

# Assignment 16.1

Employee table created

Create table employee\_table (emp\_name int, skill string, dep int, sal int, location string) row format delimited fields terminated by ',';

Load inpath '/home/acadgild/Downloads/emp\_details.txt' overwrite into table employee\_table;

```
Acadgild_64bit x
File Edit View VM Tabs Help
Acadgild_64bit
108 Ranbir 14000 3 NULL
109 Katrina 1000 4 NULL
110 Priyanka 2000 5 NULL
111 Tushar 500 1 NULL
112 Ajay 5000 2 NULL
113 Jubeen 1000 1 NULL
114 Madhuri 2000 2 NULL
Time taken: 0.633 seconds, Fetched: 14 row(s)
hive> load data local inpath '/home/acadgild/Downloads/emp_details.txt' overwrite into table employee_table;
Loading data to table acadgild.employee_table
Table acadgild.employee_table stats: [numFiles=1, numRows=0, totalSize=224, rawDataSize=0]
OK
Time taken: 0.895 seconds
hive> select * from employee_table;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'employee_tables'
hive> select * from employee_table;
OK
Amit Big Data 1 4000 NULL
Venkat Web Technology 2 375 NULL
Aditya DBA 1 2000 NULL
Ravinder Java 2 3950 NULL
Sunil C# 1 2250 NULL
Anil ASP 2 1800 NULL
Mihir Big Data 3 3900 NULL
Mohit Java 1 2500 NULL
Radha Big Data 2 1900 NULL
Time taken: 0.136 seconds, Fetched: 9 row(s)
hive> drop table employee_table;
OK
Time taken: 1.126 seconds
hive> create table employee_table(emp_name string, skill string, dep int, sal int, location string) row format delimited fields terminated by ',';
OK
Time taken: 0.414 seconds
hive>
```

Select \* from employee\_table;

```
Acadgild_64bit x
File Edit View VM Tabs Help
Acadgild_64bit
Amit Big Data 1 4000 NULL
Venkat Web Technology 2 375 NULL
Aditya DBA 1 2000 NULL
Ravinder Java 2 3950 NULL
Sunil C# 1 2250 NULL
Anil ASP 2 1800 NULL
Mihir Big Data 3 3900 NULL
Mohit Java 1 2500 NULL
Radha Big Data 2 1900 NULL
Time taken: 0.136 seconds, Fetched: 9 row(s)
hive> drop table employee_table;
OK
Time taken: 1.126 seconds
hive> create table employee_table(emp_name string, skill string, dep int, sal int, location string) row format delimited fields terminated by ',';
OK
Time taken: 0.414 seconds
hive> load data local inpath '/home/acadgild/Downloads/emp_details.txt' overwrite into table employee_table;
Loading data to table acadgild.employee_table
Table acadgild.employee_table stats: [numFiles=1, numRows=0, totalSize=224, rawDataSize=0]
OK
Time taken: 0.696 seconds
hive> select * from employee_table;
OK
Amit Big Data 1 4000 BBSR
Venkat Web Technology 2 375 BBSR
Aditya DBA 1 2000 BNG
Ravinder Java 2 3950 BBSR
Sunil C# 1 2250 BBSR
Anil ASP 2 1800 BNG
Mihir Big Data 3 3900 BBSR
Mohit Java 1 2500 BBSR
Radha Big Data 2 1900 BNG
Time taken: 0.138 seconds, Fetched: 9 row(s)
hive>
```

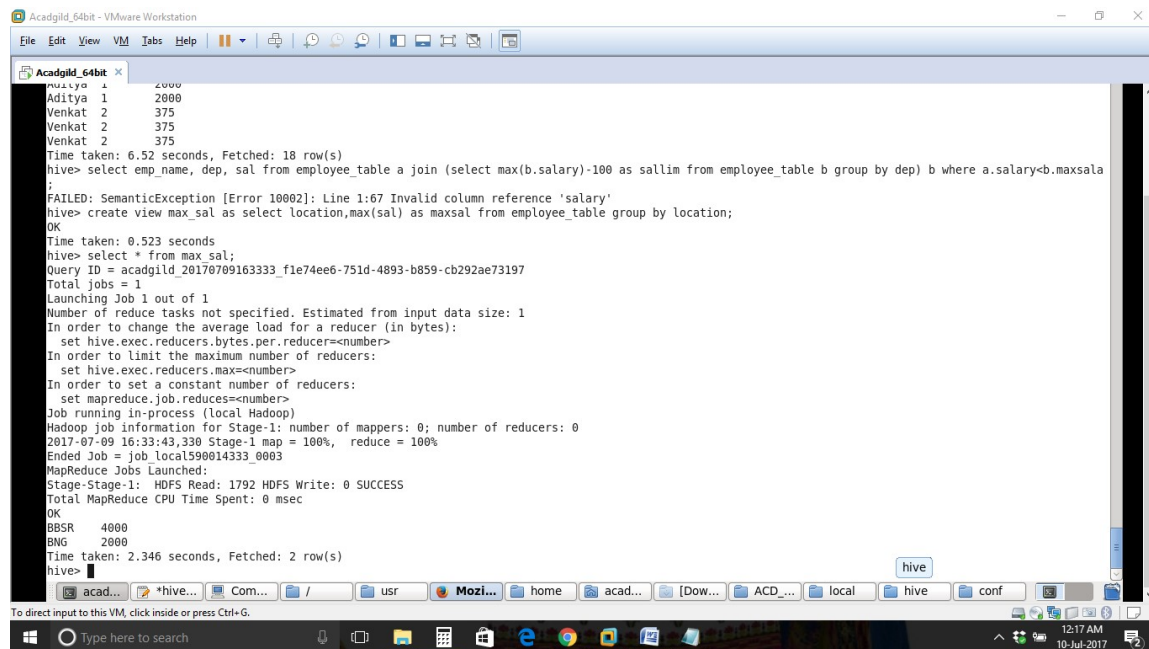
## Problem Statement 1:

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

View created with max salary in each location

Create view max\_sal as select location, max(sal) as maxsal from employee\_table group by location;

Select \* from max\_sal;



```
Acadgild_64bit x
File Edit View VM Tabs Help
Acadgild_64bit x
emp_name dep sal
Aditya 1 2000
Venkat 2 375
Venkat 2 375
Venkat 2 375
Time taken: 6.52 seconds, Fetched: 18 row(s)
hive> select emp_name, dep, sal from employee_table a join (select max(b.salary)-100 as sallim from employee_table b group by dep) b where a.salary<b.maxsala
;
FAILED: SemanticException [Error 10002]: Line 1:67 Invalid column reference 'salary'
hive> create view max_sal as select location,max(sal) as maxsal from employee_table group by location;
OK
Time taken: 0.523 seconds
hive> select * from max_sal;
Query ID = acadgild_20170709163333_f1e74ee6-751d-4893-b859-cb292ae73197
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>
Job running in-process (Local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2017-07-09 16:33:43,330 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local590014333_0003
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 1792 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
BBSR 4000
BNG 2000
Time taken: 2.346 seconds, Fetched: 2 row(s)
hive>
```

**Query to get details of employees who have salary difference is more than 100 than maximum salary in same location or unit**

Select a.empname,a.sal,b.maxsal from employee\_table a, max\_sal b where a.location = b.location and a.sal+100<b.maxsal;

```
Acadgild_64bit
Anil 1800
Time taken: 4.281 seconds, Fetched: 4 row(s)
hive> select a.emp name, a.sal,b.maxsal from employee_table a, max_sal b where a.location=b.location and a.sal+100<b.maxsal;
Query ID = acadgild_20170709163535_c5cc1344-0121-4c63-937c-187c956c16fa
Total jobs = 2
Launching Job 1 out of 2
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2017-07-09 16:35:23,760 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local206197249_0006
Launching Job 2 out of 2
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-2: number of mappers: 0; number of reducers: 0
2017-07-09 16:35:25,455 Stage-2 map = 100%, reduce = 100%
Ended Job = job_local1280228329_0007
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 3136 HDFS Write: 0 SUCCESS
Stage-Stage-2: HDFS Read: 5376 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
acac... *hive... Com... / usr Mozi... home acad... [Dow... ACD... local hive conf
```

```
Acadgild_64bit
Launching Job 1 out of 2
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2017-07-09 16:35:23,760 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local206197249_0006
Launching Job 2 out of 2
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-2: number of mappers: 0; number of reducers: 0
2017-07-09 16:35:25,455 Stage-2 map = 100%, reduce = 100%
Ended Job = job_local1280228329_0007
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 3136 HDFS Write: 0 SUCCESS
Stage-Stage-2: HDFS Read: 5376 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Mohit 2500 4000
Sunil 2250 4000
Venkat 375 4000
Anil 1800 2000
Time taken: 3.747 seconds, Fetched: 4 row(s)
hive>
acac... *hive... Com... / usr Mozi... home acad... [Dow... ACD... local hive conf
```

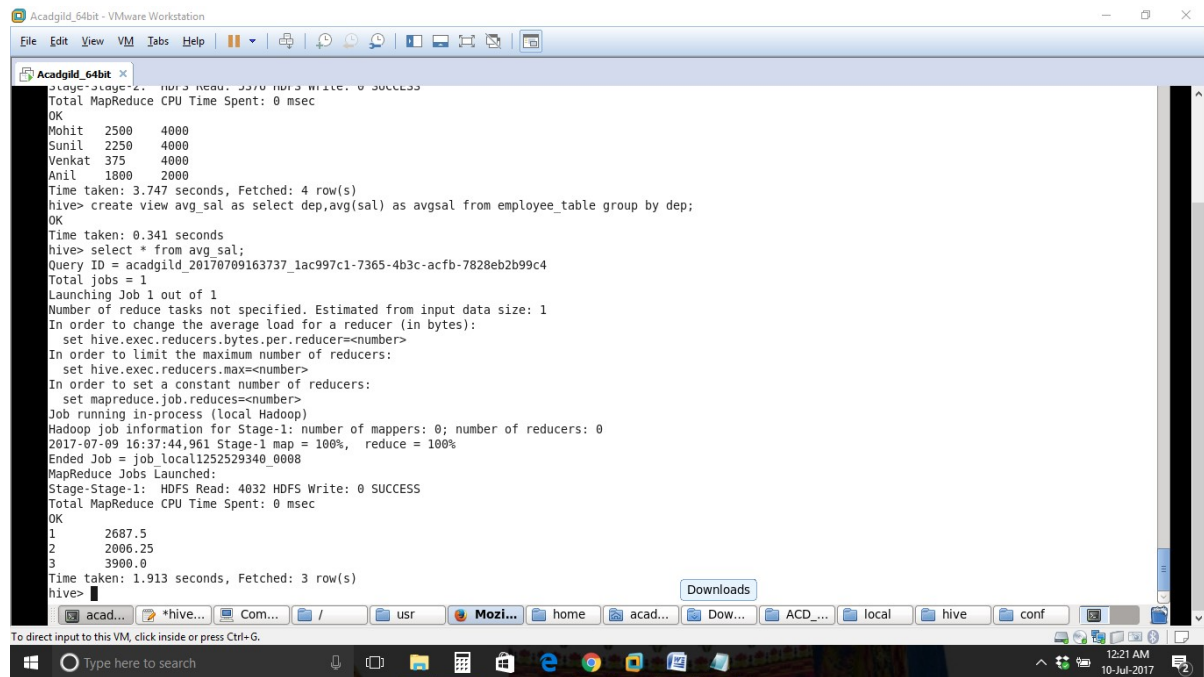
## Problem Statement 2:

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

View created for average salary details

Create view avg\_sal as select dep,avg(sal) as avgsal from employee\_table group by dep;

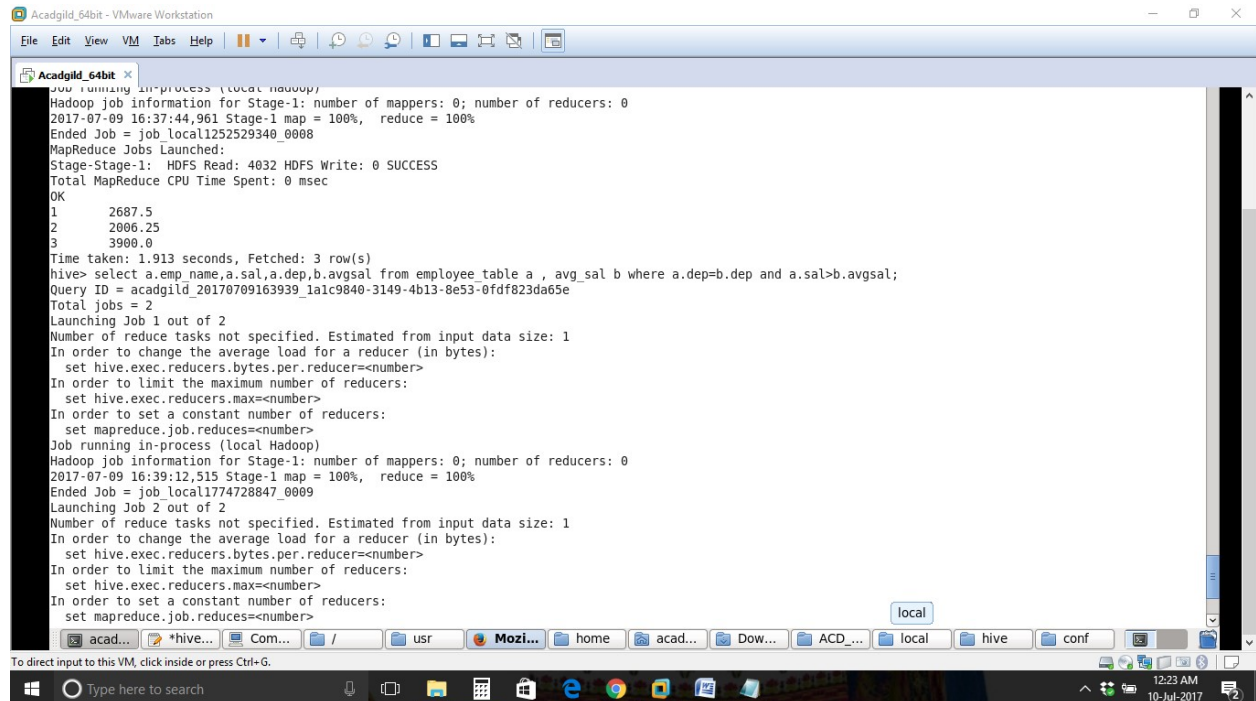
Select \* from avg\_sal;



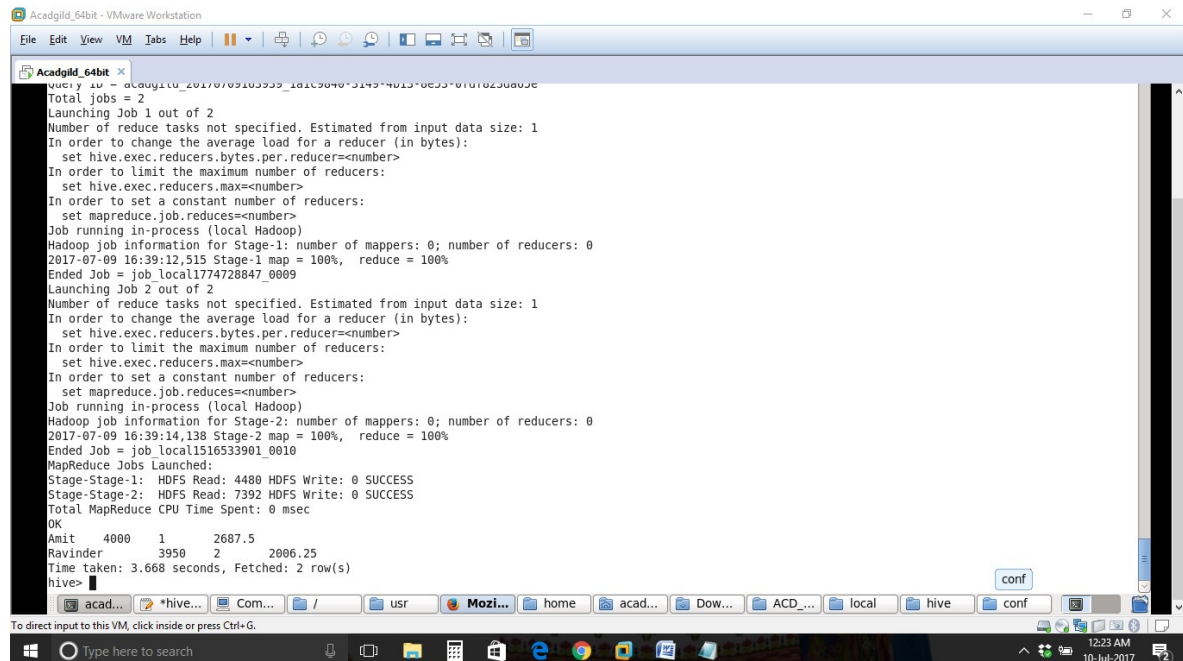
```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Stage-Stage-2: HDFS Read: 4032 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Mohit 2500 4000
Sunil 2250 4000
Venkat 375 4000
Anil 1800 2000
Time taken: 3.747 seconds, Fetched: 4 row(s)
hive> create view avg_sal as select dep,avg(sal) as avgsal from employee_table group by dep;
OK
Time taken: 0.341 seconds
hive> select * from avg_sal;
Query ID = acadgild_20170709163737_1ac997c1-7365-4b3c-acfb-7828eb2b99c4
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2017-07-09 16:37:44,961 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local1252529340_0008
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 4032 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1 2687.5
2 2006.25
3 3900.0
Time taken: 1.913 seconds, Fetched: 3 row(s)
hive>
```

Query to list employee whose salary is more than average salary in each department

Select a.emp\_name,a.sal,a.dep,b.avg\_sal from employee\_table a, avg\_sal b where a.dep=b.dep and a.sal>b.avg\_sal;



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
You running in-process (local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2017-07-09 16:37:44,961 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local1252529340_0008
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 4032 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1      2687.5
2      2006.25
3      3900.0
Time taken: 1.913 seconds, Fetched: 3 row(s)
hive> select a.emp_name,a.sal,a.dep,b.avg_sal from employee_table a , avg_sal b where a.dep=b.dep and a.sal>b.avg_sal;
Query ID = acadgild_20170709163939_1a1c9840-3149-4b13-8e53-0fd823da65e
Total jobs = 2
Launching Job 1 out of 2
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2017-07-09 16:39:12,515 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local1774728847_0009
Launching Job 2 out of 2
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>
local
acad... *hive... Com... / usr Mozi... home acad... Dow... ACD... local hive conf
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:23 AM 10-Jul-2017
```



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Query ID = acadgild_20170709163939_1a1c9840-3149-4b13-8e53-0fd823da65e
Total jobs = 2
Launching Job 1 out of 2
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2017-07-09 16:39:12,515 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local1774728847_0009
Launching Job 2 out of 2
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>
Job running in-process (local Hadoop)
Hadoop job information for Stage-2: number of mappers: 0; number of reducers: 0
2017-07-09 16:39:14,138 Stage-2 map = 100%, reduce = 100%
Ended Job = job_local1516533901_0010
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 4480 HDFS Write: 0 SUCCESS
Stage-Stage-2: HDFS Read: 7392 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Amit 4000 1 2687.5
Ravinder 3950 2 2006.25
Time taken: 3.668 seconds, Fetched: 2 row(s)
hive>
conf
acad... *hive... Com... / usr Mozi... home acad... Dow... ACD... local hive conf
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:23 AM 10-Jul-2017
```