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:-

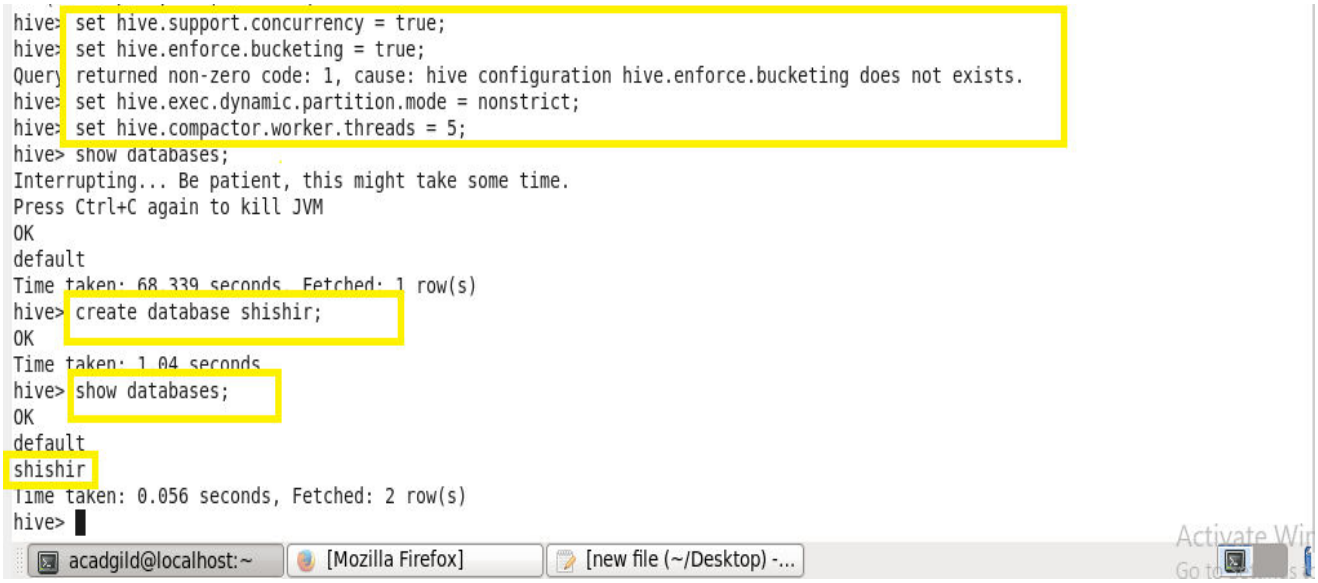
Transactions are provided at the row-level in Hive 0.14. The different row-level transactions available in Hive 0.14 are as follows:

- Insert
- Delete
- Update

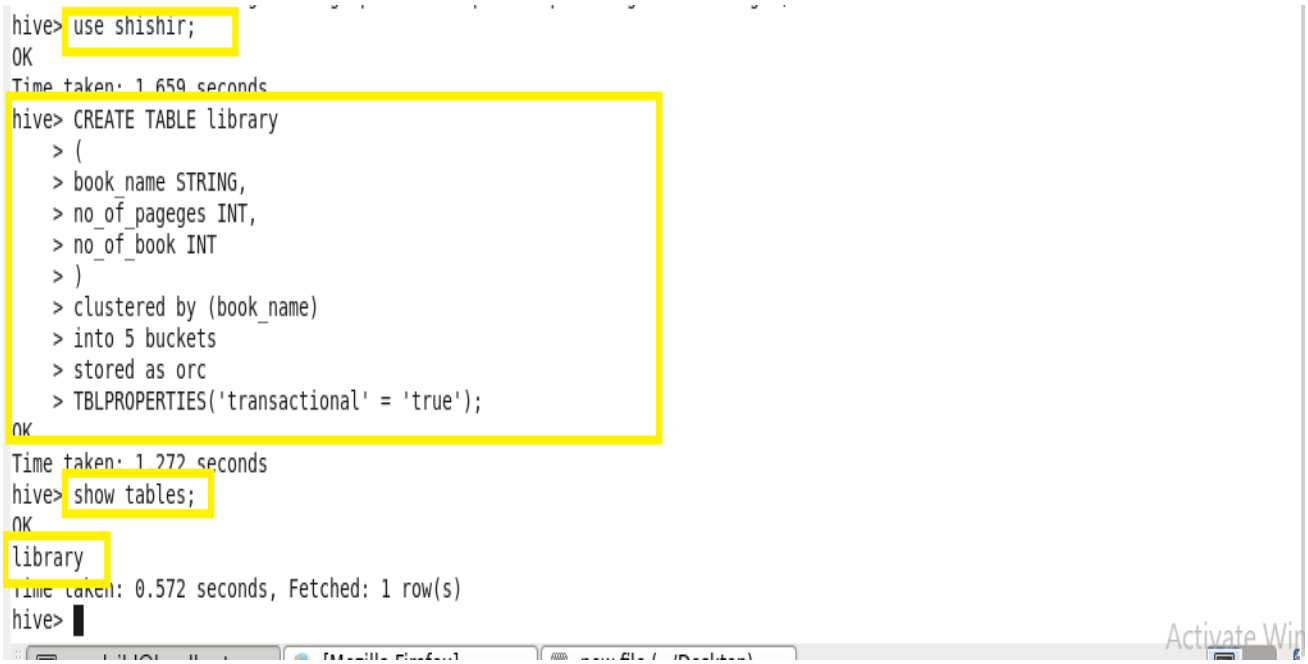
There are numerous limitations with the present transactions available in Hive 0.14. ORC is the file format supported by Hive transaction. It is now essential to have ORC file format for performing transactions in Hive. The table needs to be bucketed in order to support transactions.

Row-level Transactions:-

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

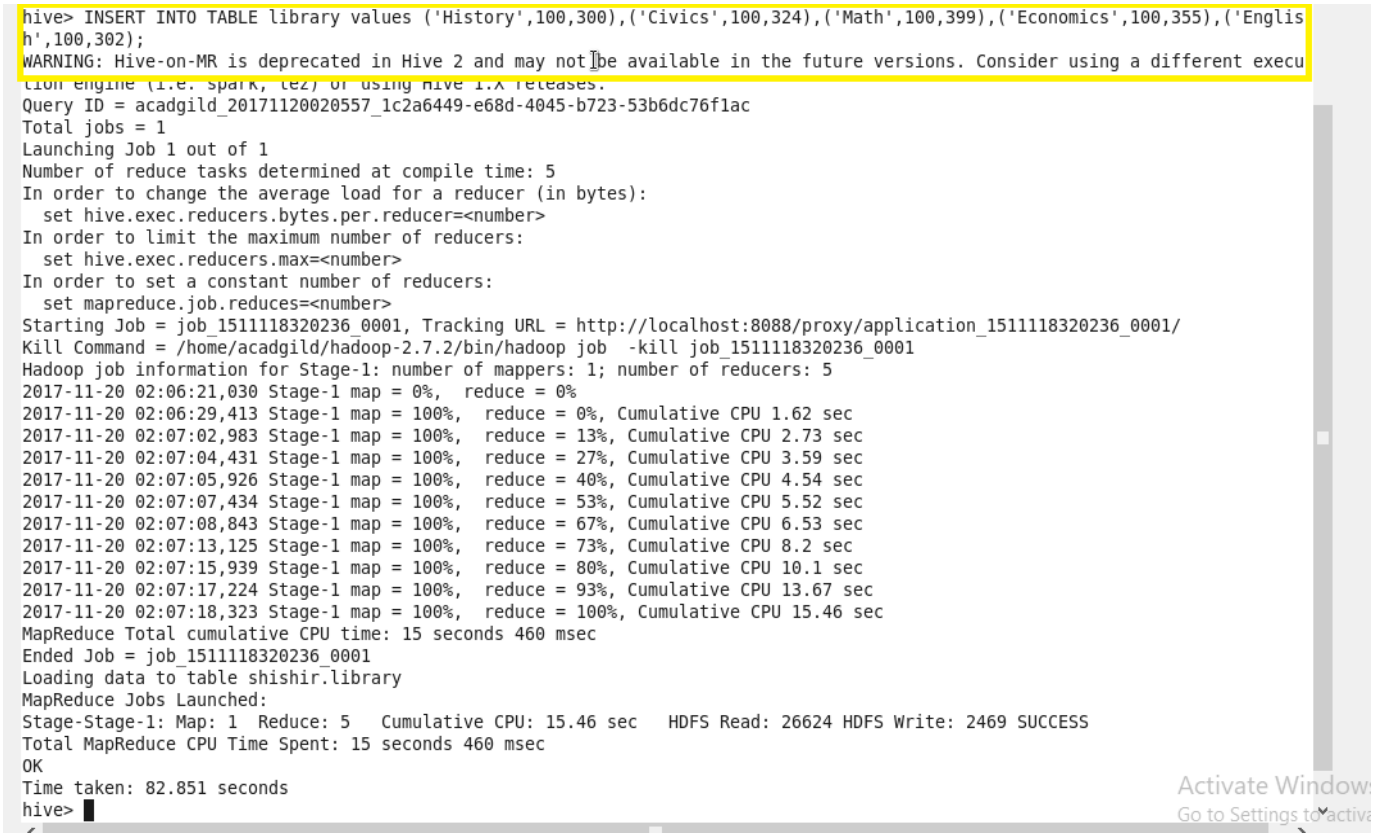


Creating a Table That Supports Hive Transactions:-



```
hive> use shishir;
OK
Time taken: 1.659 seconds
hive> CREATE TABLE library
    > (
    > book_name STRING,
    > no_of_pageges INT,
    > no_of_book INT
    > )
    > clustered by (book_name)
    > into 5 buckets
    > stored as orc
    > TBLPROPERTIES('transactional' = 'true');
OK
Time taken: 1.272 seconds
hive> show tables;
OK
library
Time taken: 0.572 seconds, Fetched: 1 row(s)
hive>
```

Inserting Data into a Hive Table



```
hive> INSERT INTO TABLE library values ('History',100,300),('Civics',100,324),('Math',100,399),('Economics',100,355),('Englis
h',100,302);
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_20171120020557_1c2a6449-e68d-4045-b723-53b6dc76f1ac
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_1511118320236_0001, Tracking URL = http://localhost:8088/proxy/application_1511118320236_0001/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1511118320236_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2017-11-20 02:06:21,030 Stage-1 map = 0%,  reduce = 0%
2017-11-20 02:06:29,413 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.62 sec
2017-11-20 02:07:02,983 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 2.73 sec
2017-11-20 02:07:04,431 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 3.59 sec
2017-11-20 02:07:05,926 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 4.54 sec
2017-11-20 02:07:07,434 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 5.52 sec
2017-11-20 02:07:08,843 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 6.53 sec
2017-11-20 02:07:13,125 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 8.2 sec
2017-11-20 02:07:15,939 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 10.1 sec
2017-11-20 02:07:17,224 Stage-1 map = 100%,  reduce = 93%, Cumulative CPU 13.67 sec
2017-11-20 02:07:18,323 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.46 sec
MapReduce Total cumulative CPU time: 15 seconds 460 msec
Ended Job = job_1511118320236_0001
Loading data to table shishir.library
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 15.46 sec   HDFS Read: 26624 HDFS Write: 2469 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 460 msec
OK
Time taken: 82.851 seconds
hive>
```

```
hive> select * from library;
OK
Economics       100     355
History 100     300
English 100     302
Math    100     399
Civics  100     324
Time taken: 0.229 seconds, Fetched: 5 row(s)
hive>
```

From the above image, we can see that the data has been inserted successfully into the table.

Now if we try to re-insert the same data again, it will be appended to the previous data as shown below:

```
hive> INSERT INTO TABLE library values ('History',100,300),('Civics',100,324),('Math',100,399),('Economics',100,355),('English',100,302);
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_20171120021107_8673715d-fb84-49ea-97a3-afdc5eb660da
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_1511118320236_0002, Tracking URL = http://localhost:8088/proxy/application_1511118320236_0002/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1511118320236_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2017-11-20 02:11:19,789 Stage-1 map = 0%,  reduce = 0%
2017-11-20 02:11:29,413 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.18 sec
2017-11-20 02:11:54,340 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 3.18 sec
2017-11-20 02:11:59,769 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 4.17 sec
2017-11-20 02:12:01,176 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 5.21 sec
2017-11-20 02:12:02,615 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 6.25 sec
2017-11-20 02:12:04,029 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 8.86 sec
2017-11-20 02:12:07,987 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 10.46 sec
2017-11-20 02:12:09,263 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 12.0 sec
2017-11-20 02:12:10,369 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.08 sec
MapReduce Total cumulative CPU time: 15 seconds 80 msec
Ended Job = job_1511118320236_0002
Loading data to table shishir.library
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 15.08 sec   HDFS Read: 26394 HDFS Write: 2479 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 80 msec
OK
Time taken: 64.497 seconds
hive>
```

```
hive> select * from library;
OK
Economics       100     355
History 100     300
Economics       100     355
History 100     300
English 100     302
Math    100     399
Civics  100     324
English 100     302
Math    100     399
Civics  100     324
Time taken: 0.206 seconds, Fetched: 10 row(s)
hive>
```

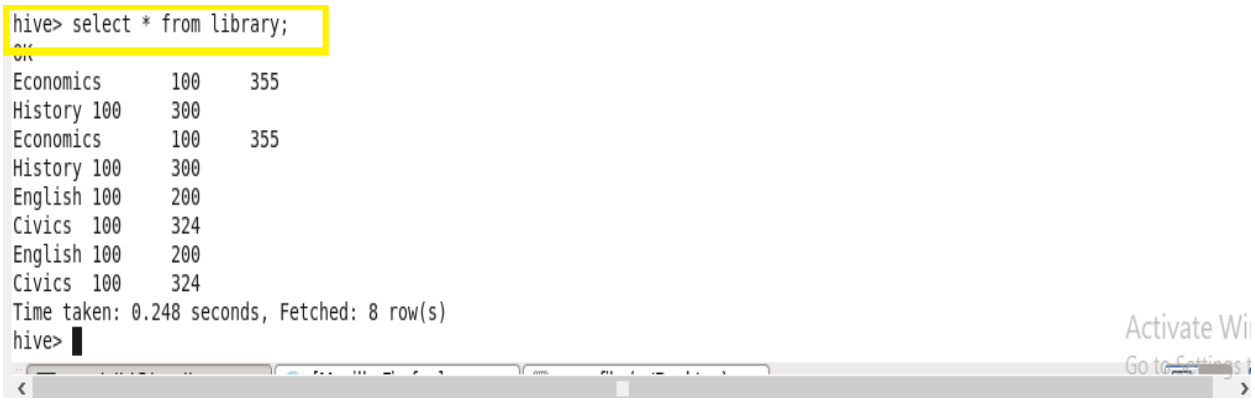
Updating the Data in Hive Table

```
hive> UPDATE library set book name = 'Hindi' WHERE book name = 'English';
FAILED: SemanticException [Error 10302]: Updating values of bucketing columns is not supported.  Column book_name.
hive> UPDATE library set no of book = 200  WHERE book name= 'English';
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_20171120021855_7d8d6b8c-08da-4ff1-ace0-6cf8f957d56a
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_1511118320236_0003, Tracking URL = http://localhost:8088/proxy/application_1511118320236_0003/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1511118320236_0003
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-11-20 02:19:08,527 Stage-1 map = 0%,  reduce = 0%
2017-11-20 02:19:43,949 Stage-1 map = 20%,  reduce = 0%, Cumulative CPU 4.72 sec
2017-11-20 02:19:46,884 Stage-1 map = 40%,  reduce = 0%, Cumulative CPU 7.44 sec
2017-11-20 02:19:50,108 Stage-1 map = 80%,  reduce = 0%, Cumulative CPU 12.45 sec
2017-11-20 02:19:54,747 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 13.12 sec
2017-11-20 02:20:17,530 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 14.09 sec
2017-11-20 02:20:18,894 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 16.06 sec
2017-11-20 02:20:21,642 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 18.9 sec
2017-11-20 02:20:23,080 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 20.52 sec
2017-11-20 02:20:24,203 Stage-1 map = 100%,  reduce = 93%, Cumulative CPU 21.28 sec
2017-11-20 02:20:25,297 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 22.34 sec
MapReduce Total cumulative CPU time: 22 seconds 340 msec
Ended Job = job_1511118320236_0003
Loading data to table shishir.library
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 22.34 sec   HDFS Read: 57565 HDFS Write: 1009 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 340 msec
OK
Time taken: 91.163 seconds
```

```
hive> select * from library;
OK
Economics       100     355
History 100     300
Economics       100     355
History 100     300
English 100     200
Math    100     399
Civics  100     324
English 100     200
Math    100     399
Civics  100     324
Time taken: 0.209 seconds, Fetched: 10 row(s)
hive>
```

Deleting a Row from Hive Table:

```
hive> delete from library WHERE book_name = 'Math';
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_20171120022755_6bf0bf20-c94d-4634-ba32-43811ab2d570
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_1511118320236_0004, Tracking URL = http://localhost:8088/proxy/application_1511118320236_0004/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1511118320236_0004
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2017-11-20 02:28:08,328 Stage-1 map = 0%,  reduce = 0%
2017-11-20 02:28:45,437 Stage-1 map = 20%,  reduce = 0%, Cumulative CPU 4.45 sec
2017-11-20 02:28:46,778 Stage-1 map = 40%,  reduce = 0%, Cumulative CPU 6.83 sec
2017-11-20 02:28:48,054 Stage-1 map = 60%,  reduce = 0%, Cumulative CPU 11.39 sec
2017-11-20 02:28:49,382 Stage-1 map = 80%,  reduce = 0%, Cumulative CPU 11.83 sec
2017-11-20 02:28:50,835 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 12.24 sec
2017-11-20 02:29:16,603 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 13.24 sec
2017-11-20 02:29:18,007 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 15.19 sec
2017-11-20 02:29:19,424 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 17.26 sec
2017-11-20 02:29:20,845 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 18.02 sec
2017-11-20 02:29:22,162 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 19.58 sec
2017-11-20 02:29:23,472 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 21.26 sec
MapReduce Total cumulative CPU time: 21 seconds 260 msec
Ended Job = job_1511118320236_0004
Loading data to table shishir.library
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 21.26 sec   HDFS Read: 55665 HDFS Write: 764 SUCCESS
Total MapReduce CPU Time Spent: 21 seconds 260 msec
OK
Time taken: 89.812 seconds
hive>
```



As we see that records related to book_name = Math is deleted from the table.

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

Submitted By:-
Shishir Jha