**Assignment 7.1**

**Problem Statement:**

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

**hive> create table if not exists employee**

**(**

**emp\_name string,**

**skill string,**

rating **int,**

**city string**

**)**

**row format delimited fields terminated by ',';**

OK

Time taken: 0.129 seconds

**hive> LOAD DATA LOCAL INPATH '/home/acadgild/hive/emp\_details.txt'INTO TABLE employee;**

Loading data to table custom.employee

Table custom.employee stats: [numFiles=1, totalSize=159]

OK

Time taken: 0.92 seconds

**hive> 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: 0.11 seconds, Fetched: 8 row(s)

**hive> SELECT skill,COUNT(emp\_name) FROM employee group by skill;**

Query ID = acadgild\_20170917194545\_afe00b29-fff1-4a20-86b4-0285478eaef1

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\_1505655198300\_0001, Tracking URL = http://localhost:8088/proxy/application\_1505655198300\_0001/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1505655198300\_0001

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-11-17 19:46:22,621 Stage-1 map = 0%, reduce = 0%

2017-11-17 19:46:42,135 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.51 sec

2017-11-17 19:47:03,111 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.47 sec

MapReduce Total cumulative CPU time: 5 seconds 470 msec

Ended Job = job\_1505655198300\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.47 sec HDFS Read: 389 HDFS Write: 52 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 470 msec

OK

**ASP 1**

**Big Data 2**

**C# 1**

**DBA 1**

**Java 2**

**Web Technology 1**

Time taken: 85.494 seconds, Fetched: 6 row(s)

hive>