Hive Oprations Assignment

Problem Statement

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

Below are steps followed to complete the assignment.

Steps:

1) First, we have to create a table named employee, with the fields corresponding to the data in the emp_details data file. The command used is as below.

```
CREATE TABLE employee
(
emp_name STRING,
skill STRING,
exp INT,
location STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';
```

Screenshot of Mobaxterm for creating the table employee:

```
hive> CREATE TABLE employee
    > (
    > emp_name STRING,
    > skill STRING,
    > exp INT,
    > location STRING
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 0.081 seconds
```

2) Next we have to load the contents of emp_details.txt. I stored the file in the path ‘/home/acadgild/hive/emp_details.txt'

   So we are loading the contents of the emp_details.txt into the table employee using the below commands.

```
LOAD DATA
LOCAL INPATH '/home/acadgild/hive/emp_details.txt'
INTO TABLE employee;
```

Screenshot of Mobaxterm for loading emp_details.txt into temperature_data:

```
hive> LOAD DATA
    > LOCAL INPATH '/home/acadgild/hive/emp_details.txt'
    > INTO TABLE employee;
Loading data to table default.employee
Table default.employee stats: [numFiles=1, totalSize=159]
OK
```

3) We can check the contents of employee table using the following command
   SELECT * from employee;

4) Now we can find the number of employees corresponding to each skill from the table 'employee' using the below command.

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

Output :

ASP 1
Big Data 2
C# 1
DBA 1
Java 2
Web Technology 1

Screenshot of Mobaxterm

```
hive> SELECT skill,count(*) from employee GROUP BY skill;
Query ID = acadgild_20171030133232_f9028243-5c9d-4a61-8710-ab58a67ffec1
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_1509349381134_0003, Tracking URL = http://localhost:8088/proxy/application_1509349381134_0003/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1509349381134_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-10-30 13:32:30,391 Stage-1 map = 0%,  reduce = 0%
2017-10-30 13:32:38,382 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.36 sec
2017-10-30 13:32:47,352 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.04 sec
MapReduce Total cumulative CPU time: 3 seconds 40 msec
Ended Job = job_1509349381134_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.04 sec   HDFS Read: 379 HDFS Write: 52 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 40 msec
OK
ASP     1
Big Data        2
C#      1
DBA     1
Java    2
Web Technology  1
Time taken: 29.816 seconds, Fetched: 6 row(s)
hive>
```