# Big Data Hadoop Training

Session 15 Assignment 1 Solution:

**Calculate the number of employees corresponding to each skill from the table 'employee' which is loaded in the Demo.**

1. I have placed the dataset in HDFS @ /tmp/hive/BDDS15A1 directory.

Now, create a table employee inside a database say, BDDS15A1.

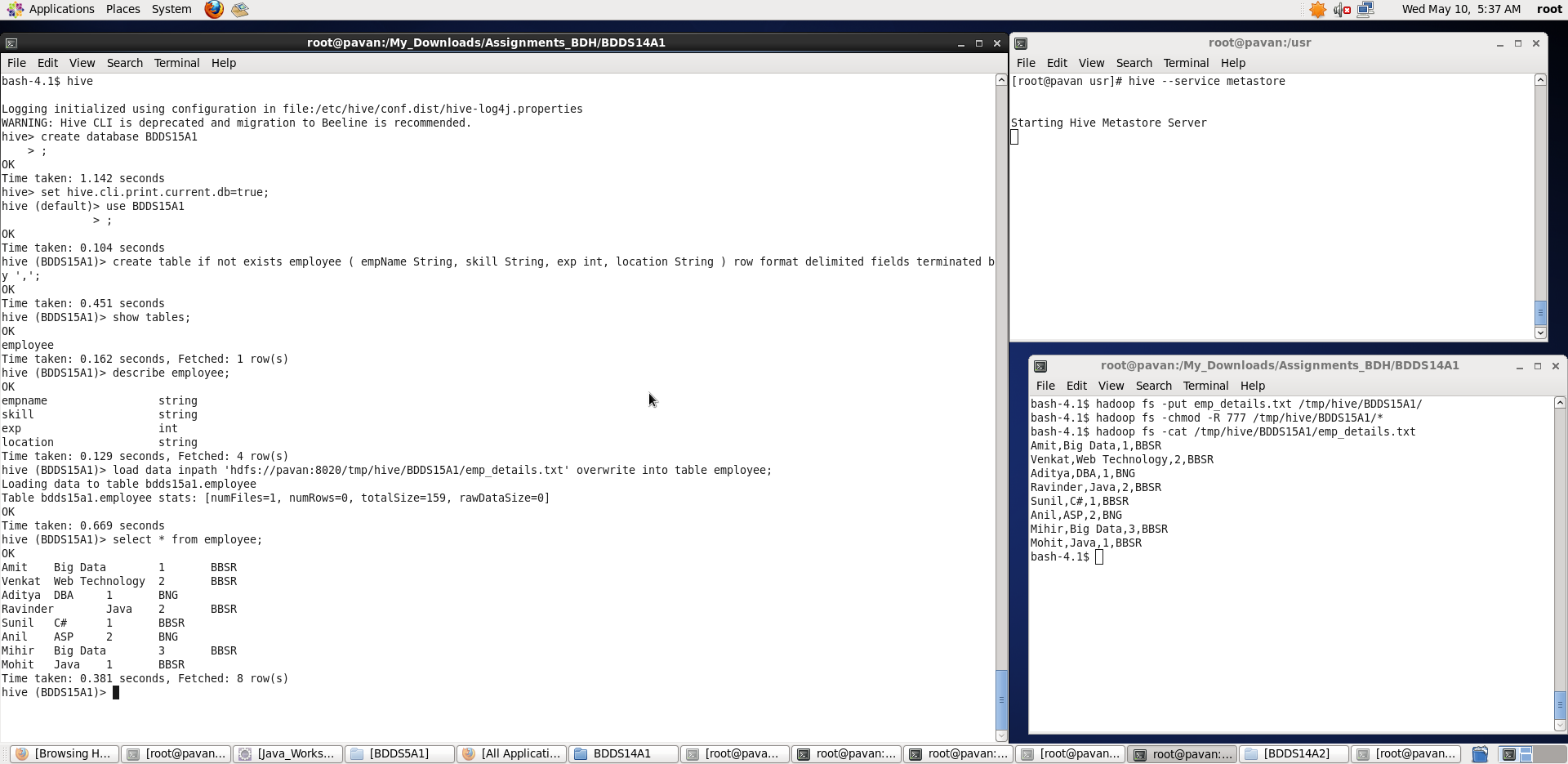
**hive> create table if not exists employee ( empName String, skill String, exp int, location String ) row format delimited fields terminated by ‘,’;**

We can see a table being created. Use **show** and **describe** commands to see the table.

Now, load the dataset into the table:

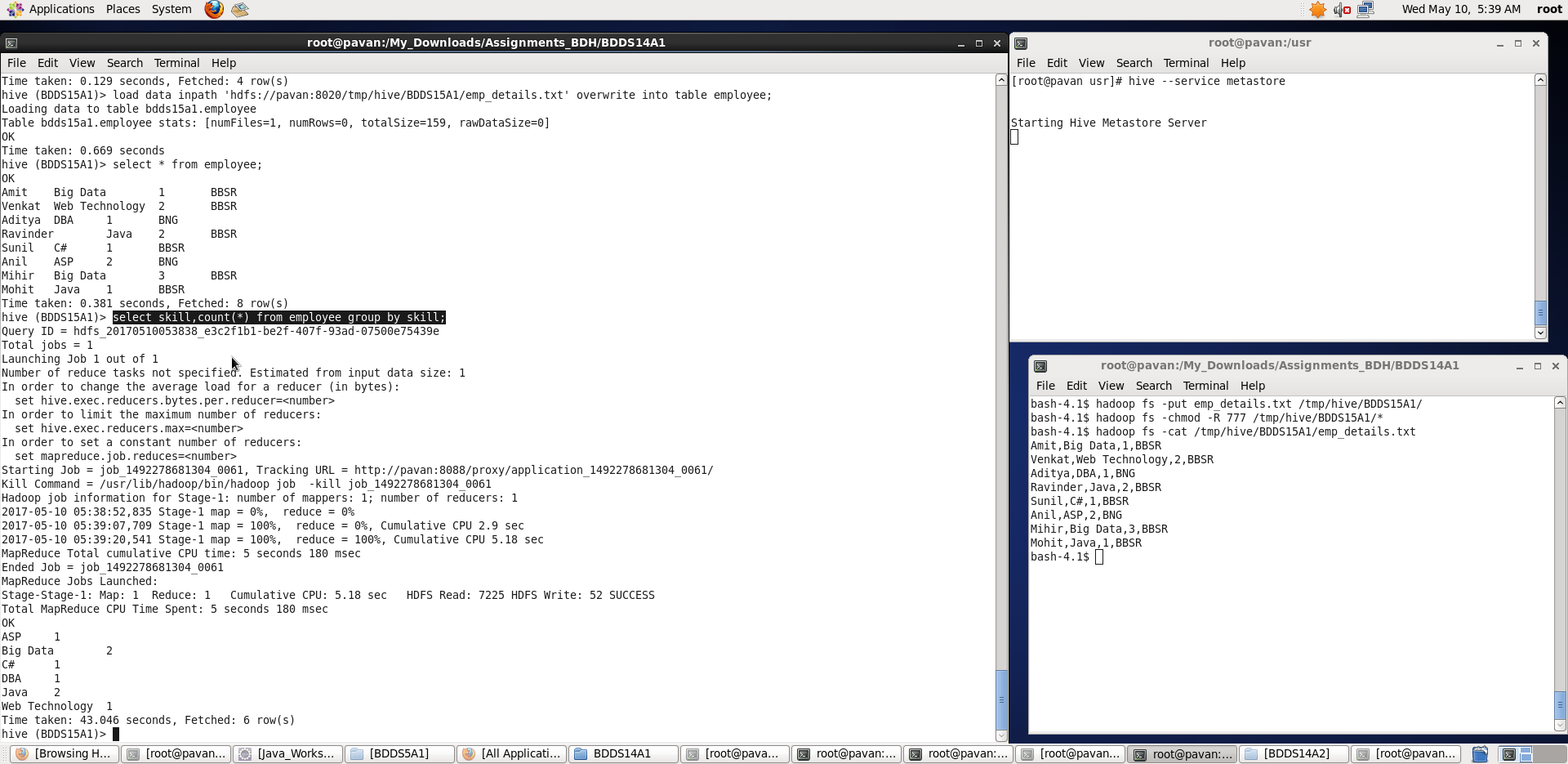
**hive> load data inpath ‘hdfs://pavan:8020/tmp/hive/BDDS15A1/emp\_details.txt’ overwrite into table employee;**

**hive> select \* from employee; -- Used to view the entire contents of the table**

****

Once, we have successfully, loaded the data into table, we will group the data based on **skill** to calculate the number of employees corresponding to each skill and use **count()** function.

**hive>** select skill,count(\*) from employee group by skill;

****

We can see a MapReduce job being created and executed and finally, we found the number of employees corresponding to each skill.

Thus, we have performed required tasks on given dataset.