'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,'Madhur
i',200,'IT_Dept',2);
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_20171116001324_9a3fb9e5-f3f3-464f-ba95-38b9285e2fbe
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_1510770641633_0002, Tracking URL = http://localhost:8088/proxy/application_1510770641633_0002/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510770641633_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2017-11-16 00:13:57,947 Stage-1 map = 0%,  reduce = 0%
2017-11-16 00:14:18,896 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.25 sec
2017-11-16 00:15:20,311 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.25 sec
2017-11-16 00:15:38,180 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 5.24 sec
2017-11-16 00:15:40,387 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 6.65 sec
2017-11-16 00:15:44,375 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 8.99 sec
2017-11-16 00:15:46,390 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 10.29 sec
2017-11-16 00:16:08,857 Stage-1 map = 100%,  reduce = 70%, Cumulative CPU 15.29 sec
2017-11-16 00:16:13,485 Stage-1 map = 100%,  reduce = 78%, Cumulative CPU 23.95 sec
2017-11-16 00:16:15,506 Stage-1 map = 100%,  reduce = 82%, Cumulative CPU 31.93 sec
2017-11-16 00:16:17,563 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 32.55 sec
2017-11-16 00:16:19,409 Stage-1 map = 100%,  reduce = 91%, Cumulative CPU 33.36 sec
2017-11-16 00:16:21,421 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 34.68 sec
MapReduce Total cumulative CPU time: 34 seconds 680 msec
Ended Job = job_1510770641633_0002
Loading data to table custom.employee_details_8_3
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 34.68 sec   HDFS Read: 30788 HDFS Write: 5154 SUCCESS
Total MapReduce CPU Time Spent: 34 seconds 680 msec
OK
Time taken: 179.96 seconds
hive>
```

```
hive> select * from employee_details_8_3;
OK
105     Pawan   250     Admin   5
110     Priyanka        222     Admin   5
105     Pawan   250     Admin   5
110     Priyanka        222     Admin   5
106     Aamir   25      Finance 1
101     Amitabh 256     Finance 1
111     Tushar  500     Finance 1
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
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
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
109     Katrina 100     Network_Team    4
104     Anubhav 50      Network_Team    4
114     Madhuri 200     IT_Dept 2
Time taken: 0.463 seconds, Fetched: 28 row(s)
hive>
```

Updating the Data in Hive Table

```
hive> UPDATE employee_details_8_3 set emp_id = 115 where emp_id = 114;
FAILED: SemanticException [Error 10302]: Updating values of bucketing columns is not supported.  Column emp_id.
hive>
```

```
hive> UPDATE college set emp_name = 'Lata' where emp_id = 114;
FAILED: SemanticException [Error 10001]: Table not found custom.college
hive> UPDATE employee_details_8_3 set emp_name = 'Lata' where emp_id = 114;
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_20171116002603_720a6603-a32b-4427-be94-026961bad261
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_1510770641633_0003, Tracking URL = http://localhost:8088/proxy/application_1510770641633_0003/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510770641633_0003
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-11-16 00:26:36,275 Stage-1 map = 0%,  reduce = 0%
2017-11-16 00:27:36,370 Stage-1 map = 0%,  reduce = 0%
2017-11-16 00:28:36,837 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 15.19 sec
2017-11-16 00:28:45,799 Stage-1 map = 40%,  reduce = 0%, Cumulative CPU 24.23 sec
2017-11-16 00:28:55,721 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 40.78 sec
2017-11-16 00:29:56,555 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 40.78 sec
2017-11-16 00:30:12,529 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 42.6 sec
2017-11-16 00:30:16,458 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 44.94 sec
2017-11-16 00:30:20,419 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 45.83 sec
2017-11-16 00:30:22,389 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 47.1 sec
2017-11-16 00:30:34,119 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 52.95 sec
2017-11-16 00:30:36,163 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 55.76 sec
2017-11-16 00:30:38,915 Stage-1 map = 100%,  reduce = 93%, Cumulative CPU 58.42 sec
2017-11-16 00:30:40,776 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 61.74 sec
MapReduce Total cumulative CPU time: 1 minutes 1 seconds 740 msec
Ended Job = job_1510770641633_0003
Loading data to table custom.employee_details_8_3
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 61.74 sec   HDFS Read: 68430 HDFS Write: 1209 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 1 seconds 740 msec
OK
Time taken: 279.832 seconds
hive>
```

```
hive> select * from employee_details_8_3;
OK
105     Pawan   250     Admin   5
110     Priyanka        222     Admin   5
105     Pawan   250     Admin   5
110     Priyanka        222     Admin   5
106     Aamir   25      Finance 1
101     Amitabh 256     Finance 1
111     Tushar  500     Finance 1
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
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
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     Lata    200     IT_Dept 2
109     Katrina 100     Network_Team    4
104     Anubhav 50      Network_Team    4
114     Lata    200     IT_Dept 2
Time taken: 0.506 seconds, Fetched: 28 row(s)
hive>
```

Deleting a Row from Hive Table

```
hive> delete from employee_details_8_3 where emp_id=114;
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_20171116003455_30613979-9701-4662-b3fe-0704343af785
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_1510770641633_0004, Tracking URL = http://localhost:8088/proxy/application_1510770641633_0004/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510770641633_0004
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-11-16 00:35:26,504 Stage-1 map = 0%,  reduce = 0%
2017-11-16 00:36:28,059 Stage-1 map = 0%,  reduce = 0%
2017-11-16 00:37:25,286 Stage-1 map = 20%,  reduce = 0%, Cumulative CPU 18.13 sec
2017-11-16 00:37:27,346 Stage-1 map = 60%,  reduce = 0%, Cumulative CPU 24.67 sec
2017-11-16 00:37:29,853 Stage-1 map = 80%,  reduce = 0%, Cumulative CPU 28.04 sec
2017-11-16 00:37:35,816 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 31.94 sec
2017-11-16 00:38:36,492 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 31.94 sec
2017-11-16 00:38:50,356 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 32.84 sec
2017-11-16 00:38:52,538 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 37.65 sec
2017-11-16 00:39:10,385 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 41.32 sec
2017-11-16 00:39:12,321 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 46.65 sec
2017-11-16 00:39:13,578 Stage-1 map = 100%,  reduce = 97%, Cumulative CPU 51.96 sec
2017-11-16 00:39:15,407 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 52.62 sec
MapReduce Total cumulative CPU time: 52 seconds 620 msec
Ended Job = job_1510770641633_0004
Loading data to table custom.employee_details_8_3
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 52.62 sec   HDFS Read: 63425 HDFS Write: 818 SUCCESS
Total MapReduce CPU Time Spent: 52 seconds 620 msec
OK
Time taken: 263.656 seconds
hive>
```

```
hive> select * from employee_details_8_3;
OK
105     Pawan   250     Admin   5
110     Priyanka        222     Admin   5
105     Pawan   250     Admin   5
110     Priyanka        222     Admin   5
106     Aamir   25      Finance 1
101     Amitabh 256     Finance 1
111     Tushar  500     Finance 1
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
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
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
109     Katrina 100     Network_Team    4
104     Anubhav 50      Network_Team    4
Time taken: 0.53 seconds, Fetched: 26 row(s)
hive>
```

As we see that records related to emp_id=114 is deleted from the table.

This is how the transactions or row-wise operations are performed in Hive.

Submitted By:-

Hardik Kaushik