Session 08: Advanced Hive

Assignment 1

**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

**Solution—**

Loading the data file

1 Amit 100 DNA

2 Sumit 200 DNA

3 Yadav 300 DNA

4 Sunil 500 FCS

5 Kranti 100 FCS

6 Mahoor 200 FCS

Creating table :-



**Hive>** CREATE TABLE emp\_det

(

id INT,

name STRING,

salary INT,

unit STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t';

Loading data into table:-

**hive>** load data local inpath '/home/acadgild/Acadgild/Assignment8\_1/emp\_dataset.txt'

into table emp\_det;



Now starting with the problem statements:-

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

**Logic:** Finding next higher salary in a unit using LEAD function. Then using this data to find difference between salary and LEAD(salary)

**Hive>** select a.id, a.name, a.salary, a.unit from

(SELECT id, name, salary, unit, LEAD(salary) OVER(PARTITION BY unit ORDER BY salary) as leadsal

FROM emp\_det) a

where a.leadsal-a.salary <=100

and a.leadsal is not null;

**Output-**

1 Amit 100 DNA

2 Sumit 200 DNA

5 Kranti 100 FCS

**Screenshot:-**





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

**Logic:** Finding avg salary of each unit. Then joining this table with emp\_det to find employees with salary > than unit’s avg salary.

**hive>** set hive.auto.convert.join.noconditionaltask =false;

**hive>** select a.id, a.name, a.salary, a.unit from emp\_det a join

(select unit, avg(salary) as avgsal from emp\_det group by unit) b

on (a.unit=b.unit)

where a.salary > b.avgsal;

**Output –**

3 Yadav 300 DNA

4 Sunil 500 FCS

**Screenshot:-**



