Session​ ​ 08:Advanced​ ​ Hive

Assignment​ ​ 1

Big​ ​ Data​ ​ and​ ​ Hadoop​ ​ Development

Table​ ​ of​ ​ Contents

1.​ ​ Introduction​ ​ ..........................................................................................................................................3

2.​ ​ Objective​ ​ ...............................................................................................................................................3

3.​ ​ Prerequisites:​ ​ ........................................................................................................................................3

4.​ ​ Associated​ ​ Data​ ​ Files​ ​ ............................................................................................................................3

5.​ ​ Problem​ ​ Statement...............................................................................................................................3

6.​ ​ Expected​ ​ Output...................................................................................................................................3

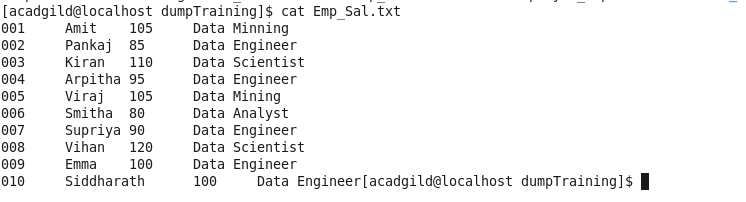
7.​ ​ Approximate​ ​ Time​ ​ to​ ​ Complete​ ​ Task​ ​ ...................................................................................................3

**1. Introduction**

This​ ​ assignment​ ​ will​ ​ help​ ​ you​ ​ to​ ​ consolidate​ ​ the​ ​ concepts​ ​ learnt​ ​ in​ ​ the​ ​ session.

**2. Problem**​ ​ **Solution**

**Dataset**​ ​ **:**​ ​​ **Emp\_Sal**

****

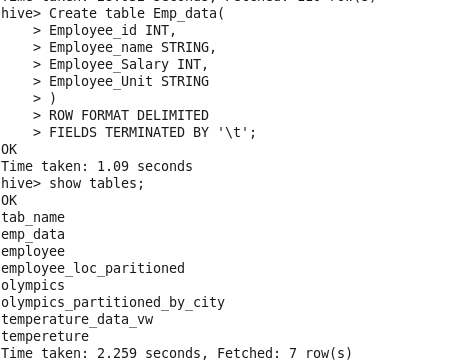
**Dataset**​ ​ **Details**

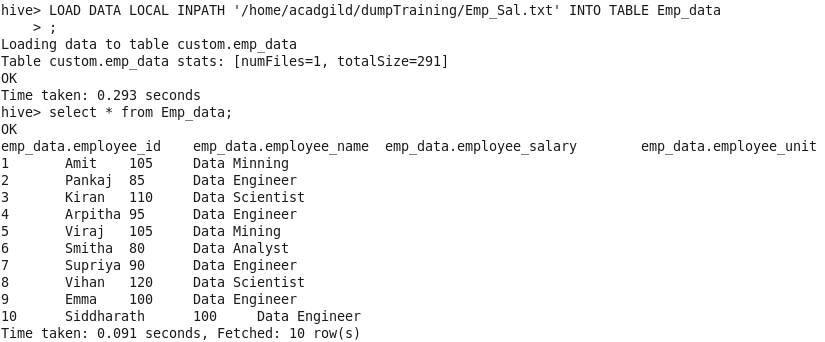
**Column1:**​ ​​ Employee\_Id

**Column2:**​ ​​ Employee\_Name

**Column3:**​ ​​ Employee\_Salary

**Column4:**​ ​​ Employee\_Unit​ ​ (Department\_Name)





● Get​ ​ the​ ​ list​ ​ of​ ​ employees​ ​ who​ ​ receive​ ​ a​ ​ salary​ ​ less​ ​ than​ ​ 100,​ ​ compared​ ​ to​ ​ their

immediate​ ​ employee​ ​ with​ ​ higher​ ​ salary​ ​ in​ ​ the​ ​ same​ ​ unit

SELECT

Employee\_name,Employee\_Salary,LEAD(Employee\_Salary)

OVER

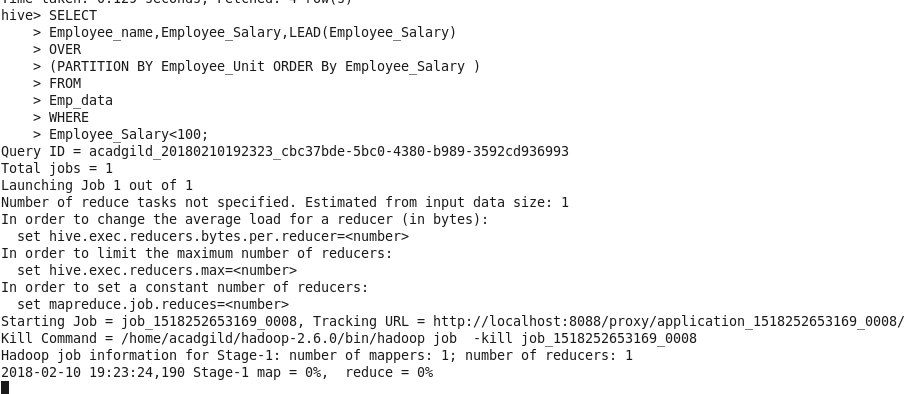
(PARTITION BY Employee\_Unit ORDER By Employee\_Salary )

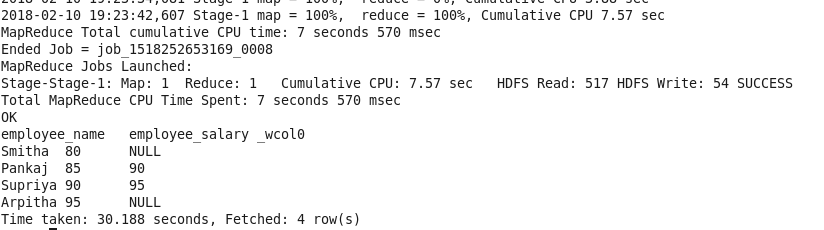
FROM

Emp\_data

WHERE

Employee\_Salary<100;





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

SELECT

Employee\_id,Employee\_name,Employee\_Salary,Employee\_Unit ,AVGSALARY

From

(SELECT Employee\_id, Employee\_name,Employee\_Unit,Employee\_Salary,AVG(Employee\_Salary) OVER (PARTITION BY Employee\_Unit) AS AVGSALARY

FROM Emp\_data) temp WHERE Employee\_Salary > AVGSALARY ;