

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

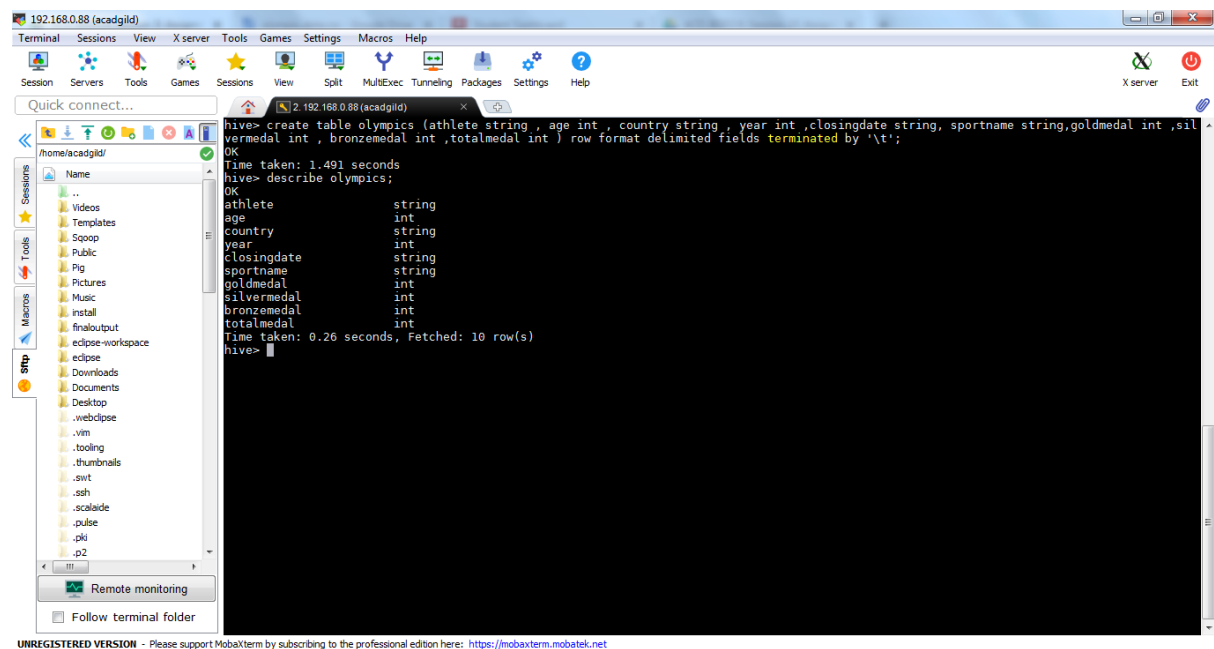
Gold Medals: No. of Gold medals

Silver Medals: No. of Silver medals

Bronze Medals: No. of Bronze medals

Total Medals: Consists of total no. of medals

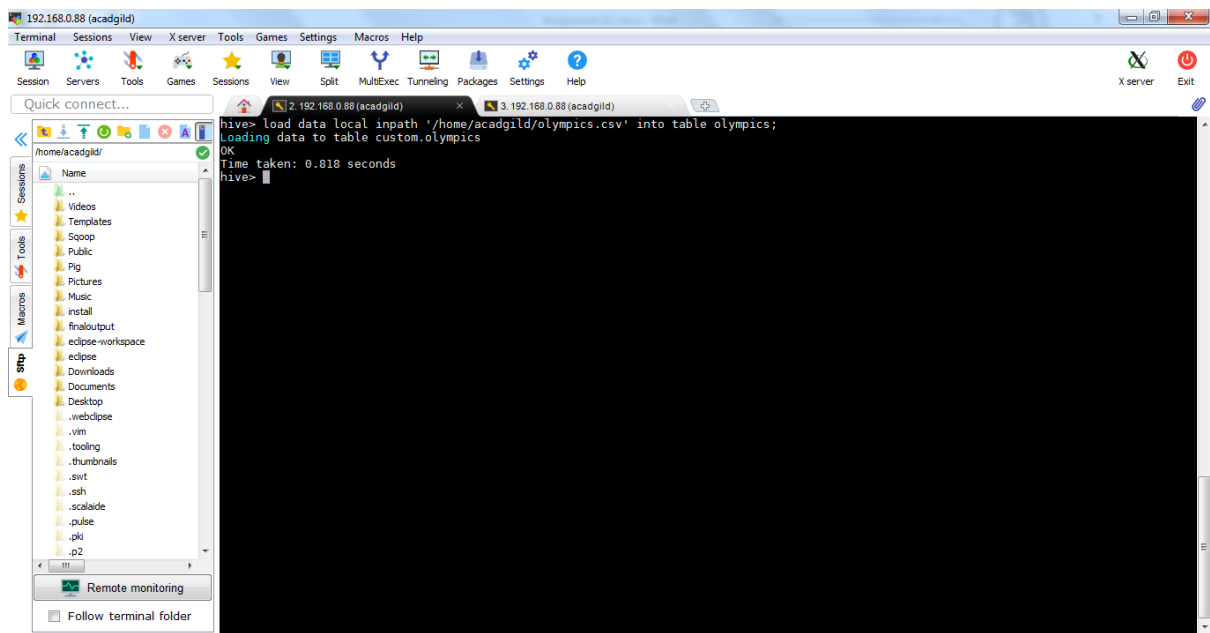
Creating and Loading data into table :



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

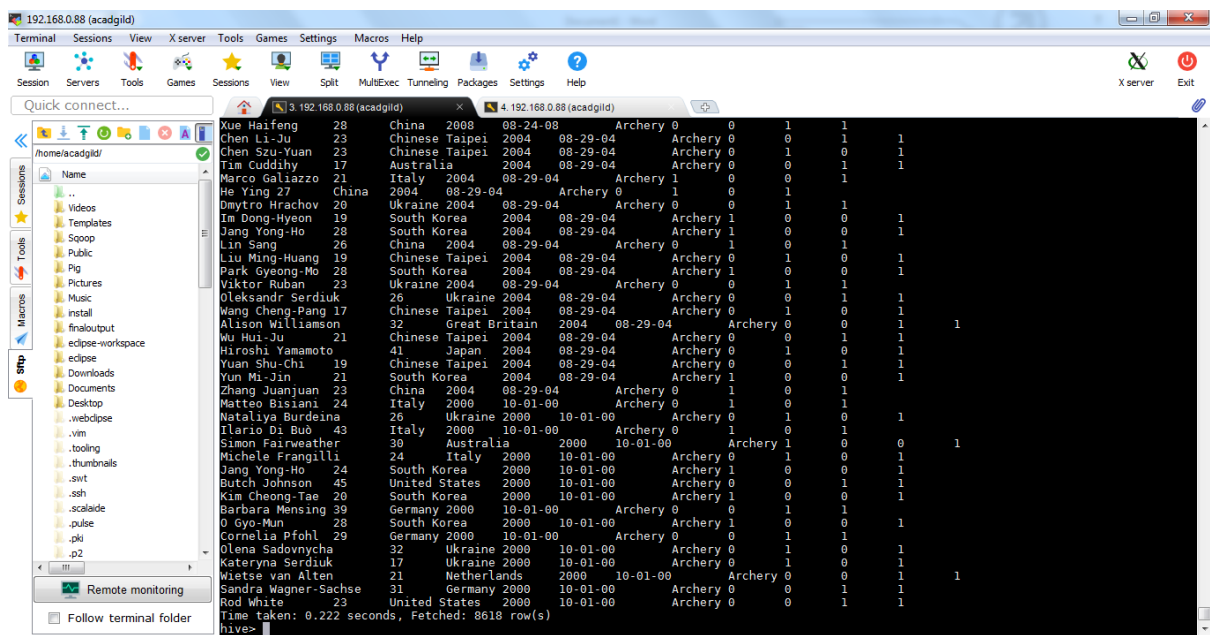
```
hive> create table olympics (athlete string , age int , country string , year int ,closingdate string, sportname string,goldmedal int ,silvermedal int , bronzemedal int ,totalmedal int ) row format delimited fields terminated by '\t';
OK
Time taken: 1.491 seconds
hive> describe olympics;
OK
athlete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



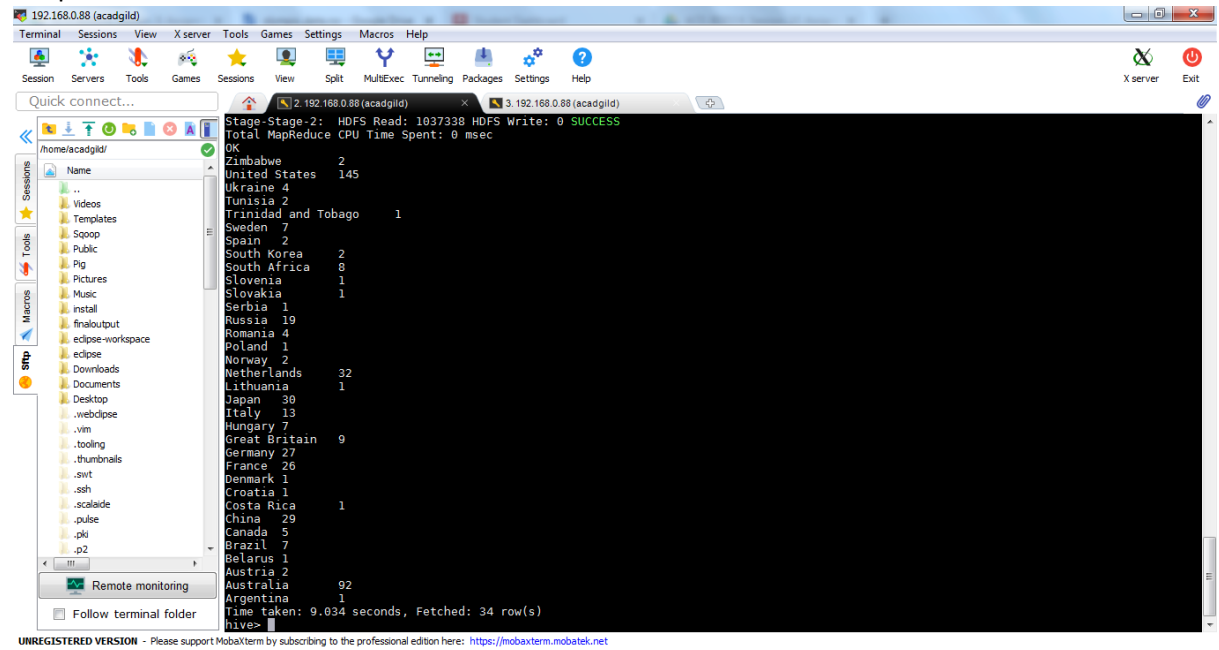
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

5. Problem Statement

Task 1

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

Output:-



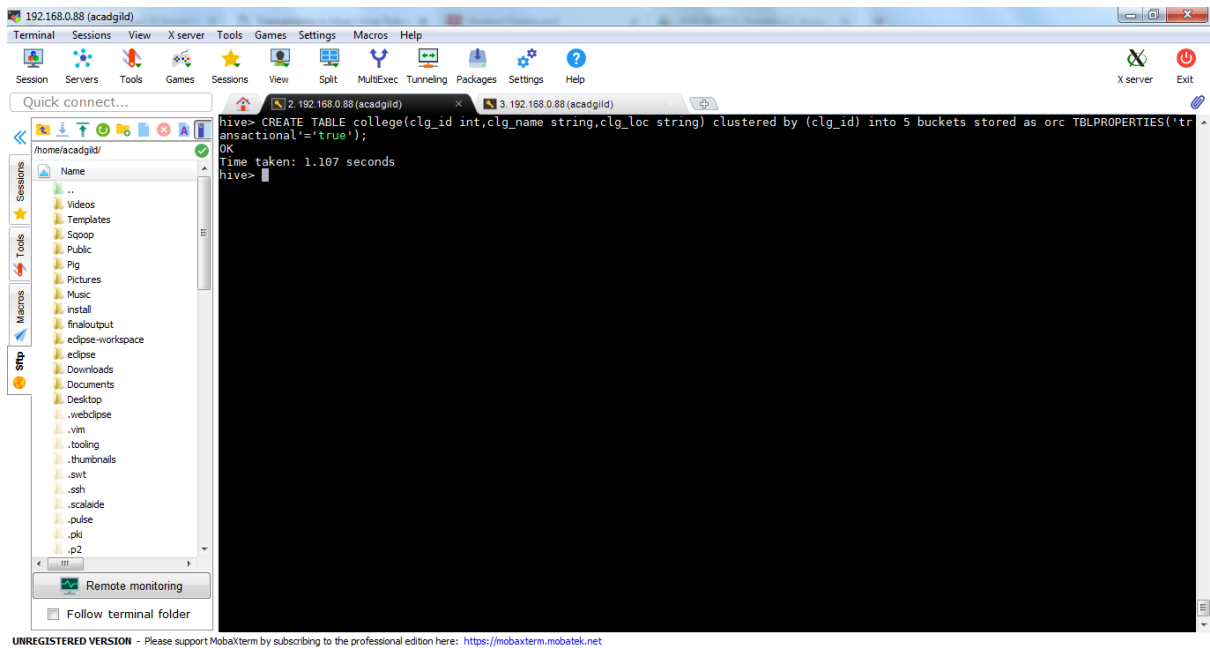
```

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: 9.034 seconds, Fetched: 34 row(s)
hive>

```

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

Output:-



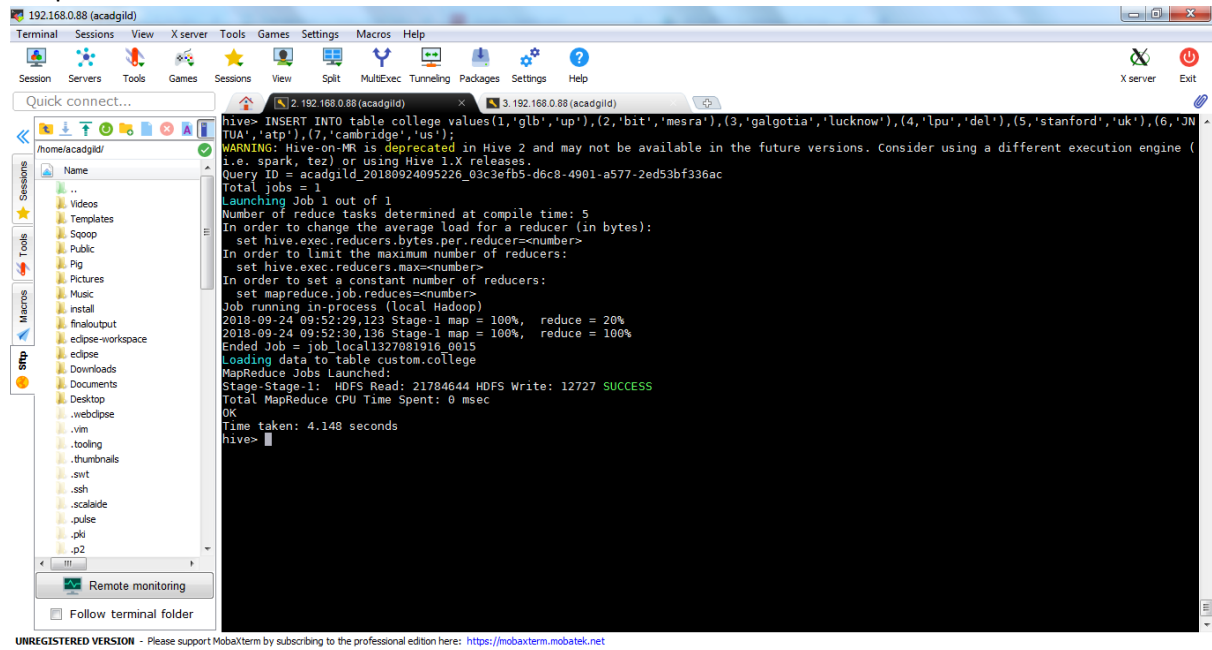
```

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.

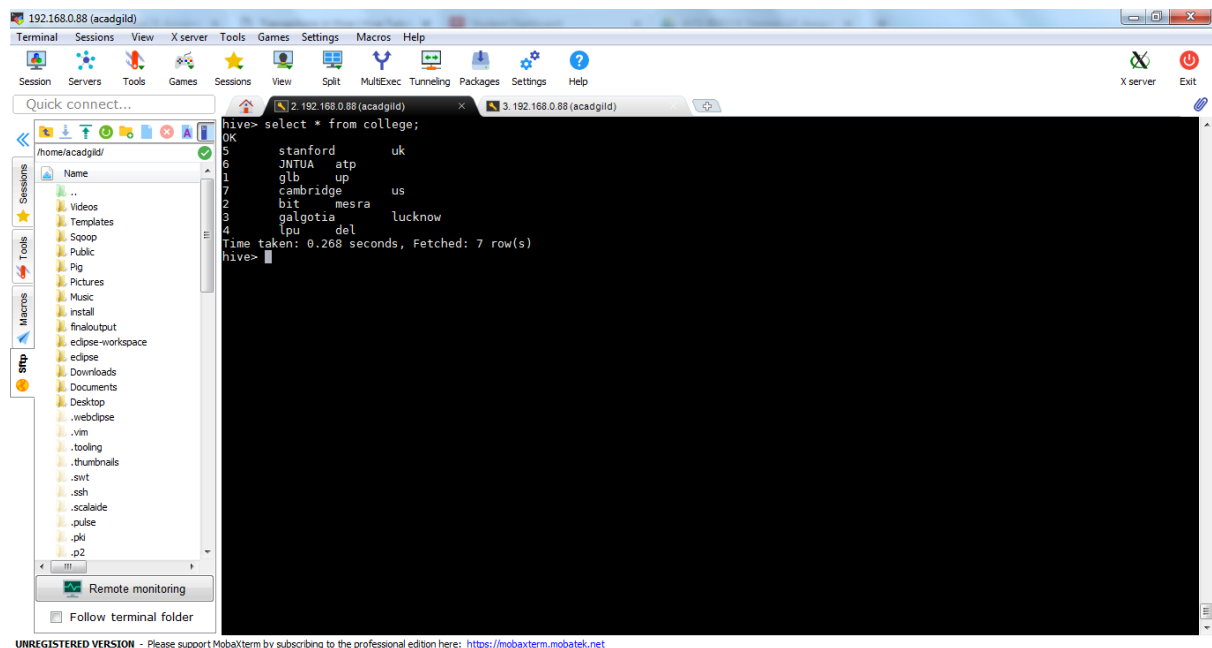
Output:-



The screenshot shows a MobaXterm terminal window with a file explorer on the left. The terminal displays the output of a Hive MR job. The command executed is `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');`. The output includes a warning about Hive-on-MR being deprecated, query ID, total jobs (1), and job launch details. The job completed successfully with a time taken 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.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.



The screenshot shows a MobaXterm terminal window with a file explorer on the left. The terminal displays the output of a Hive query. The command executed is `hive> select * from college;`. The output shows 7 rows of data with columns for country, institution, and city. The time taken for the query is 0.268 seconds.

```
hive> select * from college;
OK
6
stanford      uk
JNTUA         atp
glb           up
cambridge     us
bit           mesra
galgotia      lucknow
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 '\t';

→ 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";

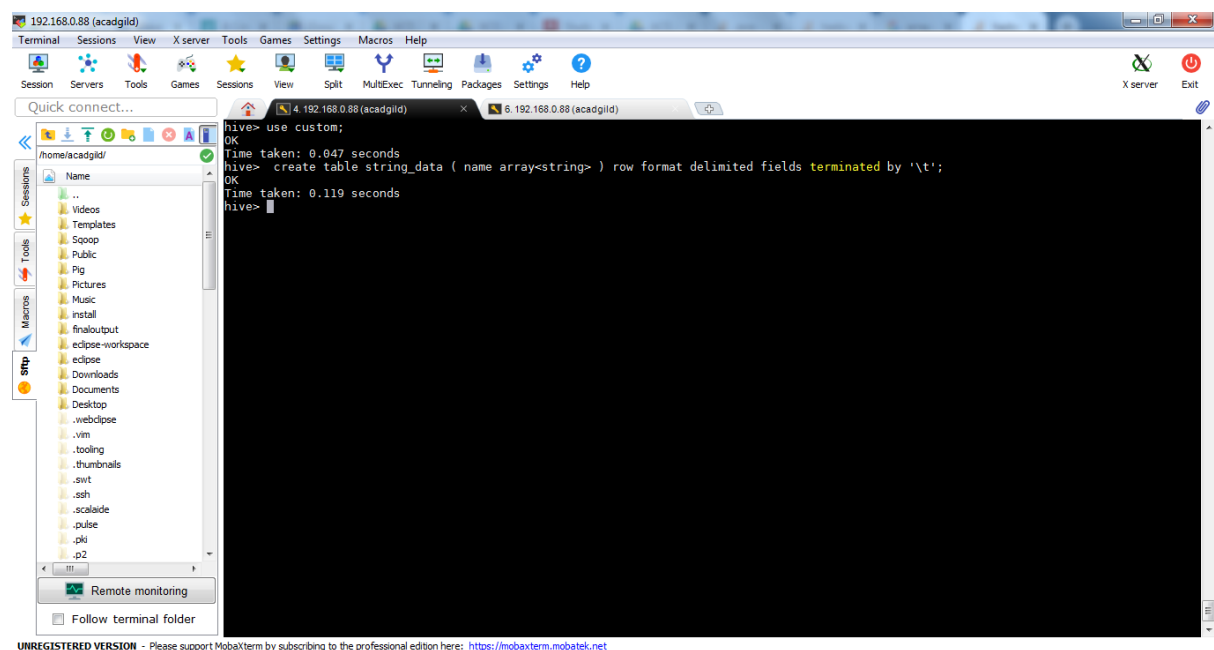
→ create temporary function to execute command

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

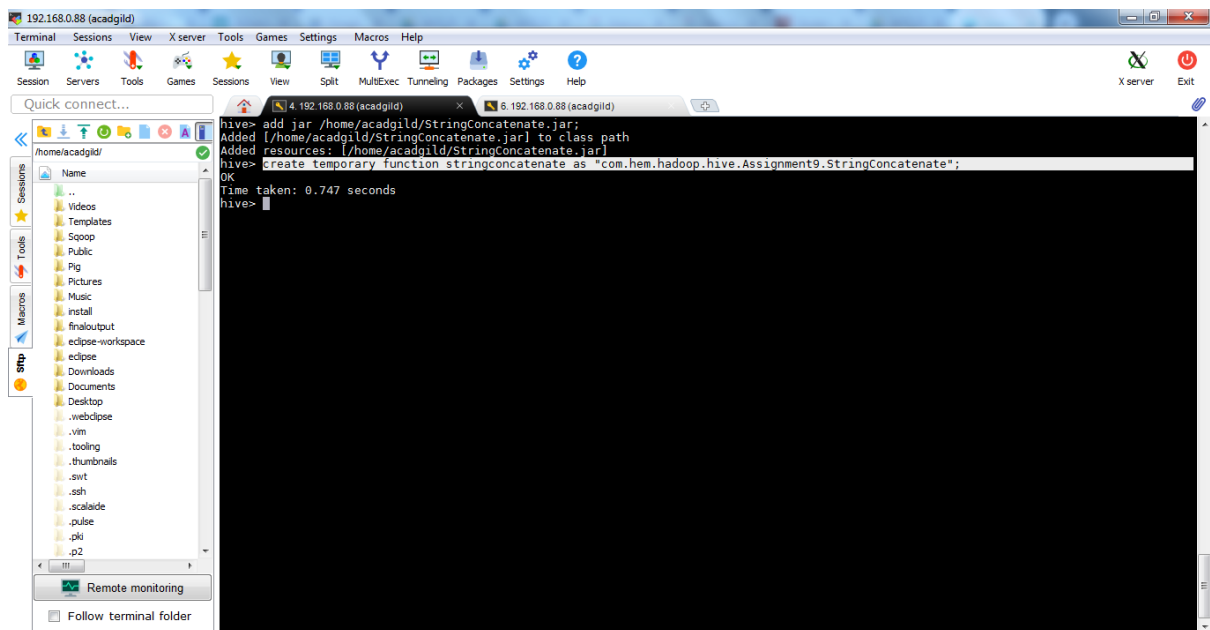


The screenshot shows a MobaXterm terminal window with a file explorer on the left. The terminal displays the following commands and output:

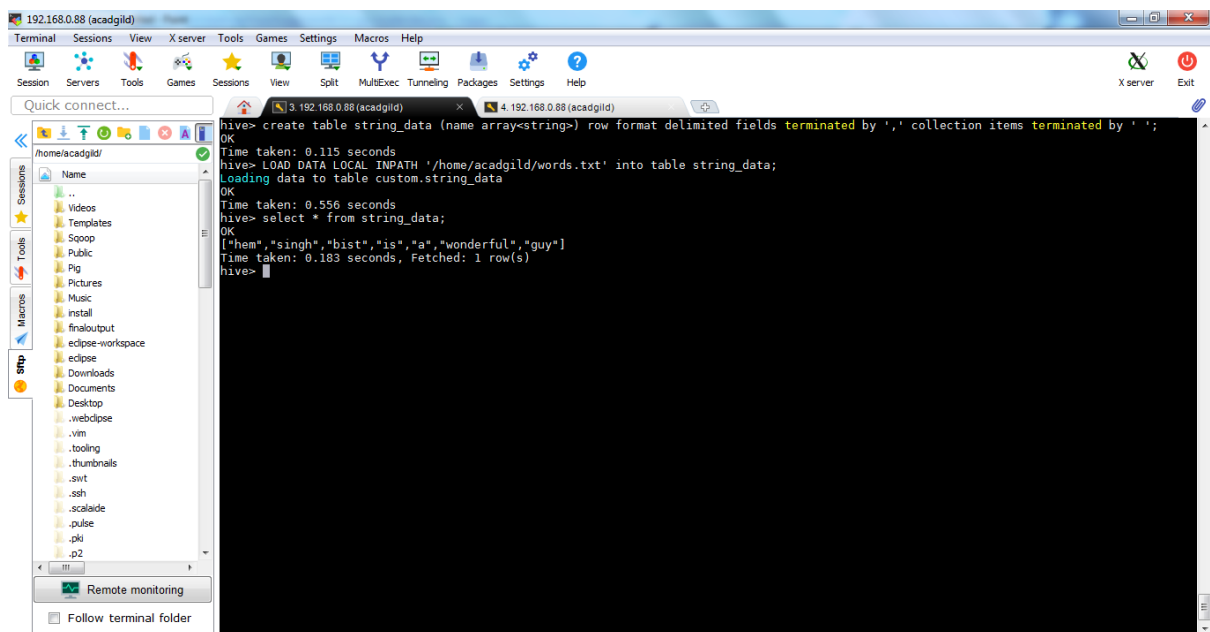
```
hive> use custom;
OK
Time taken: 0.047 seconds
hive> create table string_data ( name array<string> ) row format delimited fields terminated by '\t';
OK
Time taken: 0.119 seconds
hive>
```

The file explorer on the left shows the directory structure of the user's home folder, including subdirectories like Videos, Templates, Sqoop, Public, Pig, Pictures, Music, Install, finaloutput, eclipse-workspace, eclipse, Downloads, Documents, Desktop, .webclipse, .vim, .tooling, .thumbnails, .swt, .ssh, .scalaide, .pulse, .plk, and .p2.

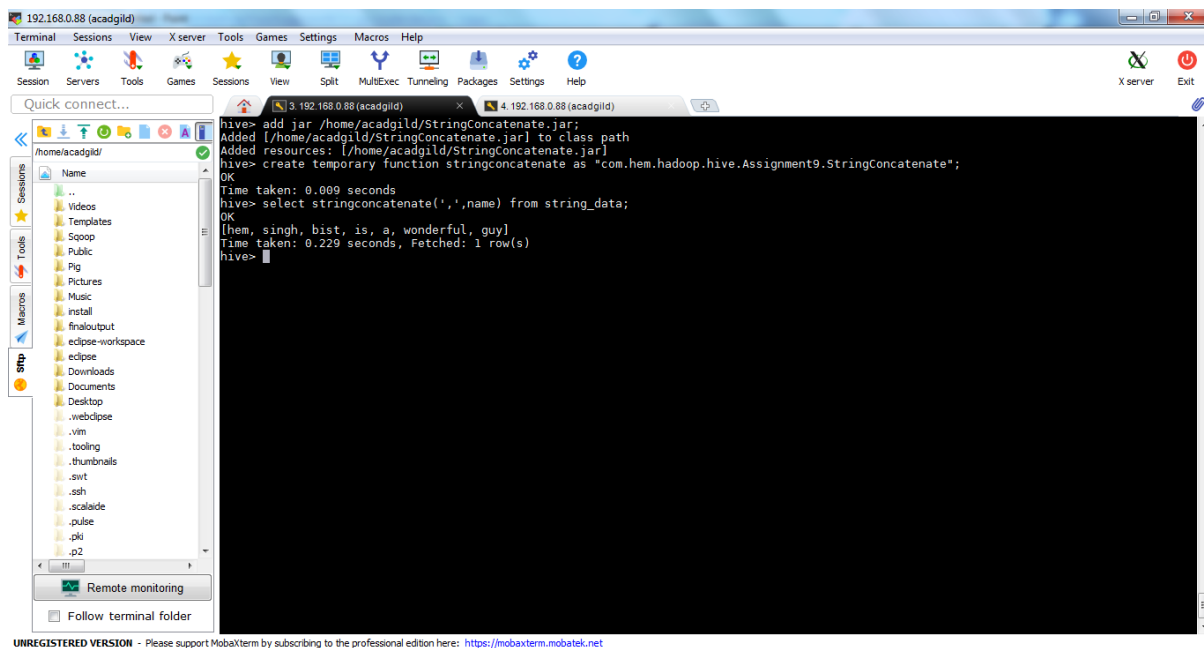
At the bottom of the terminal window, there is a message: "UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>"



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



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;
```

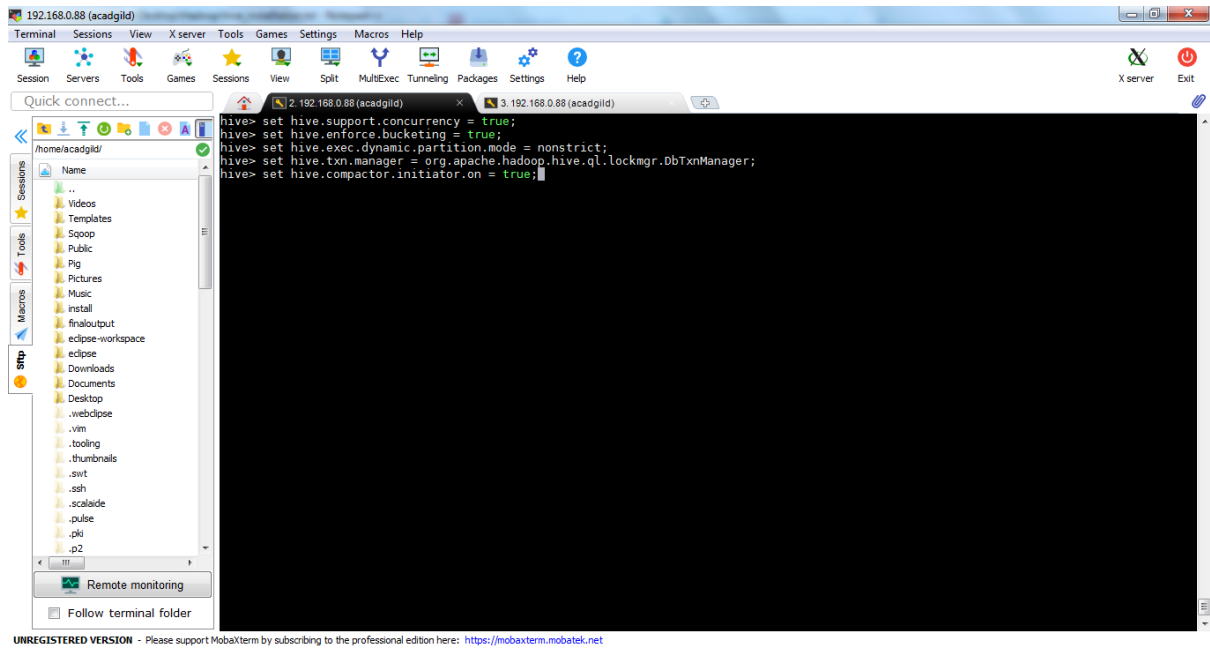
```
hive>set hive.exec.dynamic.partition.mode = nonstrict;
```

```
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:-



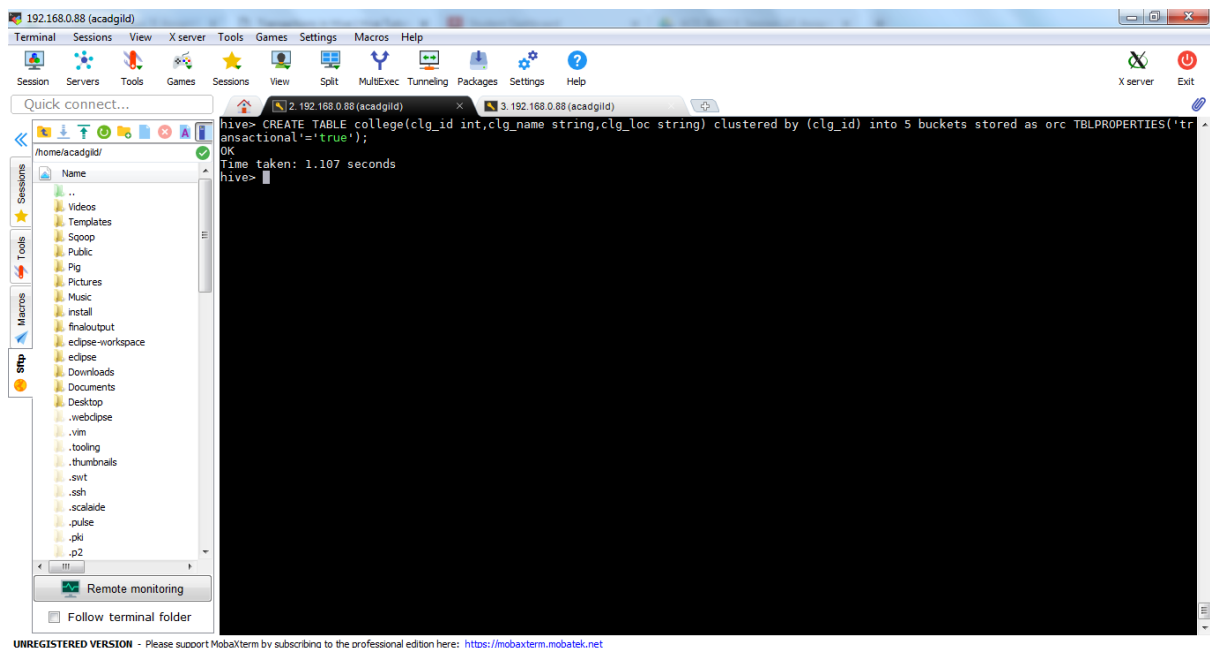
A screenshot of the MobaXterm application window. The title bar shows the IP address 192.168.0.88 and the session name (acadgild). The interface includes a menu bar (Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, Help), a toolbar with icons for Session, Servers, Tools, Games, Sessions, View, Split, MultiExec, Tunneling, Packages, Settings, and Help, and a sidebar with buttons for Quick connect..., Sessions, Tools, and Macros. The main terminal pane shows a Hive prompt (hive>) and the following commands: set hive.support.concurrency = true; set hive.enforce.bucketing = true; set hive.exec.dynamic.partition.mode = nonstrict; set hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager; set hive.compactor.initiator.on = true;. The left sidebar shows a file tree for /home/acadgild/ with various folders like Videos, Templates, Sqoop, Public, Pig, Pictures, Music, Install, finaloutput, eclipse-workspace, eclipse, Downloads, Documents, Desktop, .webclipse, .vim, .tooling, .thumbnails, .svt, .ssh, .scalade, .pulse, .pki, and .p2. At the bottom, there is a message: "UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>".

```
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;
```

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:-



A screenshot of the MobaXterm application window, similar to the one above. The main terminal pane shows the execution of the Hive 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');. The output shows "OK" and "Time taken: 1.107 seconds". The prompt then returns to hive>. The interface elements (menu bar, toolbar, sidebar, and footer message) are identical to the previous screenshot.

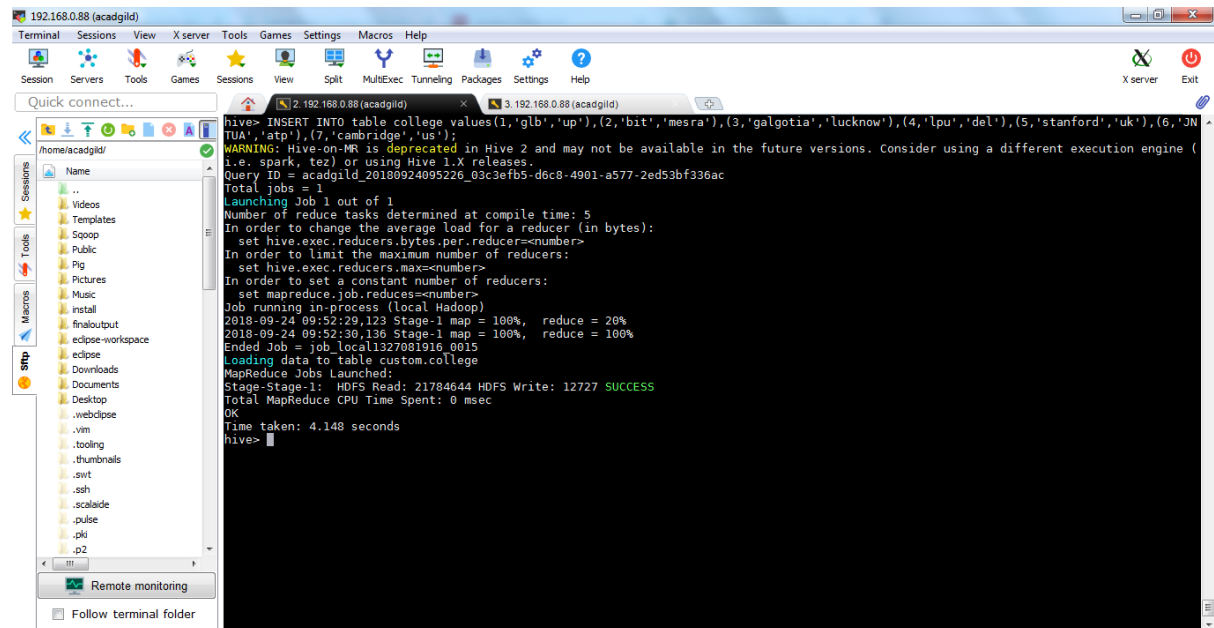
```
hive> CREATE TABLE college(clg_id int,clg_name string,clg_loc string) clustered by (clg_id) into 5 buckets stored as orc TBLPROPERTIES('transactional'='true');
OK
Time taken: 1.107 seconds
hive>
```


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 terminal window with a file explorer on the left. The terminal displays the following output for the Hive command:

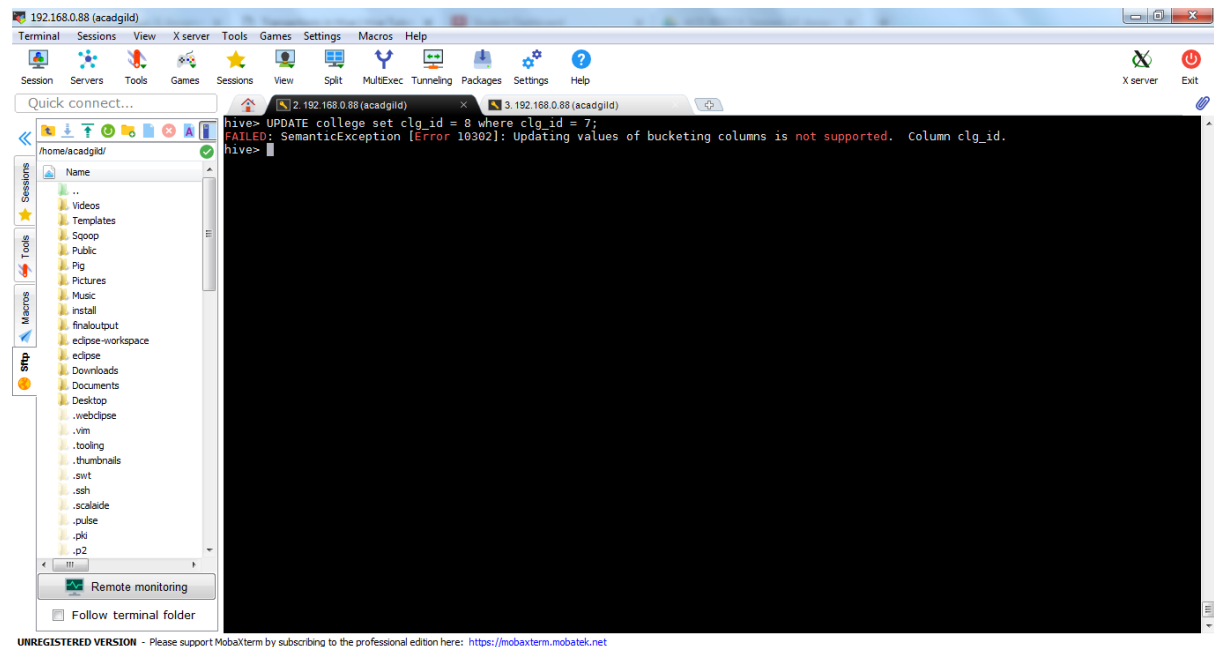
```
hive> INSERT INTO table college values(1,'glb','up'),(2,'bit','mesra'),(3,'galgotia','lucknow'),(4,'lpu','del'),(5,'stanford','uk'),(6,'JNTUA','atp'),(7,'cambridge','us');
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution 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_local1827081016_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;

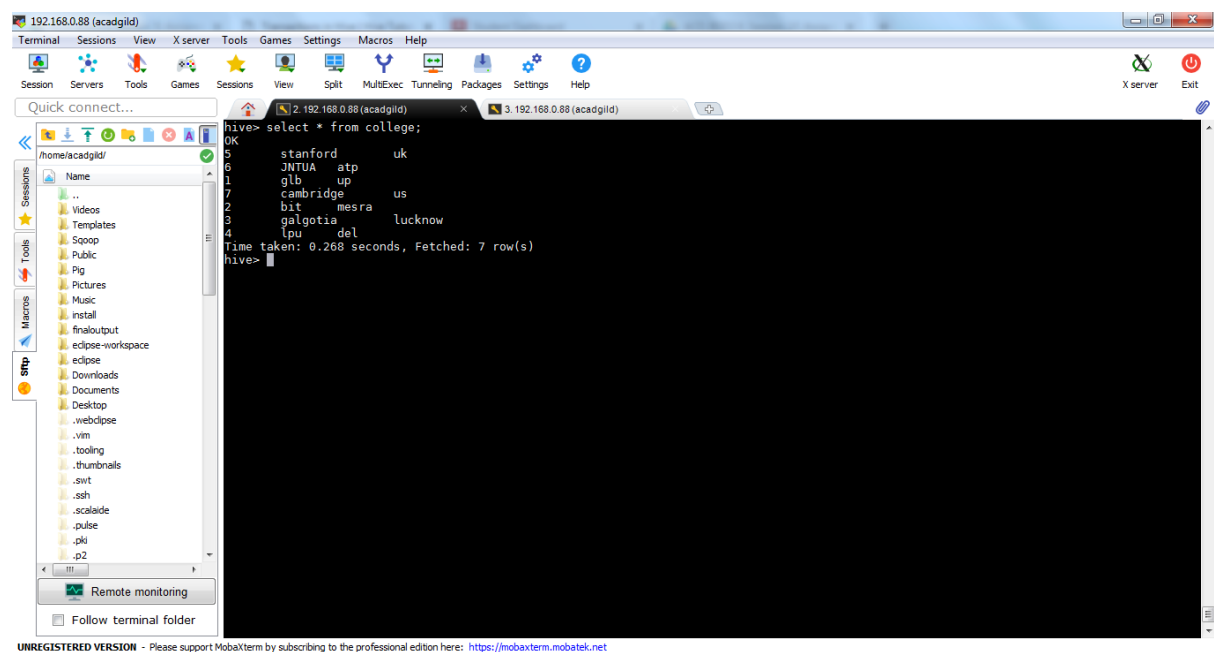
Output-:



The screenshot shows the MobaXterm interface with a terminal window. The terminal displays the following text:

```
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>
```

The interface includes a sidebar with a file explorer showing various folders like Videos, Templates, and Downloads. The top menu bar includes options like Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, and Help.



The screenshot shows the MobaXterm interface with a terminal window. The terminal displays the following text:

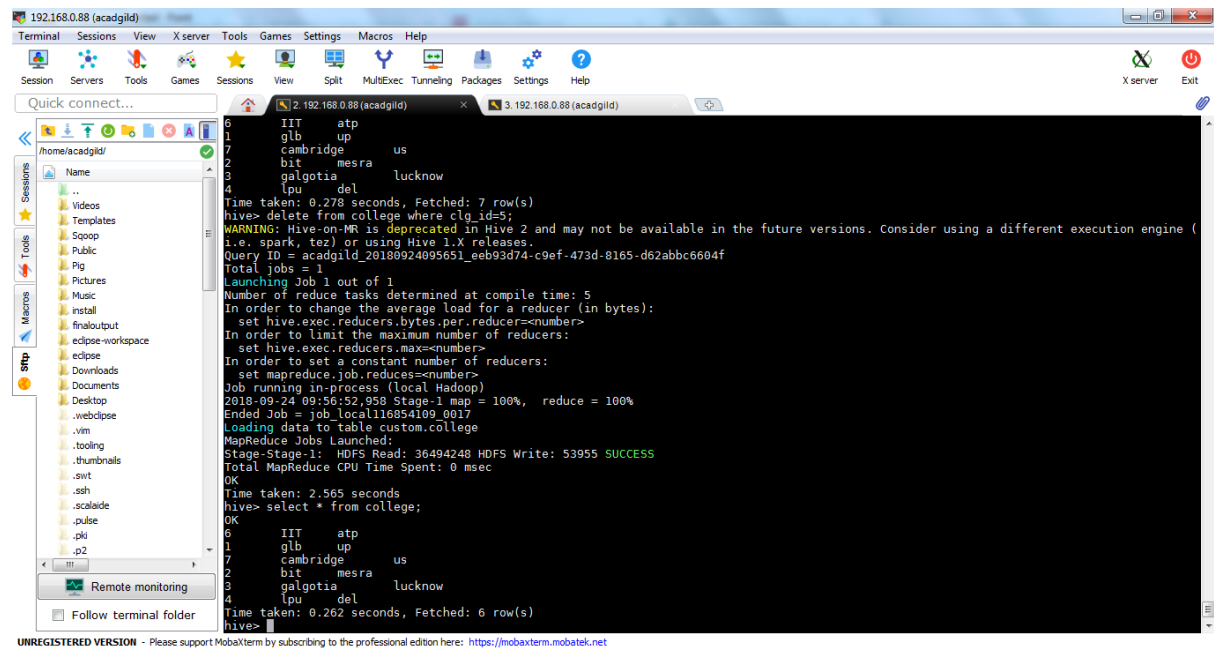
```
hive> select * from college;  
OK  
5      stanford      uk  
6      JNTUA      atp  
1      glb      up  
7      cambridge      us  
2      bit      mesra  
3      galgotia      lucknow  
4      lpu      del  
Time taken: 0.268 seconds, Fetched: 7 row(s)  
hive>
```

The interface is similar to the previous screenshot, showing the same sidebar and menu bar.

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 MultExec Tunneling Packages Settings Help
Quick connect...
/home/acadgild/
Sessions
Name
Videos
Templates
Spoon
Public
Pig
Pictures
Music
Install
finaloutput
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.vim
.tooling
.thumbnails
.ssh
.scalade
.pulse
.pli
.p2
Remote monitoring
Follow terminal folder
2 192.168.0.88 (acadgild)
3 192.168.0.88 (acadgild)
6 IIT atp
7 glb up
7 cambridge us
2 bit mesra
3 galgotia lucknow
4 lpu del
Time taken: 0.278 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_20180924095051_eeb93d74-c9ef-473d-8165-d62abb6604f
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 09:56:52.958 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local116894109_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>
```