

## Problem Statement

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

## Solution:

Creating directory **Assignment7\_1** in directory **ankita**

Copying **emp\_details.txt** to directory **Assignment7\_1/employee**

Showing **employee** contents

```
[acadgild@localhost ankita]$ cd /home/acadgild/ankita/Assignment7_1
[acadgild@localhost Assignment7_1]$ cp /home/acadgild/Downloads/emp_details.txt /home/acadgild/ankita/Assignment7_1/employee
[acadgild@localhost Assignment7_1]$ cat employee;
Amit,Big Data,1,BBSR
Venkat,Web Technology,2,BBSR
Aditya,DBA,1,BNG
Ravinder,Java,2,BBSR
Sunil,C#,1,BBSR
Anil,ASP,2,BNG
Mihir,Big Data,3,BBSR
Mohit,Java,1,BBSR
```

Starting **Hive**

```
[acadgild@localhost Assignment7_1]$ sudo service mysqld start
Starting mysqld: [ OK ]
[acadgild@localhost Assignment7_1]$ hive
```

Showing databases and Creating Database employee and using database employee and Creating Table employee

**hive> use employee**

**hive>set hive.cli.print.current.db=true;**

```
hive> show databases;
OK
default
Time taken: 3.057 seconds, Fetched: 1 row(s)
hive> create database employee;
OK
Time taken: 0.432 seconds
hive> show databases;
OK
default
employee
Time taken: 0.173 seconds, Fetched: 2 row(s)
hive> use employee;
OK
Time taken: 0.252 seconds
hive> set hive.cli.print.current.db=true;
hive (employee)> CREATE TABLE if not exists employee
> (
>   ename string,
>   skill string,
>   rank int,
>   place string
> )
> ROW FORMAT DELIMITED fields terminated by ','
> lines terminated by '\n'
> stored as textfile;
OK
Time taken: 1.931 seconds
```

Load the data from 'employee' into table employee.

```
hive (employee)> LOAD DATA LOCAL INPATH '/home/acadgild/ankita/Assignment7_1/employee' INTO table employee;
Loading data to table employee.employee
Table employee.employee stats: [numFiles=1, totalSize=159]
OK
Time taken: 5.278 seconds
```

Showing data of table employee

```
hive (employee)> select * from employee;
OK
Amit      Big Data      1      BBSR
Venkat    Web Technology 2      BBSR
Aditya    DBA           1      BNG
Ravinder  Java          2      BBSR
Sunil     C#            1      BBSR
Anil      ASP           2      BNG
Mihir     Big Data      3      BBSR
Mohit     Java          1      BBSR
Time taken: 1.596 seconds, Fetched: 8 row(s)
hive (employee)>
```

Calculating the number of employees corresponding to each skill from the table 'employee'

```
Acadgild_64bit [Running] - Oracle VM VirtualBox
Help
Mohit Java 1 BBSR
Time taken: 1.596 seconds, Fetched: 8 row(s)
hive (employee)> select skill, count(ename)
> from employee
> group by skill;
Query ID = acadgild_20171029140707_edf6356b-5627-40a9-bc61-53c5efc2a327
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
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>
Starting Job = job_1509251323495_0001, Tracking URL = http://localhost:8088/proxy/application_1509251323495_0001/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1509251323495_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-10-29 14:08:18,259 Stage-1 map = 0%, reduce = 0%
2017-10-29 14:09:07,928 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.3 sec
2017-10-29 14:09:52,112 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 7.86 sec
2017-10-29 14:10:02,024 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 12.55 sec
MapReduce Total cumulative CPU time: 12 seconds 550 msec
Ended Job = job_1509251323495_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 12.55 sec HDFS Read: 384 HDFS Write: 52 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 550 msec
OK
ASP      1
Big Data  2
C#        1
DBA       1
Java      2
Web Technology 1
Time taken: 183.561 seconds, Fetched: 6 row(s)
hive (employee)>
```