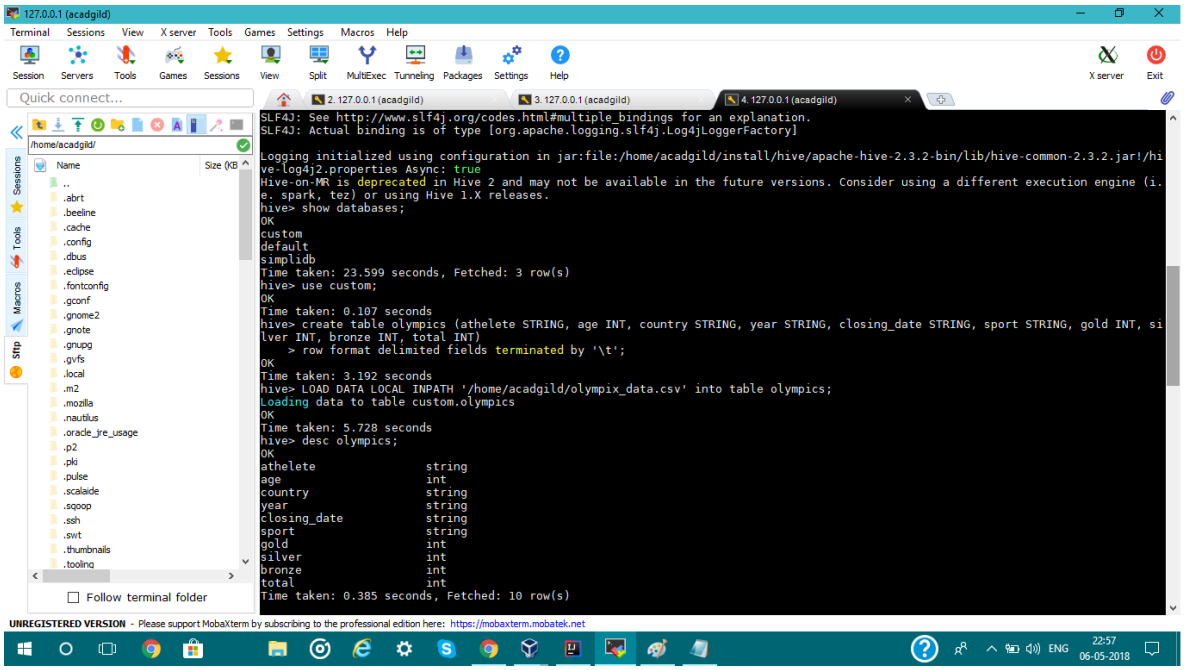
Assignment 9

Creation of table for Olympics table



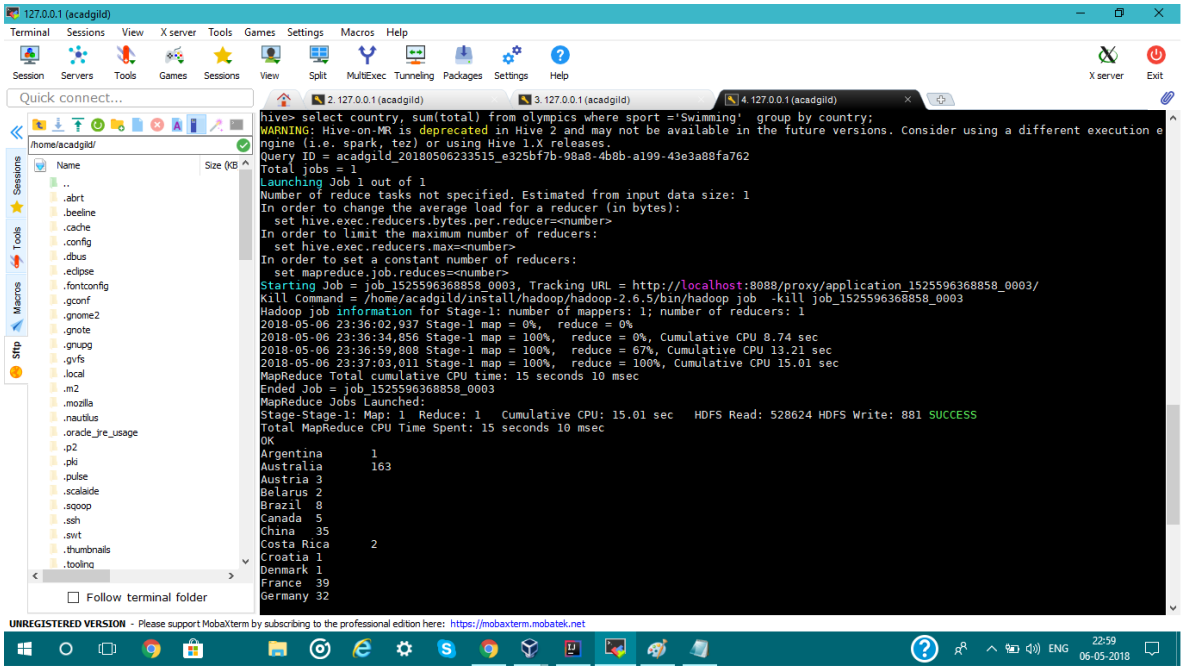
Command used: create table olympic (athelete STRING, age INT, country STRING, year STRING, closing_date STRING, sport STRING, gold INT, silver INT, bronze INT, total INT)

row format delimited fields terminated by '\t';

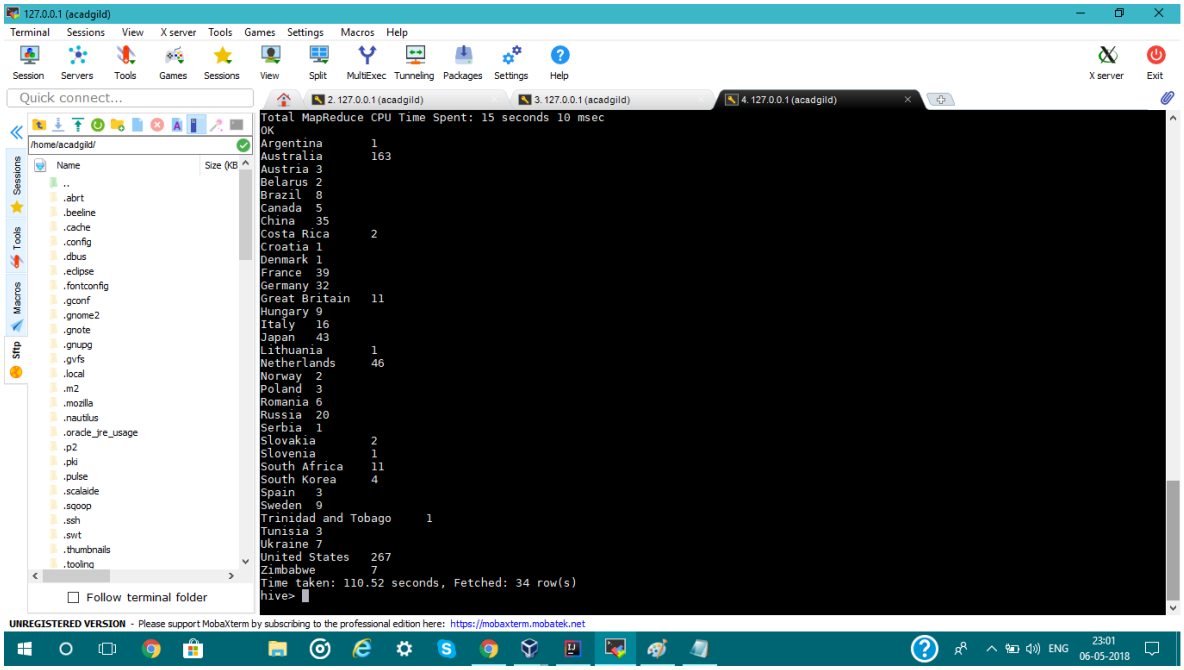
Task1

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

Command used: select country, sum(total) from Olympics where sport='Swimming' group by country;

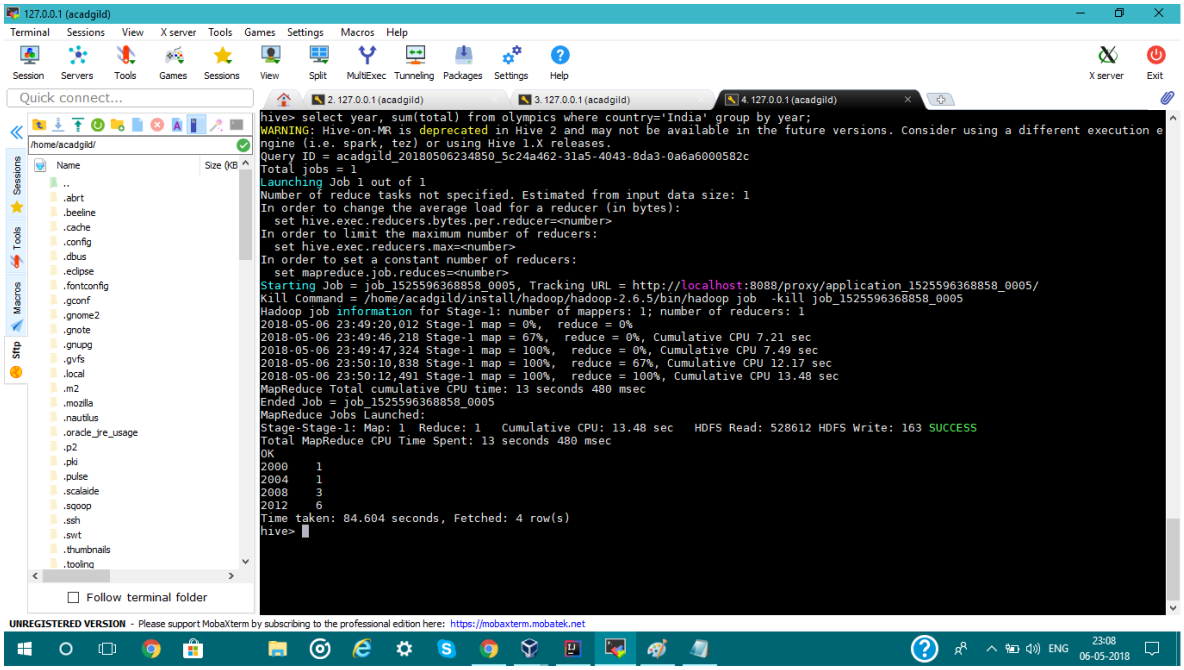


Output



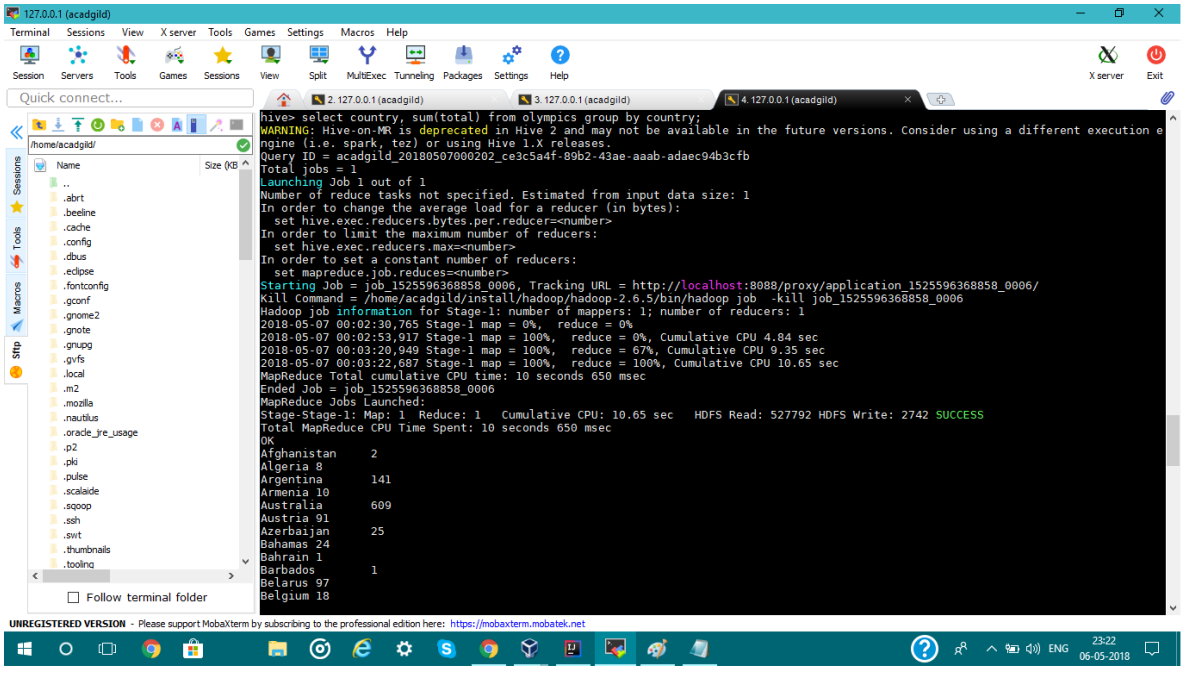
b. Write a Hive program to find the number of medals that India won year wise.

Command used: select year, sum (total) from Olympics where country = 'India' group by year;

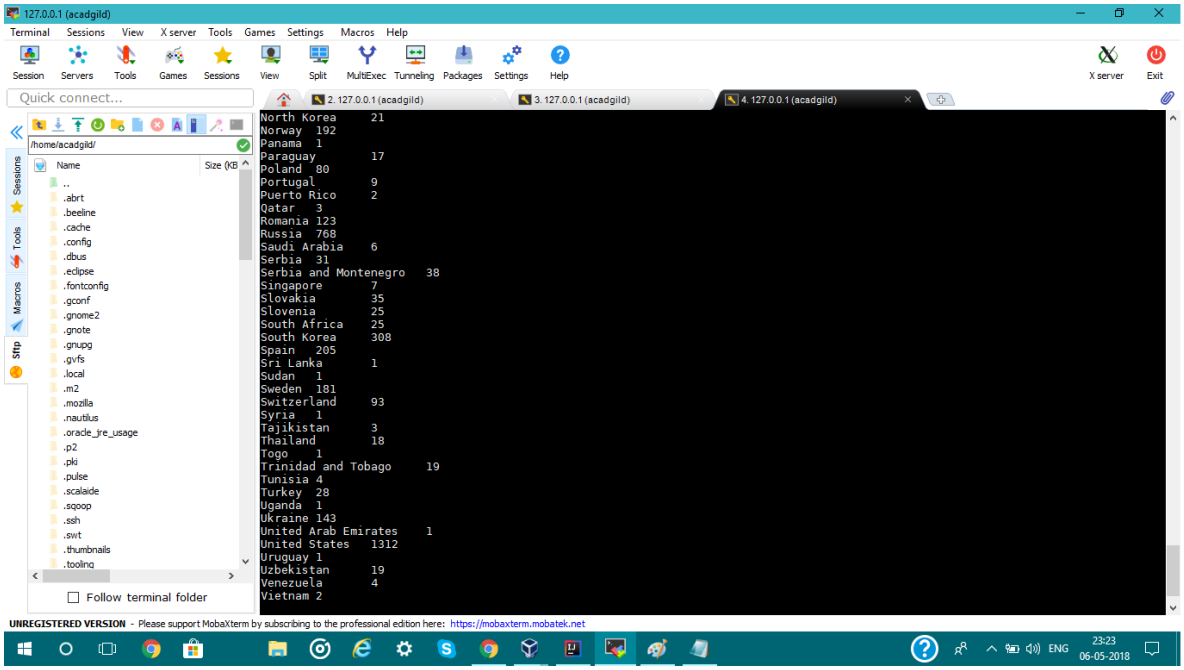


c. Write a Hive Program to find the total number of medals each country won.

Command used: select country, sum (total) from Olympics group by country;



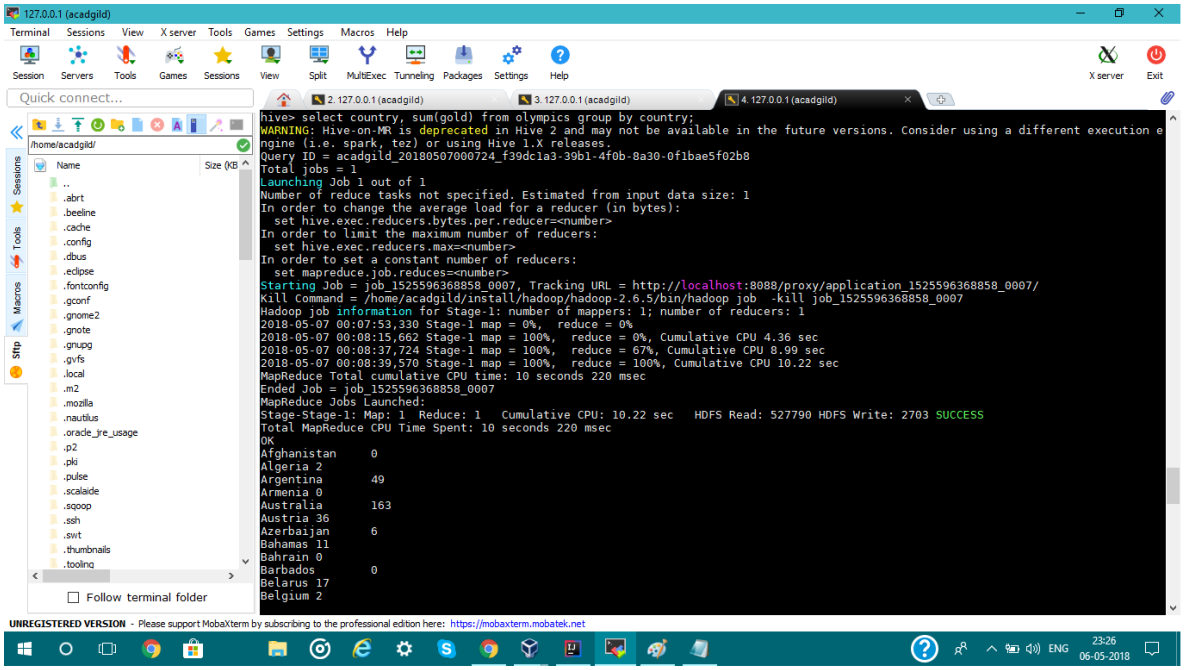
Output



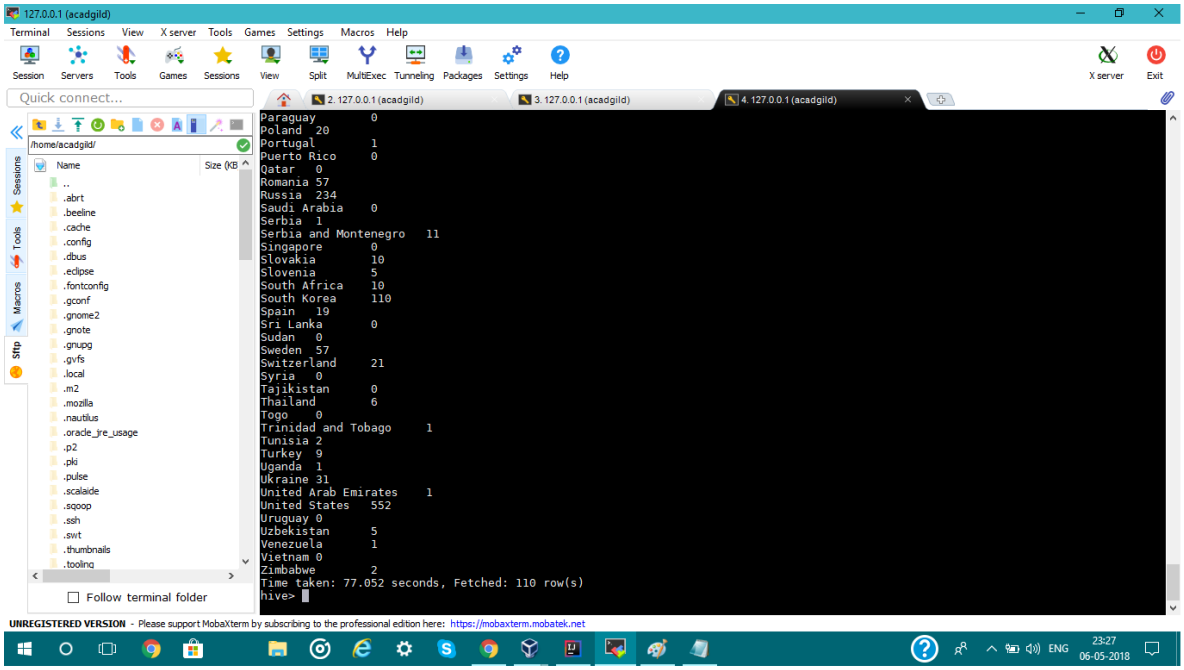
---

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

Command used: select country, sum (gold) from Olympics group by country;



Output



Task2

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 used: customer.txt

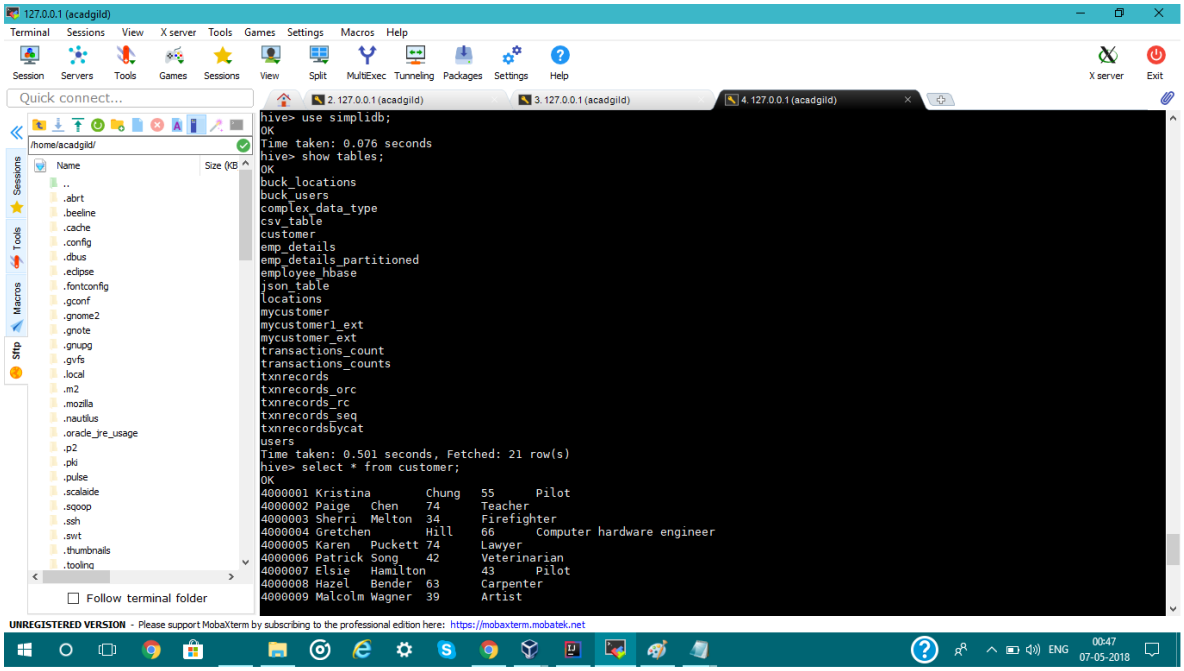
Command used: select concat_ws(‘ SEP ’, collect_list(fname)) from customer;

Explanation: Here, the function concat_ws is used to concat the elements of an array.

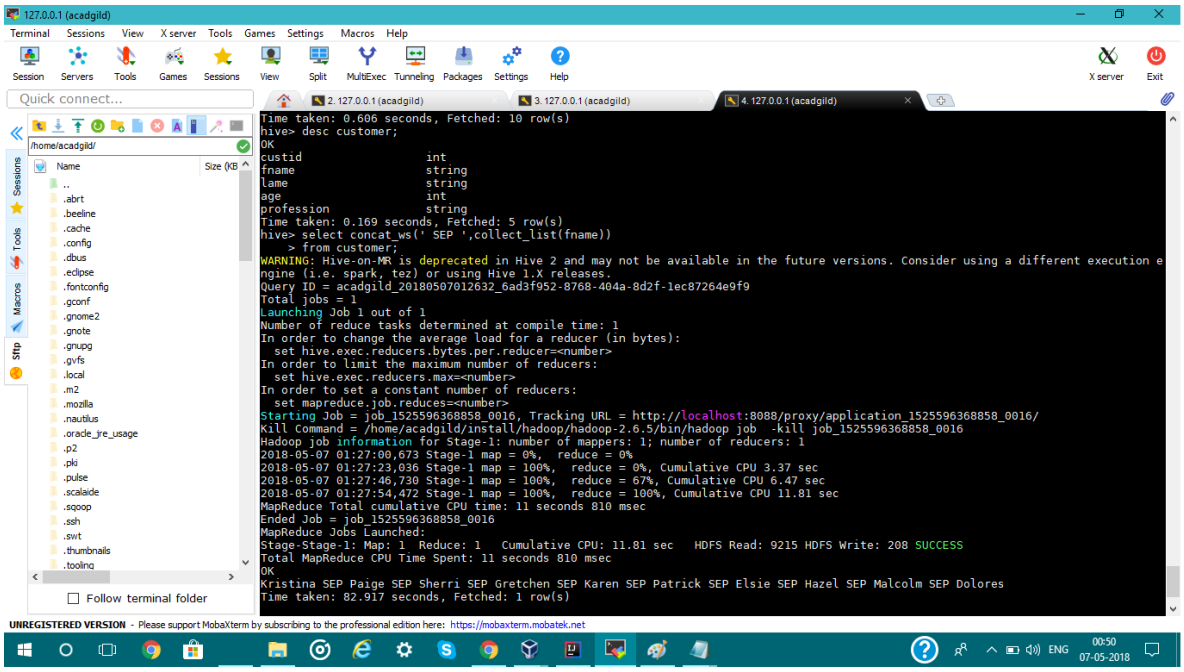
Syntax: concat_ws(‘delimiter’, array).

This function takes two arguments in case of an array: first is the delimiter for the concatenation and second is the array of elements for which the concatenation is to be done.

For this task, the delimiter is the string, ‘SEP’ with white spaces before and after; and the column for which the array is to be generated via collect_list is ‘fname’

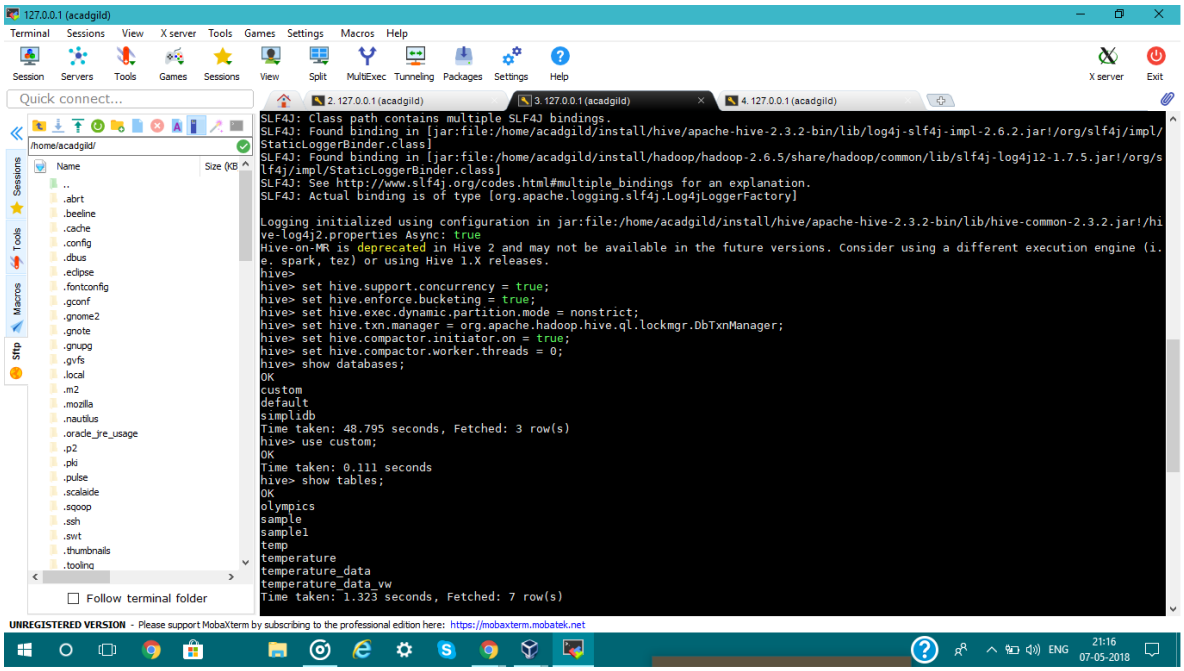


Output

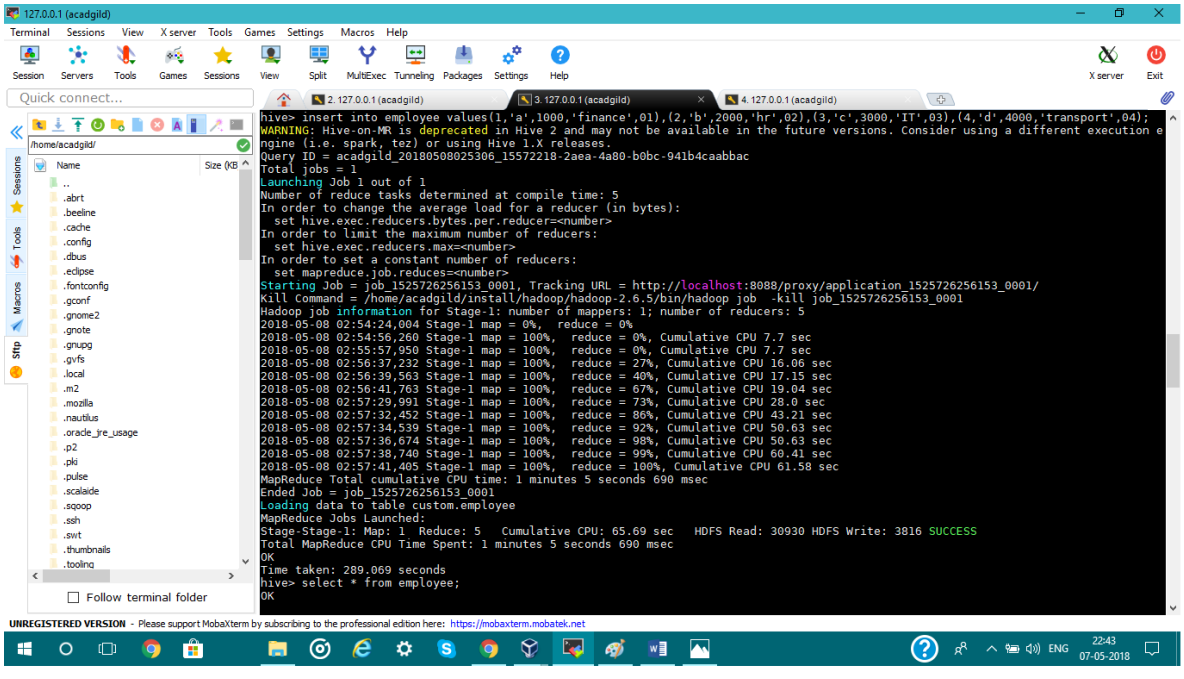


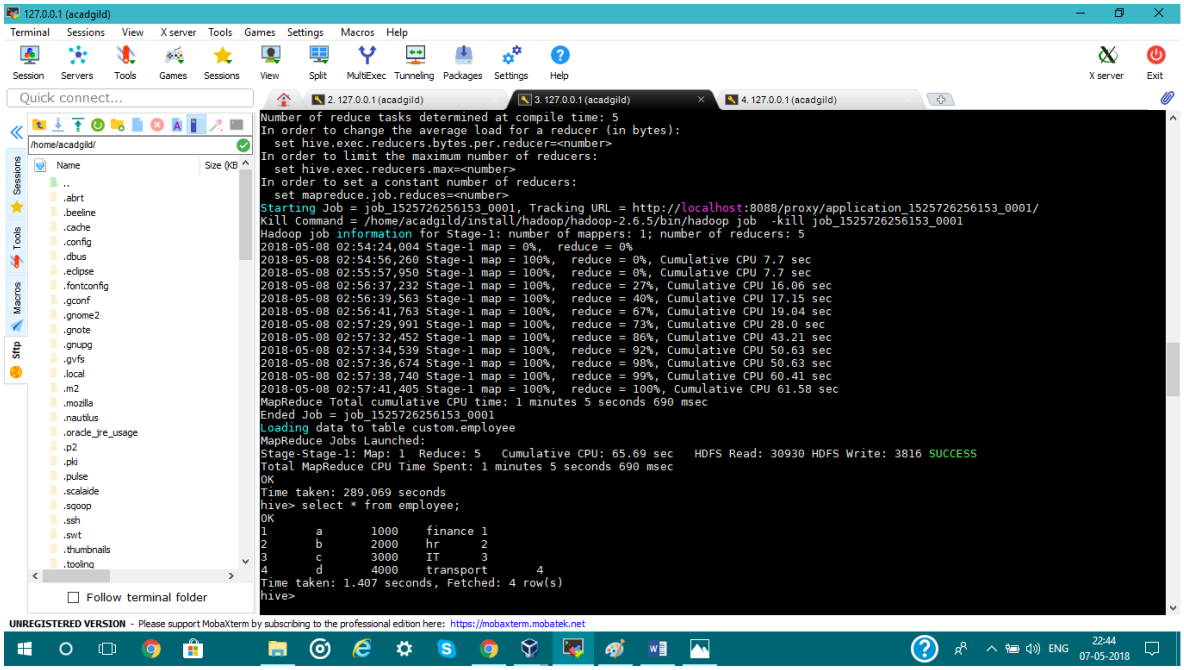
Task3

a. Setting up of properties

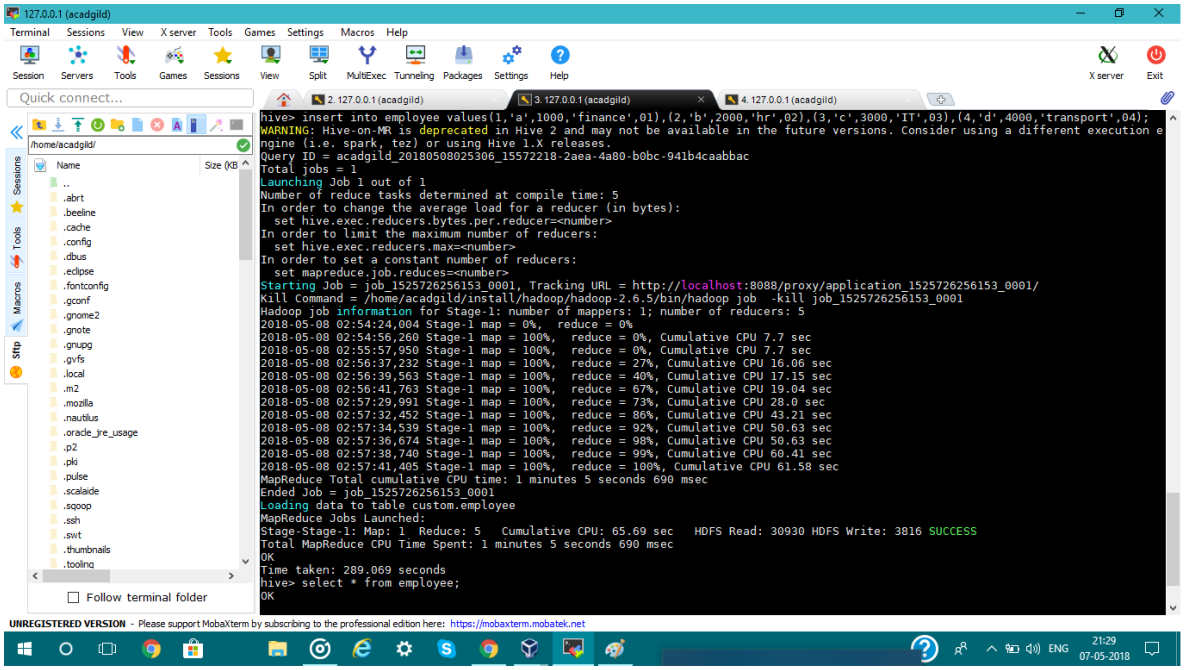


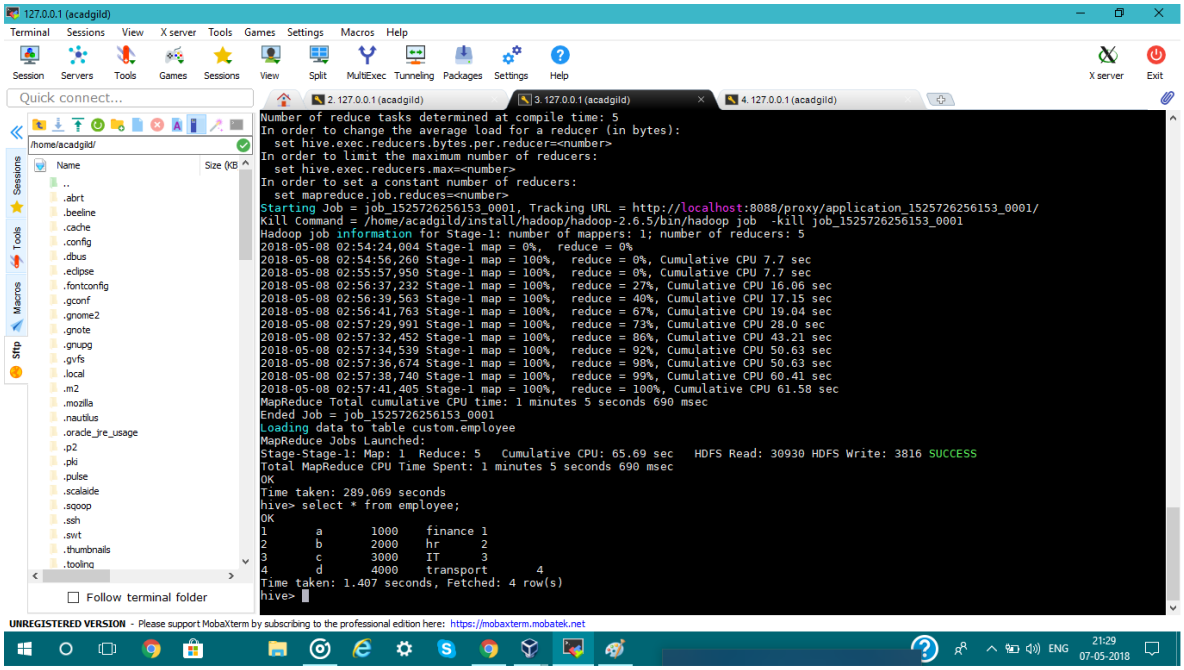
b. Creation of table that supports Hive Transaction



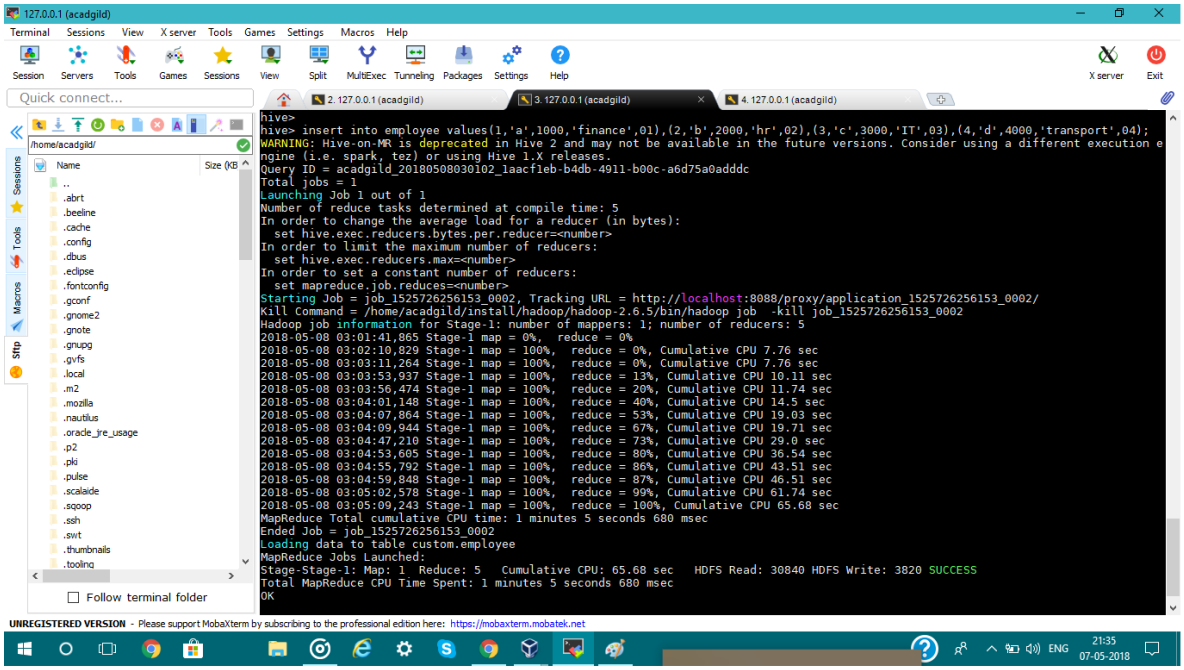


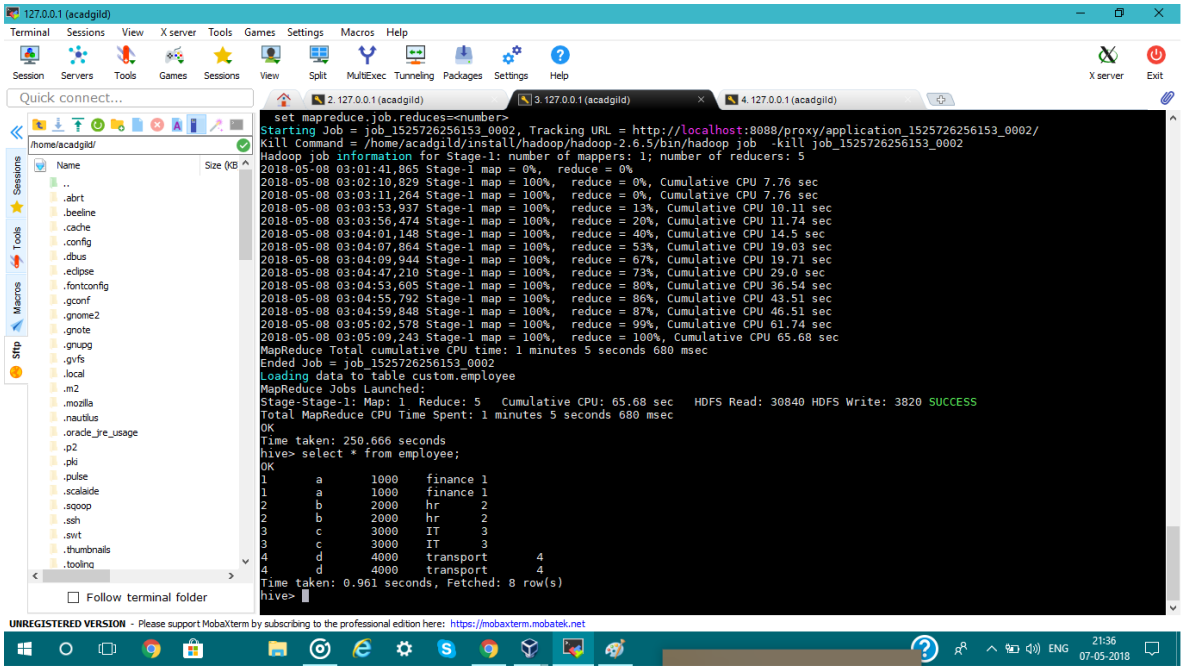
c. Inserting data into Hive table



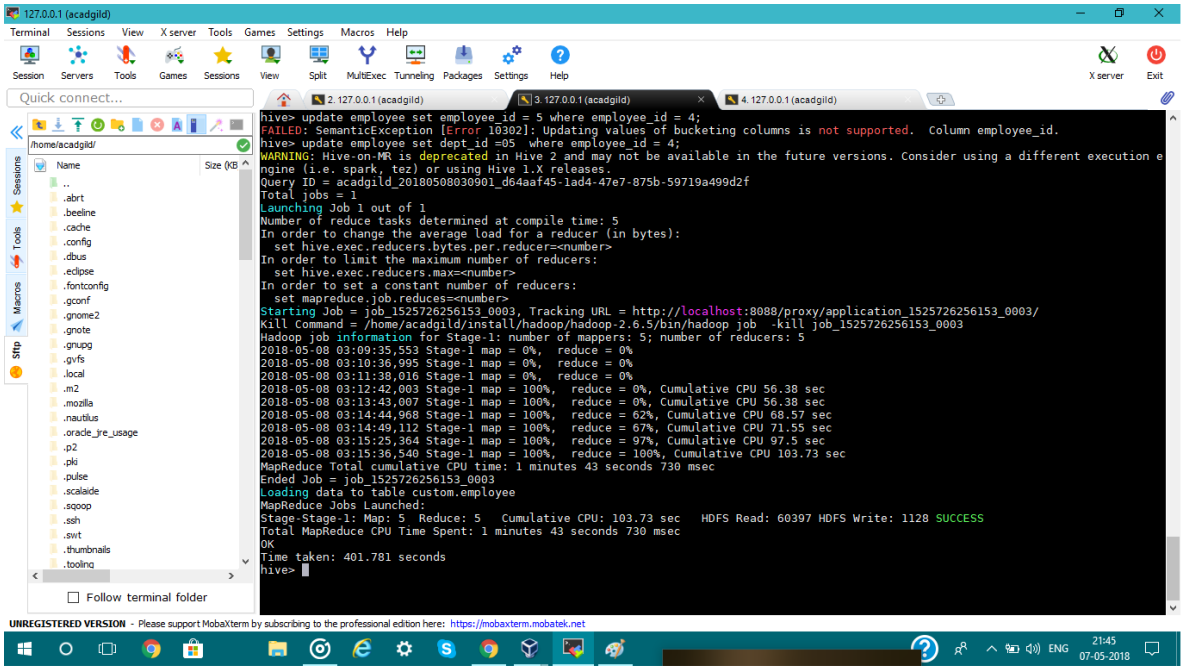


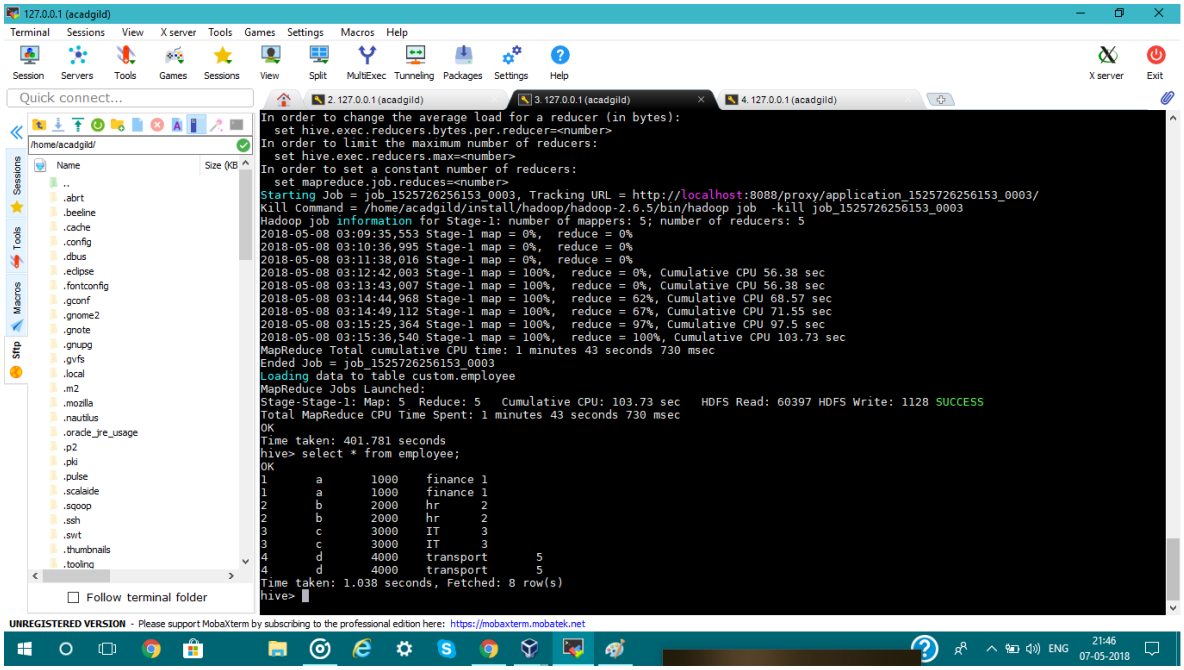
d. Re-inserting of data



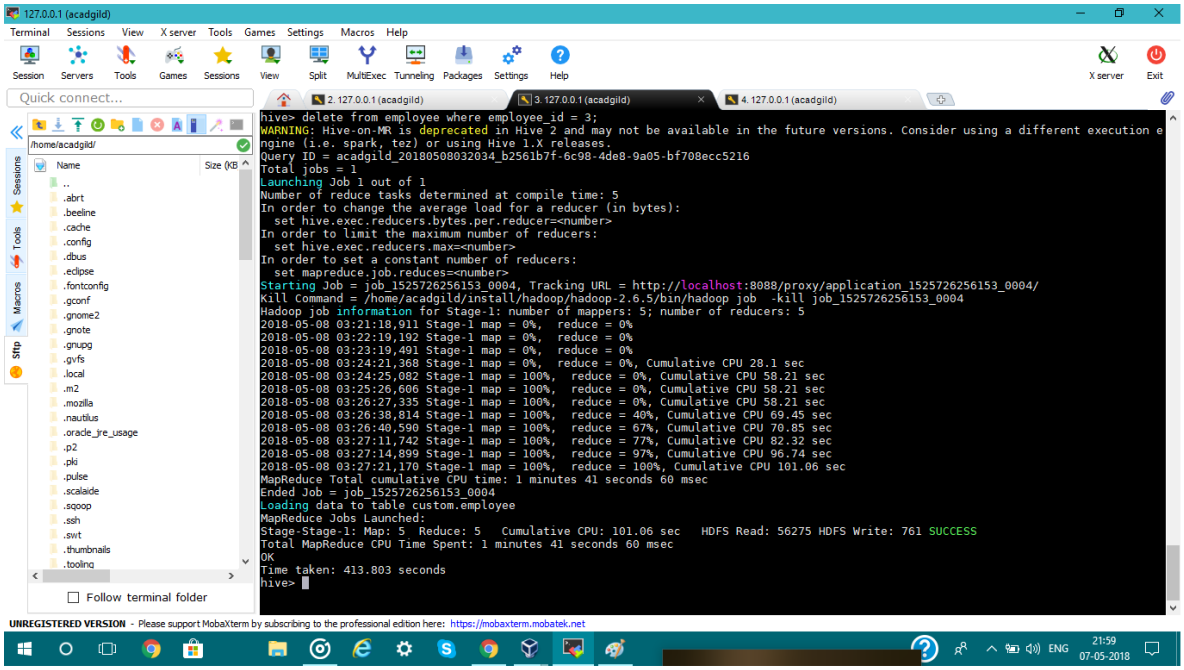


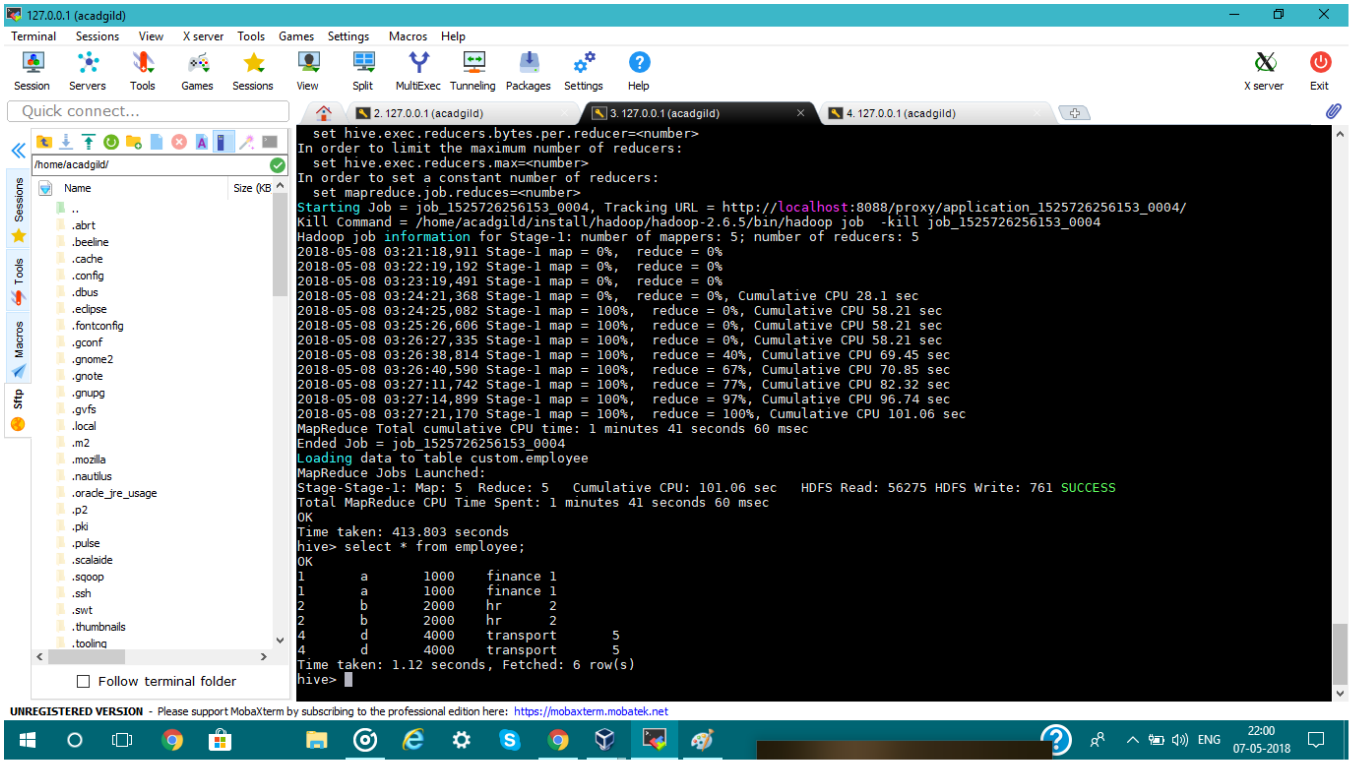
e. Updating a column (bucketed and non-bucketed)





f. Deleting a row in Hive table




127.0.0.1 (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. 127.0.0.1 (acadgild)
3. 127.0.0.1 (acadgild)
4. 127.0.0.1 (acadgild)
/home/acadgild/
Name Size (KB
Follow terminal folder
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_1525726256153_0004, Tracking URL = http://localhost:8088/proxy/application_1525726256153_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525726256153_0004
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-05-08 03:21:18,911 Stage-1 map = 0%, reduce = 0%
2018-05-08 03:22:19,192 Stage-1 map = 0%, reduce = 0%
2018-05-08 03:23:19,491 Stage-1 map = 0%, reduce = 0%
2018-05-08 03:24:21,368 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 28.1 sec
2018-05-08 03:24:25,082 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 58.21 sec
2018-05-08 03:25:26,606 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 58.21 sec
2018-05-08 03:26:27,335 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 58.21 sec
2018-05-08 03:26:38,814 Stage-1 map = 100%, reduce = 40%, Cumulative CPU 69.45 sec
2018-05-08 03:26:40,590 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 70.85 sec
2018-05-08 03:27:11,742 Stage-1 map = 100%, reduce = 77%, Cumulative CPU 82.32 sec
2018-05-08 03:27:14,899 Stage-1 map = 100%, reduce = 97%, Cumulative CPU 96.74 sec
2018-05-08 03:27:21,170 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 101.06 sec
MapReduce Total cumulative CPU time: 1 minutes 41 seconds 60 msec
Ended Job = job_1525726256153_0004
Loading data to table custom.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 101.06 sec HDFS Read: 56275 HDFS Write: 761 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 41 seconds 60 msec
OK
Time taken: 413.803 seconds
hive> select * from employee;
OK
1 a 1000 finance 1
1 a 1000 finance 1
2 b 2000 hr 2
2 b 2000 hr 2
4 d 4000 transport 5
4 d 4000 transport 5
Time taken: 1.12 seconds, Fetched: 6 row(s)
hive>
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
ENG
22:00
07-05-2018