BIG DATA DEVELOPMENT

ACADGILD

Session 07: Hive Operations

Assignment 1

Big Data and Hadoop Development

Downloaded dataset from given link:

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.

hive> create table employee (name String, skill String, dept_id int, location String) row format delimited fields terminated by ',';

OK

Time taken: 11.173 seconds

hive> load data local inpath '/home/cloudera/Downloads/emp_details.txt' into table employee;

Loading data to table default.employee

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

OK

Time taken: 0.946 seconds

hive> select skill,count(name) from employee group by skill;

```
hive> create table employee(name String, skill String, dept_id int, location String) row format delimited fields terminated by ',';
OK
Time taken: 11.173 seconds
hive> load data local inpath '/home/cloudera/Downloads/emp_details.txt' into table employee;
Loading data to table default.employee
Table default.employee stats: [numFiles=1, totalSize=159]
OK
Time taken: 0.946 seconds
hive> select skill,count(name) from employee group by skill;
```

Output: -

ASP   1

Big Data   2

C#   1

DBA   1

Java   2

Web Technology 1

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.43 sec   HDFS Read: 7433 HDFS Write: 52 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 430 msec
OK
ASP     1
Big Data        2
C#      1
DBA     1
Java    2
Web Technology  1
Time taken: 27.525 seconds, Fetched: 6 row(s)
```