**Problem Objective**

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

Code/Script

-- create a database custom

create database custom;

-- entering to database custom

use custom;

-- create table employee

create table employee

(name string,skill string,no int,location String)

row format delimited

fields terminated by ',';

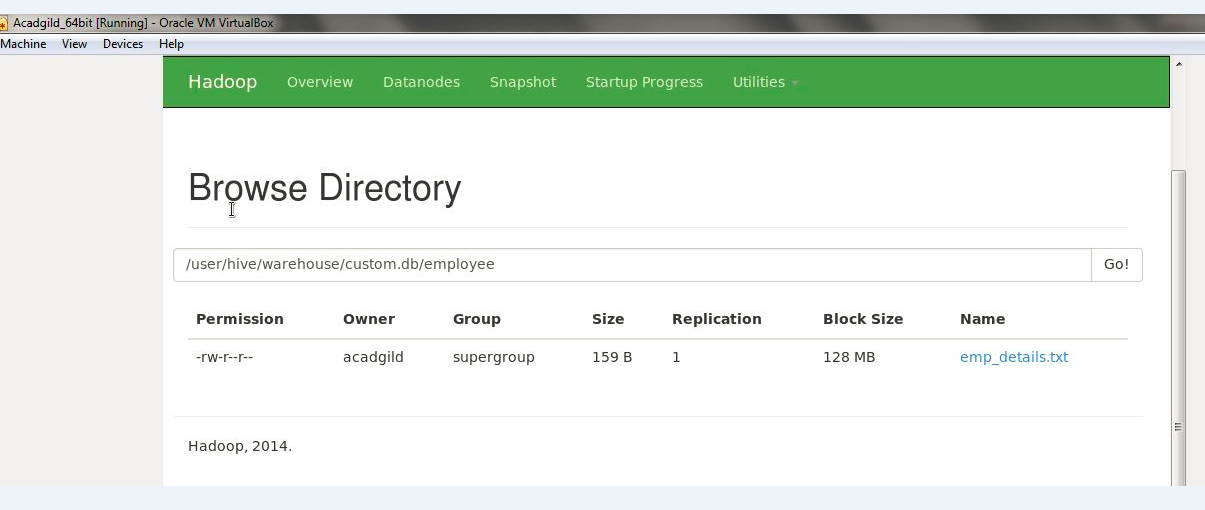
-- load data into employee

load data local inpath '/home/acadgild/Downloads/emp\_details.txt' into table employee;

-- count employees group by skills

select skill,count(1) from employee group by skill;

Output



Log/Steps

[acadgild@localhost ~]$hive> create table employee

> (name string,skill string,no int,location String)

> row format delimited

> fields terminated by ',';

OK

Time taken: 1.974 seconds

hive> load data local inpath '/home/acadgild/Downloads/emp\_details.txt' into table employee;

Loading data to table custom.employee

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

OK

Time taken: 7.434 seconds

hive> select skill,count(1) from employee group by skill;

Query ID = acadgild\_20170401233838\_fe330701-3eac-4aa7-a6f1-c3a06176a7c2

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\_1491017400247\_0017, Tracking URL = http://localhost:8088/proxy/application\_1491017400247\_0017/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1491017400247\_0017

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

2017-04-01 23:45:47,171 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:46:48,903 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:47:38,673 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.33 sec

2017-04-01 23:48:39,345 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.33 sec

2017-04-01 23:49:10,253 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 11.65 sec

2017-04-01 23:49:24,784 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 19.49 sec

MapReduce Total cumulative CPU time: 19 seconds 490 msec

Ended Job = job\_1491017400247\_0017

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 20.13 sec HDFS Read: 389 HDFS Write: 52 SUCCESS

Total MapReduce CPU Time Spent: 20 seconds 130 msec

OK

ASP 1

Big Data 2

C# 1

DBA 1

Java 2

Web Technology 1

Time taken: 665.467 seconds, Fetched: 6 row(s)

hive>