

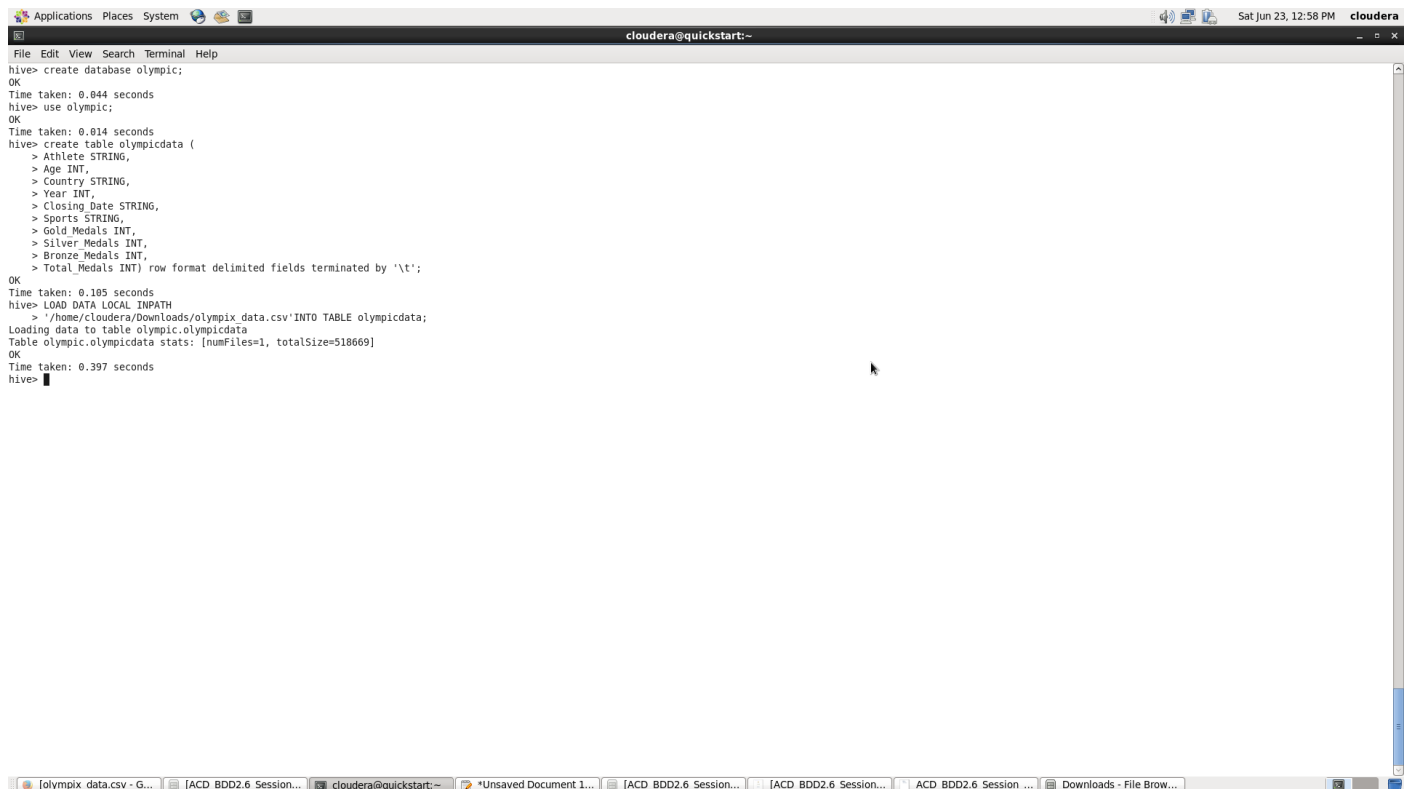
Session 9: Advance Hive

DATE SET DESCRIPTION The data set consists of the following fields.

Athlete: This field consists of the athlete name Age: This field consists of athlete ages Country: This fields consists of the country names which participated in Olympics Year: This field consists of the year Closing Date: This field consists of the closing date of ceremony Sport: Consists of the sports name Gold Medals: No. of Gold medals Silver Medals: No. of Silver medals Bronze Medals: No. of Bronze medals Total Medals: Consists of total no. of medals

Task 1:

- Created Olympic database & created Olympicdata table
 - Inseted data into table
 - Processed data as show in below pictures
1. Write a Hive program to find the number of medals won by each country in swimming.



```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
hive> create database olympic;
OK
Time taken: 0.044 seconds
hive> use olympic;
OK
Time taken: 0.014 seconds
hive> create table olympicdata (
  > Athlete STRING,
  > Age INT,
  > Country STRING,
  > Year INT,
  > Closing Date STRING,
  > Sports STRING,
  > Gold Medals INT,
  > Silver Medals INT,
  > Bronze Medals INT,
  > Total Medals INT) row format delimited fields terminated by '\t';
OK
Time taken: 0.185 seconds
hive> LOAD DATA LOCAL INPATH
  > '/home/cloudera/Downloads/olympix_data.csv' INTO TABLE olympicdata;
Loading data to table olympic.olympicdata
Table olympic.olympicdata stats: [numFiles=1, totalSize=518669]
OK
Time taken: 0.397 seconds
hive>
```

Session 9: Advance Hive

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select country,sum(total_medals) from olympCiddata where sports = 'Swimming' group by country;
FAILED: SemanticException [Error 10001]: Line 1:39 Table not found 'olympCiddata'
hive> select country,sum(total_medals) from olympicdata where sports = 'Swimming' group by country;
Query ID = cloudera_20180623131212_cbd713c4-081f-4869-9dd4-162690bc78ad
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified, Estimated from input data size: 1
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_1529769661844_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-23 13:12:43,505 Stage-1 map = 0%, reduce = 0%
2018-06-23 13:12:55,548 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.45 sec
2018-06-23 13:13:08,714 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.45 sec
MapReduce Total cumulative CPU time: 4 seconds 750 msec
Ended Job = job_1529769661844_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.75 sec HDFS Read: 528499 HDFS Write: 386 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 750 msec
OK
Argentina      1
Australia     163
Austria        3
Belarus        2
Brazil         8
Canada         5
China          35
Costa Rica     2
Croatia        1
Denmark        1
France         39
Germany        32
Great Britain  11
Hungary        9
Italy          16
Japan          43
Lithuania      1
Netherlands    46
Norway         2
Poland         3
Romania        6
Russia         20
Serbia         1
Slovakia       2
Slovenia       1
South Africa   11
South Korea    4
Spain          3
Sweden         9
Trinidad and Tobago 1
Tunisia        3
Ukraine       7

hive see table definiti... ACD_BDD2.6_Session... cloudera@quickstart:~ *Unsaved Document 1... ACD_BDD2.6_Session... ACD_BDD2.6_Session... cloudera@quickstart:~
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified, Estimated from input data size: 1
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_1529769661844_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-23 13:12:43,505 Stage-1 map = 0%, reduce = 0%
2018-06-23 13:12:55,548 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.45 sec
2018-06-23 13:13:08,714 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.45 sec
MapReduce Total cumulative CPU time: 4 seconds 750 msec
Ended Job = job_1529769661844_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.75 sec HDFS Read: 528499 HDFS Write: 386 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 750 msec
OK
Argentina      1
Australia     163
Austria        3
Belarus        2
Brazil         8
Canada         5
China          35
Costa Rica     2
Croatia        1
Denmark        1
France         39
Germany        32
Great Britain  11
Hungary        9
Italy          16
Japan          43
Lithuania      1
Netherlands    46
Norway         2
Poland         3
Romania        6
Russia         20
Serbia         1
Slovakia       2
Slovenia       1
South Africa   11
South Korea    4
Spain          3
Sweden         9
Trinidad and Tobago 1
Tunisia        3
Ukraine       7
United States  267
Zimbabwe       7
Time taken: 40.084 seconds, Fetched: 34 row(s)
hive>
```

2. Write a Hive program to find the number of medals that India won year wise.

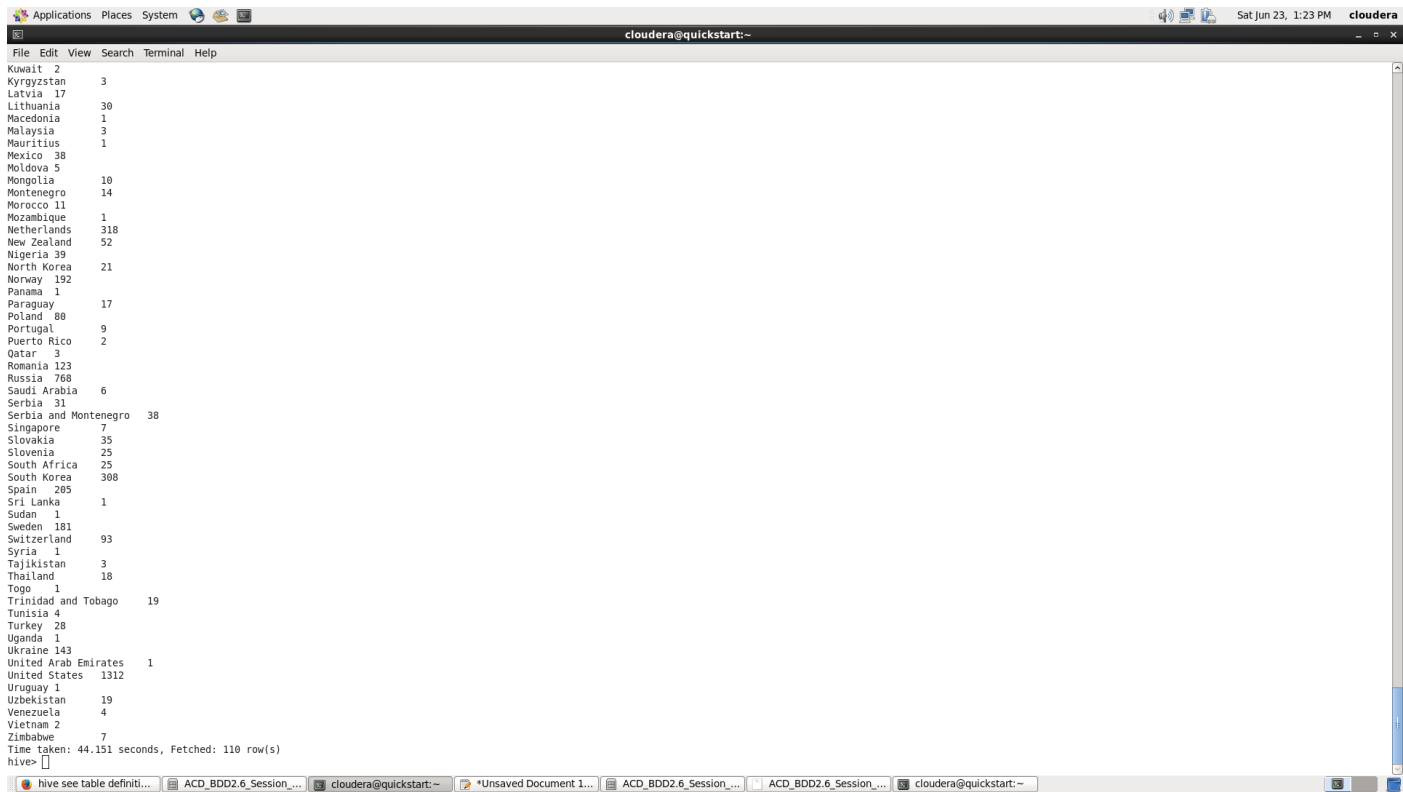
Session 9: Advance Hive

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
Costa Rica 2
Croatia 1
Denmark 1
France 39
Germany 32
Great Britain 11
Hungary 9
Italy 16
Japan 43
Lithuania 1
Netherlands 46
Norway 2
Poland 3
Romania 6
Russia 20
Serbia 1
Slovakia 2
Slovenia 1
South Africa 11
South Korea 4
Spain 3
Sweden 9
Trinidad and Tobago 1
Tunisia 3
Ukraine 7
United States 267
Zimbabwe 7
Time taken: 40.084 seconds, Fetched: 34 row(s)
hive> select year,sum(total_medals) from olympicdata where country = 'India' group by year order by year;
Query ID = cloudera_20180623132121_3c7f6409-1e51-4334-90d4-8fe9be845af8
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
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_1529769661844_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0021
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-23 13:21:28,893 Stage-1 map = 0%, reduce = 0%
2018-06-23 13:21:39,762 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.46 sec
2018-06-23 13:21:52,741 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.76 sec
MapReduce Total cumulative CPU time: 4 seconds 760 msec
Ended Job = job_1529769661844_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.76 sec HDFS Read: 528483 HDFS Write: 28 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 760 msec
OK
2009 1
2004 1
2008 3
2012 6
Time taken: 39.036 seconds, Fetched: 4 row(s)
hive>
```

3. Write a Hive Program to find the total number of medals each country won.

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select country,sum(total_medals) from olympicdata group by country;
Query ID = cloudera_20180623132222_7e649fb9-52af-42da-83fe-e282b164fd2a
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified, Estimated from input data size: 1
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_1529769661844_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0022
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-23 13:22:53,274 Stage-1 map = 0%, reduce = 0%
2018-06-23 13:23:04,065 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.13 sec
2018-06-23 13:23:22,459 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.48 sec
MapReduce Total cumulative CPU time: 4 seconds 480 msec
Ended Job = job_1529769661844_0022
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.48 sec HDFS Read: 527425 HDFS Write: 1315 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 480 msec
OK
Afghanistan 2
Algeria 8
Argentina 141
Armenia 10
Australia 609
Austria 91
Azerbaijan 25
Bahamas 24
Bahrain 1
Barbados 1
Belarus 97
Belgium 18
Botswana 1
Brazil 221
Bulgaria 41
Cameroon 20
Canada 370
Chile 22
China 530
Chinese Taipei 20
Colombia 13
Costa Rica 2
Croatia 81
Cuba 188
Cyprus 1
Czech Republic 81
Denmark 89
Dominican Republic 5
Ecuador 1
Egypt 8
Eritrea 1
Estonia 18
Ethiopia 29
Finland 118
```

Session 9: Advance Hive



The screenshot shows a terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output displays a list of countries and the number of gold medals they won, sorted in descending order. The data is as follows:

Country	Gold Medals
Kuwait	2
Kyrgyzstan	3
Latvia	17
Lithuania	30
Macedonia	1
Malaysia	3
Mauritius	1
Mexico	38
Moldova	5
Mongolia	10
Montenegro	14
Morocco	11
Mozambique	1
Netherlands	318
New Zealand	52
Nigeria	39
North Korea	21
Norway	192
Panama	1
Paraguay	17
Poland	88
Portugal	9
Puerto Rico	2
Qatar	3
Romania	123
Russia	768
Saudi Arabia	6
Serbia	31
Serbia and Montenegro	38
Singapore	7
Slovakia	35
Slovenia	25
South Africa	25
South Korea	388
Spain	285
Sri Lanka	1
Sudan	1
Sweden	181
Switzerland	93
Syria	1
Tajikistan	3
Thailand	18
Togo	1
Trinidad and Tobago	19
Tunisia	4
Turkey	28
Uganda	1
Ukraine	143
United Arab Emirates	1
United States	1312
Uruguay	1
Uzbekistan	19
Venezuela	4
Vietnam	2
Zimbabwe	7

Time taken: 44.151 seconds, Fetched: 110 row(s)
hive>

The terminal window also shows a taskbar at the bottom with several open applications: 'hive see table definiti...', 'ACD_BDD2.6_Session...', 'cloudera@quickstart:~', '*Unsaved Document 1...', 'ACD_BDD2.6_Session...', 'ACD_BDD2.6_Session...', and 'cloudera@quickstart:~'.

4. Write a Hive program to find the number of gold medals each country won.

Session 9: Advance Hive

```
File Edit View Search Terminal Help
hive> select country,sum(gold medals) from olympicdata group by country;
Query ID = cloudera_20180623132424_9a5cc161-df60-4c97-84b3-6c5c19995512
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified, Estimated from input data size: 1
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.reducers=<number>
Starting Job = job_1529769661844_0023, Tracking URL = http://quickstart.cloudera:8080/proxy/application_1529769661844_0023/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0023
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-23 13:24:38,021 Stage-1 map = 0%, reduce = 0%
2018-06-23 13:24:50,713 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.12 sec
2018-06-23 13:25:02,649 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.39 sec
MapReduce Total cumulative CPU time: 4 seconds 390 msec
Ended Job = job_1529769661844_0023
MapReduce Jobs Launched:
  Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.39 sec HDFS Read: 527423 HDFS Write: 1276 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 390 msec
OK
Afghanistan 0
Algeria 2
Argentina 49
Armenia 0
Australia 163
Austria 36
Azerbaijan 6
Bahamas 11
Bahrain 0
Barbados 0
Belarus 17
Belgium 2
Botswana 0
Brazil 46
Bulgaria 0
Cameroon 20
Canada 168
Chile 3
China 234
Chinese Taipei 2
Colombia 2
Costa Rica 0
Croatia 35
Cuba 57
Cyprus 0
Czech Republic 14
Denmark 46
Dominican Republic 3
Ecuador 0
Egypt 1
Eritrea 0
Estonia 6
Ethiopia 13
Finland 11
Kuwait 0
Kyrgyzstan 0
Latvia 3
Lithuania 5
Macedonia 0
Malaysia 0
Mauritius 0
Mexico 19
Moldova 0
Mongolia 2
Montenegro 0
Morocco 2
Mozambique 1
Netherlands 101
New Zealand 18
Nigeria 6
North Korea 6
Norway 97
Panama 1
Paraguay 0
Poland 20
Portugal 1
Puerto Rico 0
Qatar 0
Romania 57
Russia 234
Saudi Arabia 0
Serbia 1
Serbia and Montenegro 11
Singapore 0
Slovakia 10
Slovenia 5
South Africa 10
South Korea 110
Spain 19
Sri Lanka 0
Sudan 0
Sweden 57
Switzerland 21
Syria 0
Tajikistan 0
Thailand 6
Togo 0
Trinidad and Tobago 1
Tunisia 2
Turkey 9
Uganda 1
Ukraine 31
United Arab Emirates 1
United States 552
Uruguay 0
Uzbekistan 5
Venezuela 1
Vietnam 0
Zimbabwe 2
Time taken: 39.742 seconds, Fetched: 110 row(s)
hive>
```

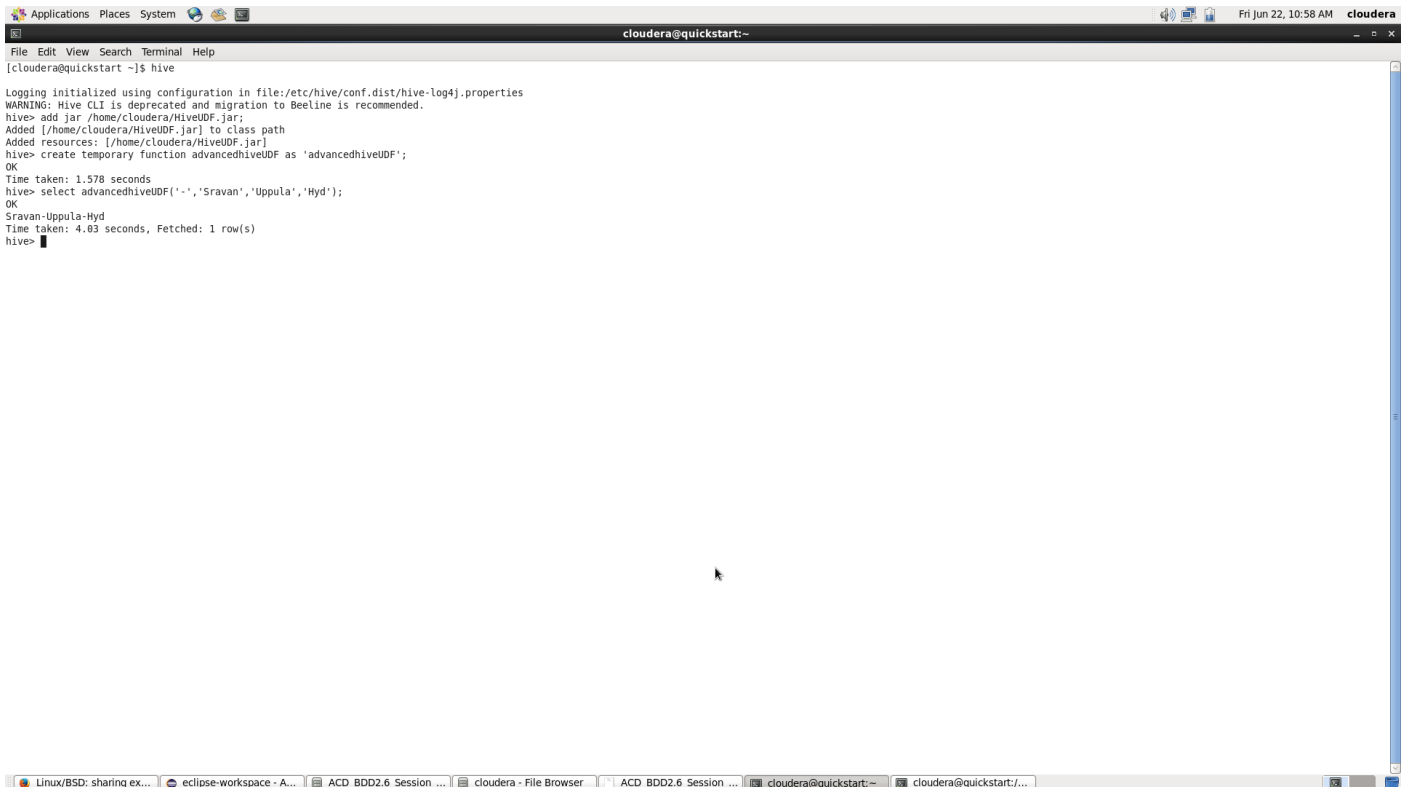
Task 2:

Write a hive UDF that implements functionality of string concat_ws(string SEP, array<string>). This UDF will accept two arguments, one string and one array of string. It will return a single string where all the elements of the array are separated by the SEP.

- Created UDF for concat_ws example, example is given & executed\

Session 9: Advance Hive

- Attached jar & java file for the same in Github



```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
[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> add jar /home/cloudera/HiveUDF.jar;
Added [/home/cloudera/HiveUDF.jar] to class path
Added resources: [/home/cloudera/HiveUDF.jar]
hive> create temporary function advancedhiveUDF as 'advancedhiveUDF';
OK
Time taken: 1.578 seconds
hive> select advancedhiveUDF('-', 'Sravan', 'Uppula', 'Hyd');
OK
Sravan-Uppula-Hyd
Time taken: 4.83 seconds, Fetched: 1 row(s)
hive>
```



```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> select advancedhiveUDF('-', city, mytemp) from tempUDF;
FAILED: SemanticException [Error 10011]: Line 1:7 Invalid function 'advancedhive
UDF'
hive> add jar /home/cloudera/HiveUDF.jar;
Added [/home/cloudera/HiveUDF.jar] to class path
Added resources: [/home/cloudera/HiveUDF.jar]
hive> create temporary function advancedhiveUDF as 'advancedhiveUDF';
OK
Time taken: 0.263 seconds
hive> select advancedhiveUDF('-', city, mytemp) from tempUDF;
OK
hyderabad-this-is-acadgild
banglore-this-is-hive
chennai-this-is-udf
Time taken: 1.229 seconds, Fetched: 3 row(s)
hive>
```

Session 9: Advance Hive

Task 3:

Transactions in Hive

Transactions in Hive are introduced in Hive 0.13, but they only partially fulfill the ACID properties like atomicity, consistency, durability, at the partition level. Here, Isolation can be provided by turning on one of the locking mechanisms available with zookeeper or in memory.

But in Hive 0.14, new API's have been added to completely fulfill the ACID properties while performing any transaction.

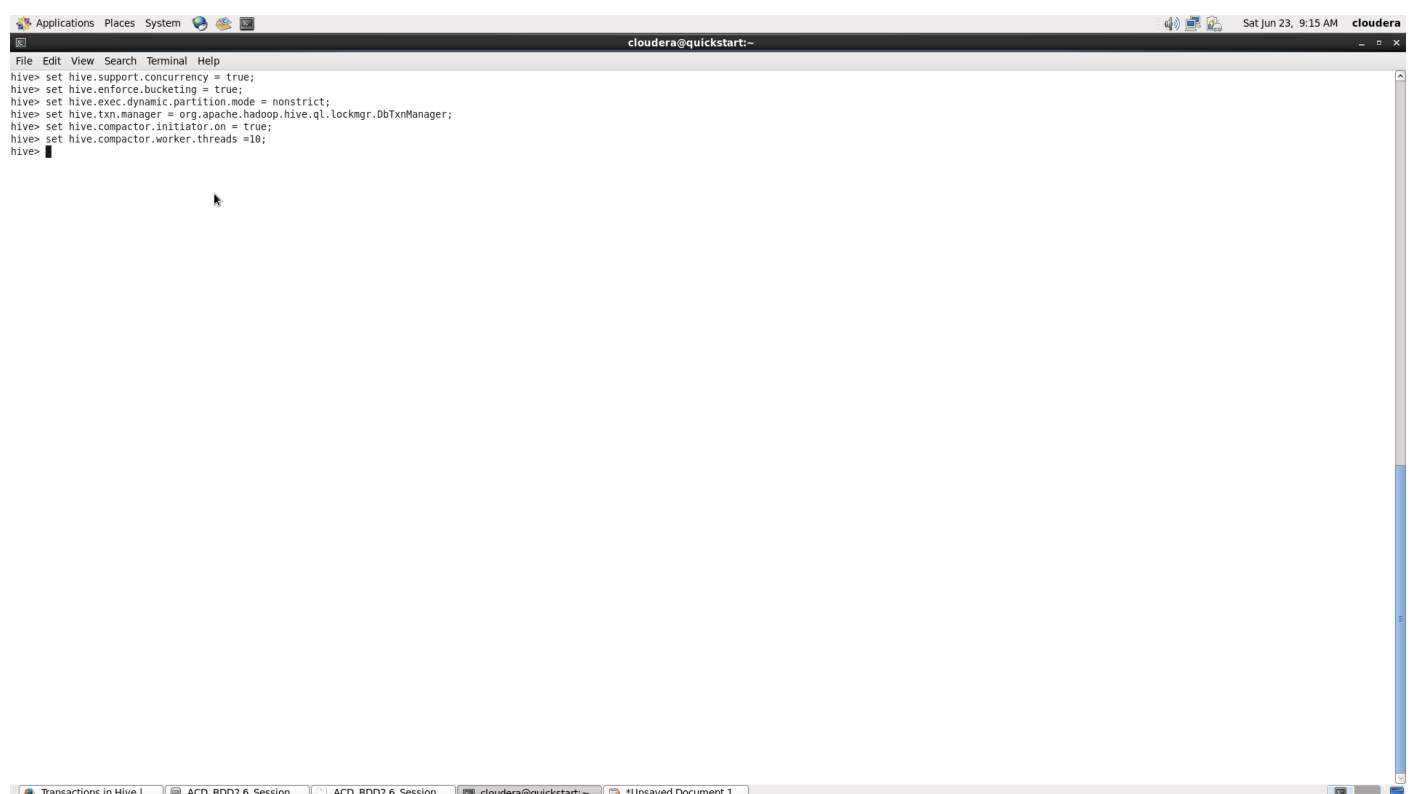
Transactions are provided at the row-level in Hive 0.14. The different row-level transactions available in Hive 0.14 are as follows:

1. Insert
2. Delete
3. Update

Row-level Transactions Available in Hive 0.14

Let's perform some row-level transactions available in Hive 0.14. Before creating a Hive table that supports transactions, the transaction features present in Hive needs to be turned on, as by default they are turned off.

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



```
cloudera@quickstart:~$ 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 =10;
hive>
```

The screenshot shows a terminal window titled 'cloudera@quickstart:~'. The terminal displays a series of Hive configuration commands being entered at the 'hive>' prompt. The commands are: 'set hive.support.concurrency = true;', 'set hive.enforce.bucketing = true;', 'set hive.exec.dynamic.partition.mode = nonstrict;', 'set hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;', 'set hive.compactor.initiator.on = true;', and 'set hive.compactor.worker.threads =10;'. The terminal window has a standard Ubuntu-style top bar showing 'Applications Places System' and a date/time of 'Sat Jun 23, 9:15 AM'. The bottom of the image shows a window manager taskbar with several open tabs, including 'Transactions in Hive', 'ACD_BDD2.6_Session...', and '*Unsaved Document 1...'.

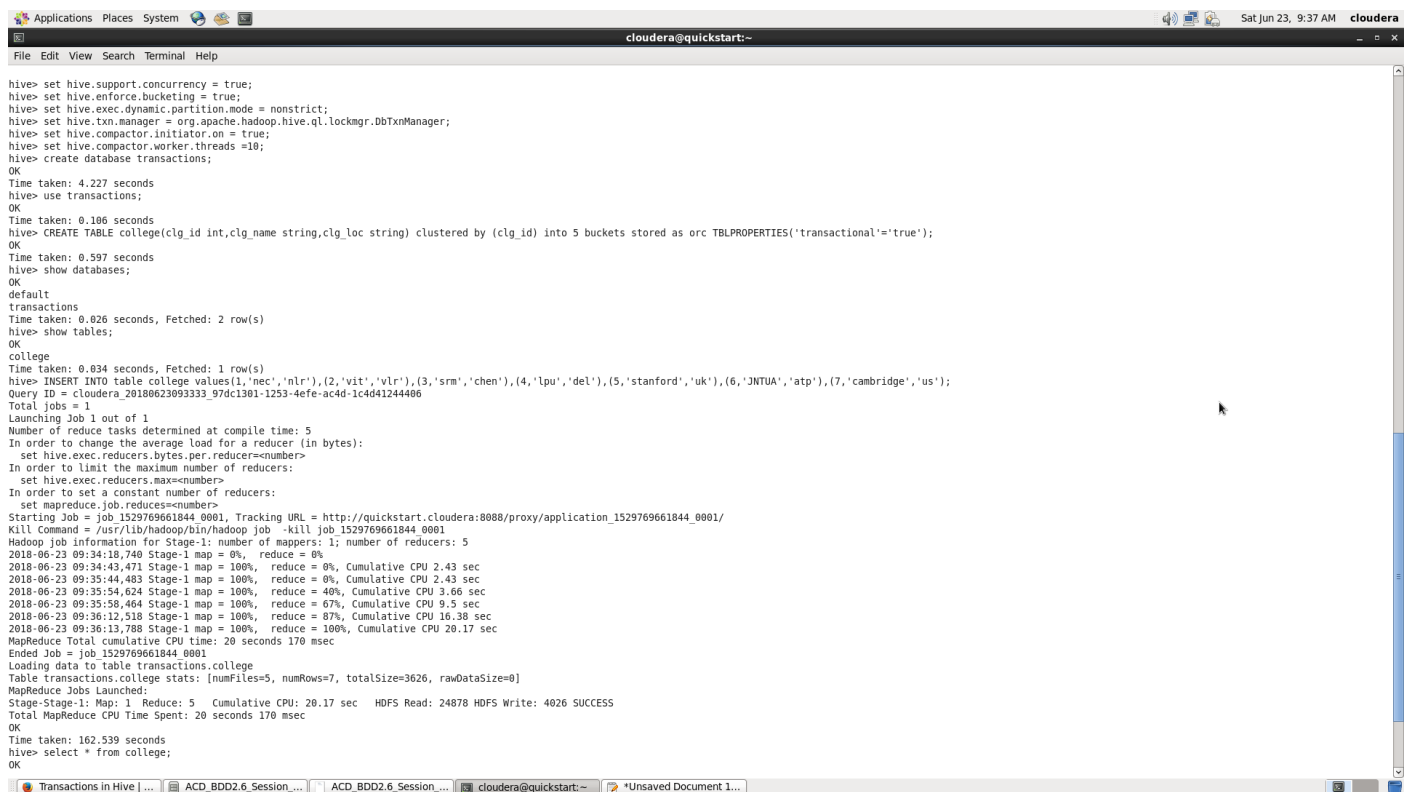
Session 9: Advance Hive

Creating a Table That Supports Hive Transactions

```
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');
```

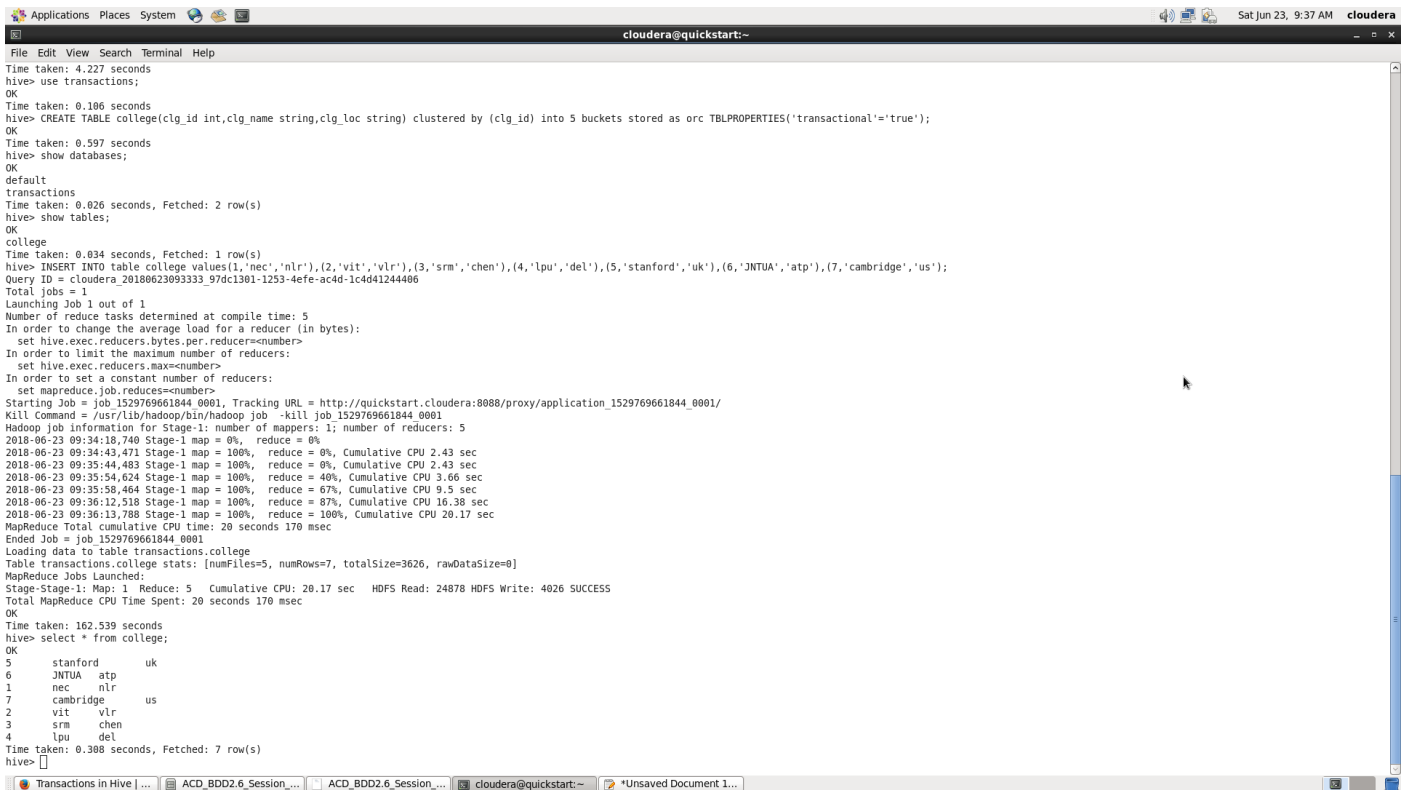
Inserting Data into a Hive Table

```
INSERT INTO table college
values(1,'nec','nlr'),(2,'vit','vlr'),(3,'srm','chen'),(4,'lpu','del'),(5,'stanford','u
k'),(6,'JNTUA','atp'),(7,'cambridge','us');
```



```
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 =10;
hive> create database transactions;
OK
Time taken: 4.227 seconds
hive> use transactions;
OK
Time taken: 0.106 seconds
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: 0.597 seconds
hive> show databases;
OK
default
transactions
Time taken: 0.026 seconds, Fetched: 2 row(s)
hive> show tables;
OK
college
Time taken: 0.034 seconds, Fetched: 1 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_20180623093333_97dc1301-1253-4efe-ac4d-1c4d41244406
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_1529769661844_0001, Tracking URL = http://quickstart.cloudera:8080/proxy/application_1529769661844_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-06-23 09:34:18,740 Stage-1 map = 0%, reduce = 0%
2018-06-23 09:34:43,471 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.43 sec
2018-06-23 09:35:44,483 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.43 sec
2018-06-23 09:35:54,624 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 3.66 sec
2018-06-23 09:35:58,464 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 9.5 sec
2018-06-23 09:36:12,518 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 16.38 sec
2018-06-23 09:36:13,788 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 20.17 sec
MapReduce Total cumulative CPU time: 20 seconds 170 msec
Ended Job = job_1529769661844_0001
Loading data to table transactions.college
Table transactions.college stats: [numFiles=5, numRows=7, totalSize=3626, rawDataSize=0]
MapReduce Jobs Launched:
Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 20.17 sec HDFS Read: 24878 HDFS Write: 4026 SUCCESS
Total MapReduce CPU Time Spent: 20 seconds 170 msec
OK
Time taken: 162.539 seconds
hive> select * from college;
OK
```


Session 9: Advance Hive



```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
Time taken: 4.227 seconds
hive> use transactions;
OK
Time taken: 0.106 seconds
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: 0.597 seconds
hive> show databases;
OK
default
transactions
Time taken: 0.026 seconds, Fetched: 2 row(s)
hive> show tables;
OK
college
Time taken: 0.834 seconds, Fetched: 1 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_20180623093333_97dc1301-1253-4efe-ac4d-1c4d41244406
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_1529769661844_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-06-23 09:34:18,740 Stage-1 map = 0%, reduce = 0%
2018-06-23 09:34:43,471 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.43 sec
2018-06-23 09:35:44,483 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.43 sec
2018-06-23 09:35:54,024 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 3.66 sec
2018-06-23 09:35:58,464 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 9.5 sec
2018-06-23 09:36:12,518 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 16.38 sec
2018-06-23 09:36:13,788 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 20.17 sec
MapReduce Total Cumulative CPU time: 20 seconds 170 msec
Ended Job = job_1529769661844_0001
Loading data to table transactions.college
Table transactions.college stats: [numFiles=5, numRows=7, totalSize=3626, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 20.17 sec HDFS Read: 2487B HDFS Write: 4026 SUCCESS
Total MapReduce CPU Time Spent: 20 seconds 170 msec
OK
Time taken: 162.539 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: 0.308 seconds, Fetched: 7 row(s)
hive>
```

Updating the Data in Hive Table

```
UPDATE college set clg_id = 8 where clg_id = 7;
```

Now let's perform the update operation on Non bucketed column

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

Session 9: Advance Hive

```
Applications Places System cloudera@quickstart:~
cloudera@quickstart:~
File Edit View Search Terminal Help
1 nec nlr us
2 vit vlr
3 srm chen
4 lpu del
Time taken: 0.16 seconds, Fetched: 7 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_20180623095555_6f1d5815-11f8-486b-933d-875fb3858085
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_1529769661844_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-06-23 09:55:43,801 Stage-1 map = 0%, reduce = 0%
2018-06-23 09:55:54,770 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.96 sec
2018-06-23 09:56:39,747 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 3.71 sec
2018-06-23 09:56:47,176 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 7.08 sec
2018-06-23 09:56:48,380 Stage-1 map = 100%, reduce = 47%, Cumulative CPU 9.86 sec
2018-06-23 09:56:51,019 Stage-1 map = 100%, reduce = 60%, Cumulative CPU 11.05 sec
2018-06-23 09:56:52,292 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 14.17 sec
2018-06-23 09:56:54,851 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 15.38 sec
2018-06-23 09:56:56,038 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 16.56 sec
2018-06-23 09:56:57,175 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 17.71 sec
MapReduce Total cumulative CPU time: 17 seconds 710 msec
Ended Job = job_1529769661844_0006
Loading data to table transactions.college
Table transactions.college stats: [numFiles=21, numRows=14, totalSize=14201, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 17.71 sec HDFS Read: 25353 HDFS Write: 4033 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 710 msec
OK
Time taken: 89.432 seconds
hive> select * from college;
OK
5 stanford uk
5 stanford uk
6 JNTUA atp
1 nec nlr
6 JNTUA atp
1 nec nlr
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: 0.134 seconds, Fetched: 14 row(s)
hive>
```

Deleting a Row from Hive Table

```
delete from college where clg_id=5;
```

Session 9: Advance Hive

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
Time taken: 0.068 seconds, Fetched: 14 row(s)
hive> UPDATE college set clg_name = 'IIT' where clg_id = 6;
Query ID = cloudera_20180623103030_56ca68da-3e80-4bcb-8b74-e56e7bffd39
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_1529769661844_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0017
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-06-23 10:30:40,314 Stage-1 map = 0%, reduce = 0%
2018-06-23 10:31:28,929 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 2.45 sec
2018-06-23 10:31:31,557 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 4.9 sec
2018-06-23 10:31:35,365 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 7.42 sec
2018-06-23 10:31:36,608 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.35 sec
2018-06-23 10:32:22,224 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 16.58 sec
2018-06-23 10:32:24,100 Stage-1 map = 100%, reduce = 57%, Cumulative CPU 19.03 sec
2018-06-23 10:32:25,328 Stage-1 map = 100%, reduce = 77%, Cumulative CPU 21.12 sec
2018-06-23 10:32:26,440 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 24.22 sec
MapReduce Total cumulative CPU time: 24 seconds 220 msec
Ended Job = job_1529769661844_0017
Loading data to table transactions.college
Table transactions.college stats: [numFiles=56, numRows=14, totalSize=37848, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 24.22 sec HDFS Read: 95870 HDFS Write: 999 SUCCESS
Total MapReduce CPU Time Spent: 24 seconds 220 msec
OK
Time taken: 120.811 seconds
hive> select * from college;
OK
5      stanford      uk
5      stanford      uk
6      IIT            atp
1      nec            nlr
6      IIT            atp
1      nec            nlr
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: 0.112 seconds, Fetched: 14 row(s)
hive> delete from college where clg_id=5;
Query ID = cloudera_20180623103232_b88811e5-e926-40e6-8f61-2898fa3585b7
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_1529769661844_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0018
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-06-23 10:33:09,741 Stage-1 map = 0%, reduce = 0%
2018-06-23 10:34:00,706 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 2.5 sec
2018-06-23 10:34:03,255 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 5.01 sec
2018-06-23 10:34:05,765 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 10.02 sec
2018-06-23 10:34:06,908 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.49 sec
2018-06-23 10:34:54,973 Stage-1 map = 100%, reduce = 20%, Cumulative CPU 14.53 sec
2018-06-23 10:34:56,323 Stage-1 map = 100%, reduce = 97%, Cumulative CPU 23.04 sec
2018-06-23 10:34:58,537 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 24.1 sec
MapReduce Total cumulative CPU time: 24 seconds 100 msec
Ended Job = job_1529769661844_0018
Loading data to table transactions.college
Table transactions.college stats: [numFiles=57, numRows=12, totalSize=38372, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 24.1 sec HDFS Read: 92347 HDFS Write: 788 SUCCESS
Total MapReduce CPU Time Spent: 24 seconds 100 msec
OK
Time taken: 124.883 seconds
hive> select * from college;
OK
6      IIT            atp
1      nec            nlr
6      IIT            atp
1      nec            nlr
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: 0.179 seconds, Fetched: 12 row(s)
hive>
```

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
1      nec            nlr
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: 0.112 seconds, Fetched: 14 row(s)
hive> delete from college where clg_id=5;
Query ID = cloudera_20180623103232_b88811e5-e926-40e6-8f61-2898fa3585b7
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_1529769661844_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1529769661844_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1529769661844_0018
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-06-23 10:33:09,741 Stage-1 map = 0%, reduce = 0%
2018-06-23 10:34:00,706 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 2.5 sec
2018-06-23 10:34:03,255 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 5.01 sec
2018-06-23 10:34:05,765 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 10.02 sec
2018-06-23 10:34:06,908 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.49 sec
2018-06-23 10:34:54,973 Stage-1 map = 100%, reduce = 20%, Cumulative CPU 14.53 sec
2018-06-23 10:34:56,323 Stage-1 map = 100%, reduce = 97%, Cumulative CPU 23.04 sec
2018-06-23 10:34:58,537 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 24.1 sec
MapReduce Total cumulative CPU time: 24 seconds 100 msec
Ended Job = job_1529769661844_0018
Loading data to table transactions.college
Table transactions.college stats: [numFiles=57, numRows=12, totalSize=38372, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 24.1 sec HDFS Read: 92347 HDFS Write: 788 SUCCESS
Total MapReduce CPU Time Spent: 24 seconds 100 msec
OK
Time taken: 124.883 seconds
hive> select * from college;
OK
6      IIT            atp
1      nec            nlr
6      IIT            atp
1      nec            nlr
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: 0.179 seconds, Fetched: 12 row(s)
hive>
```