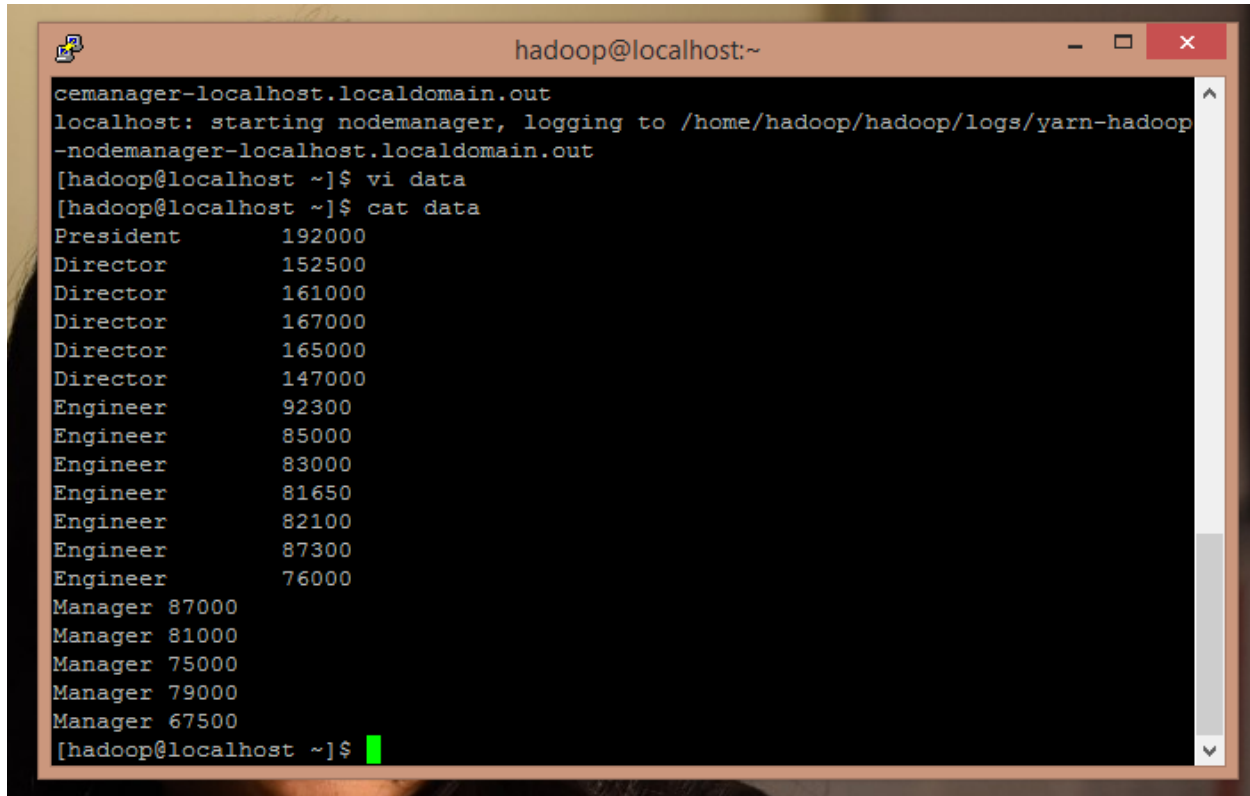


Name: renu Mariam mathew
Student id:16376
Hw#5

Step 1: The data from the notes



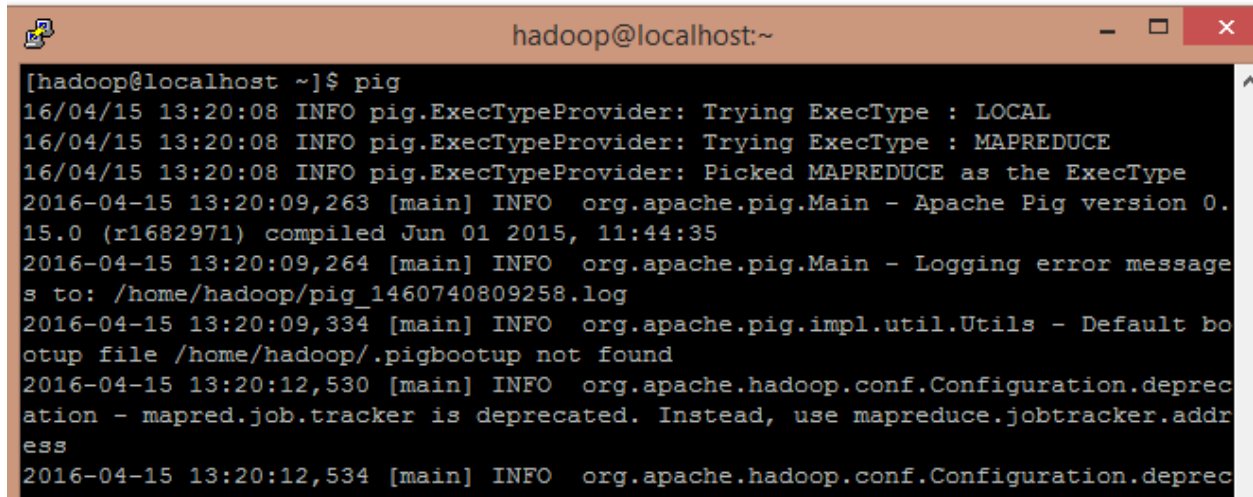
```
hadoop@localhost:~  
cemanager-localhost.localdomain.out  
localhost: starting nodemanager, logging to /home/hadoop/hadoop/logs/yarn-hadoop  
-nodemanager-localhost.localdomain.out  
[hadoop@localhost ~]$ vi data  
[hadoop@localhost ~]$ cat data  
President      192000  
Director      152500  
Director      161000  
Director      167000  
Director      165000  
Director      147000  
Engineer      92300  
Engineer      85000  
Engineer      83000  
Engineer      81650  
Engineer      82100  
Engineer      87300  
Engineer      76000  
Manager 87000  
Manager 81000  
Manager 75000  
Manager 79000  
Manager 67500  
[hadoop@localhost ~]$
```

Step 2: starting the pig shell

Name: renu Mariam mathew

Student id:16376

Hw#5

A terminal window titled 'hadoop@localhost:~' with standard window controls. The terminal shows the execution of the 'pig' command. The output includes log messages from the Pig execution engine, such as 'Trying ExecType : LOCAL', 'Trying ExecType : MAPREDUCE', and 'Picked MAPREDUCE as the ExecType'. It also displays the Apache Pig version (0.15.0) and compilation details. A warning message indicates that 'mapred.job.tracker' is deprecated and should be replaced with 'mapreduce.jobtracker.address'. The terminal text is as follows:

```
[hadoop@localhost ~]$ pig
16/04/15 13:20:08 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
16/04/15 13:20:08 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
16/04/15 13:20:08 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2016-04-15 13:20:09,263 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2015, 11:44:35
2016-04-15 13:20:09,264 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1460740809258.log
2016-04-15 13:20:09,334 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootstrap not found
2016-04-15 13:20:12,530 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2016-04-15 13:20:12,534 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

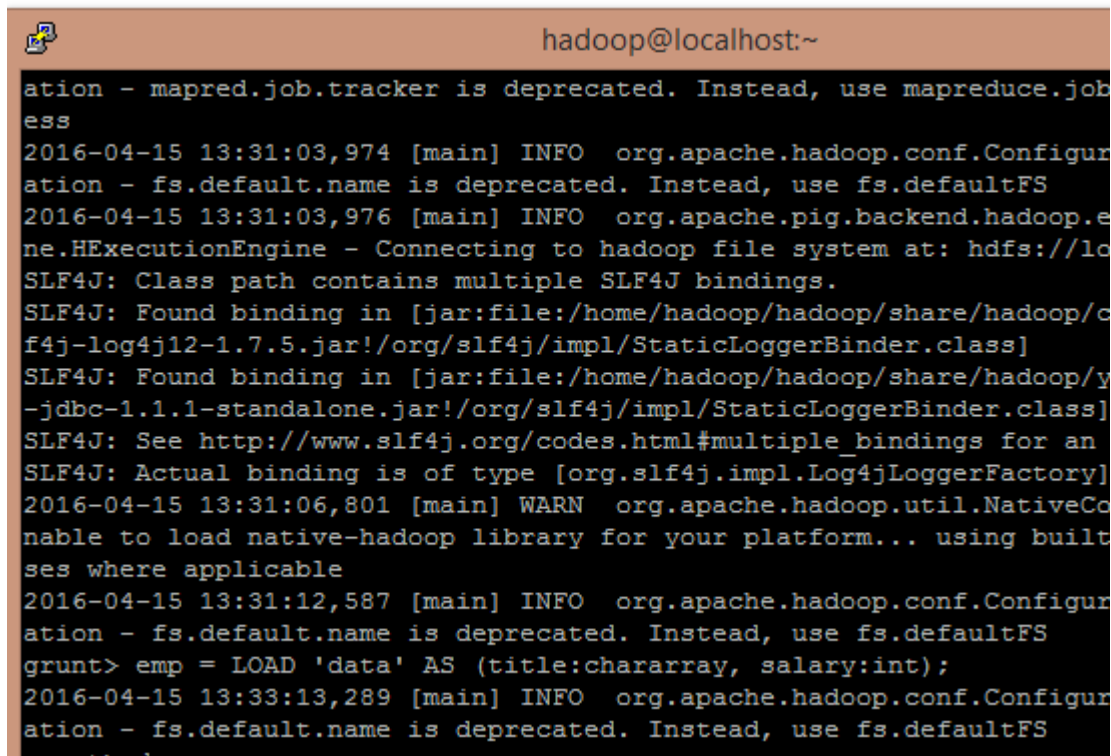
Step 3

Emp = Load 'data' AS (title:chararray,salary:int);

Name: renu Mariam mathew

Student id:16376

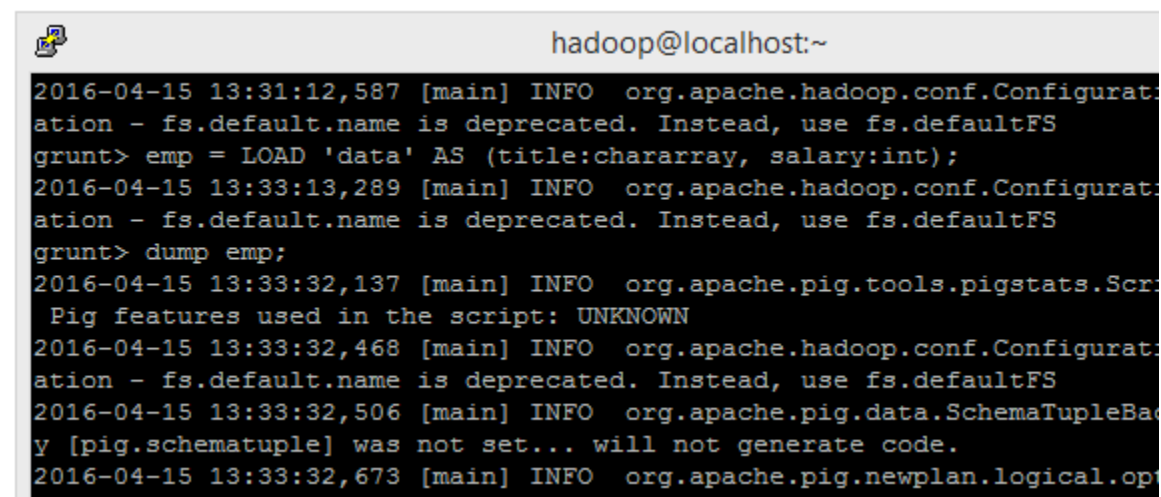
Hw#5



```
hadoop@localhost:~  
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.job  
ess  
2016-04-15 13:31:03,974 [main] INFO org.apache.hadoop.conf.Configur  
ation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2016-04-15 13:31:03,976 [main] INFO org.apache.pig.backend.hadoop.e  
ne.HExecutionEngine - Connecting to hadoop file system at: hdfs://lo  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/home/hadoop/hadoop/share/hadoop/c  
f4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/home/hadoop/hadoop/share/hadoop/y  
-jdbc-1.1.1-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
2016-04-15 13:31:06,801 [main] WARN org.apache.hadoop.util.NativeCo  
nable to load native-hadoop library for your platform... using built  
ses where applicable  
2016-04-15 13:31:12,587 [main] INFO org.apache.hadoop.conf.Configur  
ation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> emp = LOAD 'data' AS (title:chararray, salary:int);  
2016-04-15 13:33:13,289 [main] INFO org.apache.hadoop.conf.Configur  
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

Step 4:

Dump emp;

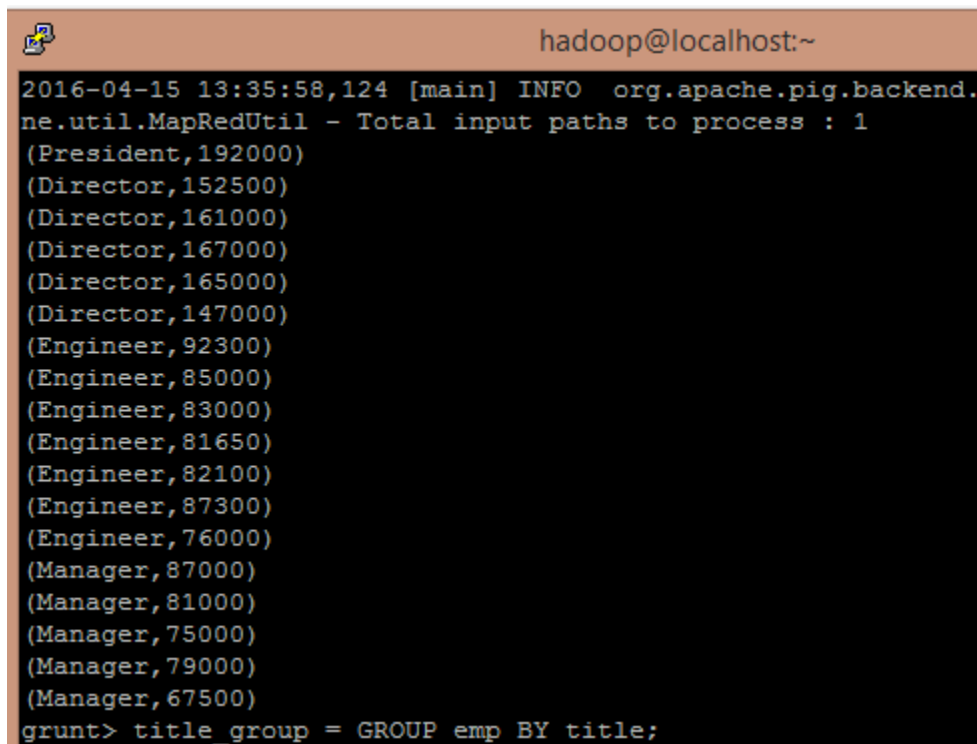


```
hadoop@localhost:~  
2016-04-15 13:31:12,587 [main] INFO org.apache.hadoop.conf.Configurat  
ation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> emp = LOAD 'data' AS (title:chararray, salary:int);  
2016-04-15 13:33:13,289 [main] INFO org.apache.hadoop.conf.Configurat  
ation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> dump emp;  
2016-04-15 13:33:32,137 [main] INFO org.apache.pig.tools.pigstats.Scri  
Pig features used in the script: UNKNOWN  
2016-04-15 13:33:32,468 [main] INFO org.apache.hadoop.conf.Configurat  
ation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2016-04-15 13:33:32,506 [main] INFO org.apache.pig.data.SchemaTupleBa  
y [pig.schematuple] was not set... will not generate code.  
2016-04-15 13:33:32,673 [main] INFO org.apache.pig.newplan.logical.op
```

Name: renu Mariam mathew
Student id:16376
Hw#5

Step 5:

Title_group = group emp by title:

A terminal window with a black background and orange title bar. The title bar contains a small icon and the text 'hadoop@localhost:~'. The terminal shows the output of a Pig script, listing employee titles and salaries in descending order. The output is as follows:

```
2016-04-15 13:35:58,124 [main] INFO org.apache.pig.backend.  
ne.util.MapRedUtil - Total input paths to process : 1  
(President,192000)  
(Director,152500)  
(Director,161000)  
(Director,167000)  
(Director,165000)  
(Director,147000)  
(Engineer,92300)  
(Engineer,85000)  
(Engineer,83000)  
(Engineer,81650)  
(Engineer,82100)  
(Engineer,87300)  
(Engineer,76000)  
(Manager,87000)  
(Manager,81000)  
(Manager,75000)  
(Manager,79000)  
(Manager,67500)  
grunt> title_group = GROUP emp BY title;
```

Step 6:

Top_salaries = FOREACH title_group (sorted = order emp by
salary desc; highest_paid = limit sorted 3; GENERATE
group,highest_paid);

Name: renu Mariam mathew
Student id:16376
Hw#5

```
Details at logfile: /home/hadoop/pig_1460741457120.log
grunt> top_salaries = FOREACH title_group {sorted = ORDER emp BY salary DESC; highest_paid = LIMIT sorted 3; GENERATE group,highest_paid
;};
grunt>
```

STEP 7: DUMP top_salaries;

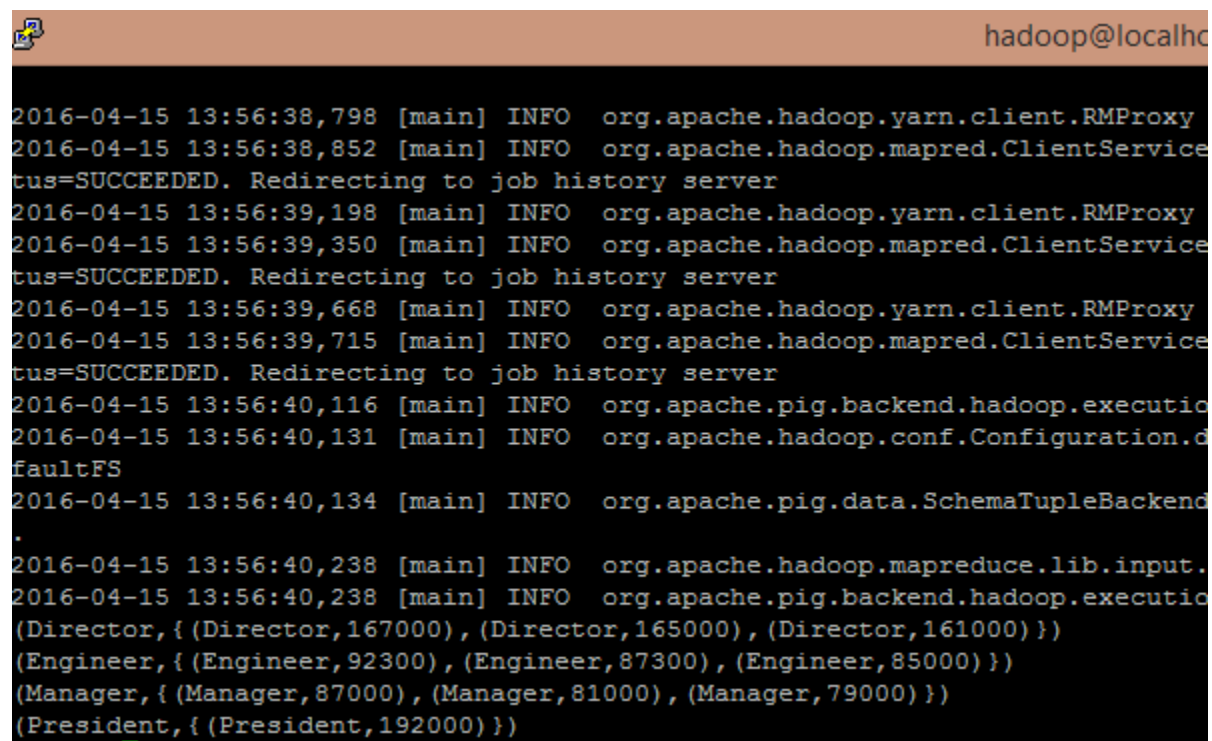
```
Details at logfile: /home/hadoop/pig_1460741457120.log
grunt> top_salaries = FOREACH title_group {sorted = ORDER emp BY
;};
grunt> dump top_salaries;
2016-04-15 13:53:33,320 [main] INFO org.apache.pig.tools.pigstat
2016-04-15 13:53:33,544 [main] INFO org.apache.hadoop.conf.Confi
faultFS
2016-04-15 13:53:33,547 [main] WARN org.apache.pig.data.SchemaTu
2016-04-15 13:53:33,549 [main] INFO org.apache.pig.newplan.logic
pKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitC
erOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten,
2016-04-15 13:53:33,732 [main] INFO org.apache.pig.backend.hadoo
```

```
hadoop@localho

2016-04-15 13:56:38,798 [main] INFO org.apache.hadoop.yarn.client.RMPProxy
2016-04-15 13:56:38,852 [main] INFO org.apache.hadoop.mapred.ClientService
tus=SUCCEEDED. Redirecting to job history server
2016-04-15 13:56:39,198 [main] INFO org.apache.hadoop.yarn.client.RMPProxy
2016-04-15 13:56:39,350 [main] INFO org.apache.hadoop.mapred.ClientService
tus=SUCCEEDED. Redirecting to job history server
2016-04-15 13:56:39,668 [main] INFO org.apache.hadoop.yarn.client.RMPProxy
2016-04-15 13:56:39,715 [main] INFO org.apache.hadoop.mapred.ClientService
tus=SUCCEEDED. Redirecting to job history server
2016-04-15 13:56:40,116 [main] INFO org.apache.pig.backend.hadoop.executio
2016-04-15 13:56:40,131 [main] INFO org.apache.hadoop.conf.Configuration.d
faultFS
2016-04-15 13:56:40,134 [main] INFO org.apache.pig.data.SchemaTupleBackend
.
2016-04-15 13:56:40,238 [main] INFO org.apache.hadoop.mapreduce.lib.input.
2016-04-15 13:56:40,238 [main] INFO org.apache.pig.backend.hadoop.executio
(Director,{ (Director,167000) , (Director,165000) , (Director,161000) })
(Engineer,{ (Engineer,92300) , (Engineer,87300) , (Engineer,85000) })
(Manager,{ (Manager,87000) , (Manager,81000) , (Manager,79000) })
(President,{ (President,192000) })
```

Name: renu Mariam mathew
Student id:16376
Hw#5

The final output the top 3 salary of each job category :

A screenshot of a terminal window with a title bar that says 'hadoop@localho'. The terminal displays a series of log messages from the Hadoop ecosystem. The messages include timestamps, log levels (INFO), and class names. The final part of the log shows a list of job categories and their top three salaries: Director (167000, 165000, 161000), Engineer (92300, 87300, 85000), Manager (87000, 81000, 79000), and President (192000).

```
2016-04-15 13:56:38,798 [main] INFO org.apache.hadoop.yarn.client.RMProxy
2016-04-15 13:56:38,852 [main] INFO org.apache.hadoop.mapred.ClientService
tus=SUCCEEDED. Redirecting to job history server
2016-04-15 13:56:39,198 [main] INFO org.apache.hadoop.yarn.client.RMProxy
2016-04-15 13:56:39,350 [main] INFO org.apache.hadoop.mapred.ClientService
tus=SUCCEEDED. Redirecting to job history server
2016-04-15 13:56:39,668 [main] INFO org.apache.hadoop.yarn.client.RMProxy
2016-04-15 13:56:39,715 [main] INFO org.apache.hadoop.mapred.ClientService
tus=SUCCEEDED. Redirecting to job history server
2016-04-15 13:56:40,116 [main] INFO org.apache.pig.backend.hadoop.executio
2016-04-15 13:56:40,131 [main] INFO org.apache.hadoop.conf.Configuration.d
faultFS
2016-04-15 13:56:40,134 [main] INFO org.apache.pig.data.SchemaTupleBackend
.
2016-04-15 13:56:40,238 [main] INFO org.apache.hadoop.mapreduce.lib.input.
2016-04-15 13:56:40,238 [main] INFO org.apache.pig.backend.hadoop.executio
(Director,{ (Director,167000) , (Director,165000) , (Director,161000) })
(Engineer,{ (Engineer,92300) , (Engineer,87300) , (Engineer,85000) })
(Manager,{ (Manager,87000) , (Manager,81000) , (Manager,79000) })
(President,{ (President,192000) })
```