

Assignment 9

Task 1:

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

<https://drive.google.com/open?id=0ByJLBTmJojzV1czX3NhaOR3bTQ>

Create Table:

CREATE TABLE olympic

```
acadgild@localhost:~/hive
hive> CREATE TABLE olympic
> (
>   athlete STRING,
>   age INT,
>   country STRING,
>   year STRING,
>   closing STRING,
>   sport STRING,
>   gold INT,
>   silver INT,
>   bronze INT,
>   total INT
> )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY '\t';
OK
Time taken: 3.33 seconds
```

Load the data into table:

LOAD DATA LOCAL INPATH '/home/acadgild/hive/olympix_data.csv'

INTO TABLE olympic;

```
acadgild@localhost:~
hive> LOAD DATA LOCAL INPATH '/home/acadgild/hive/olympix_data.csv'
> INTO TABLE olympic;
Loading data to table default.olympic
OK
Time taken: 3.876 seconds
hive> █
```

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

Query:

select country,SUM(total) from olympic where sport='Swimming' Group By country;

Output:

```
acadmild@localhost:~$ 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 execution engine (i.e. spark, tez) or using Hive
1.X releases.
Query ID = acadmild_20180219202038_c07439fa-1dae-4d67-b3bc-e7424fb9fa30
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.pct.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_1519042378766_0004, Tracking URL = http://localhost:8088/proxy/application_1519042378766_0004/
Kill Command = /home/acadmild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1519042378766_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-02-19 20:21:07,990 Stage-1 map = 0%, reduce = 0%
2018-02-19 20:21:23,997 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.32 sec
2018-02-19 20:21:38,441 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.03 sec
MapReduce Total cumulative CPU time: 8 seconds 30 msec
Ended Job = job_1519042378766_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.03 sec HDFS Read: 1580529 HDFS Write: 881 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 30 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

acadmild@localhost:~$ 2018-02-19 20:21:33,997 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.32 sec
2018-02-19 20:21:38,441 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.03 sec
MapReduce Total cumulative CPU time: 8 seconds 30 msec
Ended Job = job_1519042378766_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.03 sec HDFS Read: 1580529 HDFS Write: 881 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 30 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: 61.581 seconds, Fetched: 34 row(s)
hive>
```

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

Query:

select year,SUM(total) from olympic where country='India' Group By year;

Output:

```
acavgild@localhost:~$
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 execution engine (i.e. spark, tez) or using Hive
1.X releases.
Query ID = acavgild_20180219203221_f97f4bee-c22c-440e-991e-f94731028222
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_1519042378766_0005, Tracking URL = http://localhost:8088/proxy/application_1519042378766_0005/
Kill Command = /home/acavgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1519042378766_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-02-19 20:32:41,407 Stage-1 map = 0%, reduce = 0%
2018-02-19 20:32:54,951 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.17 sec
2018-02-19 20:33:10,299 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.67 sec
MapReduce Total cumulative CPU time: 7 seconds 670 msec
Ended Job = job_1519042378766_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.67 sec HDFS Read: 1580525 HDFS Write: 163 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 670 msec
OK
2000 1
2004 1
2008 3
2012 6
Time taken: 49.489 seconds, Fetched: 4 row(s)
hive>
```

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

Query:

select country,SUM(total) from olympic Group By country;

Output:

```
acavgild@localhost:~$
hive> select country,SUM(total) 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 execution engine (i.e. spark, tez) or using Hive
1.X releases.
Query ID = acavgild_20180219203702_78cbcf13-bba1-4107-b7b0-6855a52dbfe2
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_1519042378766_0006, Tracking URL = http://localhost:8088/proxy/application_1519042378766_0006/
Kill Command = /home/acavgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1519042378766_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-02-19 20:37:22,851 Stage-1 map = 0%, reduce = 0%
2018-02-19 20:37:38,227 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.41 sec
2018-02-19 20:37:52,478 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.95 sec
MapReduce Total cumulative CPU time: 6 seconds 950 msec
Ended Job = job_1519042378766_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.95 sec HDFS Read: 1579705 HDFS Write: 2760 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 950 msec
OK
NULL NULL
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
```

```

acadgild@localhost:~
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
El Salvador 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

```

```

acadgild@localhost:~
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

```

```

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: 51.769 seconds, Fetched: 111 row(s)
hive> █

```

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

Query:

select country,SUM(gold) from olympic Group By country;

Output:

```
acdgild@localhost:~$ 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 execution engine (i.e. spark, tez) or using Hive
1.X releases.
Query ID = acdgild_20180219204849_3927075a-011a-4b0b-af8c-31e3c3822bc7
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_1519042378766_0007, Tracking URL = http://localhost:8080/proxy/application_1519042378766_0007/
Kill Command = /home/acdgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1519042378766_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-02-19 20:49:08,239 Stage-1 map = 0%, reduce = 0%
2018-02-19 20:49:21,316 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.96 sec
2018-02-19 20:49:34,271 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.65 sec
MapReduce Total cumulative CPU time: 5 seconds 650 msec
Ended Job = job_1519042378766_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.65 sec HDFS Read: 1579703 HDFS Write: 2721 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 650 msec
OK
NULL      NULL
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
```

```
acsdgild@localhost:~  
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  
  
acsdgild@localhost:~  
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: 46.35 seconds, Fetched: 111 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.

Solution:

We have companies list and its company website URL, but the 'www' and the remaining domain are separated. In our output we try to achieve the output as below,

Dataset:

Cat > comp.txt

```
acadgild@localhost:~  
login as: acadgild  
acadgild@192.168.85.129's password:  
Last login: Sun Feb 18 22:16:37 2018 from 192.168.85.1  
[acadgild@localhost ~]$ cat > comp.txt  
1,wipro,www,wipro.com  
2,infosys,www,infosys.com  
3,google,www,google.com  
4,apple,www,apple.com  
5,walmart,www,walmart.com  
^C  
[acadgild@localhost ~]$ cat comp.txt  
1,wipro,www,wipro.com  
2,infosys,www,infosys.com  
3,google,www,google.com  
4,apple,www,apple.com  
5,walmart,www,walmart.com  
[acadgild@localhost ~]$ █
```

Create Database:

Create database company;

```
acadgild@localhost:~  
hive> create database company;  
Sun Feb 18 22:24:57 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.  
Sun Feb 18 22:24:58 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.  
Sun Feb 18 22:24:59 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.  
Sun Feb 18 22:24:59 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.  
Sun Feb 18 22:25:05 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.  
Sun Feb 18 22:25:05 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.  
Sun Feb 18 22:25:05 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.  
OK  
Time taken: 14.75 seconds  
hive> █
```

Show databases;

Use company;

```

acadgild@localhost:~
hive> show databases;
OK
advanced
company
custom
default
fortune
fortunel
olympics
retail
Time taken: 0.354 seconds, Fetched: 8 row(s)
hive> use company;
OK
Time taken: 0.073 seconds
hive> █

```

Create Table and Load the data:

-> CREATE TABLE company1(rank int,company_name string,website string,protocal string)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

-> LOAD DATA LOCAL INPATH '/home/acadgild/comp.txt'

INTO TABLE company1;

```

acadgild@localhost:~
hive> CREATE TABLE company1(rank int,company_name string,website string,protocal string)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
OK
Time taken: 1.585 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/comp.txt'
> INTO TABLE company1;
Loading data to table company.company1
OK
Time taken: 2.029 seconds
hive> █

```

View the data in the table company1:

Select * from company1;

```

acadgild@localhost:~
hive> SELECT * FROM company1;
OK
1      wipro   www      wipro.com
2      infosys www      infosys.com
3      google  www      google.com
4      apple   www      apple.com
5      walmart www      walmart.com
Time taken: 2.926 seconds, Fetched: 5 row(s)
hive> █

```


HIVE UDF java code:

```
package concatws;

import org.apache.hadoop.hive.q1.exec.UDF;
import org.apache.hadoop.hive.q1.exec.Description;
@Description(name = "concatws", value = "_FUNC_(string SEP, array<string>) -  
RETURN_TYPE(String)\n" + "Description: Concatenate two strings, separated by the  
seperator",  
extended = "Example:\n"  
    + "> SELECT CONCAT_WS (website, '.', protocol) FROM src;\n"  
    + "www.walmart.com")

public class hiveUDF extends UDF

{
    public String evaluate(String param1, String[] param2)

    {
        String Output = "";
        if(param1==null && param2==null)
        {
            return null;
        }
        for(int i = 0; i < param2.length; i++)
        {
            Output+= param2[i];
        }
        return(param1.concat(Output));
    }
}
```

After that we are adding JAR created from the JAVA class which is defining the UDF using below syntax-

Adding the jar function:

Add jar /home/acadgild/udf.jar;

```
hive> add jar /home/acadgild/udf.jar;
Added [/home/acadgild/udf.jar] to class path
Added resources: [/home/acadgild/udf.jar]
hive>
```

creating a temporary function "CONCAT_WS"

```
CREATE TEMPORARY FUNCTION CONCAT_WS AS 'concatws.hiveUDF';
```

```
acadgild@localhost:~  
hive> CREATE TEMPORARY FUNCTION CONCAT_WS AS 'concatws.hiveUDF';  
OK  
Time taken: 0.027 seconds  
hive>
```

After that we run below query to take one column (company_name) input as String and another array(website,'.',protocol) as Array of Strings and concatenate them,

Select rank,company_name,CONCAT_WS(website,'.',protocol) from company1;

```
acadgild@localhost:~  
hive> SELECT rank,company_name,CONCAT_WS(website,'.',protocol) from company1;  
OK  
1      wipro    www.wipro.com  
2      infosys www.infosys.com  
3      google  www.google.com  
4      apple   www.apple.com  
5      walmart www.walmart.com  
Time taken: 1.171 seconds, Fetched: 5 row(s)  
hive>
```

Task3:

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.

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

```

acsdgild@localhost:~
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.q1.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 hi
ve.compactor.worker.threads expects INT type value.
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');

```

acsdgild@localhost:~
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');
Sun Feb 18 23:01:42 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit op
tion isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'f
alse'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for s
erver certificate verification.
Sun Feb 18 23:01:42 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit op
tion isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'f
alse'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for s
erver certificate verification.
Sun Feb 18 23:01:42 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit op
tion isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'f
alse'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for s
erver certificate verification.
Sun Feb 18 23:01:42 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit op
tion isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'f
alse'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for s
erver certificate verification.
Sun Feb 18 23:01:42 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit op
tion isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'f
alse'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for s
erver certificate verification.
Sun Feb 18 23:01:43 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit op
tion isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'f
alse'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for s
erver certificate verification.
Sun Feb 18 23:01:43 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended.
According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit op
tion isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'f
alse'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for s
erver certificate verification.
OK
Time taken: 2.133 seconds

```

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','uk'),(6,'JNTUA','atp'),(7,
'cambridge','us');

```

acadgild@localhost:~
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');
Sun Feb 18 23:04:23 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+
requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertif
icate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate ve
rification.
WARNING: Hive-on-HR 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_20180218230422_1f989999-e3ac-4b0e-80eb-ffa7ae61f88
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>
Sun Feb 18 23:04:24 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+
requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertif
icate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate ve
rification.
Starting Job = job_1518963991621_0002, Tracking URL = http://localhost:8088/proxy/application_1518963991621_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1518963991621_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-02-18 23:04:49,636 Stage-1 map = 0%, reduce = 0%
2018-02-18 23:05:05,582 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.64 sec
2018-02-18 23:05:49,646 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 5.85 sec
2018-02-18 23:05:54,750 Stage-1 map = 100%, reduce = 27%, Cumulative CPU 7.9 sec
2018-02-18 23:05:58,544 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 9.93 sec
2018-02-18 23:05:59,945 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 11.97 sec
2018-02-18 23:06:01,193 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 14.03 sec
2018-02-18 23:06:07,578 Stage-1 map = 100%, reduce = 73%, Cumulative CPU 17.24 sec
2018-02-18 23:06:11,230 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 20.38 sec
2018-02-18 23:06:13,642 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 25.65 sec
2018-02-18 23:06:14,718 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 28.92 sec
MapReduce Total cumulative CPU time: 28 seconds 920 msec
Ended Job = job_1518963991621_0002
Loading data to table company.college
MapReduce Jobs Launched:
Stage-1: Map: 1 Reduce: 5 Cumulative CPU: 28.92 sec HDFS Read: 27087 HDFS Write: 4001 SUCCESS
Total MapReduce CPU Time Spent: 28 seconds 920 msec
OK
Time taken: 114.287 seconds
hive>

```

View the data:

select * from college

```

acadgild@localhost:~
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: 0.396 seconds, Fetched: 7 row(s)
hive> █

```

Updating the Data in Hive Table

UPDATE college set clg_id = 8 where clg_id = 7;

```

acadgild@localhost:~
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;

```

acadgild@localhost:~
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_20180218230851_a527ed27-74da-462c-9826-a0224f201ab4
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_1518963991621_0003, Tracking URL = http://localhost:8088/proxy/application_1518963991621_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1518963991621_0003
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-02-18 23:09:15,107 Stage-1 map = 0%, reduce = 0%
2018-02-18 23:10:15,220 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 2.79 sec
2018-02-18 23:10:28,635 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 20.61 sec
2018-02-18 23:10:29,767 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 22.91 sec
2018-02-18 23:10:30,854 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 24.03 sec
2018-02-18 23:10:31,996 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 25.23 sec
2018-02-18 23:11:26,097 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 35.56 sec
2018-02-18 23:11:34,829 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 42.79 sec
2018-02-18 23:11:35,886 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 43.85 sec
MapReduce Total cumulative CPU time: 43 seconds 850 msec
Ended Job = job_1518963991621_0003
Loading data to table company.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 43.85 sec HDFS Read: 52089 HDFS Write: 937 SUCCESS
Total MapReduce CPU Time Spent: 43 seconds 850 msec
OK
Time taken: 168.122 seconds
hive>

```

View the data:

Select *from college;

```

acadgild@localhost:~
hive> select * from college;
OK
5      stanford      uk
6      IIT           atp
1      nec           nlr
7      cambridge     us
2      vit           vlr
3      sim           chen
4      lpu           del
Time taken: 0.568 seconds, Fetched: 7 row(s)
hive>

```

Deleting a Row from Hive Table

Delete from college where clg_id=5;

```

acsgild@localhost:~
hive> delete from college where c1g_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 = acsgild_20180218231739_66c6aalf-540f-48f5-8a84-0d80768dd68c
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_1518963991621_0004, Tracking URL = http://localhost:8088/proxy/application/1518963991621_0004/
Kill Command = /home/acsgild/install/hadoop/hadoop-2.9.0/bin/hadoop job -kill job_1518963991621_0004
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-02-18 23:17:59,758 Stage-1 map = 0%, reduce = 0%
2018-02-18 23:18:59,990 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 7.7 sec
2018-02-18 23:19:05,117 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 12.69 sec
2018-02-18 23:19:06,414 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 16.47 sec
2018-02-18 23:19:09,032 Stage-1 map = 80%, reduce = 0%, Cumulative CPU 20.63 sec
2018-02-18 23:19:10,352 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 21.11 sec
2018-02-18 23:20:01,710 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 23.38 sec
2018-02-18 23:20:04,262 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 29.55 sec
2018-02-18 23:20:09,209 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 31.73 sec
2018-02-18 23:20:14,246 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 38.86 sec
2018-02-18 23:20:15,440 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 40.64 sec
MapReduce Total cumulative CPU time: 40 seconds 640 msec
Ended Job = job_1518963991621_0004
Loading data to table company.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 40.64 sec HDFS Read: 50313 HDFS Write: 752 SUCCESS
Total MapReduce CPU Time Spent: 40 seconds 640 msec
OK
Time taken: 159.596 seconds
hive>

```

View the data:

Select * from college;

```

acsgild@localhost:~
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: 0.386 seconds, Fetched: 6 row(s)
hive>

```