**Session 8: ADVANCED HIVE**

Assignment 8.1

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 8.1**–

Get a list of employees who receive a salary less than 100, compared to their immediate employee with higher salary in the same unit

List of all employees who draw higher salary than the average salary of that department

Contents

[Introduction 2](#_Toc499152752)

[Problem Statement 2](#_Toc499152753)

[Dataset 2](#_Toc499152754)

[Prerequisite – Create Database and Table 2](#_Toc499152755)

[Table 3](#_Toc499152756)

[HIVE QL 3](#_Toc499152757)

[Task 1 4](#_Toc499152758)

[HIVE QL 4](#_Toc499152759)

[Required Output 5](#_Toc499152760)

[Task 2 5](#_Toc499152761)

[HIVE QL 5](#_Toc499152762)

[Required Output 5](#_Toc499152763)

# Introduction

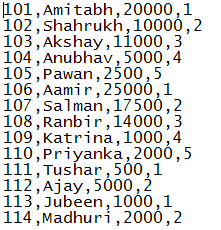
In this assignment, I’m going to write HIVE QL to achieve the task,

# Problem Statement

Get a list of employees who receive a salary less than 100, compared to their immediate employee with higher salary in the same unit

List of all employees who draw higher salary than the average salary of that department

# Dataset



# Prerequisite – Create Database and Table

Using existing database **emp\_details,**

## Table –

We are creating a table name called **emp** and we have columns as **emp\_id, emp\_name, sal** and **dept.**

### HIVE QL

***CREATE TABLE emp***

***(***

***Emp\_id int,***

***Emp\_name string,***

***Sal int,***

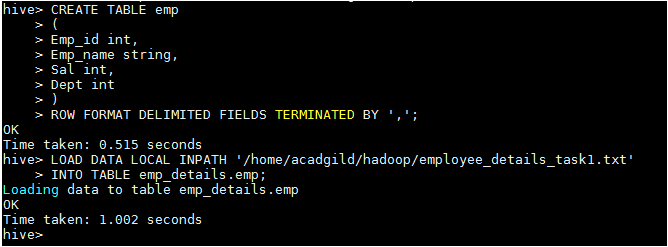
***Dept int***

***)***

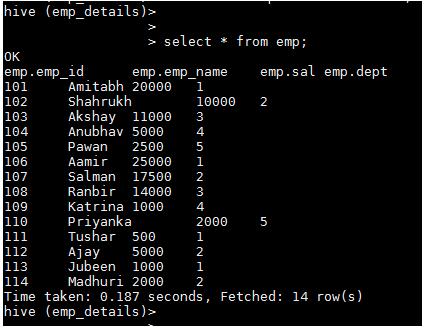
***ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';***

***LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/employee\_details\_task1.txt'***

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



***Select \* From emp;***

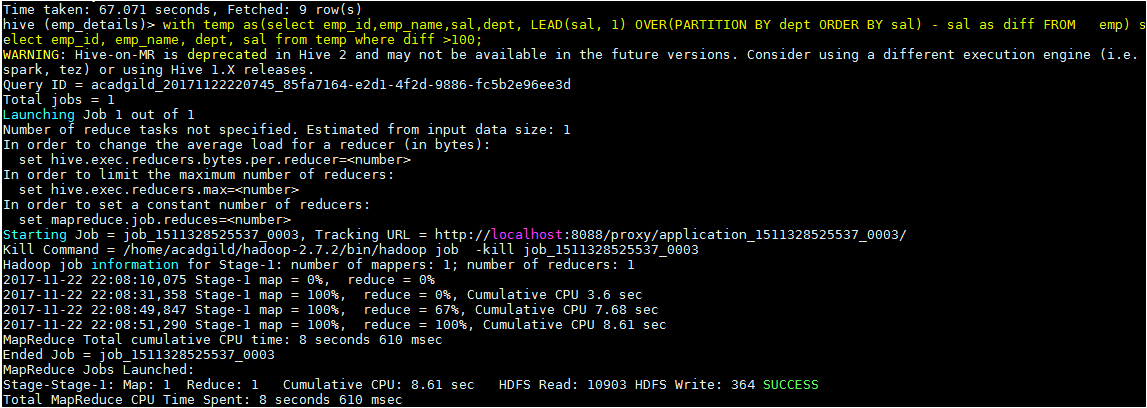


# Task 1

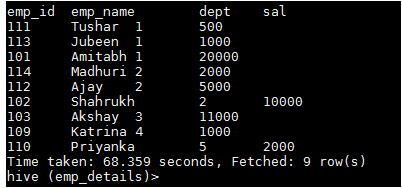
Get a list of employees who receive a salary less than 100, compared to their immediate employee with higher salary in the same unit

### HIVE QL

***with temp as(select emp\_id,emp\_name,sal,dept, LEAD(sal, 1) OVER(PARTITION BY dept ORDER BY sal) - sal as diff FROM emp) select emp\_id, emp\_name, dept, sal from temp where diff >100;***



### Required Output

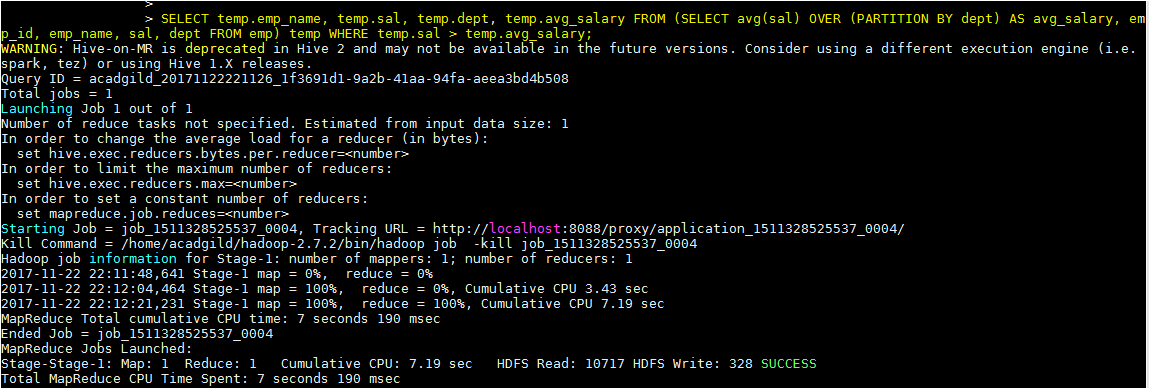


# Task 2

List of all employees who draw higher salary than the average salary of that department

### HIVE QL

***SELECT temp.emp\_name, temp.sal, temp.dept, temp.avg\_salary FROM (SELECT avg(sal) OVER (PARTITION BY dept) AS avg\_salary, emp\_id, emp\_name, sal, dept FROM emp) temp WHERE temp.sal > temp.avg\_salary;***



### Required Output

