

Bigdata Assignment 3.1

Associated Data Files -

<https://drive.google.com/file/d/0Bxr27gVaXO5sQWV4UUpOXzNuZDA/view?usp=sharing>

Problem Statement

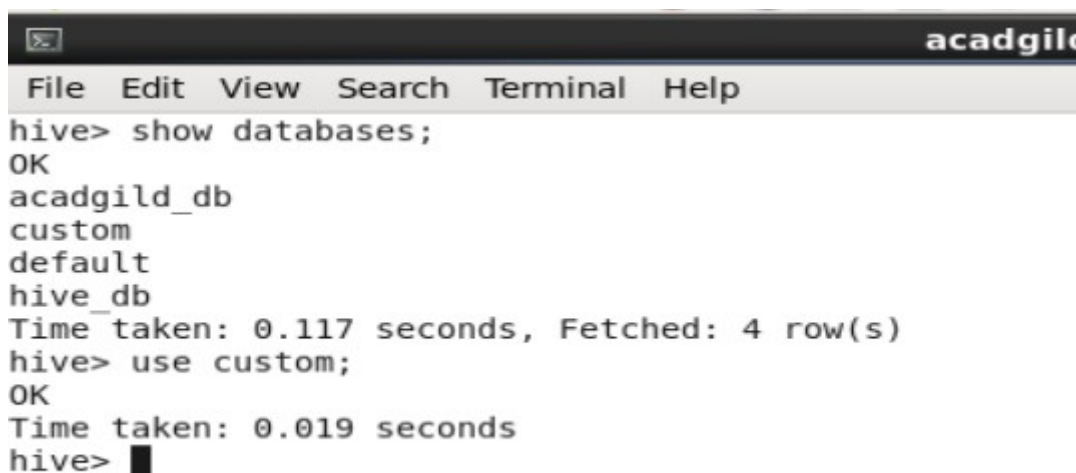
Calculate the number of employees corresponding to each skill from the table 'employee' which is loaded in the Demo.

Solution -

- Displayed the contents of the emp_details.txt file
cat emp_details.txt;

```
[acadgild@localhost HIVE]$ cat emp_details.txt
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
[acadgild@localhost HIVE]$ █
```

- Displayed the available databases and used the 'custom' database.
show databases;
use custom;



```
acadgild
File Edit View Search Terminal Help
hive> show databases;
OK
acadgild_db
custom
default
hive_db
Time taken: 0.117 seconds, Fetched: 4 row(s)
hive> use custom;
OK
Time taken: 0.019 seconds
hive> █
```

- Created the table

employee and loaded the data from local file system and displayed the contents of the table.

```
CREATE TABLE IF NOT EXISTS employee(name String , skill  
String, num int , location String) ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ',' ;  
LOAD DATA LOCAL INPATH 'emp_details.txt' into table employee ;  
select * from employee ;
```

```
acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
hive> CREATE TABLE IF NOT EXISTS employee(name String , skill String , num int , location String) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
OK
Time taken: 0.144 seconds
hive> LOAD DATA LOCAL INPATH 'emp_details.txt' into table employee;
Loading data to table custom.employee
OK
Time taken: 0.735 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: 0.152 seconds, Fetched: 8 row(s)
hive>
```

- Then for each skill , we counted no of employees using GROUP BY.

```
select skill , COUNT(*) from employee GROUP BY skill ;
```

```
acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
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_20180605194336_42eb0b4f-1893-4e67-b207-eac4a8b0dde6
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_1528171677641_0007, Tracking URL = http://localhost:8088/proxy/application_1528171677641_0007/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528171677641_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-05 19:43:44,931 Stage-1 map = 0%, reduce = 0%
2018-06-05 19:43:51,446 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.14 sec
2018-06-05 19:43:59,945 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.01 sec
MapReduce Total cumulative CPU time: 5 seconds 10 msec
Ended Job = job_1528171677641_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.01 sec HDFS Read: 8428 HDFS Write: 211 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 10 msec
OK
ASP      1
Big Data  2
C#        1
DBA       1
Java      2
Web Technology 1
Time taken: 24.306 seconds, Fetched: 6 row(s)
hive>
```