**Session 6: HIVE Operations**

Assignment 7.1

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 7.1**– Calculate the number of employees corresponding to each skill from the table 'employee' which is loaded in the Demo.

Contents

[Introduction 2](#_Toc497910743)

[Associated Data files 2](#_Toc497910744)

[Problem Statement 2](#_Toc497910745)

[HIVE Queries 2](#_Toc497910746)

[Task 4](#_Toc497910747)

[HIVE Query 4](#_Toc497910748)

[Expected Output 4](#_Toc497910749)

# Introduction

In this assignment, I’m going to create a table called **‘emp\_details’** with a table name called **‘employee’** which contains the columns name, technology, nos and the location. The given dataset file will be loaded in to the database ‘**emp\_details’** into the table employee.

The given dataset file is placed at **home/acadgild/hadoop.**

# 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.

# HIVE Queries

1. ***CREATE DATABASE emp\_details;***
2. ***USE emp\_details;***
3. ***CREATE TABLE employee***

***(name STRING,***

***technology STRING,***

***nos INT,***

***location STRING)***

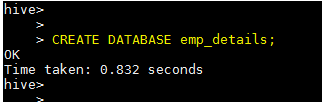
***ROW FORMAT DELIMITED***

***FIELDS TERMINATED BY ',';***

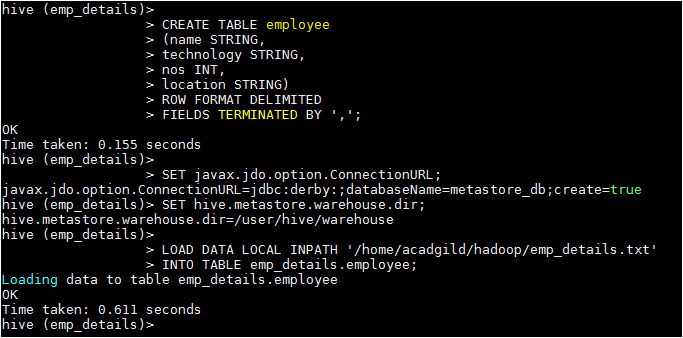
***LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/emp\_details.txt'***

***INTO TABLE emp\_details.employee;***

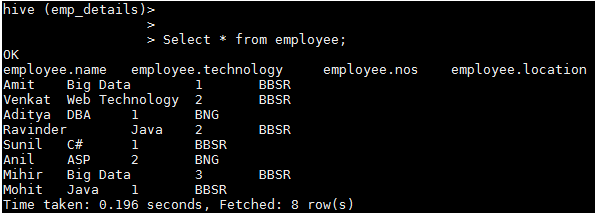
1. Creating a Database named **emp\_details**



1. Create a Table **employee** with the columns **name**, **technology** and **location** as STRING type and **nos** as integer type. We are taking the input data from the location '/home/acadgild/hadoop/emp\_details.txt'



1. Data in the table **employee**



# Task

Now, calculate the number of employees corresponding to each skill from the table **'employee'**

## HIVE Query

***Select technology, SUM(t1.nos) as No\_Of\_Employee FROM(Select technology, nos FROM employee) t1 GROUP BY technology;***



# Expected Output

