

## Assignment 8.1

Created the data **employee.txt** and loaded into **/user/acadgild/test**

```
[acadgild@localhost ~]$ cat >employee.txt
101,Amitabh,256,Finance,1
102,Shahrukh,78,IT_Dept,2
103,Akshay,110,HR,3
104,Anubhav,50,Network_Team,4
105,Pawan,250,Admin,5
106,Aamir,25,Finance,1
107,Salman,175,IT_Dept,2
108,Ranbir,142,HR,3
109,Katrina,100,Network_Team,4
110,Priyanka,222,Admin,5
111,Tushar,500,Finance,1
112,Ajay,58,IT_Dept,2
113,Jubeen,100,Finance,1
114,Madhuri,200,IT_Dept,2
```

Created a table **employee** and loaded the data **employee.txt** into it.

```
hive> create table employee
> (
>   Emp_Id INT,
>   Emp_Name STRING,
>   Salary INT,
>   Department STRING,
>   Unit INT
> )
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
OK
Time taken: 0.177 seconds
hive> LOAD DATA INPATH '/user/acadgild/test/employee.txt' OVERWRITE INTO TABLE employee;
Loading data to table sumona.employee
Table sumona.employee stats: [numFiles=1, numRows=0, totalSize=346, rawDataSize=0]
OK
Time taken: 0.818 seconds
hive> █
```

1. List of employee's who receive a salary less than 100 compared to their immediate employee with higher salary in the same unit.

```

hive> SELECT Emp_Name,Unit,Salary FROM employee WHERE Salary<100 GROUP BY Emp_Name,Unit,Salary;
Query ID = acadgild_20171111155151_c5821486-4431-4ade-84e6-7b30ebate8e0
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_1510393157006_0005, Tracking URL = http://localhost:8088/proxy/application_1510393157006_0005/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1510393157006_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-11 15:51:27,355 Stage-1 map = 0%, reduce = 0%
2017-11-11 15:51:39,814 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.19 sec
2017-11-11 15:51:53,459 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.71 sec
MapReduce Total cumulative CPU time: 5 seconds 710 msec
Ended Job = job_1510393157006_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.71 sec HDFS Read: 573 HDFS Write: 48 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 710 msec
OK
Aamir 1 25
Ajay 2 58
Anubhav 4 50
Shahrukh 2 78
Time taken: 39.493 seconds, Fetched: 4 row(s)
hive>

```

OUTPUT

- List of employee's who draw higher salary than the average salary of the department.

```

hive> SELECT e.* FROM (SELECT e.*, avg(Salary) OVER (PARTITION BY Unit) AS avgsalary FROM employee e) e WHERE e.Salary> e.avgsalary;
Query ID = acadgild_20171111204646_703bdd52-a095-45a2-8fa9-7084549d1a5d
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_1510393157006_0013, Tracking URL = http://localhost:8088/proxy/application_1510393157006_0013/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1510393157006_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-11 20:46:44,252 Stage-1 map = 0%, reduce = 0%
2017-11-11 20:46:54,508 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.44 sec
2017-11-11 20:47:09,419 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.47 sec
MapReduce Total cumulative CPU time: 5 seconds 470 msec
Ended Job = job_1510393157006_0013
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.47 sec HDFS Read: 573 HDFS Write: 220 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 470 msec
OK
101 Amitabh 256 Finance 1 220.25
111 Tushar 500 Finance 1 220.25
114 Madhuri 200 IT_Dept 2 127.75
107 Salman 175 IT_Dept 2 127.75
108 Ranbir 142 HR 3 126.0
109 Katrina 100 Network Team 4 75.0
105 Pawan 250 Admin 5 236.0
Time taken: 38.642 seconds, Fetched: 7 row(s)

```

OUTPUT