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

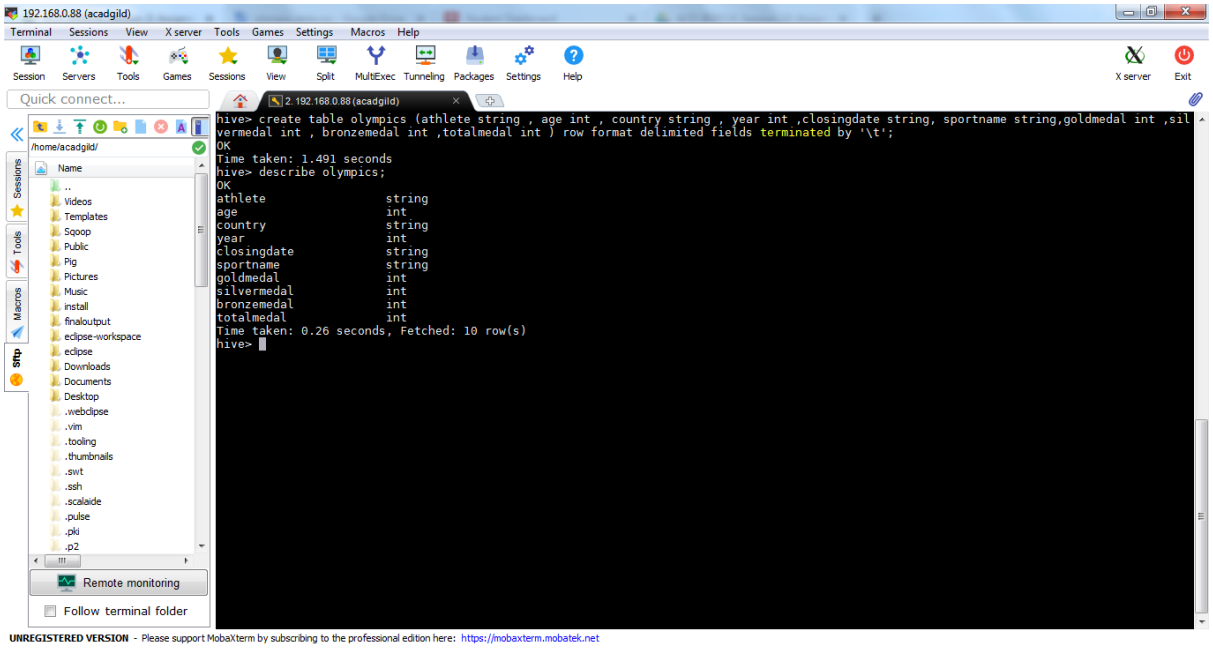
Gold Medals: No. of Gold medals

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 :



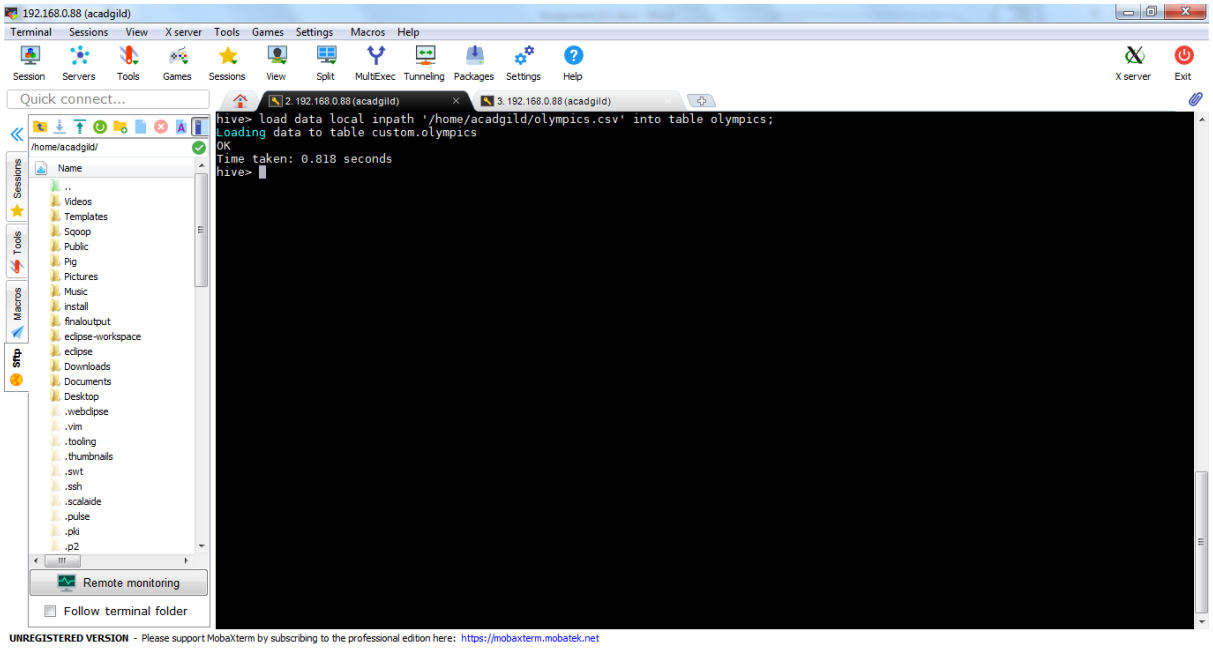
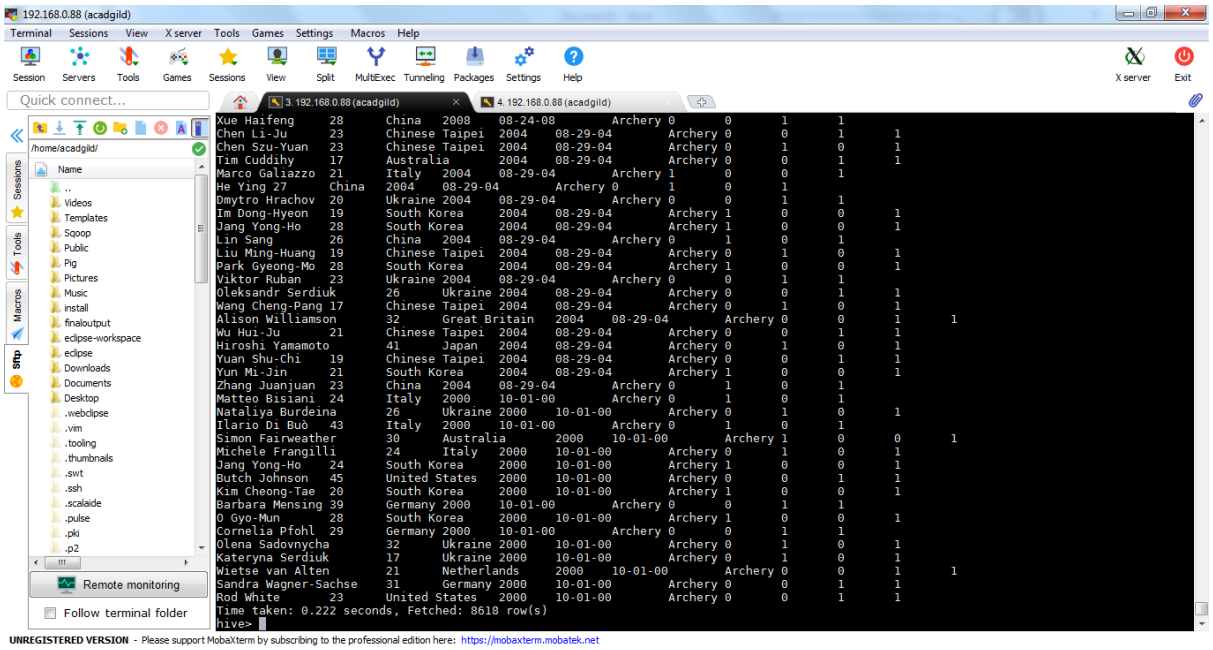


Table Content-:

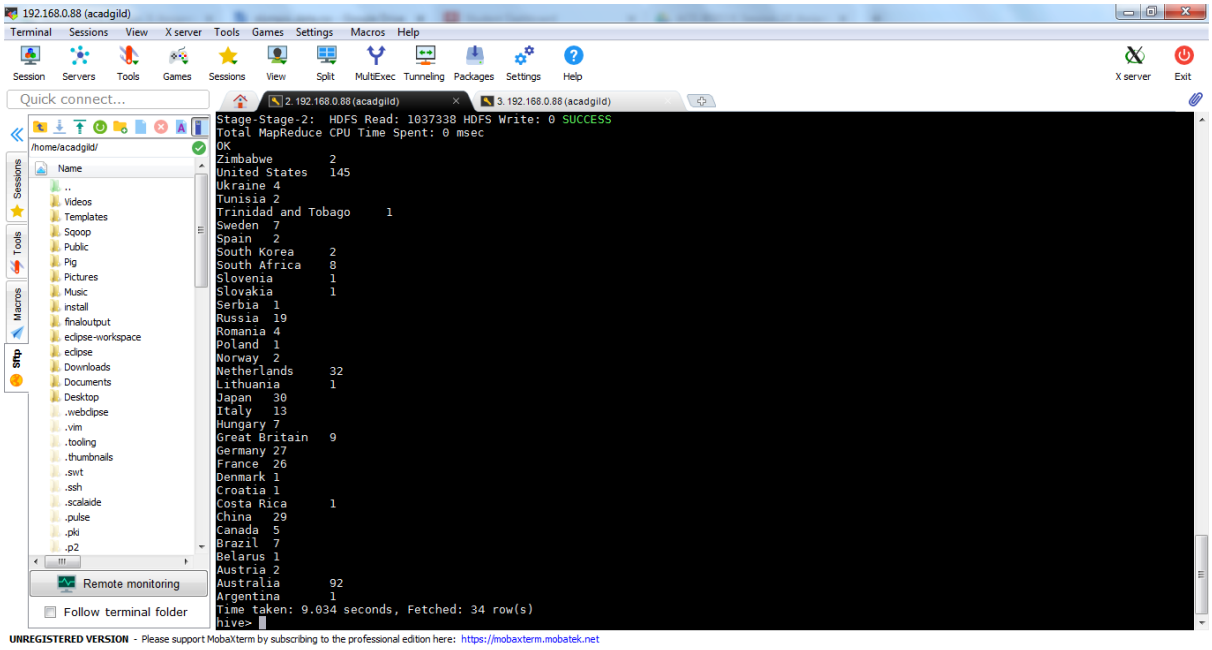


5. Problem Statement

Task 1

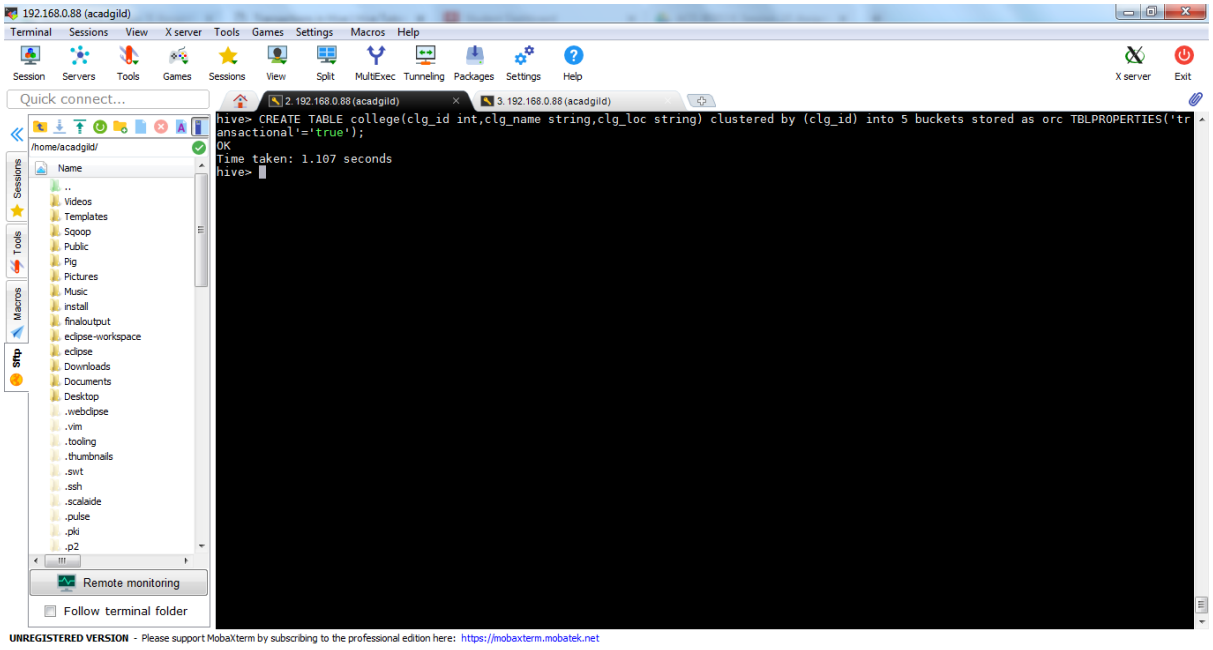
1. Write a Hive program to find the number of medals won by each country in swimming.

Output-:



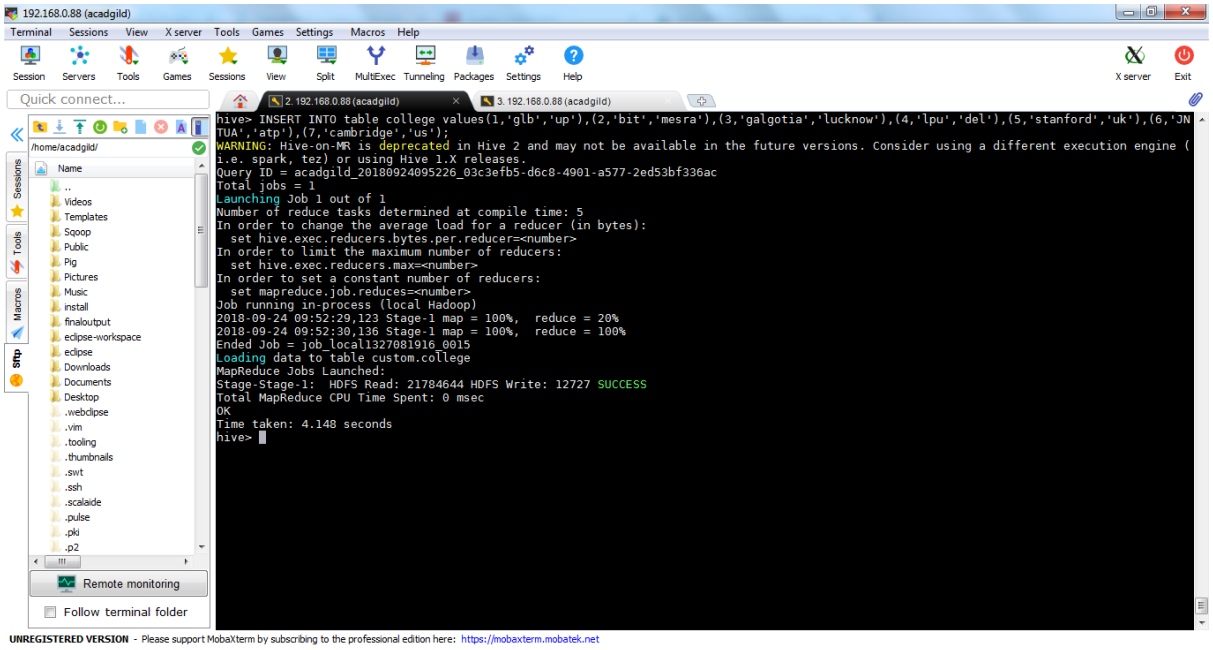
2. Write a Hive program to find the number of medals that India won year wise.

Output-:

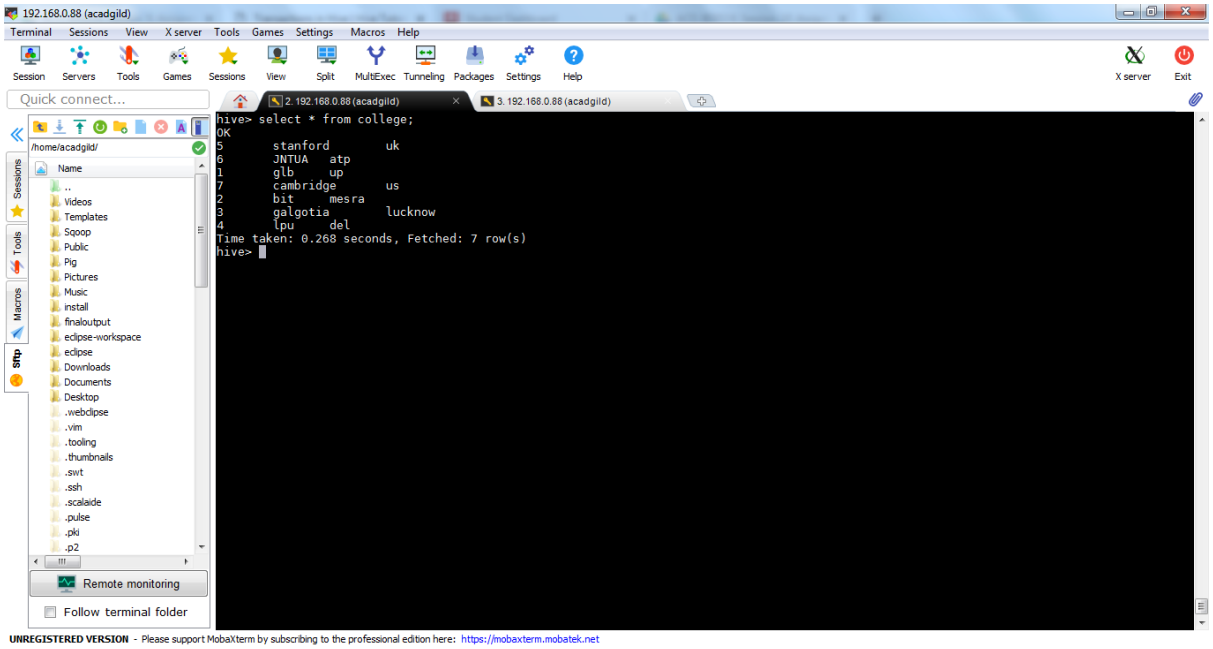


3. Write a Hive Program to find the total number of medals each country won.

Output-:



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



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;

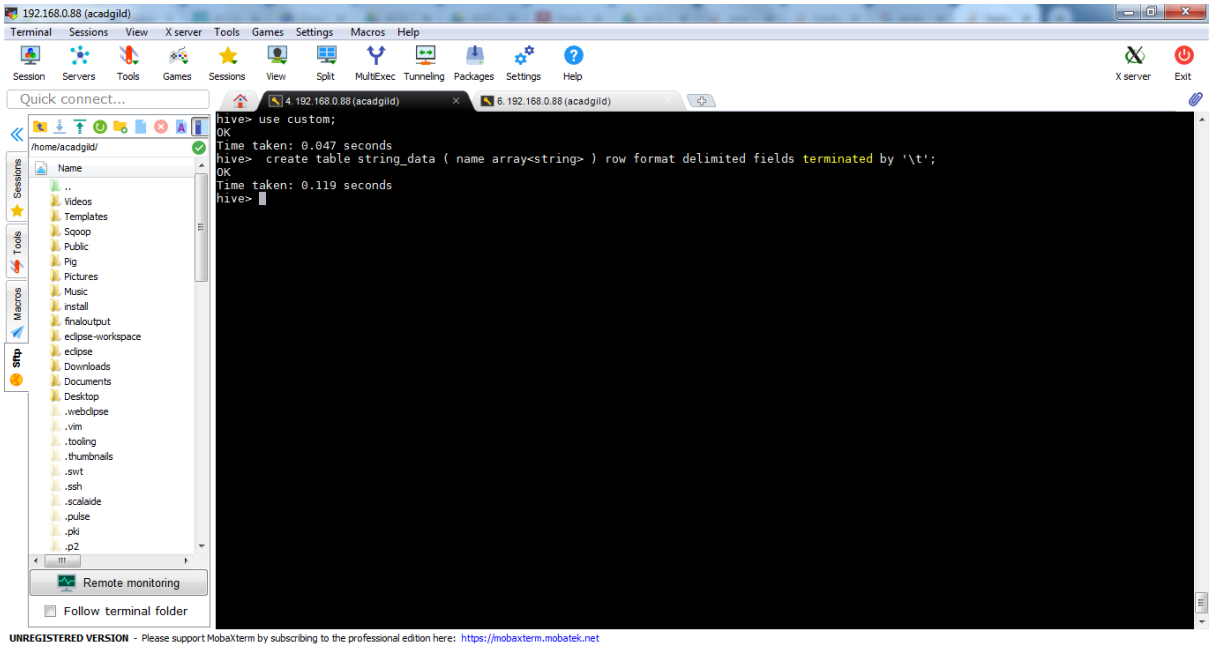
->>Quering in hive language using temporary function created above with the use of table string data

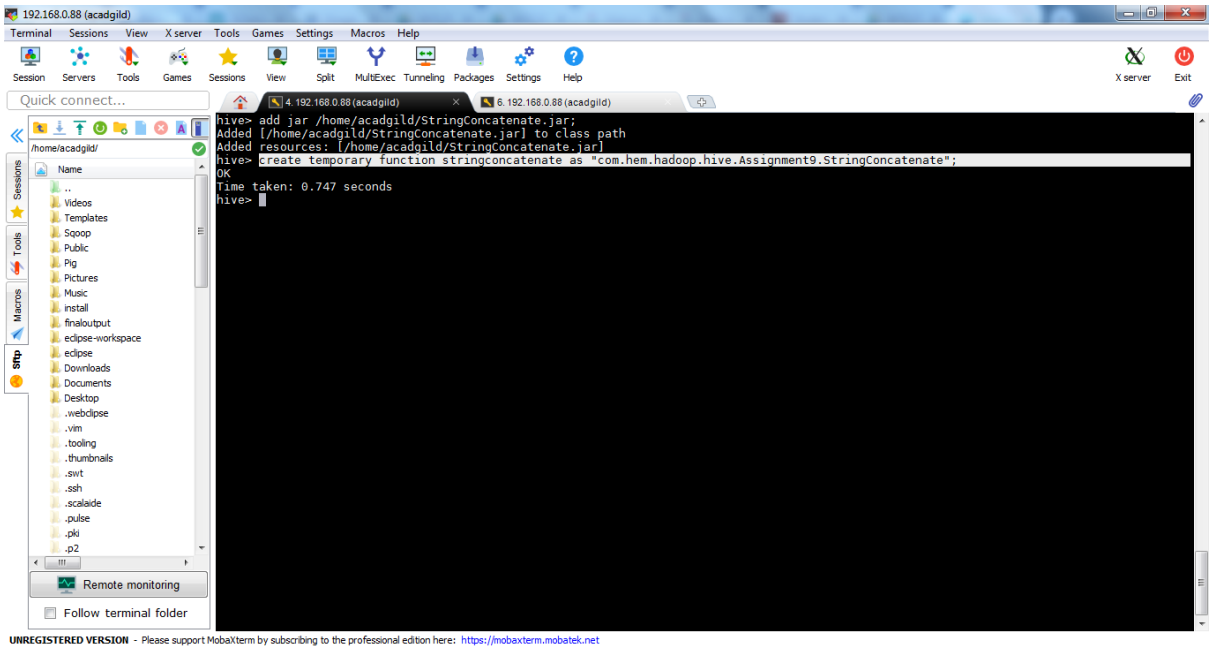
Content of words.txt

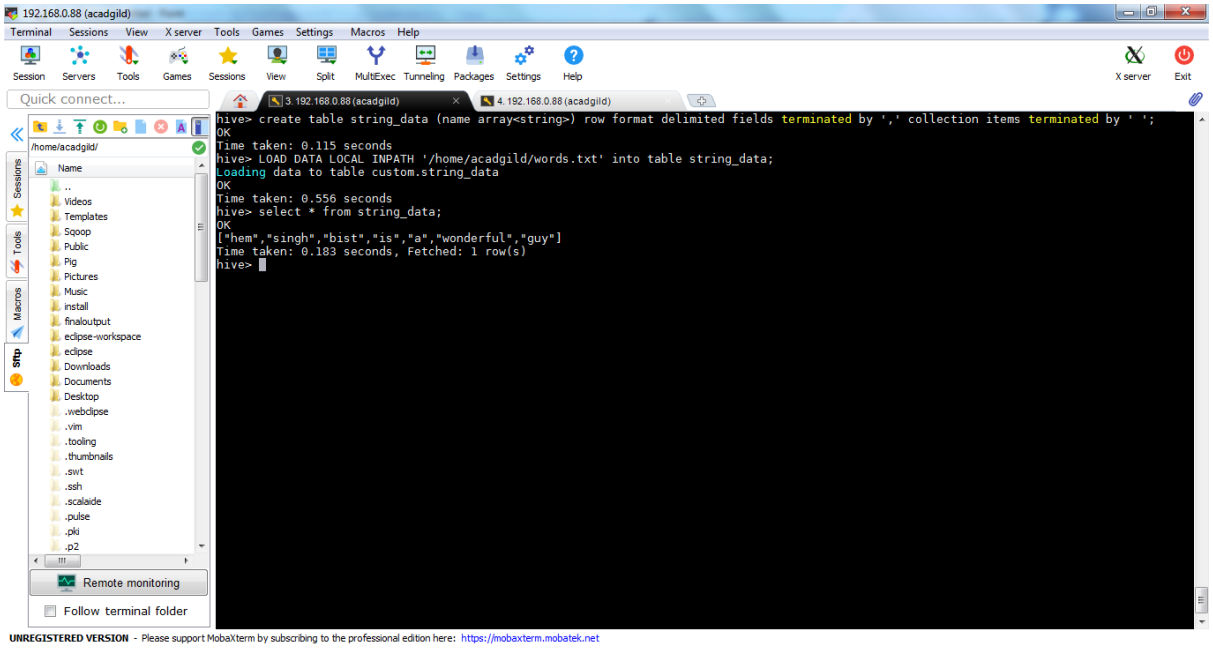
hem singh bist is a wonderful guy

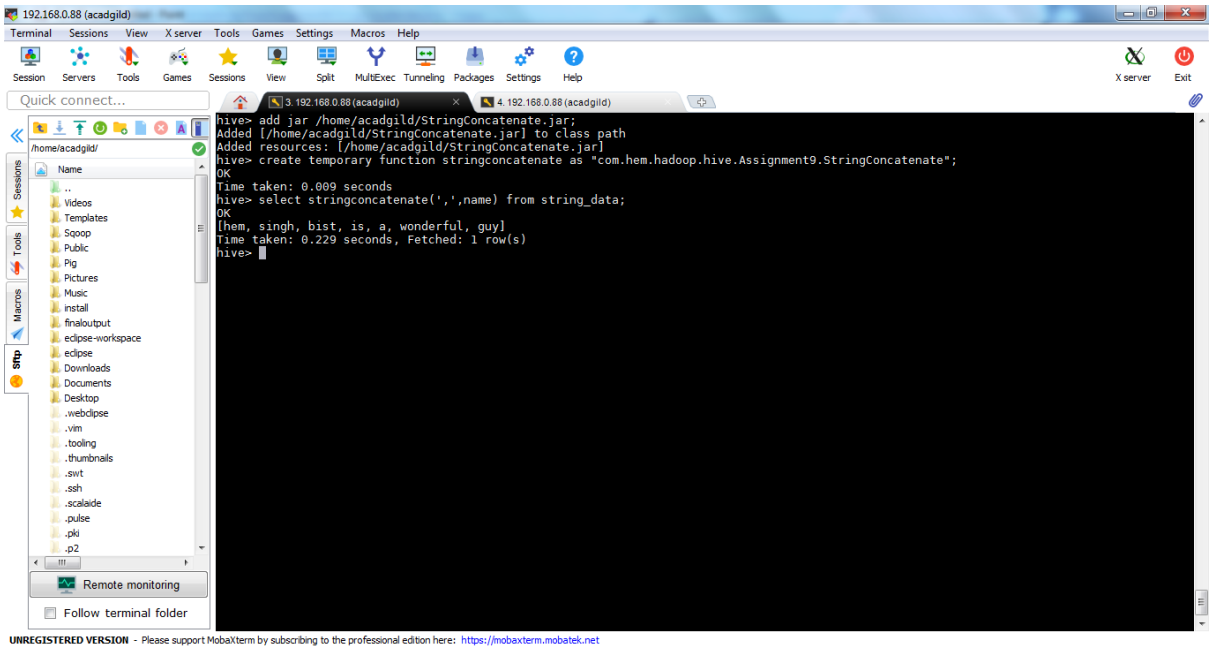
StringConcatenate Program-:

```
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 delimeter, 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);
		}

	}
}
```


192.168.0.88 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
4. 192.168.0.88 (acadgild)
6. 192.168.0.88 (acadgild)
/home/acadgild/
Name
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
finaloutput
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.vim
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
Remote monitoring
Follow terminal folder
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>
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net


192.168.0.88 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
4. 192.168.0.88 (acadgild)
6. 192.168.0.88 (acadgild)
/home/acadgild/
Name
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
finaloutput
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.vim
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
Remote monitoring
Follow terminal folder
hive> add jar /home/acadgild/StringConcatenate.jar;
Added [/home/acadgild/StringConcatenate.jar] to class path
Added resources: [/home/acadgild/StringConcatenate.jar]
hive> create temporary function stringconcatenate as "com.hem.hadoop.hive.Assignment9.StringConcatenate";
OK
Time taken: 0.747 seconds
hive>
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;

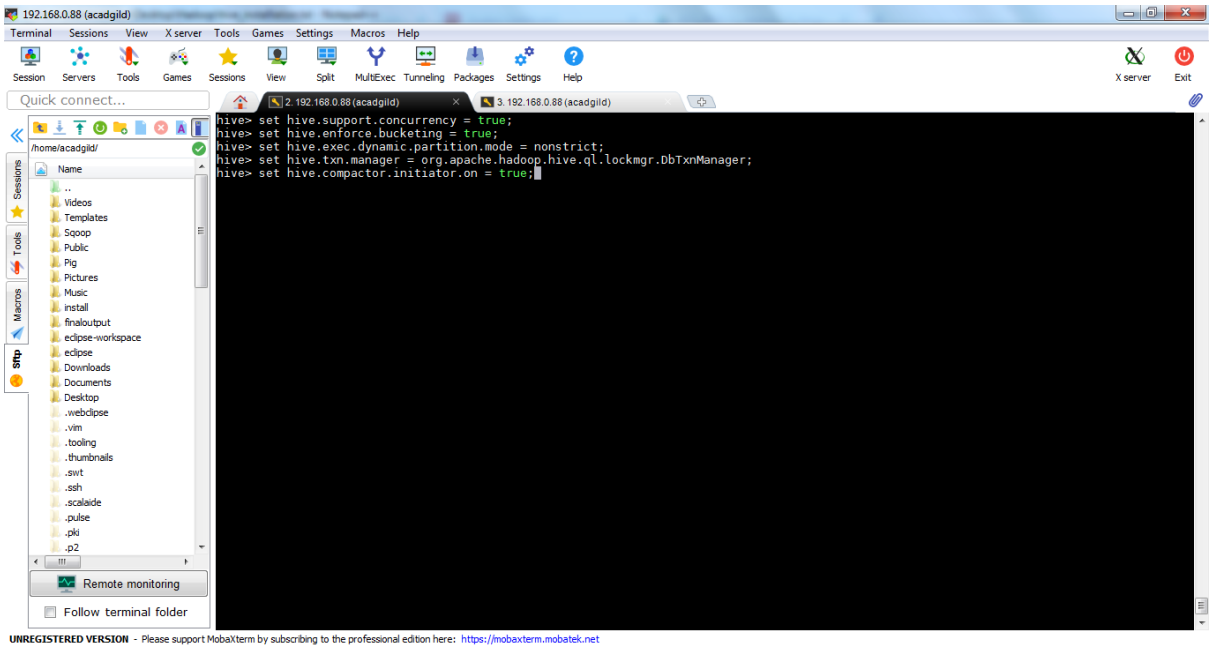
hive>set hive.exec.dynamic.partition.mode = nonstrict;

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;

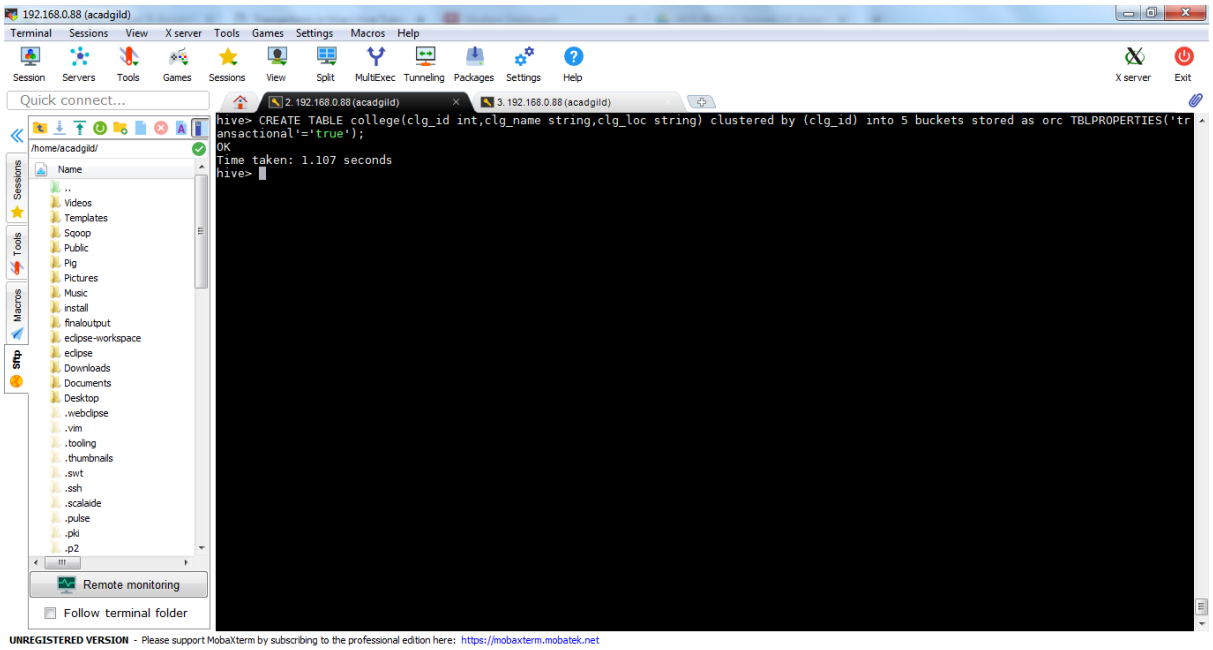
Output-:



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');

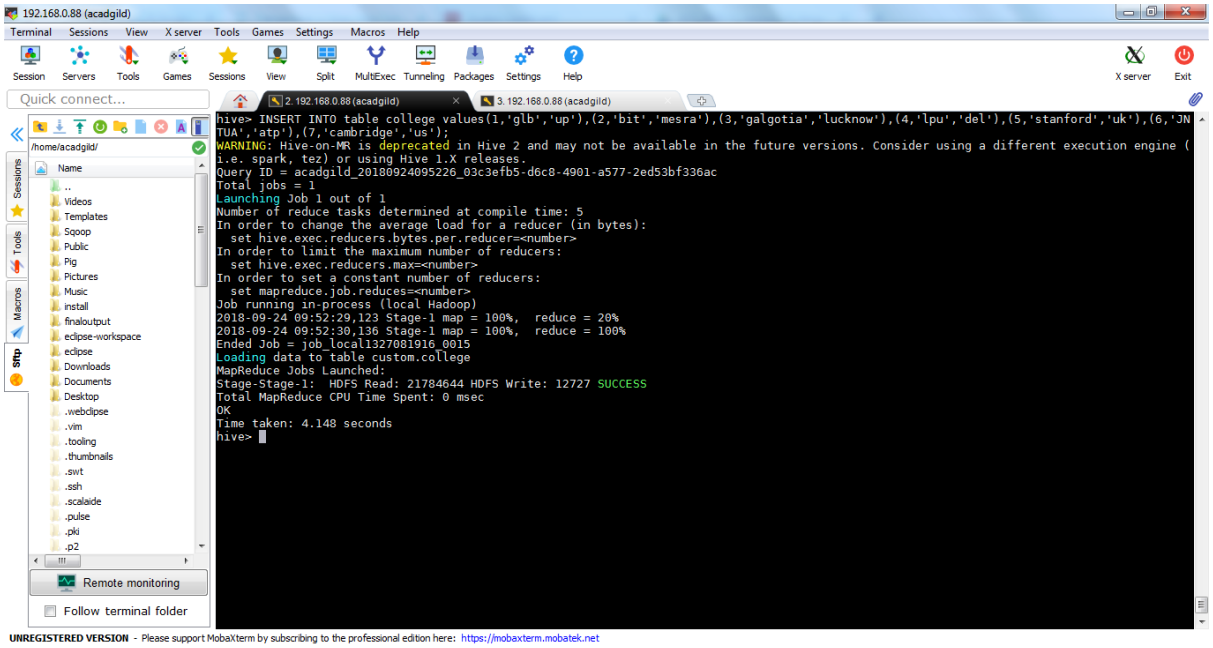
Output-:



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');

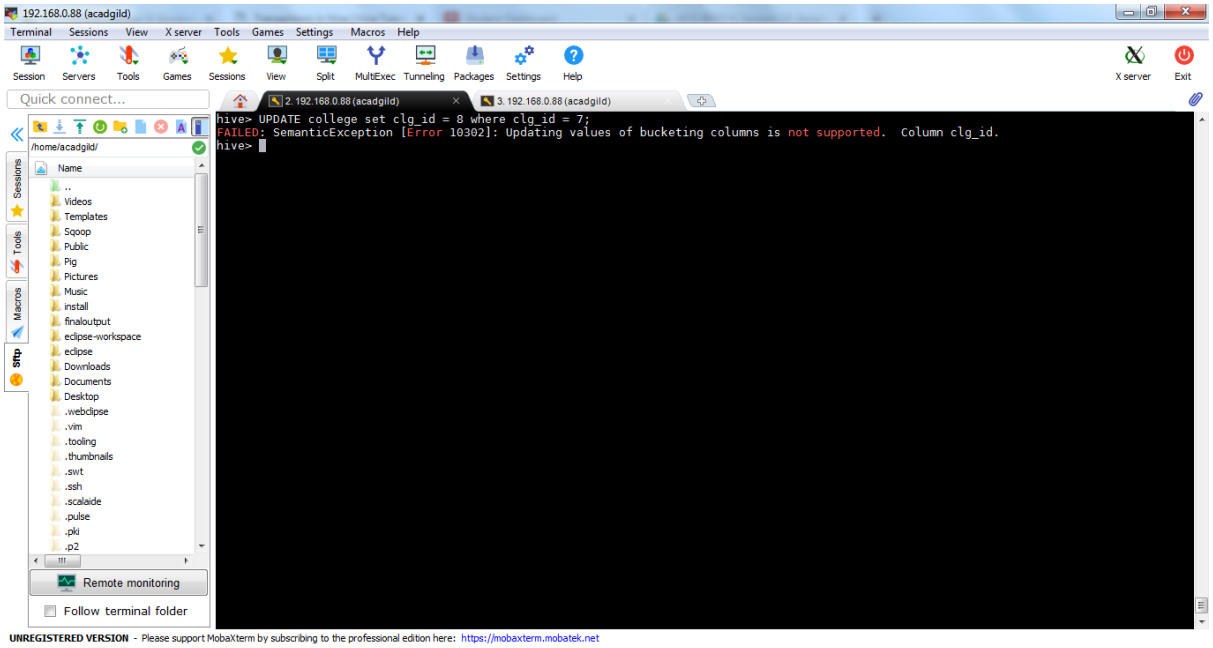
Output-:

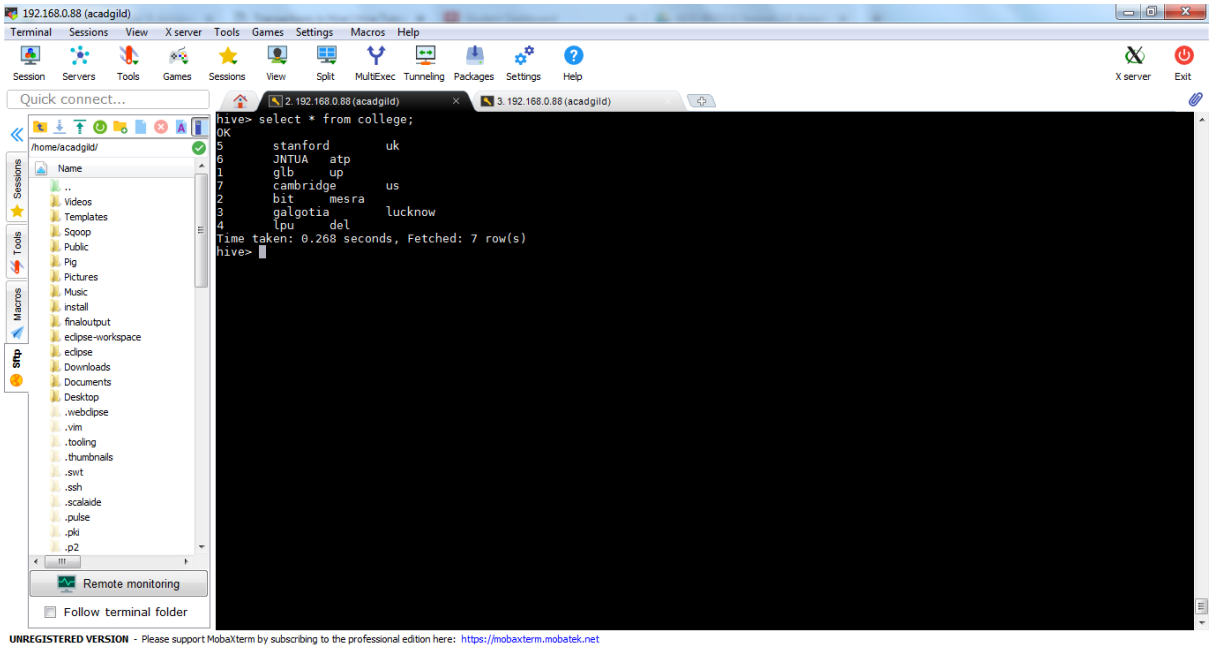


Command-:

UPDATE college set clg_id = 8 where clg_id = 7;

Output-:

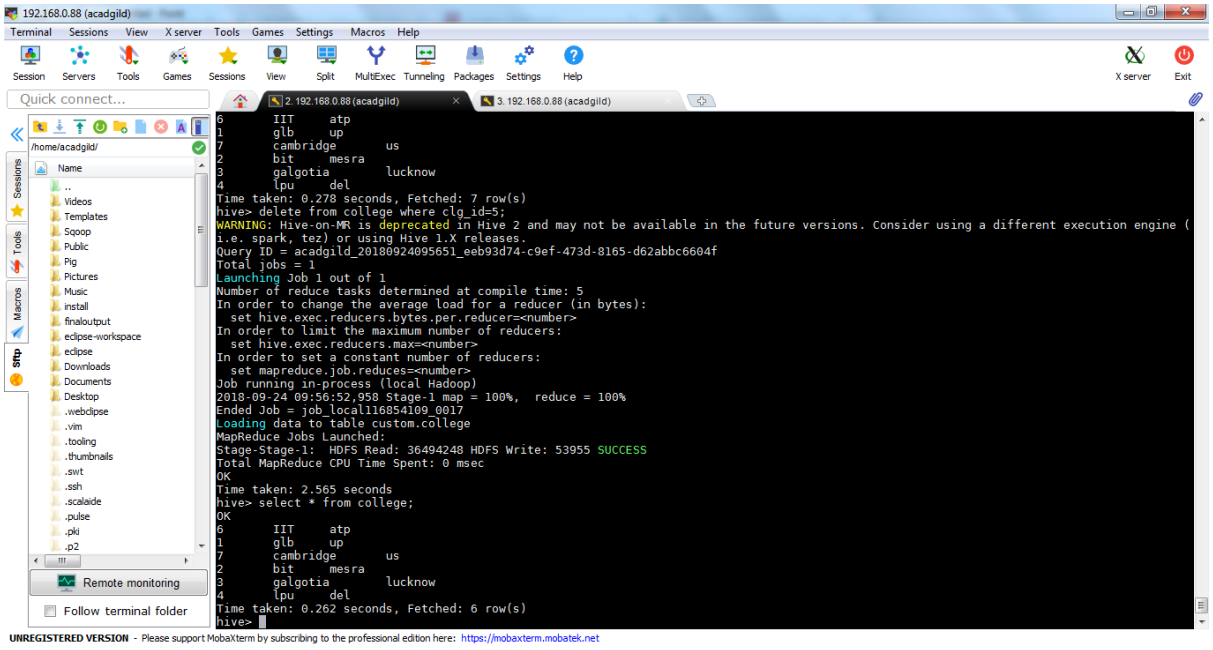




Command-:

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 MultiExec Tunneling Packages Settings Help
X server Exit
Quick connect...
2. 192.168.0.88 (acadgild)
3. 192.168.0.88 (acadgild)
/home/acadgild/
Name
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
finaloutput
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.vim
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
Remote monitoring
Follow terminal folder
6 IIT atp
1 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_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.reduces=<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