

DATE SET DESCRIPTION

The data set consists of the following fields.

Athlete: This field consists of the athlete name

Age: This field consists of athlete ages

Country: This fields consists of the country names which participated in Olympics

Year: This field consists of the year

Closing Date: This field consists of the closing date of ceremony

Sport: Consists of the sports name

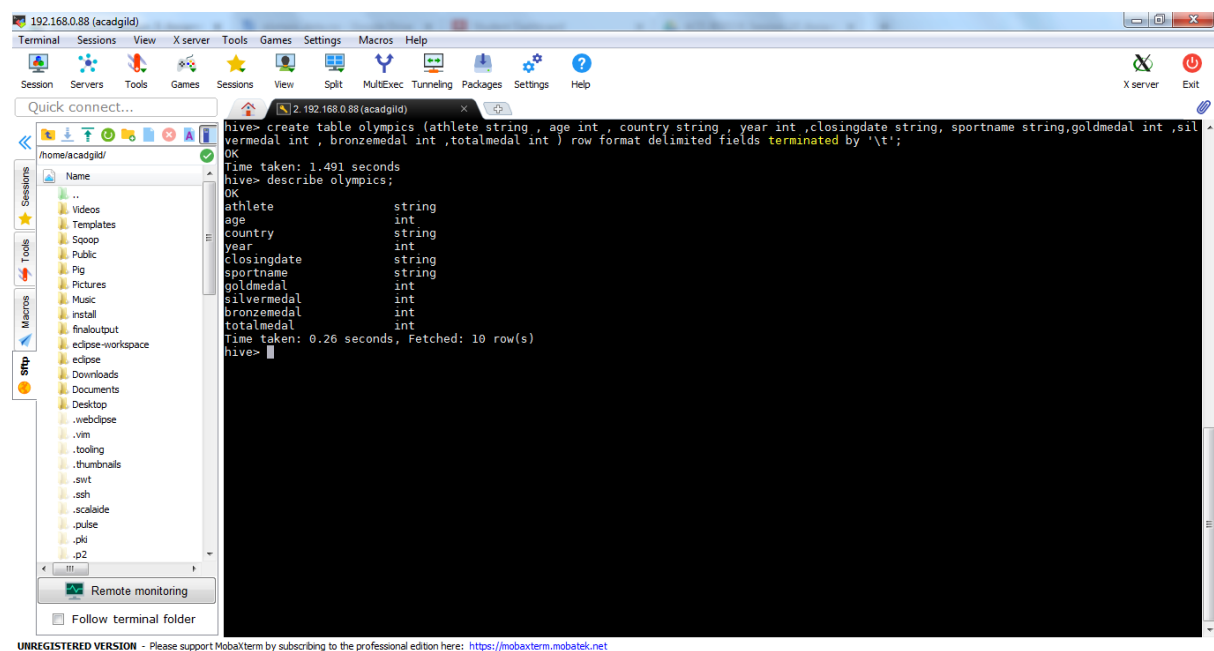
Gold Medals: No. of Gold medals

Silver Medals: No. of Silver medals

Bronze Medals: No. of Bronze medals

Total Medals: Consists of total no. of medals

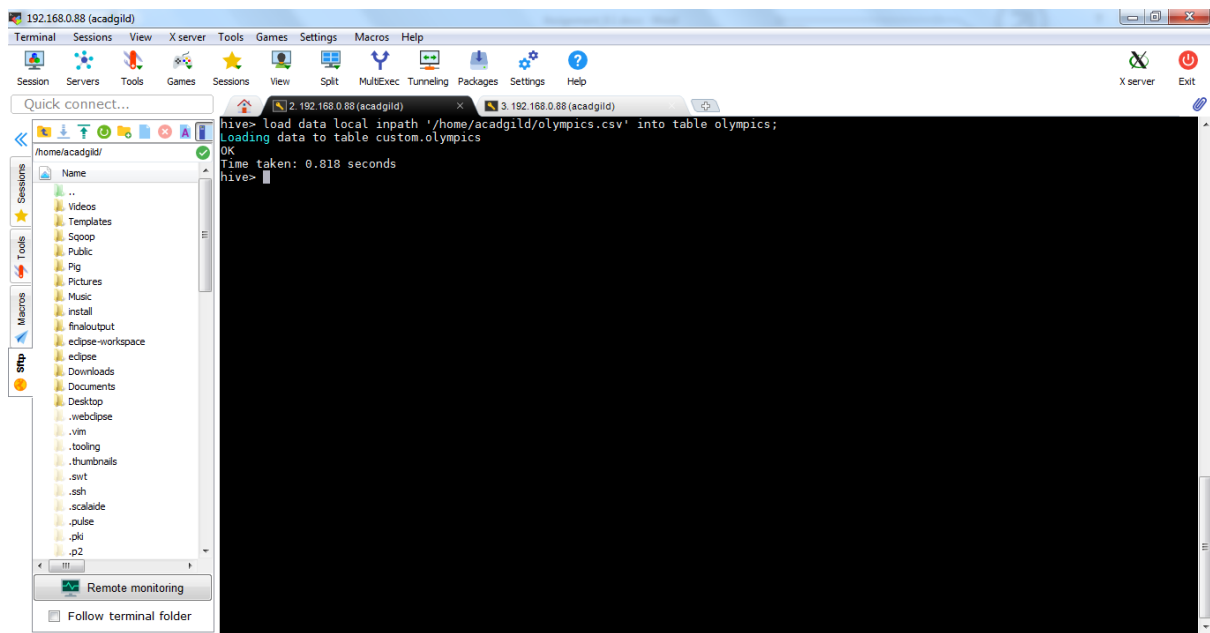
Creating and Loading data into table :



The screenshot shows a MobaXterm terminal window with a menu bar (Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, Help) and a toolbar. The terminal is connected to a remote host 192.168.0.88 (acadgild). The left sidebar shows a file explorer for the /home/acadgild/ directory. The terminal output shows the following commands and results:

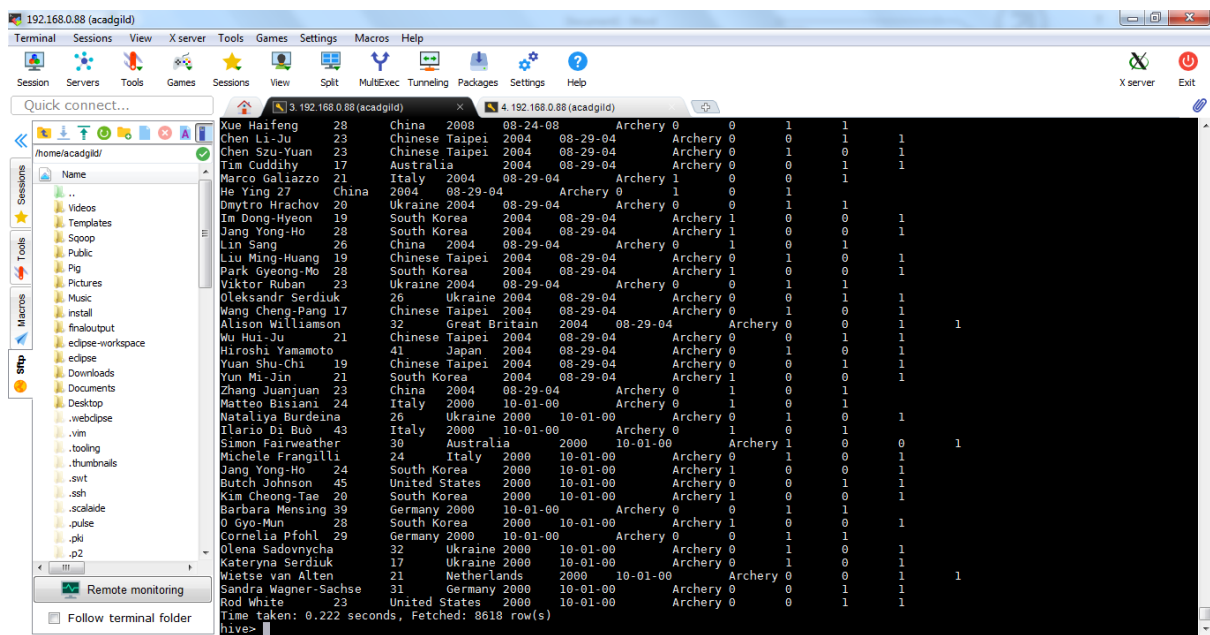
```
hive> create table olympics (athlete string , age int , country string , year int ,closingdate string, sportname string,goldmedal int ,silvermedal int , bronzemedal int ,totalmedal int ) row format delimited fields terminated by '\t';
OK
Time taken: 1.491 seconds
hive> describe olympics;
OK
athlete                string
age                     int
country                 string
year                    int
closingdate             string
sportname               string
goldmedal               int
silvermedal             int
bronzemedal             int
totalmedal              int
Time taken: 0.26 seconds, Fetched: 10 row(s)
hive>
```

At the bottom of the window, there is a footer: "UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>"



UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Table Content:-



UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

5. Problem Statement

Task 1

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

Output:-

```

Stage-Stage-2: HDFS Read: 1037338 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Zimbabwe 2
United States 145
Ukraine 4
Tunisia 2
Trinidad and Tobago 1
Sweden 7
Spain 2
South Korea 2
South Africa 8
Slovenia 1
Slovakia 1
Serbia 1
Russia 19
Romania 4
Poland 1
Norway 2
Netherlands 32
Lithuania 1
Japan 38
Italy 13
Hungary 7
Great Britain 9
Germany 27
France 26
Denmark 1
Croatia 1
Costa Rica 1
China 29
Canada 5
Brazil 7
Belarus 1
Austria 2
Australia 92
Argentina 1
Time taken: 9.834 seconds, Fetched: 34 row(s)
hive>

```

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

Output:-

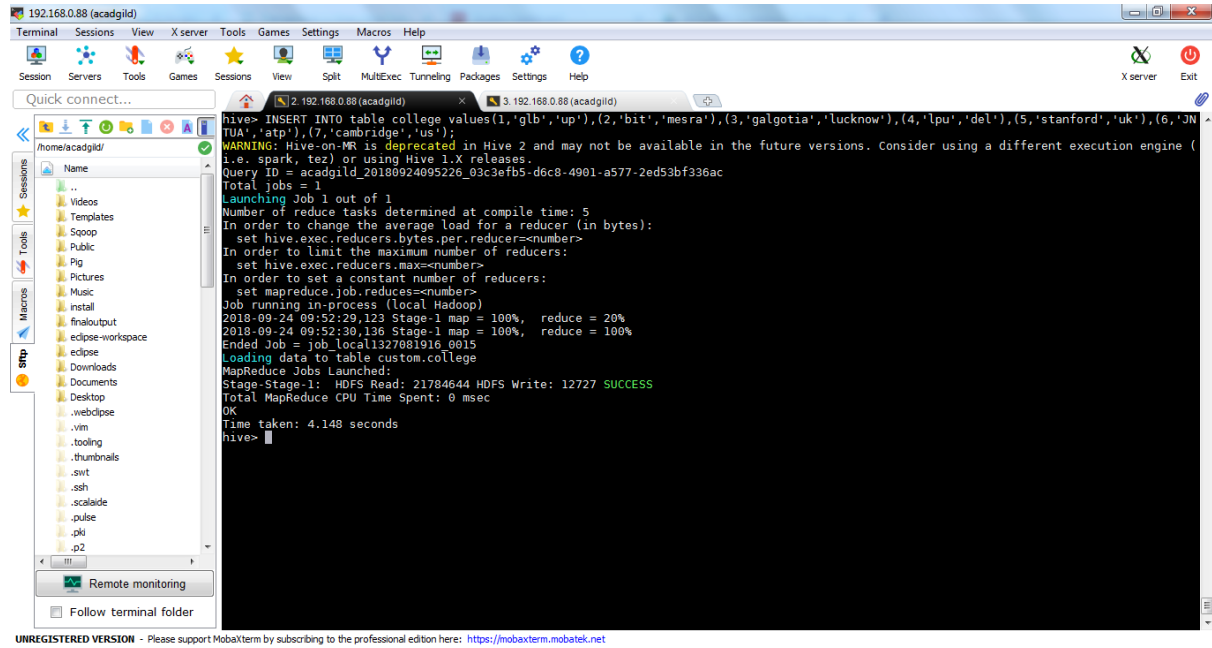
```

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: 1.107 seconds
hive>

```

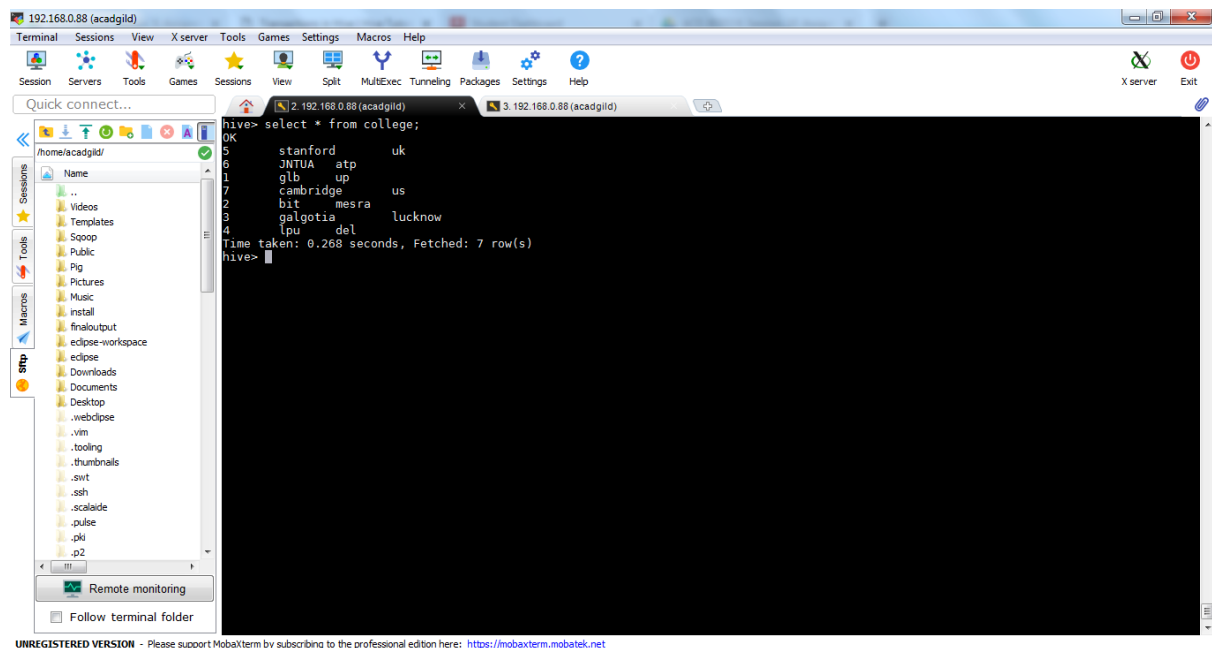
3. Write a Hive Program to find the total number of medals each country won.

Output:-



```
hive> INSERT INTO table college values(1,'glb','up'),(2,'bit','mesra'),(3,'galgotia','lucknow'),(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 execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180924095226_03c3efb5-d6c8-4901-a577-2ed53bf336ac
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>
Job running in-process (local Hadoop)
2018-09-24 09:52:29,123 Stage-1 map = 100%, reduce = 20%
2018-09-24 09:52:30,136 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local1327081916_0015
Loading data to table custom.college
MapReduce Jobs Launched:
Stage:Stage-1:  HDFS Read: 21784644 HDFS Write: 12727 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 4.148 seconds
hive>
```

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



```
hive> select * from college;
OK
5      stanford      uk
6      JNTUA      atp
7      glb      up
7      cambridge      us
2      bit      mesra
3      galgotia      lucknow
4      lpu      del
Time taken: 0.268 seconds, Fetched: 7 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.

Command-:

create table string_data (name array<string>) row format delimited fields terminated by ',' collection items terminated by ' ';

→ Create a table with one field as array

LOAD DATA LOCAL INPATH '/home/acadgild/words.txt' into table string_data;

→ Load data from local path

add jar /home/acadgild/StringConcatenate.jar;

➔ Add jar to hive run time

create temporary function stringconcatenate as
"com.hem.hadoop.hive.Assignment9.StringConcatenate";

→ create temporary function to execute command

select stringconcatenate(',',name) from string_data;

->> Querying in hive language using temporary function created above with the use of table string data

Content of words.txt

hem singh bist is a wonderful guy

StringConcatenate Program-:

```
package com.hem.hadoop.hive.Assignment9;
```

```
import java.util.List;
```

```
import org.apache.hadoop.hive.ql.exec.UDF;
```

```
import org.apache.hadoop.io.Text;
```

```
public class StringConcatenate extends UDF{
```

```
    public Text evaluate(final Text delimiter, final List<Text> param2) {
```

```
        if (param2.size() == 0) {
```

```
            return null;
```

```
        }
```

```
        if (param2.size() == 1) {
```

```
            return param2.get(0);
```

```
        } else {
```

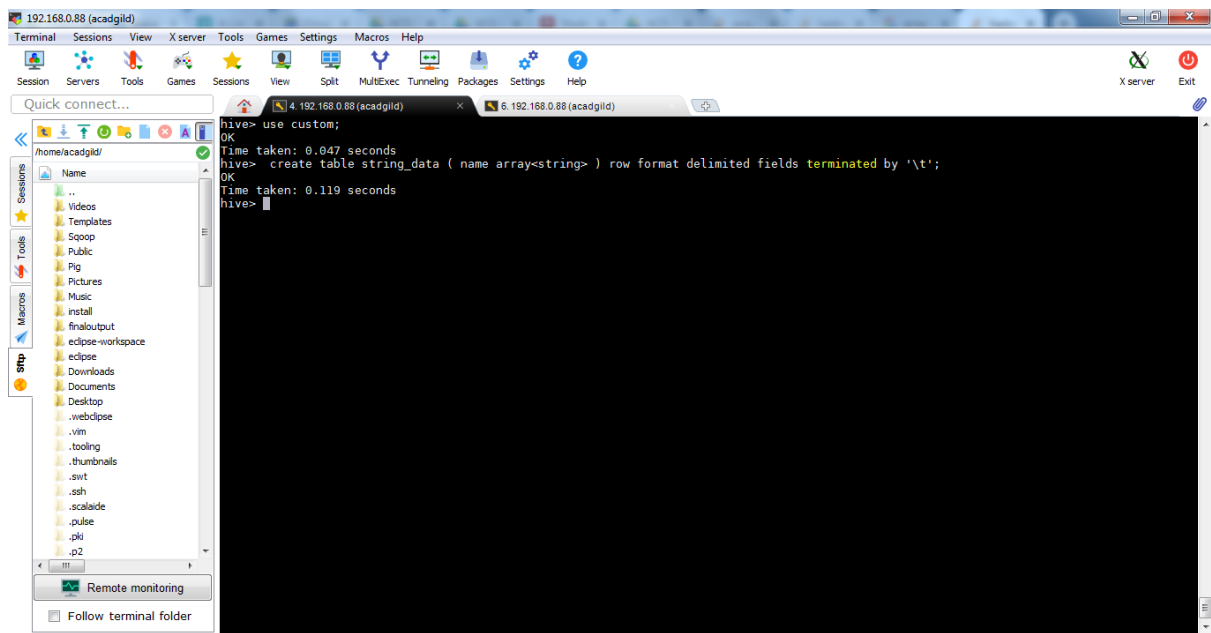
```
            String citiesCommaSeparated = String.join(",", param2.toString());
```

```
            return new Text(citiesCommaSeparated);
```

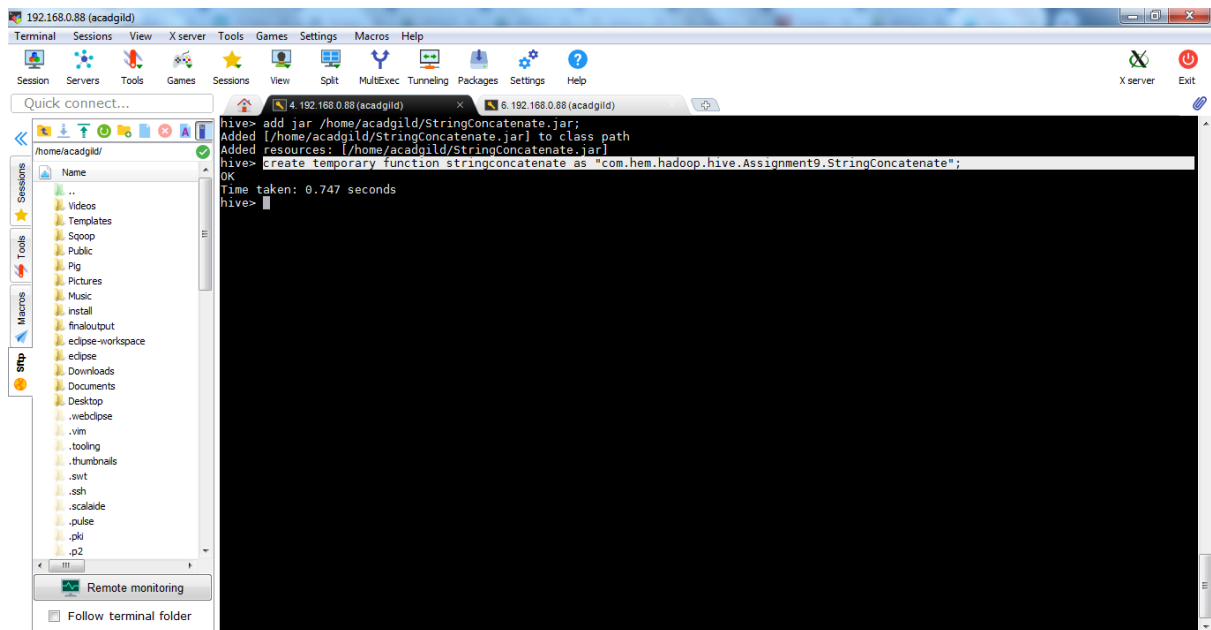
```
        }
```

```
    }
```

```
}
```



UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>



UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

The screenshot shows a MobaXterm window with a terminal session on a remote host 192.168.0.88 (acadgild). The terminal displays the following commands and output:

```
hive> create table string_data (name array<string>) row format delimited fields terminated by ',' collection items terminated by ' ';
OK
Time taken: 0.115 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/words.txt' into table string_data;
Loading data to table custom.string_data
OK
Time taken: 0.556 seconds
hive> select * from string_data;
OK
["hem","singh","bist","is","a","wonderful","guy"]
Time taken: 0.183 seconds, Fetched: 1 row(s)
hive>
```

The left sidebar shows a file explorer view of the local machine, listing various folders like Videos, Templates, Screenshot, Public, Pictures, Music, Install, finaloutput, eclipse-workspace, eclipse, Downloads, Documents, Desktop, .webclipse, .vim, .tooling, .thumbnails, .swt, .ssh, .scalaide, .pulse, .pki, and .p2.

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

The screenshot shows a MobaXterm window with a terminal session on a remote host 192.168.0.88 (acadgild). The terminal displays the following commands and output:

```
hive> add jar /home/acadgild/StringConcatenate.jar;
Added
Added /home/acadgild/StringConcatenate.jar to class path
hive> create temporary function stringconcatenate as "com.hem.hadoop.hive.Assignment9.StringConcatenate";
OK
Time taken: 0.009 seconds
hive> select stringconcatenate(',',name) from string_data;
OK
["hem,singh,bist,is,a,wonderful,guy"]
Time taken: 0.229 seconds, Fetched: 1 row(s)
hive>
```

The left sidebar shows a file explorer view of the local machine, listing various folders like Videos, Templates, Screenshot, Public, Pictures, Music, Install, finaloutput, eclipse-workspace, eclipse, Downloads, Documents, Desktop, .webclipse, .vim, .tooling, .thumbnails, .swt, .ssh, .scalaide, .pulse, .pki, and .p2.

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

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.

Command-:

```
hive>set hive.support.concurrency = true;
```

```
hive>set hive.enforce.bucketing = true;
```

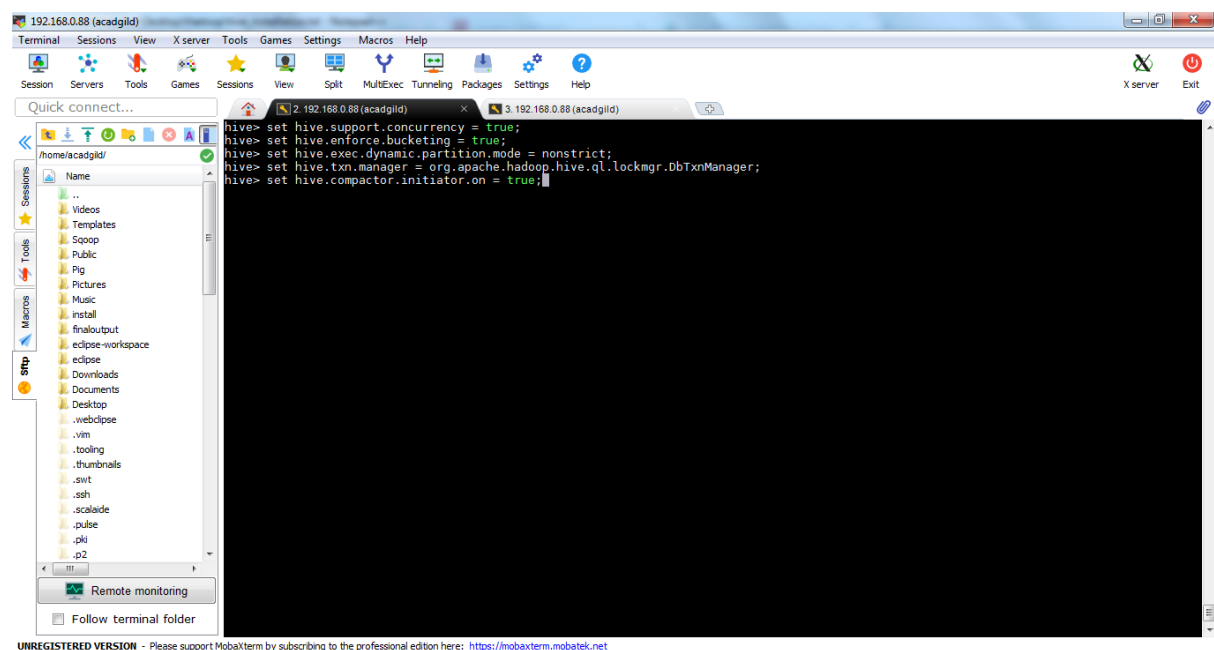
```
hive>set hive.exec.dynamic.partition.mode = nonstrict;
```

```
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;
```

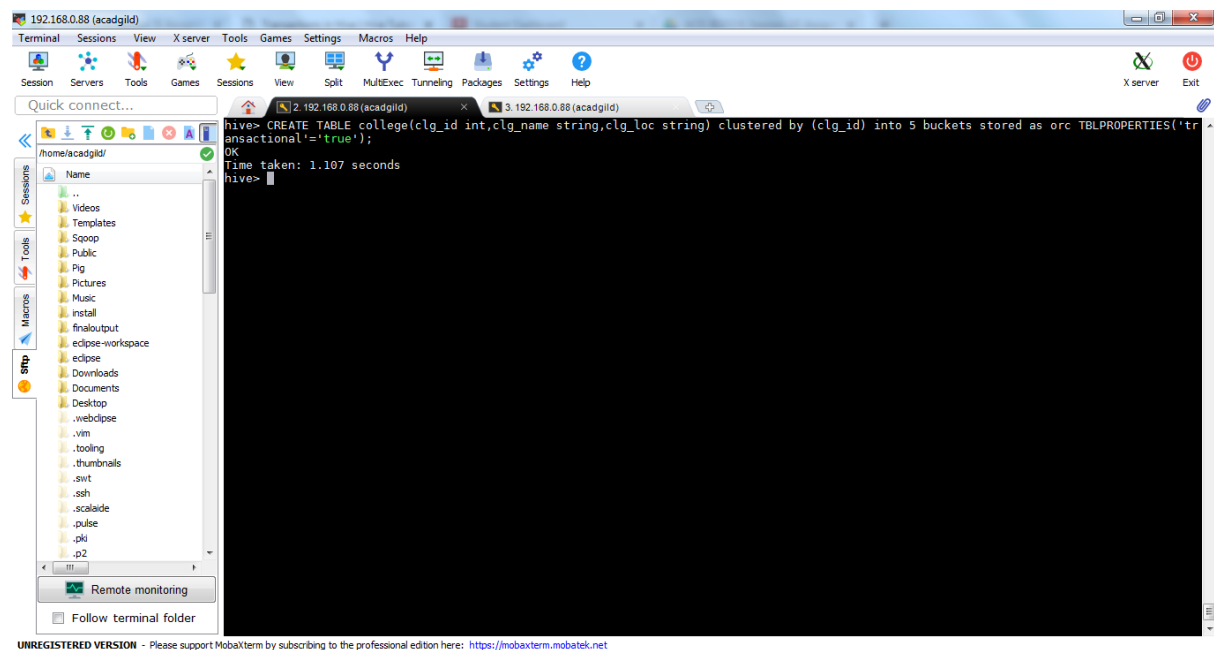
Output-:

A screenshot of a MobaXterm terminal window. The window title is '192.168.0.88 (acadgild)'. The terminal displays five Hive configuration commands entered in sequence, each on a new line. The commands are: '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;', and 'hive> set hive.compactor.initiator.on = true;'. The terminal interface includes a menu bar (Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, Help), a toolbar, and a sidebar on the left with 'Quick connect...' and 'Sessions' sections. The 'Sessions' list shows various folders like Videos, Templates, Sftp, Public, Pig, Pictures, Music, Install, finaloutput, eclipse-workspace, eclipse, Downloads, Documents, Desktop, .webclipse, .vim, .tooling, .thumbnails, .swt, .ssh, .scalaide, .pulse, .pli, and .p2. At the bottom of the window, there is a notice: 'UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net'.

Command-:

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');

Output-:



The screenshot shows a MobaXterm terminal window with a file explorer on the left. The terminal displays the following command and output:

```
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: 1.107 seconds
hive>
```

At the bottom of the window, there is a message: "UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>"

Command-:

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');

Output-:

The screenshot shows a MobaXterm window with a terminal session on a remote host (192.168.0.88). The terminal displays the execution of a Hive SQL command to insert data into a table named 'college'. The output shows the job running successfully on a local Hadoop cluster, with a total execution time of 4.148 seconds.

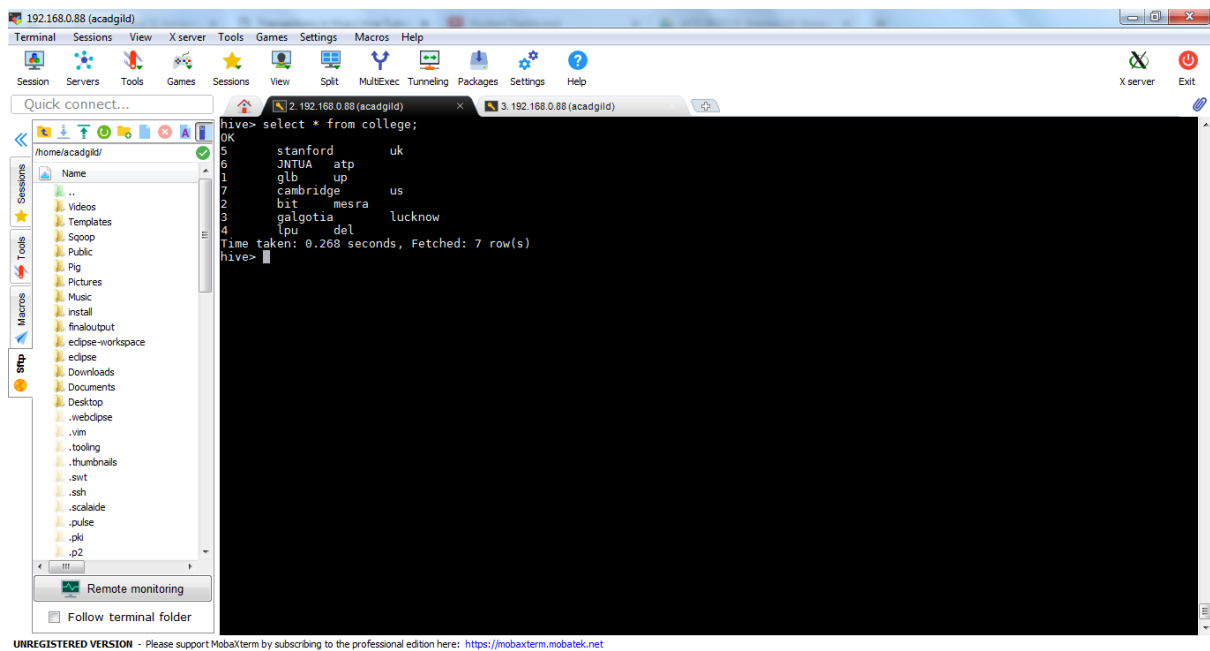
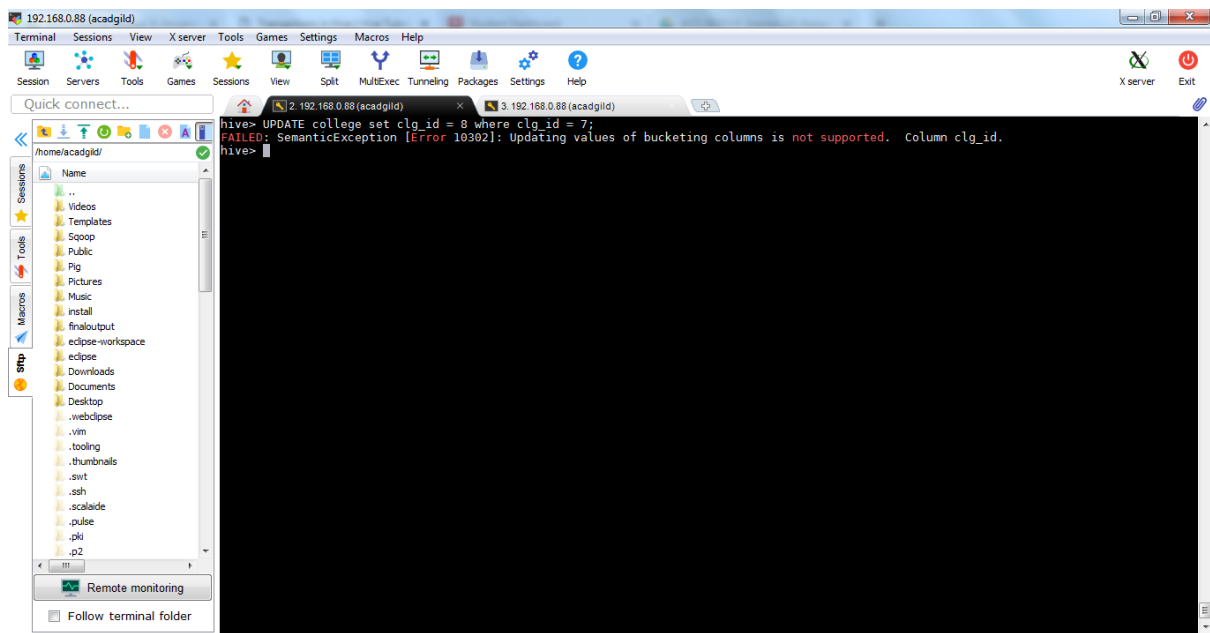
```
hive> INSERT INTO table college values(1,'glb','up'),(2,'bit','mesra'),(3,'galgotia','lucknow'),(4,'lpu','del'),(5,'stanford','uk'),(6,'JNTUA','atp'),(7,'cambridge','us');
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180924095226_03c3efb5-d6c8-4901-a577-2ed53bf336ac
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.reducers=<number>
Job running in-process (local Hadoop)
2018-09-24 09:52:29,123 Stage-1 map = 100%,  reduce = 20%
2018-09-24 09:52:30,136 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1327081916_0015
Loading data to table custom.college
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 21784644 HDFS Write: 12727 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 4.148 seconds
hive>
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Command:-

UPDATE college set clg_id = 8 where clg_id = 7;

Output:-



Command-:

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

Output-:

192.168.0.88 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MuExec Tunneling Packages Settings Help

Quick connect...

2. 192.168.0.88 (acadgild)

3. 192.168.0.88 (acadgild)

6 IIT atp
1 glb up
7 cambridge us
2 bit mesra
3 galgotia lucknow
4 lpu del
Time taken: 0.279 seconds, Fetched: 7 row(s)
hive> 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_20180924095651_eeb93d74-c9ef-473d-8165-d62abbc6604f
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.reducers=<number>
Job running in-process (local Hadoop)
2018-09-24 09:56:52,958 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local116854109_0017
Loading data to table custom.college
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 36494248 HDFS Write: 53955 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 2.565 seconds
hive> select * from college;
OK
6 IIT atp
1 glb up
7 cambridge us
2 bit mesra
3 galgotia lucknow
4 lpu del
Time taken: 0.262 seconds, Fetched: 6 row(s)
hive>

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>