

# Assignment 8.1

## Problem Statement 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

Solution:-

```
hive> CREATE TABLE employee_details
(
  emp_id INT,
  emp_name STRING,
  salary INT,
  department STRING,
  unit INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';
OK
Time taken: 1.184 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/employee_details_8.1.txt' INTO TABLE custom.employee_details;
Loading data to table custom.employee_details
OK
Time taken: 2.371 seconds
hive> select * from employee_details;
OK
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
Time taken: 0.326 seconds, Fetched: 14 row(s)
hive>
```

SELECT emp\_name,department,salary FROM employee\_details where salary < 100 GROUP BY emp\_name,department,salary;

```
hive> SELECT emp_name,department,salary FROM employee_details where salary < 100 GROUP BY emp_name,department,salary;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez)
or using Hive 1.X releases.
Query ID = acadgild_20171111220906_1dddf31b-2703-44f7-8e5f-987560b56864
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_1510407068250_0003, Tracking URL = http://localhost:8088/proxy/application_1510407068250_0003/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job_1510407068250_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-11 22:09:37,874 Stage-1 map = 0%, reduce = 0%
2017-11-11 22:09:59,953 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.73 sec
2017-11-11 22:10:21,700 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.29 sec
MapReduce Total cumulative CPU time: 8 seconds 290 msec
Ended Job = job_1510407068250_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.29 sec HDFS Read: 9749 HDFS Write: 212 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 290 msec
OK
Aamir Finance 25
Ajay IT_Dept 58
Anubhav Network_Team 50
Shahrukh IT_Dept 78
Time taken: 76.566 seconds, Fetched: 4 row(s)
hive>
```

## Problem Statement 2 :-

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

Solution:-

Suppose we want list of all employees who draw higher salary than average salary of Finance department then we ran below command :-

`select avg(salary) from employee_details where department = 'Finance';`

```
hive> select avg(salary) from employee_details where department = 'Finance';
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez)
or using Hive 1.X releases.
Query ID = acadgild_20171111222127_alld41447-da4c-4449-8ec5-ded7c7f82f8c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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_1510407068250_0004, Tracking URL = http://localhost:8088/proxy/application_1510407068250_0004/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job_1510407068250_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-11 22:21:58,997 Stage-1 map = 0%, reduce = 0%
2017-11-11 22:22:20,368 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.75 sec
2017-11-11 22:22:41,177 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.3 sec
MapReduce Total cumulative CPU time: 8 seconds 300 msec
Ended Job = job_1510407068250_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.3 sec HDFS Read: 10134 HDFS Write: 106 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 300 msec
OK
220.25
Time taken: 76.054 seconds, Fetched: 1 row(s)
hive>
```

`select * from employee_details where department = 'Finance' AND salary > 220.25;`

```
hive> select * from employee_details where department = 'Finance' AND salary > 220.25;
OK
101      Amitabh 256      Finance 1
111      Tushar  500      Finance 1
Time taken: 0.413 seconds, Fetched: 2 row(s)
hive>
```

Submitted By:-

Hardik Kaushik