ASSIGNMENT 7 HIVE

Problem Statement
Task 1
1. Write a Hive program to find the number of medals won by each country in swimming.
2. Write a Hive program to find the number of medals that India won year wise.
3. Write a Hive Program to find the total number of medals each country won.
4. Write a Hive program to find the number of gold medals each country won

EXPLANATION: CREATED A DATABASE NAMED ACADGILD. THE GIVEN DATASET 'olympix_data.csv' WAS DOWNLOADED INTO THE SYSTEM. THEN CREATED A TABLE CALLED 'DATASET.TXT' AS A TEXTFILE USING THE BELOW CODE.

COMMAND: create table dataset(name STRING, age INT, country STRING, year STRING, cldate STRING, sports STRING, gold INT, silver INT, bronze INT, total INT)
> row format delimited fields terminated by '\t' stored as textfile;

SOLUTION:
hive> create table dataset(name STRING, age INT, country STRING, year STRING, cldate STRING, sports STRING, gold INT, silver INT, bronze INT, total INT)
> row format delimited fields terminated by '\t' stored as textfile;
OK
Time taken: 0.263 seconds

EXPLANATION: THEN LOADED THE CSV FILE FROM THE PATH : '/home/acadgild/Downloads/olympix_data.csv'

COMMAND: load data local inpath '/home/acadgild/Downloads/olympix_data.csv' into table dataset;
Loading data to table acadgild.dataset

SOLUTION:
hive> load data local inpath '/home/acadgild/Downloads/olympix_data.csv' into table dataset;
Loading data to table acadgild.dataset
OK
Time taken: 0.992 seconds

TASK

1) COMMAND: select country, sum(total) from dataset where sport = "Swimming" group by country;

EXPLANATION: HERE THE OPERATION WILL SELECT THE TOTAL NUMBER OF MEDALS WON IN SWIMMING FROM THE TABLE DATASET AND DISPLAY THE OUTPUT AND GROUPS ACCORDING TO COUNTRY.

SOLUTION:
hive> select country, sum(total) from dataset where sport = "Swimming" group by country;
FAILED: SemanticException [Error 10004]: Line 1:46 Invalid table alias or column reference 'sport': (possible column names are: name, age, country, year, cldate, sports, gold, silver, bronze, total)
hive> select country, sum(total) from dataset where sports = "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_20181227202003_47a0762f-c5be-4d2c-bf5b-25e78e298be7

```
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_1545900160053_0002, Tracking URL =
http://localhost:8088/proxy/application_1545900160053_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1545900160053_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-12-27 20:20:23,117 Stage-1 map = 0%,  reduce = 0%
2018-12-27 20:20:40,331 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.26 sec
2018-12-27 20:20:56,568 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.26 sec
MapReduce Total cumulative CPU time: 7 seconds 260 msec
Ended Job = job_1545900160053_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.26 sec   HDFS Read: 528586 HDFS Write: 881
SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 260 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: 54.483 seconds, Fetched: 34 row(s)

2) COMMAND: select year, sum(total) from dataset where country = "India" group by year;

EXPLANATION: HERE THE OPERATION WILL CHECK THE CORRESPONDING YEARS WHERE INDIA HAVE WON NUMBER OF MEDALS THEN SELECTS EVERYTHING AND GROUP BY YEAR WISE AND THEN DISPLAY THE OUTPUT.

SOLUTION:
hive> select year, sum(total) from dataset 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_20181227203236_ff86977f-6316-4237-a6f0-bdc891351ffb
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_1545900160053_0003, Tracking URL =
http://localhost:8088/proxy/application_1545900160053_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1545900160053_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-12-27 20:32:59,940 Stage-1 map = 0%, reduce = 0%
2018-12-27 20:33:15,865 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.5 sec
2018-12-27 20:33:29,908 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.47 sec
MapReduce Total cumulative CPU time: 7 seconds 470 msec
Ended Job = job_1545900160053_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 7.47 sec   HDFS Read: 528542 HDFS Write: 163 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 470 msec
OK
2000 1
2004 1
2008 3
2012 6
Time taken: 54.31 seconds, Fetched: 4 row(s)

3) COMMAND: select country, sum(total) from dataset group by country;

EXPLANATION: HERE OPERATION WILL SELECT DISPLAY TOTAL NUMBER OF MEDALS WON BY EACH COUNTRY.

SOLUTION:

```
hive> select country, sum(total) from dataset 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_20181227204043_8b8f5e9c-6df6-48bf-8969-6e86fbc62228
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_1545900160053_0004, Tracking URL =
http://localhost:8088/proxy/application_1545900160053_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1545900160053_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-12-27 20:40:57,334 Stage-1 map = 0%,  reduce = 0%
2018-12-27 20:41:09,060 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.73 sec
2018-12-27 20:41:21,706 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.33 sec
MapReduce Total cumulative CPU time: 5 seconds 330 msec
Ended Job = job_1545900160053_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.33 sec   HDFS Read: 527730 HDFS Write: 2742
SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 330 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
France 318
Gabon 1
Georgia 23
Germany 629
Great Britain 322
Greece 59
Grenada 1
Guatemala 1
Hong Kong 3
Hungary 145
Iceland 15
India 11
Indonesia 22
Iran 24
Ireland 9
Israel 4
Italy 331
Jamaica 80
Japan 282
Kazakhstan 42
Kenya 39
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 80

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 308
Spain 205
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: 40.904 seconds, Fetched: 110 row(s)

4) COMMAND: select country, sum(gold) from dataset group by country;

EXPLANATION: DISPLAYS THE NUMBER OF GOLD MEDAL WON BY EACH COUNTRY.

SOLUTION:
hive> select country, sum(gold) from dataset 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_20181227204415_7a502e72-bce9-4e5d-aec2-f3e45b47aa7b
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_1545900160053_0005, Tracking URL =
http://localhost:8088/proxy/application_1545900160053_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1545900160053_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-12-27 20:44:30,304 Stage-1 map = 0%, reduce = 0%
2018-12-27 20:44:43,214 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.76 sec
2018-12-27 20:44:57,149 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.77 sec
MapReduce Total cumulative CPU time: 5 seconds 770 msec
Ended Job = job_1545900160053_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 5.77 sec   HDFS Read: 527728 HDFS Write: 2703
SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 770 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 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
Israel 1
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

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: 42.742 seconds, Fetched: 110 row(s)

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.

DATASET:

6 kevin 22000
1 shra 23000
6 kevin 22000
1 shra 23000
7 akhi 45900
2 rahu 12900
7 akhi 45900
2 rahu 12900
3 rahul 36000
3 rahul 36000
4 aka 34000
4 aka 34000

EXPLANATION:USED ARRAY STRING AS NAME AND SEPARATOR AS COMMA (,) .
SOLUTION:
hive> select CONCAT_WS(' ',name,',',name) from college;
OK
kevin , kevin
shra , shra

kevin , kevin
shra , shra
akhi , akhi
rahu , rahu
akhi , akhi
rahu , rahu
rahul , rahul
rahul , rahul
aka , aka
aka , aka

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.

EXPLANATION: HERE I CREATED A TABLE COLLEGE WITH THE FIELDS AS "ID, NAME, SALARY".

SOLUTION:

```
hive> CREATE TABLE college(id int, name string, sal int) clustered by (id) into 5 buckets stored as orc
TBLPROPERTIES('transactional'='true');
OK
Time taken: 3.154 seconds
```

1) INSERT OPERATION:

EXPLANATION: HERE I INSERTED SOME DATA INTO THE TABLE "COLLEGE" AND USING SELECT OPERATION THE DATA CAN SEEN INSERTED INTO THE TABLE.

SOLUTION:

```
hive> INSERT INTO table college
values(1,'shra','23000'),(2,'rahu','12900'),(3,'rahul','36000'),(4,'aka','34000'),(5,'ricki','10000'),(6,'shane','22000'),(7,'
akhi','45900');
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_20181227220132_812ba947-8dd4-4e94-85b1-501f9280b048
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_1545900160053_0006, Tracking URL =
http://localhost:8088/proxy/application_1545900160053_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1545900160053_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-12-27 22:02:04,284 Stage-1 map = 0%,  reduce = 0%
```

2018-12-27 22:02:33,301 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.5 sec
2018-12-27 22:03:35,086 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.33 sec
2018-12-27 22:03:36,684 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 11.58 sec
2018-12-27 22:03:38,311 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 12.21 sec
2018-12-27 22:04:07,709 Stage-1 map = 100%, reduce = 86%, Cumulative CPU 29.39 sec
2018-12-27 22:04:09,667 Stage-1 map = 100%, reduce = 92%, Cumulative CPU 33.0 sec
2018-12-27 22:04:13,533 Stage-1 map = 100%, reduce = 96%, Cumulative CPU 35.32 sec
2018-12-27 22:04:15,631 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 37.78 sec
MapReduce Total cumulative CPU time: 37 seconds 780 msec
Ended Job = job_1545900160053_0006
Loading data to table acadgild.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5   Cumulative CPU: 37.78 sec   HDFS Read: 26944 HDFS Write: 3998 SUCCESS
Total MapReduce CPU Time Spent: 37 seconds 780 msec
OK
Time taken: 169.508 seconds
hive> select * from college;
OK
5 ricki 10000
6 shane 22000
1 shra 23000
7 akhi 45900
2 rahu 12900
3 rahul 36000
4 aka 34000
Time taken: 0.871 seconds, Fetched: 7 row(s)

EXPLANATION: HERE I AGAIN DONE THE INSERT OPERATION. THIS OPERATION WILL APPEND THE SAME DATA INSIDE THE TABLE AND IT CAN BE SEEN 2 SAME ENTRIES FOR EACH DATA AND USING SELECT OPERATION DISPLAYED THE OUTPUT.

SOLUTION:
hive> INSERT INTO table college
values(1,'shra','23000'),(2,'rahu','12900'),(3,'rahul','36000'),(4,'aka','34000'),(5,'ricki','10000'),(6,'shane','22000'),(7,'akhi','45900');
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_20181227220559_39fd657c-ef70-4ec9-8c3a-2f4d87f7a65f
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_1545900160053_0007, Tracking URL = http://localhost:8088/proxy/application_1545900160053_0007/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1545900160053_0007

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-12-27 22:06:26,080 Stage-1 map = 0%, reduce = 0%
2018-12-27 22:06:44,421 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.28 sec
2018-12-27 22:07:39,955 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 5.7 sec
2018-12-27 22:07:41,613 Stage-1 map = 100%, reduce = 21%, Cumulative CPU 6.72 sec
2018-12-27 22:07:43,279 Stage-1 map = 100%, reduce = 34%, Cumulative CPU 8.3 sec
2018-12-27 22:07:44,856 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 8.75 sec
2018-12-27 22:07:49,974 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 11.22 sec
2018-12-27 22:07:51,630 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 12.7 sec
2018-12-27 22:08:07,806 Stage-1 map = 100%, reduce = 73%, Cumulative CPU 17.75 sec
2018-12-27 22:08:12,810 Stage-1 map = 100%, reduce = 77%, Cumulative CPU 21.94 sec
2018-12-27 22:08:16,016 Stage-1 map = 100%, reduce = 83%, Cumulative CPU 26.28 sec
2018-12-27 22:08:17,632 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 31.28 sec
2018-12-27 22:08:18,926 Stage-1 map = 100%, reduce = 99%, Cumulative CPU 35.15 sec
2018-12-27 22:08:20,400 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 36.08 sec
MapReduce Total cumulative CPU time: 36 seconds 80 msec
Ended Job = job_1545900160053_0007
Loading data to table acadgild.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 5   Cumulative CPU: 36.08 sec   HDFS Read: 26774 HDFS Write: 3998 SUCCESS
Total MapReduce CPU Time Spent: 36 seconds 80 msec
OK
Time taken: 144.766 seconds

hive> select * from college;
OK
5 ricki 10000
5 ricki 10000
6 shane 22000
1 shra 23000
6 shane 22000
1 shra 23000
7 akhi 45900
2 rahu 12900
7 akhi 45900
2 rahu 12900
3 rahul 36000
3 rahul 36000
4 aka 34000
4 aka 34000
Time taken: 0.6 seconds, Fetched: 14 row(s)

2) UPDATE OPERATION:

EXPLANATION: HERE THIS OPERATION IS USED TO UPDATE INSIDE THE TABLE BY REPLACING THE NAME OF THE PERSON WITH ID=6 AS KEVIN AND USING THE SELECT OPERATION IT CAN SEEN THAT THE RECORD HAS BEEN UPDATED.

SOLUTION:
hive> UPDATE college set name = 'kevin' where 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_20181227221158_0d054ccd-39d8-49cd-b781-3aa34e29c9ba
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_1545900160053_0008, Tracking URL =
http://localhost:8088/proxy/application_1545900160053_0008/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1545900160053_0008
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-12-27 22:12:19,261 Stage-1 map = 0%,  reduce = 0%
2018-12-27 22:13:19,761 Stage-1 map = 0%,  reduce = 0%
2018-12-27 22:13:51,261 Stage-1 map = 13%,  reduce = 0%, Cumulative CPU 17.46 sec
2018-12-27 22:13:53,747 Stage-1 map = 20%,  reduce = 0%, Cumulative CPU 17.46 sec
2018-12-27 22:13:56,750 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 32.06 sec
2018-12-27 22:14:53,523 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 34.07 sec
2018-12-27 22:14:57,271 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 37.27 sec
2018-12-27 22:15:00,556 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 38.36 sec
2018-12-27 22:15:02,135 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 39.71 sec
2018-12-27 22:15:13,224 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 45.85 sec
2018-12-27 22:15:14,763 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 47.7 sec
2018-12-27 22:15:16,093 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 52.16 sec
MapReduce Total cumulative CPU time: 52 seconds 160 msec
Ended Job = job_1545900160053_0008
Loading data to table acadgild.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5  Reduce: 5   Cumulative CPU: 52.16 sec   HDFS Read: 55733 HDFS Write: 988
SUCCESS
Total MapReduce CPU Time Spent: 52 seconds 160 msec
OK
Time taken: 202.226 seconds

hive> select * from college;
OK
5 ricki 10000
5 ricki 10000
6 kevin 22000
1 shra 23000
6 kevin 22000
1 shra 23000
7 akhi 45900
2 rahu 12900
7 akhi 45900
2 rahu 12900
3 rahul 36000
3 rahul 36000
```

4 aka 34000
4 aka 34000
Time taken: 0.62 seconds, Fetched: 14 row(s)

3) DELETE OPERATION:

EXPLANATION: HERE USING THIS OPERATION PERSON WITH ID=5 IS DELETED FROM THE TABLE "COLLGE" AND USING THE SELECT OPERATION IT CAN BE SEEN THE RECORD HAS BEEN DELETED FROM THE TABLE.

SOLUTION:
hive> delete from college where 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_20181227221718_ce7719af-6e77-4059-8b66-0a2ea50c0c29
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_1545900160053_0009, Tracking URL =
http://localhost:8088/proxy/application_1545900160053_0009/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1545900160053_0009
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-12-27 22:17:42,390 Stage-1 map = 0%, reduce = 0%
2018-12-27 22:18:42,804 Stage-1 map = 0%, reduce = 0%
2018-12-27 22:19:47,545 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 13.78 sec
2018-12-27 22:19:52,680 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 37.85 sec
2018-12-27 22:20:54,378 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 37.85 sec
2018-12-27 22:21:01,422 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 44.16 sec
2018-12-27 22:21:03,215 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 44.69 sec
2018-12-27 22:21:19,105 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 53.71 sec
2018-12-27 22:21:20,317 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 57.66 sec
MapReduce Total cumulative CPU time: 58 seconds 30 msec
Ended Job = job_1545900160053_0009
Loading data to table acadgild.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5   Cumulative CPU: 58.03 sec   HDFS Read: 53947 HDFS Write: 760
SUCCESS
Total MapReduce CPU Time Spent: 58 seconds 30 msec
OK
Time taken: 247.307 seconds

hive> select * from college;
OK
6 kevin 22000
1 shra 23000

6 kevin 22000
1 shra 23000
7 akhi 45900
2 rahu 12900
7 akhi 45900
2 rahu 12900
3 rahul 36000
3 rahul 36000
4 aka 34000
4 aka 34000
Time taken: 0.675 seconds, Fetched: 12 row(s)