**ACD\_BDD2.3\_Session\_8\_Assignment\_1**

**Step 1:**

**hive> create table users(id int,name string,slary int,unit string)**

**> row format delimited fields**

**> terminated by '\t';**

OK

Time taken: 29.924 seconds

**Step 2:**

**hive> load data local inpath '/home/acadgild/hive\_folder/users.txt'**

**> into table users;**

Loading data to table default.users

Table default.users stats: [numFiles=1, totalSize=101]

OK

Time taken: 39.493 seconds

**Step3:**

**hive> select \* from users;**

OK

**1 Amit 100 DNA**

**2 Sumit 200 DNA**

**3 Yadav 300 DNA**

**4 Sunil 500 FCS**

**5 Kranti 100 FCS**

**6 Mahoor 200 FCS**

Time taken: 20.01 seconds, Fetched: 6 row(s)

**Step4:**

**hive> create table locations(id int,location string)**

**> row format delimited fields**

**> terminated by '\t';**

OK

Time taken: 1.457 seconds

**Step5:**

**hive> load data local inpath '/home/acadgild/hive\_folder/locations.txt'**

**> into table locations;**

Loading data to table default.locations

Table default.locations stats: [numFiles=1, totalSize=47]

OK

Time taken: 5.347 seconds

**hive> select \* from locations;**

**OK**

**1 UP**

**2 BIHAR**

**3 MP**

**4 AP**

**5 MAHARASHTRA**

**6 GOA**

Time taken: 1.099 seconds, Fetched: 6 row(s)

**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.**

**hive> select id,name,slary,unit**

**> from**

**> ( select lead(slary) over(partition by unit order by slary) as next\_salary,**

**> id,name,unit,slary**

**> from users ) temp**

**> where (next\_salary - slary = 100);**

Query ID = acadgild\_20170924181616\_3a380449-0f28-4e25-8c13-751800453ee0

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\_1506254688213\_0001, Tracking URL = http://localhost:8088/proxy/application\_1506254688213\_0001/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1506254688213\_0001

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-10-02 18:24:56,714 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:25:58,895 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:27:00,219 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:28:00,288 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:29:01,040 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:30:01,553 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 28.39 sec

2017-10-02 18:30:15,901 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 39.55 sec

2017-10-02 18:31:16,629 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 40.46 sec

2017-10-02 18:32:17,598 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 40.46 sec

2017-10-02 18:33:18,916 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 40.46 sec

2017-10-02 18:34:19,870 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 40.46 sec

2017-10-02 18:34:56,301 Stage-1 map = 100%, reduce = 36%, Cumulative CPU 55.38 sec

2017-10-02 18:34:59,495 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 57.53 sec

2017-10-02 18:36:01,380 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 59.91 sec

2017-10-02 18:36:42,899 Stage-1 map = 100%, reduce = 77%, Cumulative CPU 118.77 sec

2017-10-02 18:36:53,625 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 124.78 sec

MapReduce Total cumulative CPU time: 2 minutes 4 seconds 780 msec

Ended Job = job\_1506254688213\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 129.23 sec HDFS Read: 312 HDFS Write: 48 SUCCESS

Total MapReduce CPU Time Spent: 2 minutes 9 seconds 230 msec

**OK**

**1 Amit 100 DNA**

**2 Sumit 200 DNA**

**5 Kranti100 FCS**

Time taken: 1259.119 seconds, Fetched: 3 row(s)

**Problem Statement - List of all employees who draw higher salary than the average salary of that department.**

**hive> select name,slary,unit**

**> from**

**> (select avg(slary) over(partition by unit order by slary ) as avg\_slary,name,unit,slary**

**> from users ) temp**

**> where slary > avg\_slary;**

Query ID = acadgild\_20170924184343\_0ca3ddef-cb59-4e12-9112-8dd0e2e3a1ab

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\_1506254688213\_0002, Tracking URL = http://localhost:8088/proxy/application\_1506254688213\_0002/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1506254688213\_0002

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-10-02 18:48:33,300 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:49:35,191 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:50:36,619 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:51:37,873 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:52:38,856 Stage-1 map = 0%, reduce = 0%

2017-10-02 18:53:40,410 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 32.47 sec

2017-10-02 18:54:20,778 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 49.46 sec

2017-10-02 18:54:37,251 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 18:55:37,524 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 18:56:38,333 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 18:57:38,565 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 18:58:39,357 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 18:59:40,278 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:00:41,200 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:01:42,048 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:02:43,971 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:03:44,803 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:04:46,394 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:05:19,843 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 55.27 sec

2017-10-02 19:05:23,253 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:06:23,556 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:07:25,262 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:08:27,179 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:09:29,395 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:10:31,338 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:11:32,135 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:12:32,814 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:13:34,389 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 55.27 sec

2017-10-02 19:13:41,426 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 68.53 sec

2017-10-02 19:13:46,794 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 70.94 sec

2017-10-02 19:14:47,022 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 73.53 sec

2017-10-02 19:15:39,046 Stage-1 map = 100%, reduce = 88%, Cumulative CPU 140.33 sec

2017-10-02 19:15:53,309 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 147.44 sec

MapReduce Total cumulative CPU time: 2 minutes 27 seconds 440 msec

Ended Job = job\_1506254688213\_0002

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 151.65 sec HDFS Read: 312 HDFS Write: 57 SUCCESS

Total MapReduce CPU Time Spent: 2 minutes 31 seconds 650 msec

**OK**

**Sumit 200 DNA**

**Yadav 300 DNA**

**Mahoor200 FCS**

**Sunil 500 FCS**

Time taken: 1966.021 seconds, Fetched: 4 row(s)