

This Data set is about Olympics. You can download the data set from the below link:

<https://drive.google.com/open?id=0ByJLBTmJojizV1czX3Nha0R3bTQ>

## DATA 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 field 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

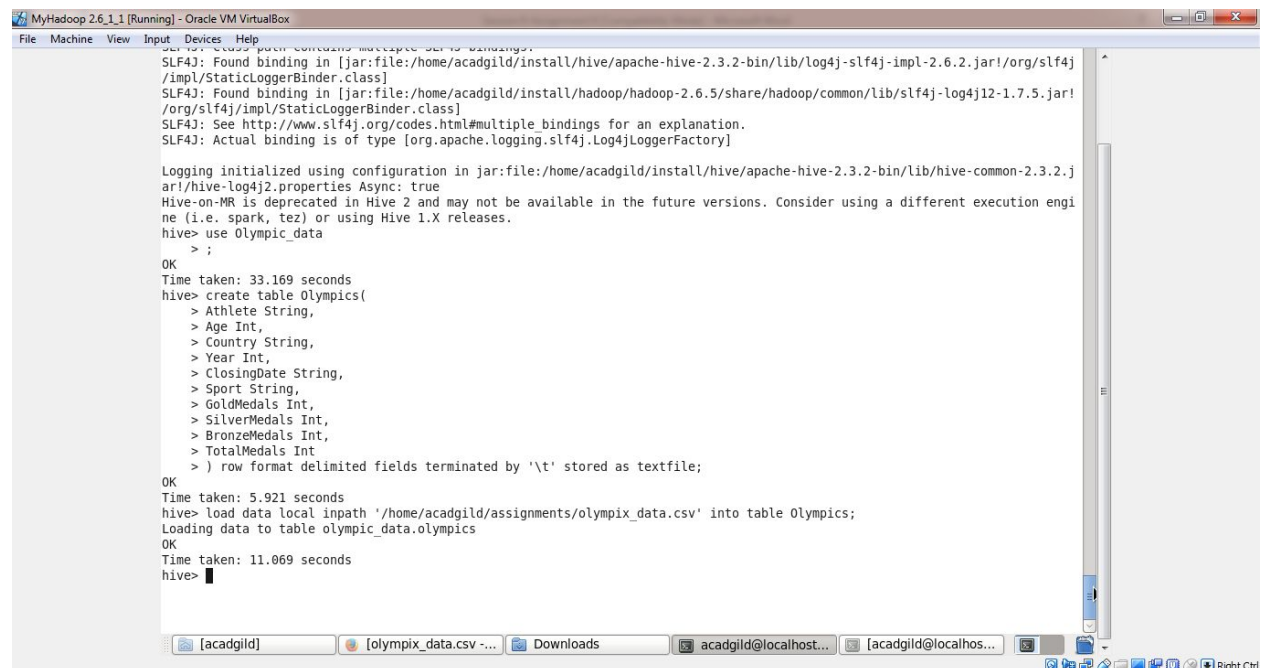
Gold Medals: No. of Gold medals

Silver Medals: No. of Silver medals

Bronze Medals: No. of Bronze medals

Total Medals: Consists of total no. of medals

Load input data into db:



```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
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.
hive> use olympic_data
> ;
OK
Time taken: 33.169 seconds
hive> create table Olympics(
  > Athlete String,
  > Age Int,
  > Country String,
  > Year Int,
  > ClosingDate String,
  > Sport String,
  > GoldMedals Int,
  > SilverMedals Int,
  > BronzeMedals Int,
  > TotalMedals Int
  > ) row format delimited fields terminated by '\t' stored as textfile;
OK
Time taken: 5.921 seconds
hive> load data local inpath '/home/acadgild/assignments/olympix_data.csv' into table Olympics;
Loading data to table olympic_data.olympics
OK
Time taken: 11.069 seconds
hive>
```

## Task 1

1. Write a Hive program to find the number of medals won by each country in swimming.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
hive> load data local inpath '/home/acadgild/assignments/olympix_data.csv' into table Olympics;
Loading data to table olympic_data.olympics
OK
Time taken: 11.069 seconds
hive> select Country, SUM( TotalMedals) from Olympics where sport = 'Swimming' Group by Country;
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_20180618220408_b1f3a8ac-a8c7-4e10-a0fb-9580f42cf645
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_1528036691222_0038, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0038/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0038
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-18 22:05:33,624 Stage-1 map = 0%, reduce = 0%
2018-06-18 22:06:29,135 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.42 sec
2018-06-18 22:07:14,222 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 15.82 sec
2018-06-18 22:07:27,283 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 22.47 sec
MapReduce Total cumulative CPU time: 22 seconds 470 msec
Ended Job = job_1528036691222_0038
MapReduce Jobs Launched:
Stage-Stage1: Map: 1 Reduce: 1 Cumulative CPU: 22.47 sec HDFS Read: 528777 HDFS Write: 881 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 470 msec
OK
Argentina      1
Australia      163
Austria        3
Belarus        2
Brazil         8
Canada         5
China          35
[acadgild]
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
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: 202.387 seconds, Fetched: 34 row(s)
hive>
```

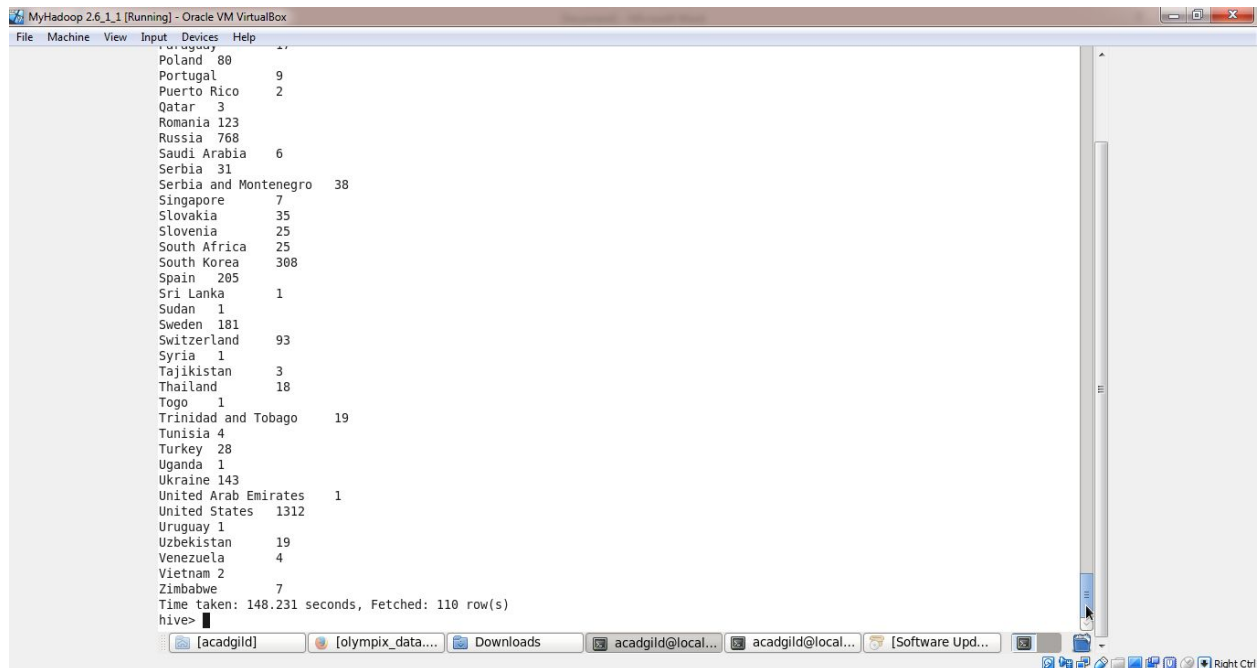
2. Write a Hive program to find the number of medals that India won year wise.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
>
> select Year, SUM( TotalMedals) from Olympics where Country = 'India' group by Year;
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_20180618221459_30879848-d234-470c-9a06-d923caca481a
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_1528036691222_0039, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0039/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0039
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-18 22:16:26,757 Stage-1 map = 0%, reduce = 0%
2018-06-18 22:17:20,335 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 13.46 sec
2018-06-18 22:18:05,279 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 16.87 sec
2018-06-18 22:18:18,676 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 24.85 sec
MapReduce Total cumulative CPU time: 24 seconds 850 msec
Ended Job = job_1528036691222_0039
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 24.85 sec HDFS Read: 528790 HDFS Write: 163 SUCCESS
Total MapReduce CPU Time Spent: 24 seconds 850 msec
OK
2000      1
2004      1
2008      3
2012      6
Time taken: 201.931 seconds, Fetched: 4 row(s)
hive>
>
>
>
```

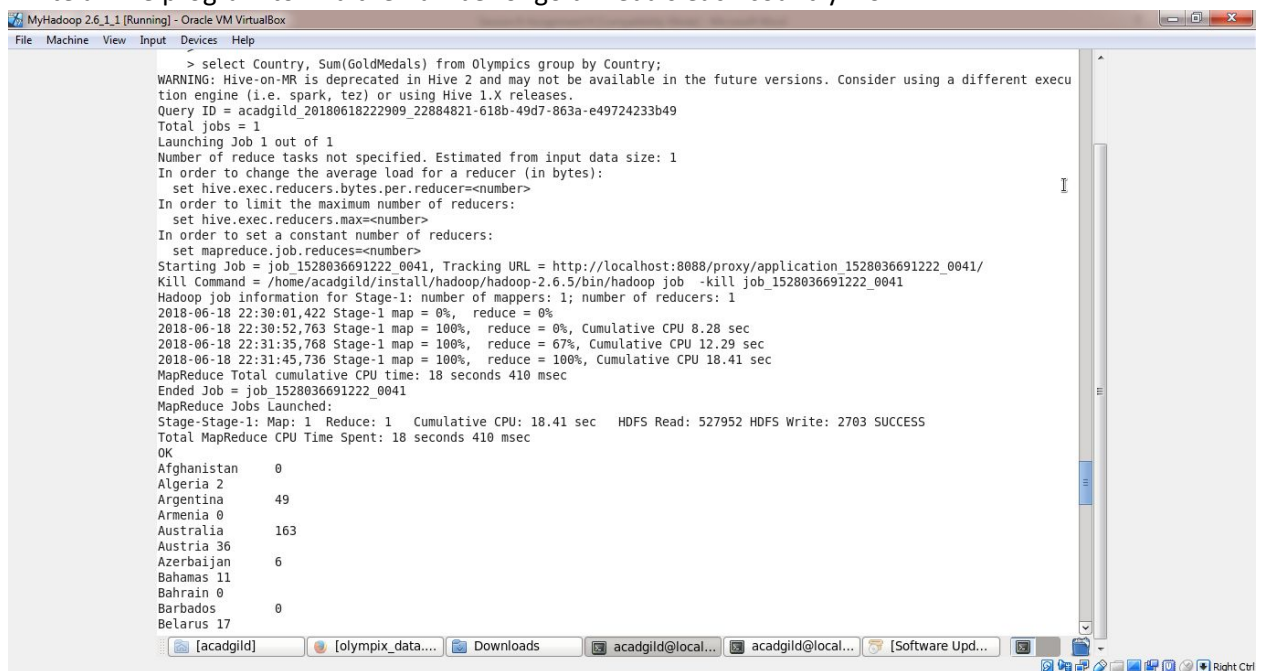
### 3. Write a Hive Program to find the total number of medals each country won.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
>
> select Country, Sum( TotalMedals ) from Olympics group by Country;
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_20180618222041_20ebfc60-86be-4eeb-8219-5f3ad22003c7
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_1528036691222_0040, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0040/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0040
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-18 22:21:36,422 Stage-1 map = 0%, reduce = 0%
2018-06-18 22:22:17,779 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.68 sec
2018-06-18 22:22:54,983 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 10.87 sec
2018-06-18 22:23:06,773 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 17.52 sec
MapReduce Total cumulative CPU time: 17 seconds 520 msec
Ended Job = job_1528036691222_0040
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 17.52 sec HDFS Read: 527954 HDFS Write: 2742 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 520 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
```



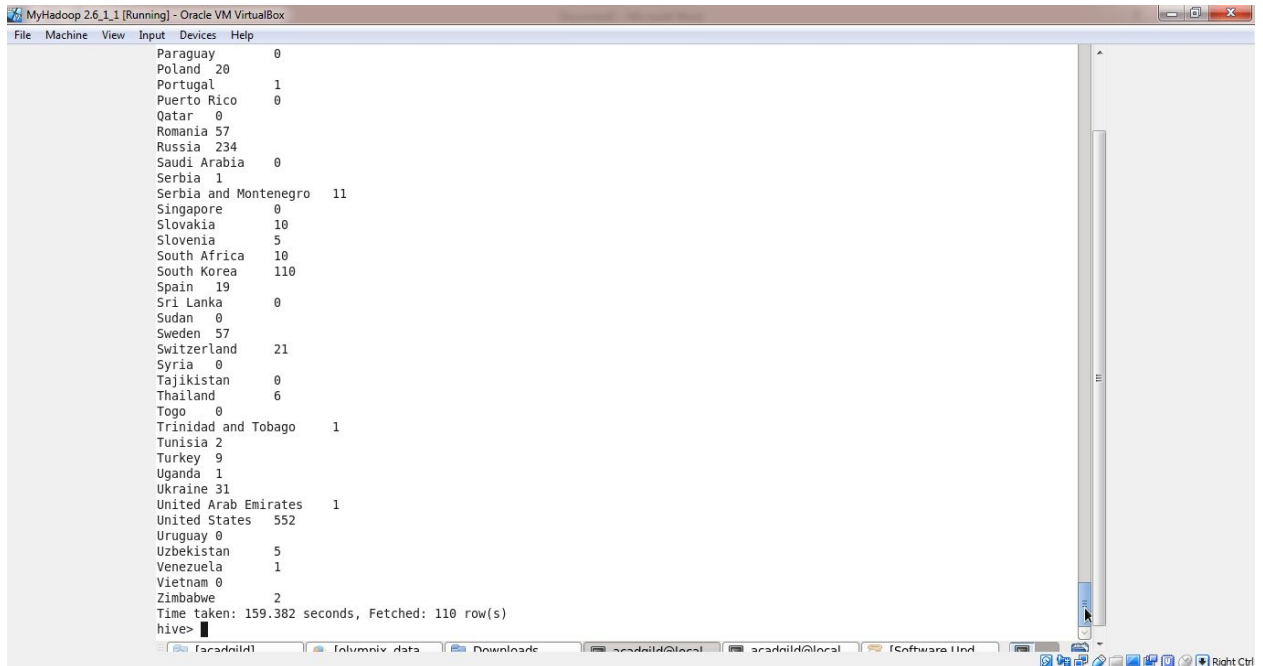


#### 4. Write a Hive program to find the number of gold medals each country won.



MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox			
File	Machine	View	Input Devices Help
			Belgium 2
			Botswana 0
			Brazil 46
			Bulgaria 8
			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
			France 108
			Gabon 0
			Georgia 6
			Germany 223
			Great Britain 124
			Greece 12
			Grenada 1
			Guatemala 0
			Hong Kong 0
			Hungary 77
			Iceland 0
			India 1
			Indonesia 5
			Iran 10
			Ireland 1

MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox			
File	Machine	View	Input Devices Help
			Italy 86
			Jamaica 24
			Japan 57
			Kazakhstan 13
			Kenya 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



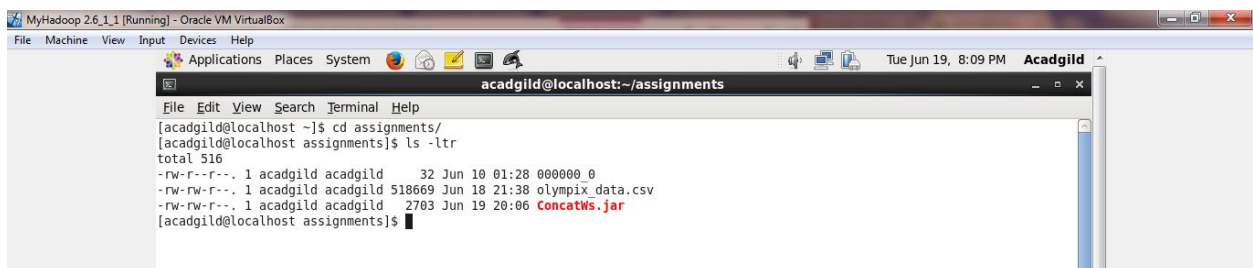
## Task 2

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 a jar for the above with name `ConcatWs.jar`.



Added jar to hive using **Add jar** command and verified if jar is added using **list jars** command.





Create a temporary function for the class Concat\_Ws with name Concat.

Run the function with some examples.



```
hive> > create temporary function Concat as 'udf.Concat_Ws';
OK
Time taken: 3.596 seconds
hive> > select Concat( "Test", "Concat", "WS", "UDF" );
OK
ConcatTestWSTestUDF
Time taken: 9.736 seconds, Fetched: 1 row(s)
hive> > select Concat( "Test", "Concat" );
OK
Concat
Time taken: 0.988 seconds, Fetched: 1 row(s)
hive> select Concat( "Concat" );
FAILED: SemanticException [Error 10015]: Line 1:7 Arguments length mismatch '"Concat"': The function CONCAT_WSneeds at least two arguments.
hive>
```

### Task 3

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.

Below properties need to be set in hive shell so that insert, update and delete operations will work perfectly.

Create a table College with columns clg\_id, clg\_name, clg\_loc columns.

Insert some values to college table as below.



```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
>
>
> 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 = 2;
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: 5.455 seconds
hive> show tables;
OK
college
olympics
Time taken: 0.821 seconds, Fetched: 2 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');
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a differ
ent execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180618230825_0de78e22-b7d8-42c7-9ffa-9291bd215a24
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_1528036691222_0042, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0042/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0042
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-06-18 23:09:41,932 Stage-1 map = 0%, reduce = 0%
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
>
>
> 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 = 2;
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: 5.455 seconds
hive> show tables;
OK
college
olympics
Time taken: 0.821 seconds, Fetched: 2 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');
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a differ
ent execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180618230825_0de78e22-b7d8-42c7-9ffa-9291bd215a24
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_1528036691222_0042, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0042/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0042
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-06-18 23:09:41,932 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:10:31,199 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.75 sec
2018-06-18 23:11:33,442 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.75 sec
2018-06-18 23:12:36,001 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.75 sec
2018-06-18 23:13:32,008 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 26.78 sec
2018-06-18 23:14:32,065 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 29.5 sec
2018-06-18 23:15:07,373 Stage-1 map = 100%, reduce = 92%, Cumulative CPU 83.05 sec
2018-06-18 23:15:20,673 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 91.26 sec
MapReduce Total cumulative CPU time: 1 minutes 35 seconds 910 msec
Ended Job = job_1528036691222_0042
Loading data to table olympic_data.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 95.91 sec HDFS Read: 27182 HDFS Write: 4026 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 35 seconds 910 msec
OK
Time taken: 465.964 seconds
hive>
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
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_1528036691222_0042, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0042/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0042
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-06-18 23:09:41,932 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:10:31,199 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.75 sec
2018-06-18 23:11:33,442 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.75 sec
2018-06-18 23:12:36,001 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.75 sec
2018-06-18 23:13:32,000 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 26.78 sec
2018-06-18 23:14:32,065 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 29.5 sec
2018-06-18 23:15:07,373 Stage-1 map = 100%, reduce = 92%, Cumulative CPU 83.05 sec
2018-06-18 23:15:20,673 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 91.26 sec
MapReduce Total cumulative CPU time: 1 minutes 35 seconds 910 msec
Ended Job = job_1528036691222_0042
Loading data to table olympic_data.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 95.91 sec HDFS Read: 27182 HDFS Write: 4026 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 35 seconds 910 msec
OK
Time taken: 465.964 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: 4.669 seconds, Fetched: 7 row(s)
hive>
```

Below is an update command on a bucketed column which shows error that it is not supported.

```
hive>
>
> UPDATE college set clg_id = 8 where clg_id = 7;
FAILED: SemanticException [Error 10302]: Updating values of bucketing columns is not supported. Column clg_id.
hive>
```

Update command on college table for column clg\_name.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
hive>
> UPDATE college set clg_name = 'IIT' where clg_id = 6;
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_20180618232020_610f7e48-b1c7-4430-b696-c1427d674c75
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_1528036691222_0043, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0043/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0043
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-06-18 23:21:46,500 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:22:49,041 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:23:51,410 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:24:53,492 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:26:01,358 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 30.2 sec
2018-06-18 23:27:06,259 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 30.2 sec
2018-06-18 23:27:41,336 Stage-1 map = 53%, reduce = 0%, Cumulative CPU 65.09 sec
2018-06-18 23:28:15,321 Stage-1 map = 87%, reduce = 0%, Cumulative CPU 92.74 sec
2018-06-18 23:28:18,837 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 94.33 sec
2018-06-18 23:29:21,705 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 94.33 sec
2018-06-18 23:30:21,986 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 94.33 sec
2018-06-18 23:31:23,786 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 94.33 sec
2018-06-18 23:32:03,249 Stage-1 map = 100%, reduce = 32%, Cumulative CPU 108.42 sec
2018-06-18 23:32:08,462 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 111.31 sec
2018-06-18 23:32:49,678 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 122.64 sec
2018-06-18 23:32:54,535 Stage-1 map = 100%, reduce = 99%, Cumulative CPU 139.41 sec
2018-06-18 23:33:15,673 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 140.82 sec
MapReduce Total cumulative CPU time: 2 minutes 28 seconds 820 msec
Ended Job = job_1528036691222_0043
Loading data to table olympic_data.college
MapReduce Jobs Launched:
```

```
hive>
> select * from college;
OK
5      stanford      uk
6      IIT      atp
1      nec      nlr
7      cambridge      us
2      vit      vlr
3      srm      chen
4      lpu      del
Time taken: 2.906 seconds, Fetched: 7 row(s)
hive>
```

Delete command on college table which has column clg\_id = 5.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

> delete from college 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 = acadgild_20180618233719_e0ca89ea-6d89-4e21-8b4a-2567557f80eb
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_1528036691222_0044, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0044/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0044
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-06-18 23:38:28,155 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:39:28,200 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:40:28,527 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:41:29,690 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:43:08,267 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 28.07 sec
2018-06-18 23:43:28,416 Stage-1 map = 13%, reduce = 0%, Cumulative CPU 37.5 sec
2018-06-18 23:43:34,216 Stage-1 map = 53%, reduce = 0%, Cumulative CPU 59.02 sec
2018-06-18 23:43:39,318 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 68.83 sec
2018-06-18 23:43:46,140 Stage-1 map = 93%, reduce = 0%, Cumulative CPU 80.13 sec
2018-06-18 23:43:48,239 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 82.3 sec
2018-06-18 23:44:50,742 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 82.3 sec
2018-06-18 23:45:52,932 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 82.3 sec
2018-06-18 23:46:56,766 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 92.29 sec
2018-06-18 23:47:00,909 Stage-1 map = 100%, reduce = 8%, Cumulative CPU 95.78 sec
2018-06-18 23:47:07,565 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 99.88 sec
2018-06-18 23:47:53,802 Stage-1 map = 100%, reduce = 99%, Cumulative CPU 124.59 sec
2018-06-18 23:48:03,044 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 134.77 sec
MapReduce Total cumulative CPU time: 2 minutes 14 seconds 770 msec
Ended Job = job_1528036691222_0044
Loading data to table olympic_data.college
MapReduce Jobs Launched:
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Hadoop job information for stage 1: number of mappers: 5, number of reducers: 5
2018-06-18 23:38:28,155 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:39:28,200 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:40:28,527 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:41:29,690 Stage-1 map = 0%, reduce = 0%
2018-06-18 23:43:08,267 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 28.07 sec
2018-06-18 23:43:28,416 Stage-1 map = 13%, reduce = 0%, Cumulative CPU 37.5 sec
2018-06-18 23:43:34,216 Stage-1 map = 53%, reduce = 0%, Cumulative CPU 59.02 sec
2018-06-18 23:43:39,318 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 68.83 sec
2018-06-18 23:43:46,140 Stage-1 map = 93%, reduce = 0%, Cumulative CPU 80.13 sec
2018-06-18 23:43:48,239 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 82.3 sec
2018-06-18 23:44:50,742 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 82.3 sec
2018-06-18 23:45:52,932 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 82.3 sec
2018-06-18 23:46:56,766 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 92.29 sec
2018-06-18 23:47:00,909 Stage-1 map = 100%, reduce = 8%, Cumulative CPU 95.78 sec
2018-06-18 23:47:07,565 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 99.88 sec
2018-06-18 23:47:53,802 Stage-1 map = 100%, reduce = 99%, Cumulative CPU 124.59 sec
2018-06-18 23:48:03,044 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 134.77 sec
MapReduce Total cumulative CPU time: 2 minutes 14 seconds 770 msec
Ended Job = job_1528036691222_0044
Loading data to table olympic_data.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 134.77 sec HDFS Read: 50588 HDFS Write: 777 SUCCESS
Total MapReduce CPU Time Spent: 2 minutes 14 seconds 770 msec
OK
Time taken: 656.264 seconds
hive>
>
> select * from college;
OK
6      IIT      atp
1      nec      nlr
7      cambridge      us
2      vit      vlr
3      srm      chen
4      lpu      del
Time taken: 10.374 seconds, Fetched: 6 row(s)
hive>
```