Big Data Hadoop ‘Session 9:Advance Hive Assignment 1’

DATASET

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

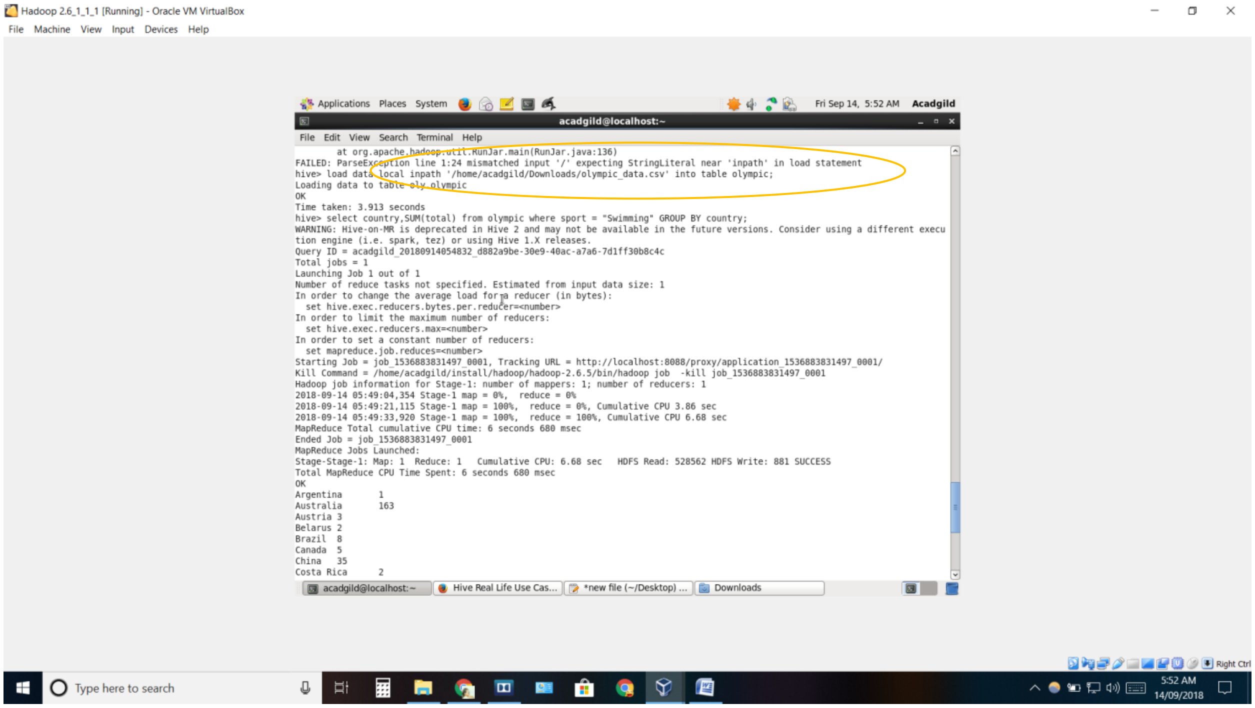
https://drive.google.com/open?id=0ByJLBTmJojjzV1czX3Nha0R3bTQ

DATE SET DESCRIPTION

The data set consists of the following fields

ete: This field consists of the athlete name
This field consists of athlete ages
ntry: This fields consists of the country names which participated in Olympics
: This field consists of the year
ing Date: This field consists of the closing date of ceremony
t: Consists of the sports name
Medals: No. of Gold medals
er Medals: No. of Silver medals
ze Medals: No. of Bronze medals
l Medals: Consists of total no. of medals

LOAD DATA in Hive olympic table

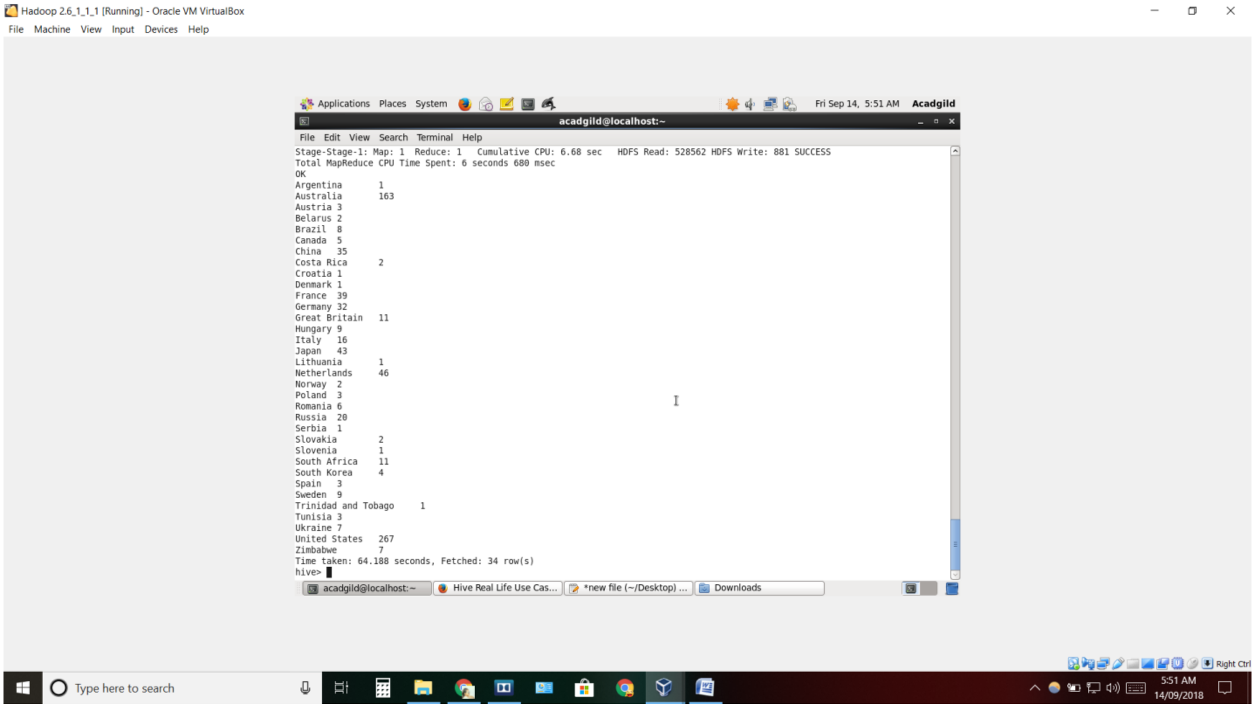

Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Fri Sep 14, 5:52 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~
Hive Real Life Use Cas...
*new file (~/Desktop) ...
Downloads
Type here to search
5:52 AM
14/09/2018

```
        at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:24 mismatched input '/' expecting StringLiteral near 'inpath' in load statement
hive> load data local inpath '/home/acadgild/Downloads/olympic_data.csv' into table olympic;
Loading data to table oly.olympic
OK
Time taken: 3.913 seconds
hive> select country,SUM(total) from olympic 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 execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180914054832_d882a9be-30e9-40ac-a7a6-7d1ff30b8c4c
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_1536883831497_0001, Tracking URL = http://localhost:8088/proxy/application_1536883831497_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1536883831497_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-09-14 05:49:04,354 Stage-1 map = 0%,  reduce = 0%
2018-09-14 05:49:21,115 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.86 sec
2018-09-14 05:49:33,920 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.68 sec
MapReduce Total cumulative CPU time: 6 seconds 680 msec
Ended Job = job_1536883831497_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.68 sec   HDFS Read: 528562 HDFS Write: 881 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 680 msec
OK
Argentina       1
Australia       163
Austria 3
Belarus 2
Brazil  8
Canada  5
China   35
Costa Rica      2
```

Task 1

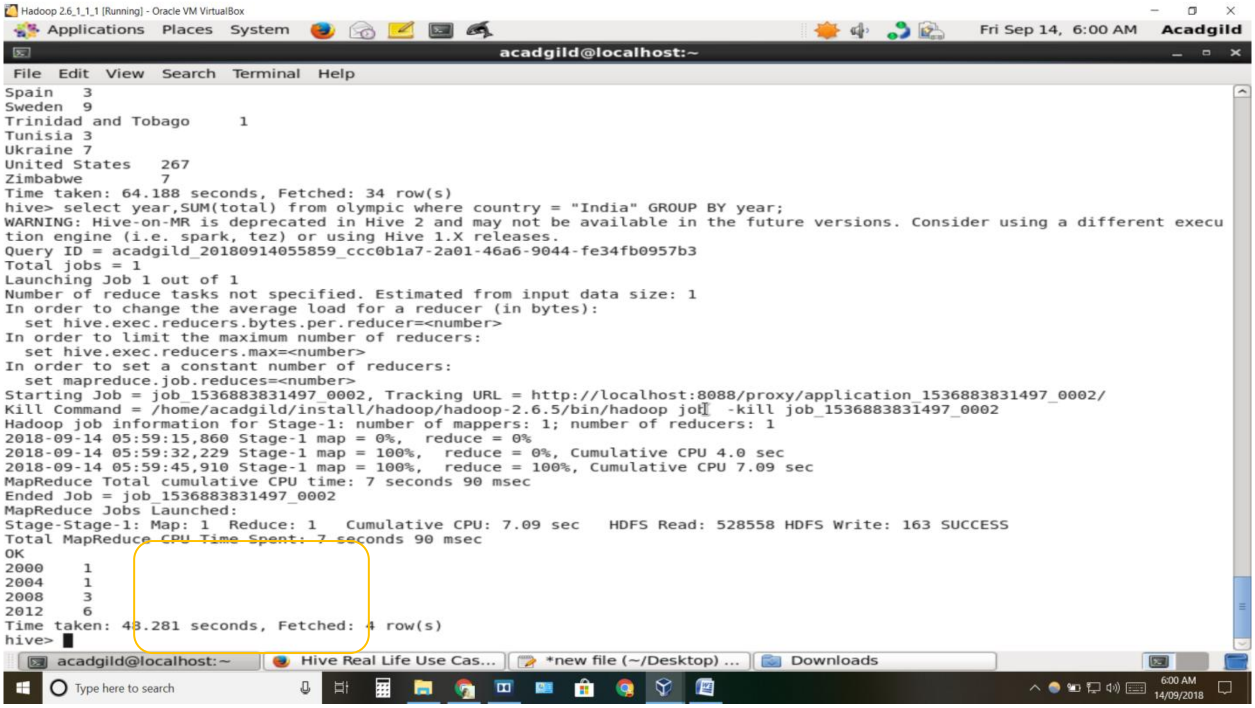
1. Write a Hive program to find the number of medals won by each country in swimming.

   select country,SUM(total) from olympic where sport = “Swimming” GROUP BY country;


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Fri Sep 14, 5:51 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~
Hive Real Life Use Cas...
*new file (~/Desktop) ...
Downloads
Right Ctrl
Type here to search
5:51 AM
14/09/2018

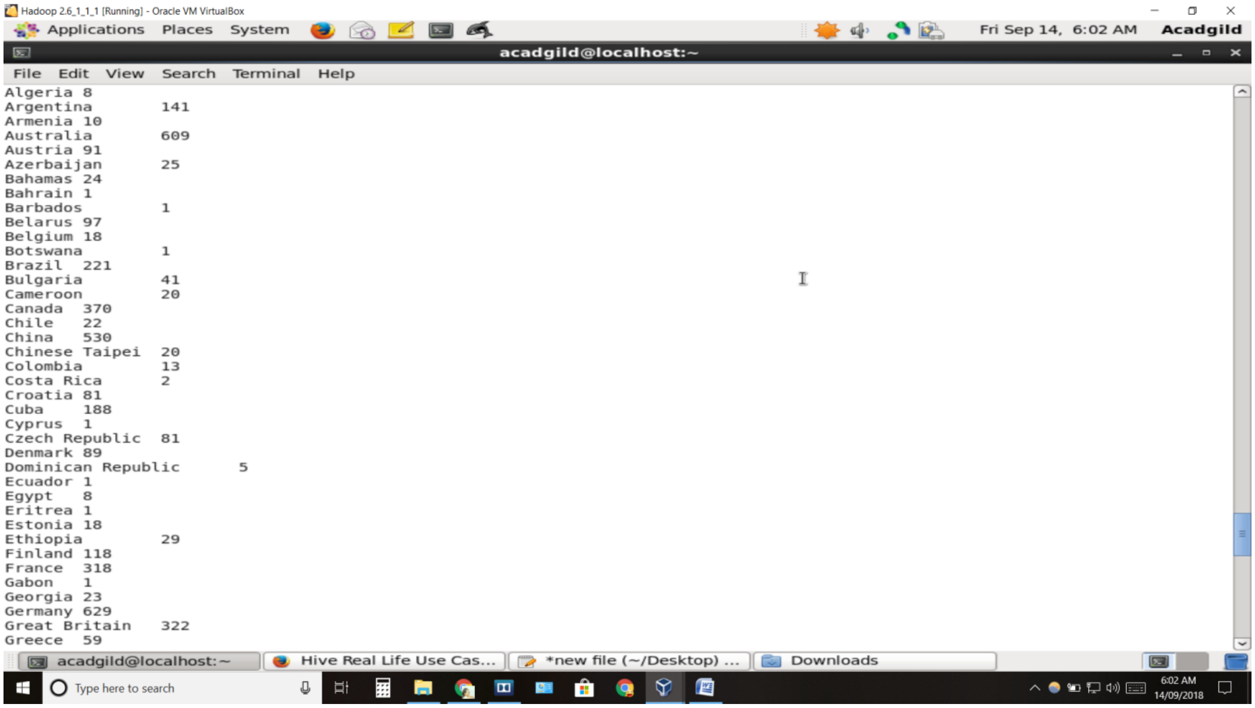
```
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.68 sec   HDFS Read: 528562 HDFS Write: 881 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 680 msec
OK
Argentina       1
Australia       163
Austria 3
Belarus 2
Brazil  8
Canada  5
China   35
Costa Rica      2
Croatia 1
Denmark 1
France  39
Germany 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: 64.188 seconds, Fetched: 34 row(s)
hive> 
```

2. Write a Hive program to find the number of medals that India won year wise.
   select year,SUM(total) from olympic where country = “India” GROUP BY year;


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Fri Sep 14, 6:00 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~
Hive Real Life Use Cas...
*new file (~/Desktop) ...
Downloads
Type here to search
6:00 AM
14/09/2018

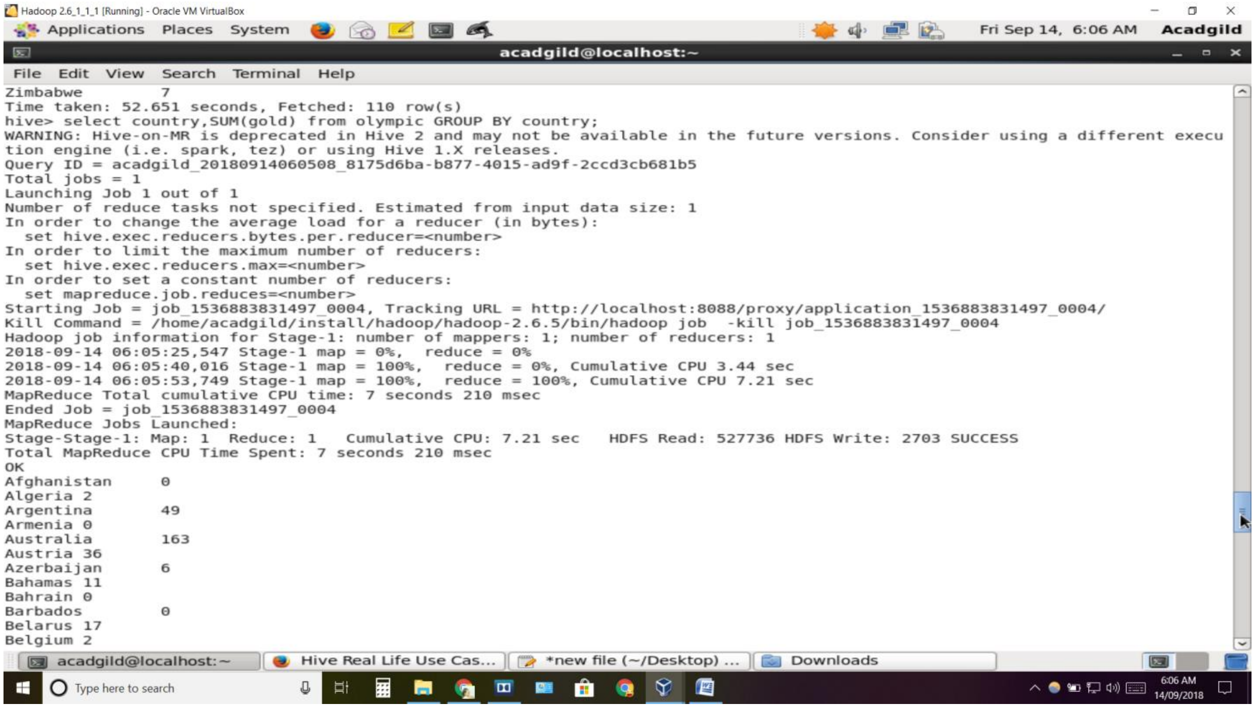
```
Spain   3
Sweden  9
Trinidad and Tobago     1
Tunisia 3
Ukraine 7
United States   267
Zimbabwe        7
Time taken: 64.188 seconds, Fetched: 34 row(s)
hive> select year,SUM(total) from olympic where country = "India" GROUP BY year;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180914055859_ccc0b1a7-2a01-46a6-9044-fe34fb0957b3
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_1536883831497_0002, Tracking URL = http://localhost:8088/proxy/application_1536883831497_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1536883831497_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-09-14 05:59:15,860 Stage-1 map = 0%,  reduce = 0%
2018-09-14 05:59:32,229 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.0 sec
2018-09-14 05:59:45,910 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.09 sec
MapReduce Total cumulative CPU time: 7 seconds 90 msec
Ended Job = job_1536883831497_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.09 sec   HDFS Read: 528558 HDFS Write: 163 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 90 msec
OK
2000    1
2004    1
2008    3
2012    6
Time taken: 48.281 seconds, Fetched: 4 row(s)
hive>
```

3. Write a Hive Program to find the total number of medals each country won.
select country,SUM(total) from olympic GROUP BY country;


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Fri Sep 14, 6:02 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~
Hive Real Life Use Cas...
*new file (~/Desktop) ...
Downloads
Type here to search
6:02 AM
14/09/2018

```
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
France  318
Gabon   1
Georgia 23
Germany 629
Great Britain   322
Greece  59
```

4. Write a Hive program to find the number of gold medals each country won.
select country,SUM(gold) from olympic GROUP BY country;


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Fri Sep 14, 6:06 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~
Hive Real Life Use Cas...
*new file (~/Desktop) ...
Downloads
Type here to search
6:06 AM
14/09/2018

```
Zimbabwe        7
Time taken: 52.651 seconds, Fetched: 110 row(s)
hive> select country,SUM(gold) from olympic 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 execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180914060508_8175d6ba-b877-4015-ad9f-2ccd3cb681b5
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_1536883831497_0004, Tracking URL = http://localhost:8088/proxy/application_1536883831497_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1536883831497_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-09-14 06:05:25,547 Stage-1 map = 0%,  reduce = 0%
2018-09-14 06:05:40,016 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.44 sec
2018-09-14 06:05:53,749 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.21 sec
MapReduce Total cumulative CPU time: 7 seconds 210 msec
Ended Job = job_1536883831497_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.21 sec   HDFS Read: 527736 HDFS Write: 2703 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 210 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
```

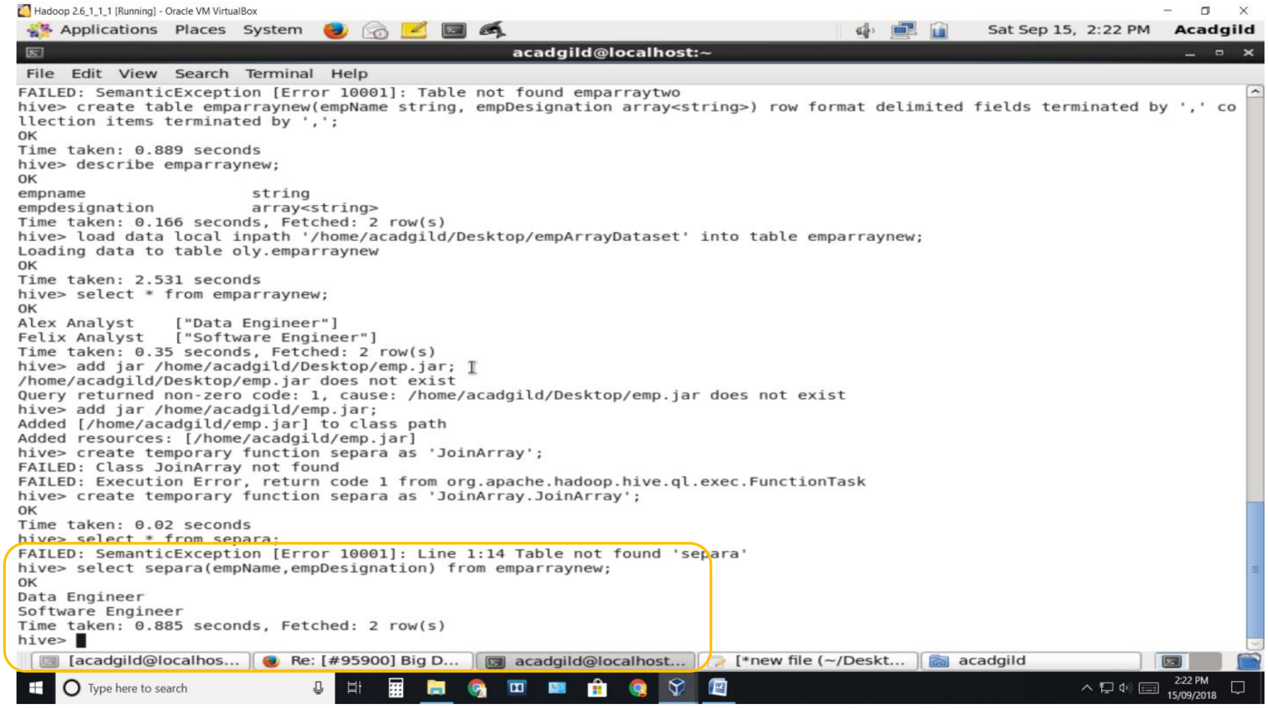
Task 2
Write a hive UDF that implements functionality of string concat_ws(string SEP, array<string>).
This UDF will accept two arguments, one string and one array of string.
It will return a single string where all the elements of the array are separated by the SEP.

Hive UDF

```
import java.util.ArrayList;
import org.apache.hadoop.hive.ql.exec.UDF;
public class JoinArray extends UDF{
public String evaluate (String separator, ArrayList<String> array) {
StringBuffer sBuffer;
if (array == null) {
return null;
}
sBuffer = new StringBuffer();
sBuffer.append(array.get(0));
for (int i=1; i < array.size(); i++) {
sBuffer.append(separator);
sBuffer.append(array.get(i));
}
return sBuffer.toString();
}
}
```

In below screen shot shows following

- Create table emparraynew;
- Describe table emparraynew;
- Load data from local in table emparrynew;
- add jar /home/acadgild/Desktop/empAr.jar;
- create temporary function separa as 'JoinArray.JoinArray';
- select separa(empName,empDesignation) from empArray;



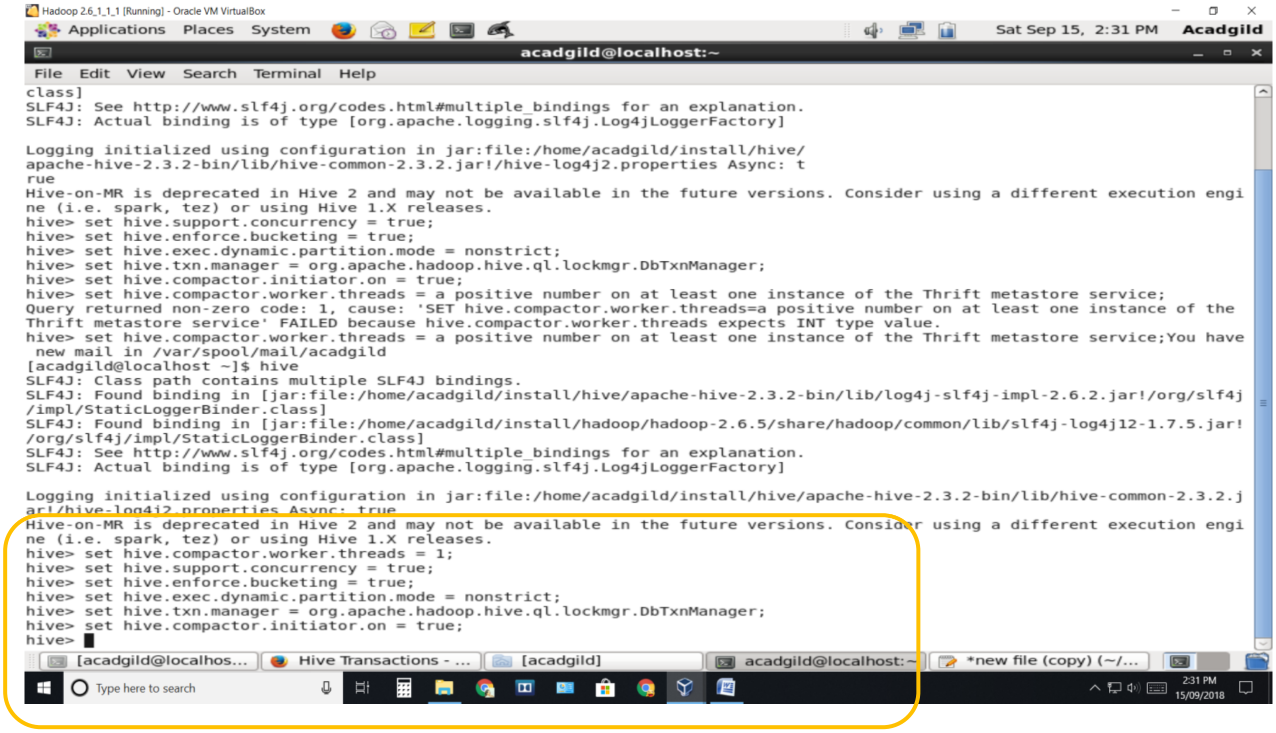
```
FAILED: SemanticException [Error 10001]: Table not found emparraytwo
hive> create table emparraynew(empName string, empDesignation array<string>) row format delimited fields terminated by ',' co
llection items terminated by ',';
OK
Time taken: 0.889 seconds
hive> describe emparraynew;
OK
empname                 string
empdesignation          array<string>
Time taken: 0.166 seconds, Fetched: 2 row(s)
hive> load data local inpath '/home/acadgild/Desktop/empArrayDataset' into table emparraynew;
Loading data to table oly.emparraynew
OK
Time taken: 2.531 seconds
hive> select * from emparraynew;
OK
Alex Analyst    ["Data Engineer"]
Felix Analyst   ["Software Engineer"]
Time taken: 0.35 seconds, Fetched: 2 row(s)
hive> add jar /home/acadgild/Desktop/emp.jar;
/home/acadgild/Desktop/emp.jar does not exist
Query returned non-zero code: 1, cause: /home/acadgild/Desktop/emp.jar does not exist
hive> add jar /home/acadgild/emp.jar;
Added [/home/acadgild/emp.jar] to class path
Added resources: [/home/acadgild/emp.jar]
hive> create temporary function separa as 'JoinArray';
FAILED: Class JoinArray not found
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.FunctionTask
hive> create temporary function separa as 'JoinArray.JoinArray';
OK
Time taken: 0.02 seconds
hive> select * from separa;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'separa'
hive> select separa(empName,empDesignation) from emparraynew;
OK
Data Engineer
Software Engineer
Time taken: 0.885 seconds, Fetched: 2 row(s)
hive>
```

Task 3

Link: https://acadgild.com/blog/transactions-in-hive/

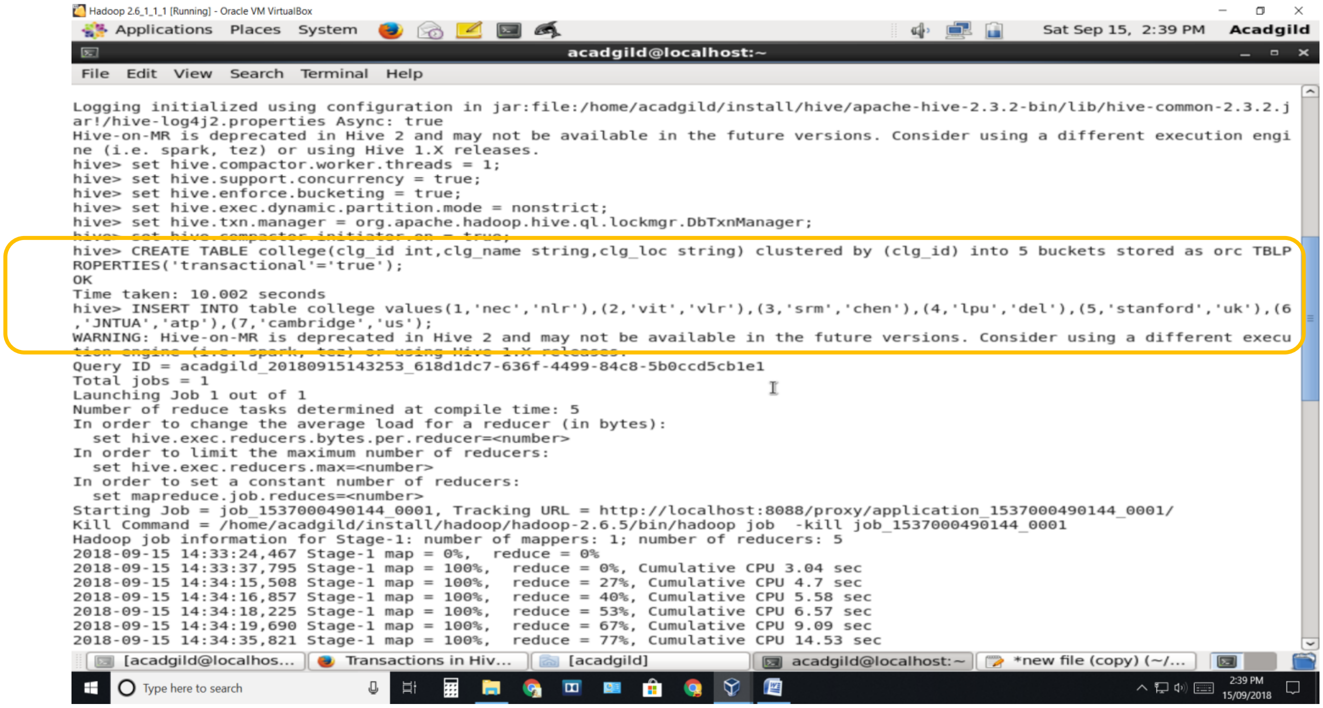
Refer the above given link for transactions in Hive and implement the operations given in the blog using your own sample data set and send us the screenshot.

Enabling ACID transactions in HIVE by using flowing tables.


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Sat Sep 15, 2:31 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhos... Hive Transactions - ... [acadgild] acadgild@localhost:~ *new file (copy) (~/...
Type here to search
2:31 PM
15/09/2018

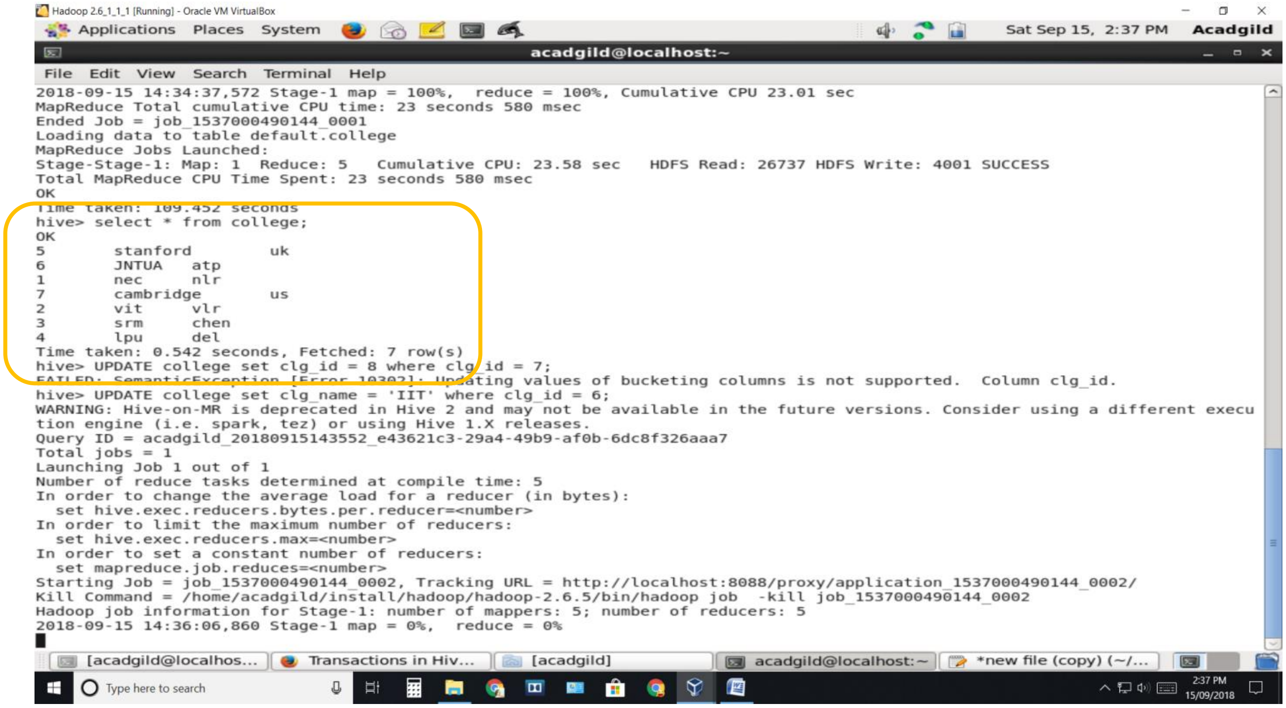
```
class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

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: t
rue
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engi
ne (i.e. spark, tez) or using Hive 1.X releases.
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 = a positive number on at least one instance of the Thrift metastore service;
Query returned non-zero code: 1, cause: 'SET hive.compactor.worker.threads=a positive number on at least one instance of the
Thrift metastore service' FAILED because hive.compactor.worker.threads expects INT type value.
hive> set hive.compactor.worker.threads = a positive number on at least one instance of the Thrift metastore service;You have
 new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
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.j
ar!/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 engi
ne (i.e. spark, tez) or using Hive 1.X releases.
hive> set hive.compactor.worker.threads = 1;
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> █
```

Create table and Insert data in table college;


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Sat Sep 15, 2:39 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhos... Transactions in Hiv... [acadgild] acadgild@localhost:~ *new file (copy) (~/...
Type here to search
2:39 PM
15/09/2018

```
Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.j
ar!/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 engi
ne (i.e. spark, tez) or using Hive 1.X releases.
hive> set hive.compactor.worker.threads = 1;
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> CREATE TABLE college(clg_id int,clg_name string,clg_loc string) clustered by (clg_id) into 5 buckets stored as orc TBLP
ROPERTIES('transactional'='true');
OK
Time taken: 10.002 seconds
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 different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180915143253_618d1dc7-636f-4499-84c8-5b0ccd5cb1e1
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_1537000490144_0001, Tracking URL = http://localhost:8088/proxy/application_1537000490144_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1537000490144_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-09-15 14:33:24,467 Stage-1 map = 0%,  reduce = 0%
2018-09-15 14:33:37,795 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.04 sec
2018-09-15 14:34:15,508 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 4.7 sec
2018-09-15 14:34:16,857 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 5.58 sec
2018-09-15 14:34:18,225 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 6.57 sec
2018-09-15 14:34:19,690 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 9.09 sec
2018-09-15 14:34:35,821 Stage-1 map = 100%,  reduce = 77%, Cumulative CPU 14.53 sec
```

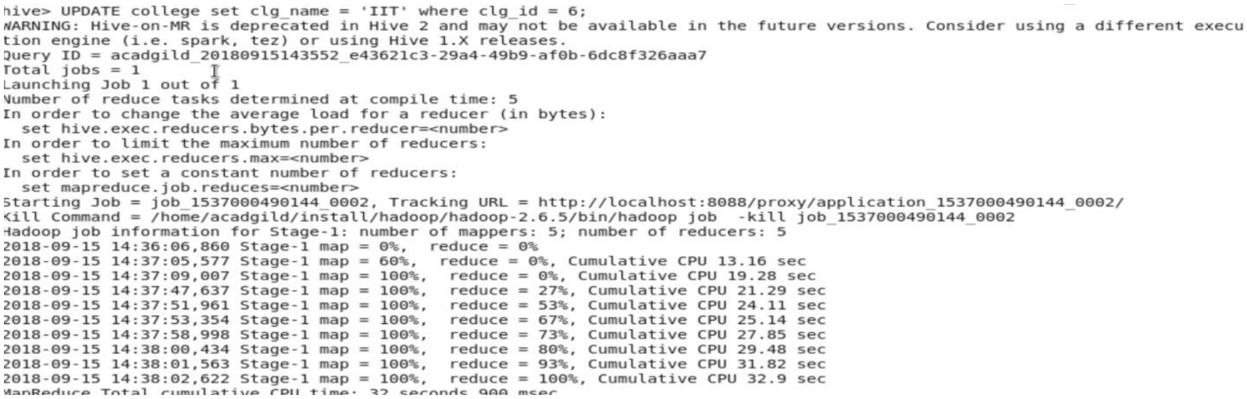
Check data in table after insertion



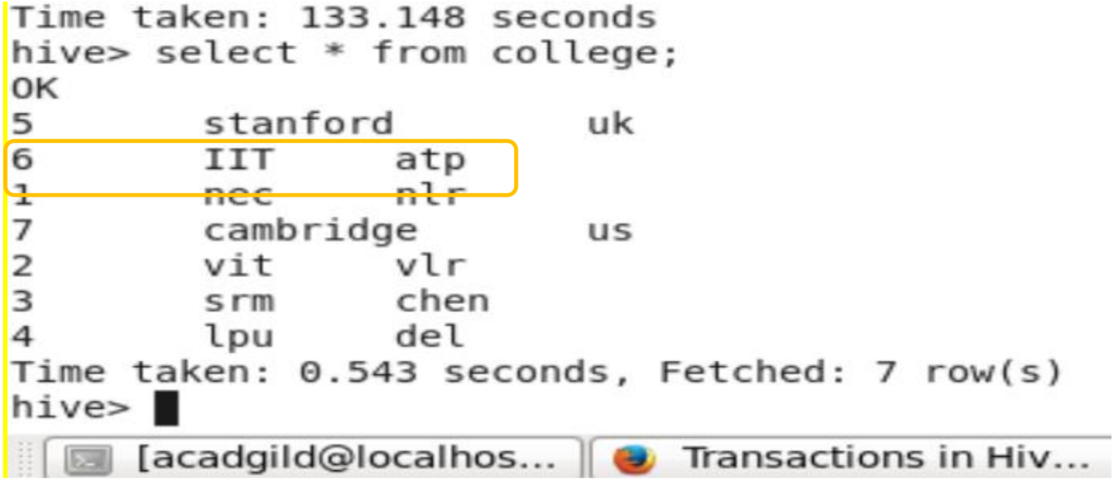
Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Sat Sep 15, 2:37 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhos... Transactions in Hiv... [acadgild] acadgild@localhost:~ *new file (copy) (~/...
Type here to search
2:37 PM 15/09/2018

```
2018-09-15 14:34:37,572 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 23.01 sec
MapReduce Total cumulative CPU time: 23 seconds 580 msec
Ended Job = job_1537000490144_0001
Loading data to table default.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 23.58 sec   HDFS Read: 26737 HDFS Write: 4001 SUCCESS
Total MapReduce CPU Time Spent: 23 seconds 580 msec
OK
Time taken: 109.452 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: 0.542 seconds, Fetched: 7 row(s)
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 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 execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180915143552_e43621c3-29a4-49b9-af0b-6dc8f326aaa7
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_1537000490144_0002, Tracking URL = http://localhost:8088/proxy/application_1537000490144_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1537000490144_0002
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-09-15 14:36:06,860 Stage-1 map = 0%,  reduce = 0%
```

Updating the Data in Hive Table College

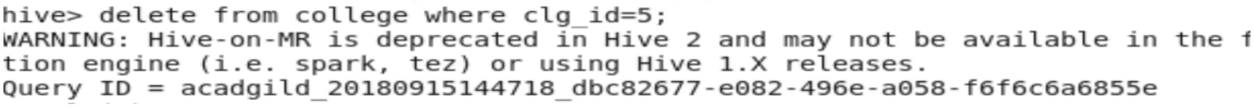


```
hive> UPDATE college set clg_name = 'IIT' where clg_id = 6;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180915143552_e43621c3-29a4-49b9-af0b-6dc8f326aaa7
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_1537000490144_0002, Tracking URL = http://localhost:8088/proxy/application_1537000490144_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1537000490144_0002
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-09-15 14:36:06,860 Stage-1 map = 0%,  reduce = 0%
2018-09-15 14:37:05,577 Stage-1 map = 60%,  reduce = 0%, Cumulative CPU 13.16 sec
2018-09-15 14:37:09,007 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 19.28 sec
2018-09-15 14:37:47,637 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 21.29 sec
2018-09-15 14:37:51,961 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 24.11 sec
2018-09-15 14:37:53,354 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 25.14 sec
2018-09-15 14:37:58,998 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 27.85 sec
2018-09-15 14:38:00,434 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 29.48 sec
2018-09-15 14:38:01,563 Stage-1 map = 100%,  reduce = 93%, Cumulative CPU 31.82 sec
2018-09-15 14:38:02,622 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 32.9 sec
MapReduce Total cumulative CPU time: 32 seconds 900 msec
```

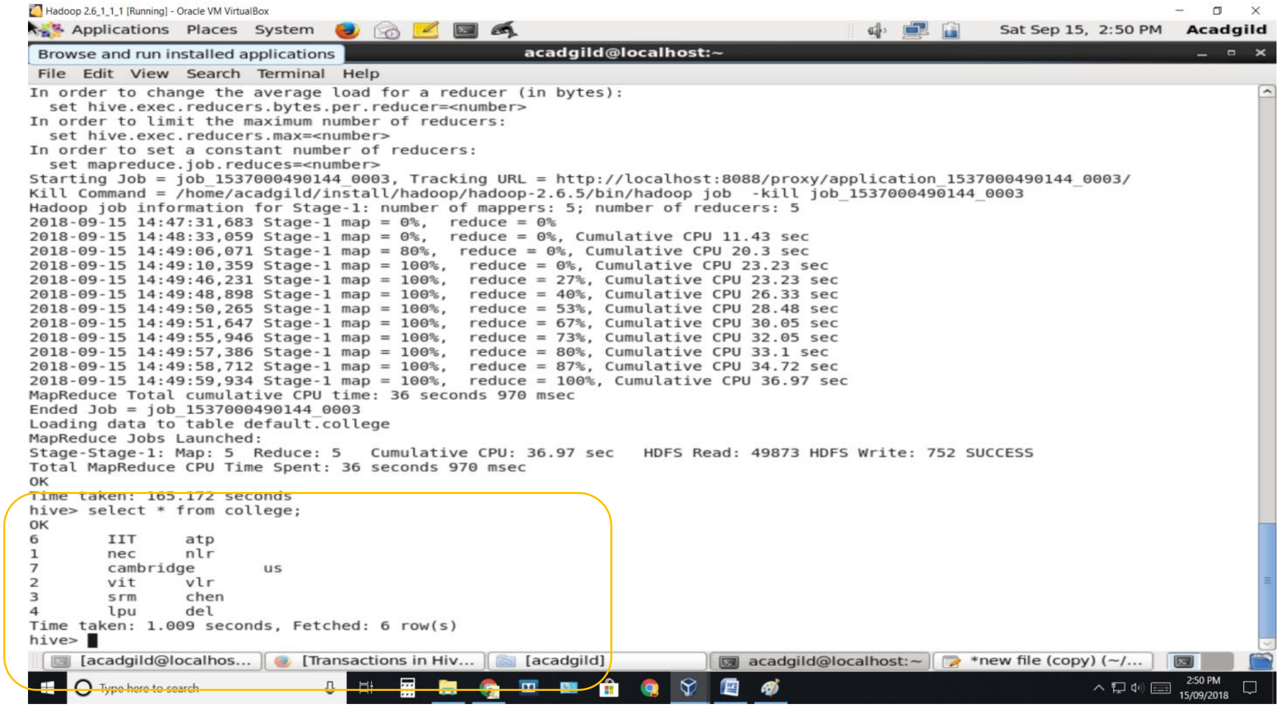


```
Time taken: 133.148 seconds
hive> select * from college;
OK
5         stanford        uk
6         IIT       atp
1         nec       nlr
7         cambridge       us
2         vit       vlr
3         srm       chen
4         lpu       del
Time taken: 0.543 seconds, Fetched: 7 row(s)
hive> █
```

Deleting a Row from Hive Table College



```
hive> delete from college where clg_id=5;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the f
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180915144718_dbc82677-e082-496e-a058-f6f6c6a6855e
```



Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
Sat Sep 15, 2:50 PM Acadgild
Browse and run installed applications
acadgild@localhost:~
File Edit View Search Terminal 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_1537000490144_0003, Tracking URL = http://localhost:8088/proxy/application_1537000490144_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1537000490144_0003
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-09-15 14:47:31,683 Stage-1 map = 0%,  reduce = 0%
2018-09-15 14:48:33,059 Stage-1 map = 0%,  reduce = 0%, Cumulative CPU 11.43 sec
2018-09-15 14:49:06,071 Stage-1 map = 80%,  reduce = 0%, Cumulative CPU 20.3 sec
2018-09-15 14:49:10,359 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 23.23 sec
2018-09-15 14:49:46,231 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 23.23 sec
2018-09-15 14:49:48,898 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 26.33 sec
2018-09-15 14:49:50,265 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 28.48 sec
2018-09-15 14:49:51,647 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 30.05 sec
2018-09-15 14:49:55,946 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 32.05 sec
2018-09-15 14:49:57,386 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 33.1 sec
2018-09-15 14:49:58,712 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 34.72 sec
2018-09-15 14:49:59,934 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 36.97 sec
MapReduce Total cumulative CPU time: 36 seconds 970 msec
Ended Job = job_1537000490144_0003
Loading data to table default.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 36.97 sec   HDFS Read: 49873 HDFS Write: 752 SUCCESS
Total MapReduce CPU Time Spent: 36 seconds 970 msec
OK
Time taken: 165.172 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: 1.009 seconds, Fetched: 6 row(s)
hive> █
```