Assignment 7.1

Problem Statement :-

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

Solution:-

```
hive> CREATE TABLE employee
    > (
    > name STRING,
    > skill STRING,
    > rating INT,
    > code STRING
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 0.802 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/emp_details.txt' INTO TABLE custom.employee;
Loading data to table custom.employee
OK
Time taken: 2.247 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: 0.277 seconds, Fetched: 8 row(s)
hive>
```

SELECT skill,count(*) FROM employee GROUP BY skill;

```
hive> SELECT skill,count(*) FROM employee GROUP BY skill;
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_20171109220513_95281069-e635-4ca9-828f-21aa9949101a
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_1510243495509_0001, Tracking URL = http://localhost:8088/proxy/application_1510243495509_0001/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510243495509_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-09 22:06:00,784 Stage-1 map = 0%,  reduce = 0%
2017-11-09 22:06:24,877 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.04 sec
2017-11-09 22:06:46,146 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.3 sec
MapReduce Total cumulative CPU time: 9 seconds 300 msec
Ended Job = job_1510243495509_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.3 sec   HDFS Read: 8485 HDFS Write: 211 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 300 msec
OK
ASP     1
Big Data        2
C#      1
DBA     1
Java    2
Web Technology  1
Time taken: 95.144 seconds, Fetched: 6 row(s)
hive>
```

Submitted By:-

Hardik Kaushik