Bigdata Assignment 3.4

- Displayed the contents of the file – **cat Employee_Salary.txt ;**

- Created a tablw with the required column and loaded the data from file into the table.

  CREATE TABLE IF NOT EXISTS employee_salary(id int , name String , salary int , dept String) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' ;
  LOAD DATA LOCAL INPATH 'Employee_Salary.txt' into table employee_salary;

- Displayed the data of the table.
  select * from employee_salary;

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.

Ans -
Each employee record is compared woth immediate employee having higher salary then it is selected if salary diff is less than 100.

WITH temp as (select id,name,salary,dept,LEAD(salary,1) OVER (PARTITION BY dept ORDER BY salary) – salary AS salary_diff FROM employee_salary) select * from temp where salary_diff < 100;


```
acadgild@localhost:~/INPUT/HIVE                              _ □ ×
File  Edit  View  Search  Terminal  Help
hive> WITH temp as (select id,name,salary,dept,LEAD(salary,1) OVER (PARTITION BY dept ORDER BY salary) - salary AS salary_dif
f FROM employee_salary) select * from temp where salary_diff < 100;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180606163502_f533429a-d5b8-467f-af47-2ea5d09fde43
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_1528279835710_0004, Tracking URL = http://localhost:8088/proxy/application_1528279835710_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1528279835710_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-06 16:35:10,738 Stage-1 map = 0%,  reduce = 0%
2018-06-06 16:35:17,183 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.82 sec
2018-06-06 16:35:25,710 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.05 sec
MapReduce Total cumulative CPU time: 6 seconds 50 msec
Ended Job = job_1528279835710_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.05 sec   HDFS Read: 11166 HDFS Write: 87 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 50 msec
OK
Time taken: 24.705 seconds
hive>
```
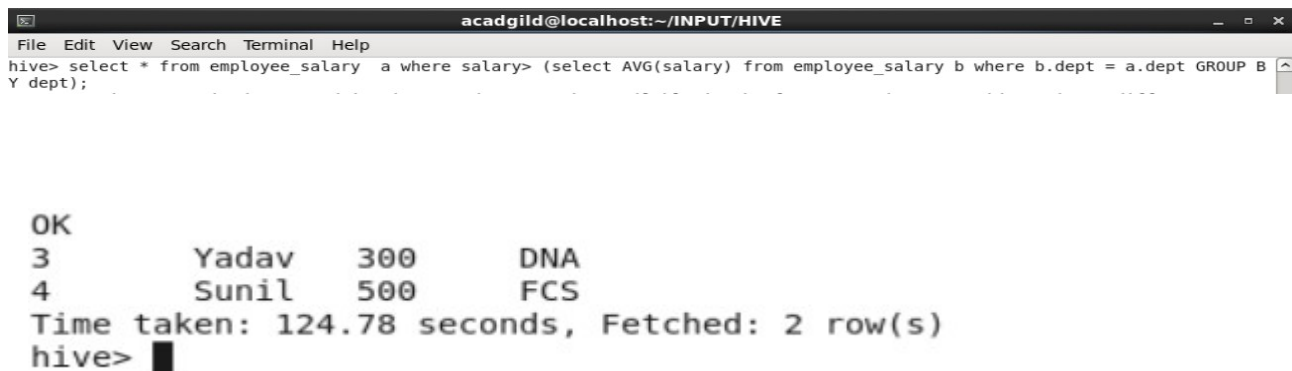
2. List of all employees who draw higher salary than the average salary of that department

Ans - seleted the employees whose salary is less than the salary of grouped employees salary of same department.

select * from employee_salary a where salary> (select AVG(salary) from employee_salary b where b.dept = a.dept GROUP BY dept) ;

```
                        acadgild@localhost:~/INPUT/HIVE                    _ □ ×
File  Edit  View  Search  Terminal  Help
hive> select * from employee_salary  a where salary> (select AVG(salary) from employee_salary b where b.dept = a.dept GROUP B ^
Y dept);


OK
3          Yadav    300        DNA
4          Sunil    500        FCS
Time taken: 124.78 seconds, Fetched: 2 row(s)
hive> █
```