

ASSIGNMENT 8.1

```
hive> create table users(id int,name string,salary int,department string)
> row format delimited fields terminated by '\t';
```

OK

Time taken: 1.254 seconds

```
[acadgild@localhost ~]$ hadoop fs -put /home/acadgild/vishnu/users.txt /vishnu
```

```
hive> load data
```

```
> inpath '/vishnu/users.txt'
```

```
> into table users;
```

Loading data to table default.users

Table default.users stats: [numFiles=1, totalSize=90]

OK

Time taken: 1.586 seconds

```
hive> select * from users
```

```
> ;
```

OK

1	Arju	100	DB
2	Sumi	200	DB
3	Vishnu	300	DB
4	Vipin	500	CS
5	Kiran	100	CS
6	Mithun	200	CS

Time taken: 0.184 seconds, Fetched: 6 row(s)

```
[acadgild@localhost ~]$ hadoop fs -put /home/acadgild/vishnu/location.txt /vishnu
```

```
hive> create table locations(id int,location string)
> row format delimited fields terminated by '\t';
```

OK

Time taken: 0.165 seconds

```
hive> load data
```

```
> inpath
```

```
> '/vishnu/location.txt'
```

```
> into table locations;
```

Loading data to table default.locations

Table default.locations stats: [numFiles=1, totalSize=62]

OK

Time taken: 0.967 seconds

```
hive> select * from locations
```

```
> ;
```

OK

1	KERALA
2	BIHAR
3	THAMILNADU

```
4    KARNADAKA
5    MAHARASHTRA
6    GOA
```

Time taken: 0.15 seconds, Fetched: 6 row(s)

Problem1

Get a list of employees who receive a salary less than 100, compared to their immediate employee with higher salary in the same unit.

```
hive> select id,name,salary,department
      > from
      > ( select lead(salary) over(partition by department order by
salary) as next_salary,
      > id,name,department,salary
      > from users ) temp
      > where (next_salary - salary = 100);
```

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.85 sec HDFS

Read: 301 HDFS Write: 43 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 850 msec

OK

```
5    Kiran      100  CS
1    Arju 100   DB
2    Sumi 200   DB
```

Time taken: 63.578 seconds, Fetched: 3 row(s)

Problem 2

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

```
hive> select name,salary,department
      > from
      > (select avg(salary) over(partition by department order by
salary ) as avg_slary,name,department,salary
```

```
> from users ) temp

> where salary > avg_salary;
```

```
Query ID =
acadgild_20171102025555_0a4c33ca-af91-49d4-8a62-ec9a69e5ee99
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_1509563670870_0008, Tracking URL =
http://localhost:8088/proxy/application_1509563670870_0008/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill
job_1509563670870_0008
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2017-11-02 02:56:09,278 Stage-1 map = 0%,   reduce = 0%
2017-11-02 02:56:23,924 Stage-1 map = 100%,   reduce = 0%, Cumulative
CPU 2.11 sec
2017-11-02 02:56:41,104 Stage-1 map = 100%,   reduce = 100%,
Cumulative CPU 6.12 sec
MapReduce Total cumulative CPU time: 6 seconds 120 msec
Ended Job = job_1509563670870_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Reduce: 1   Cumulative CPU: 6.12 sec   HDFS
Read: 301 HDFS Write: 53 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 120 msec
OK
Mithun      200   CS
Vipin       500   CS
Sumi        200   DB
Vishnu      300   DB
Time taken: 52.857 seconds, Fetched: 4 row(s)
hive>
```