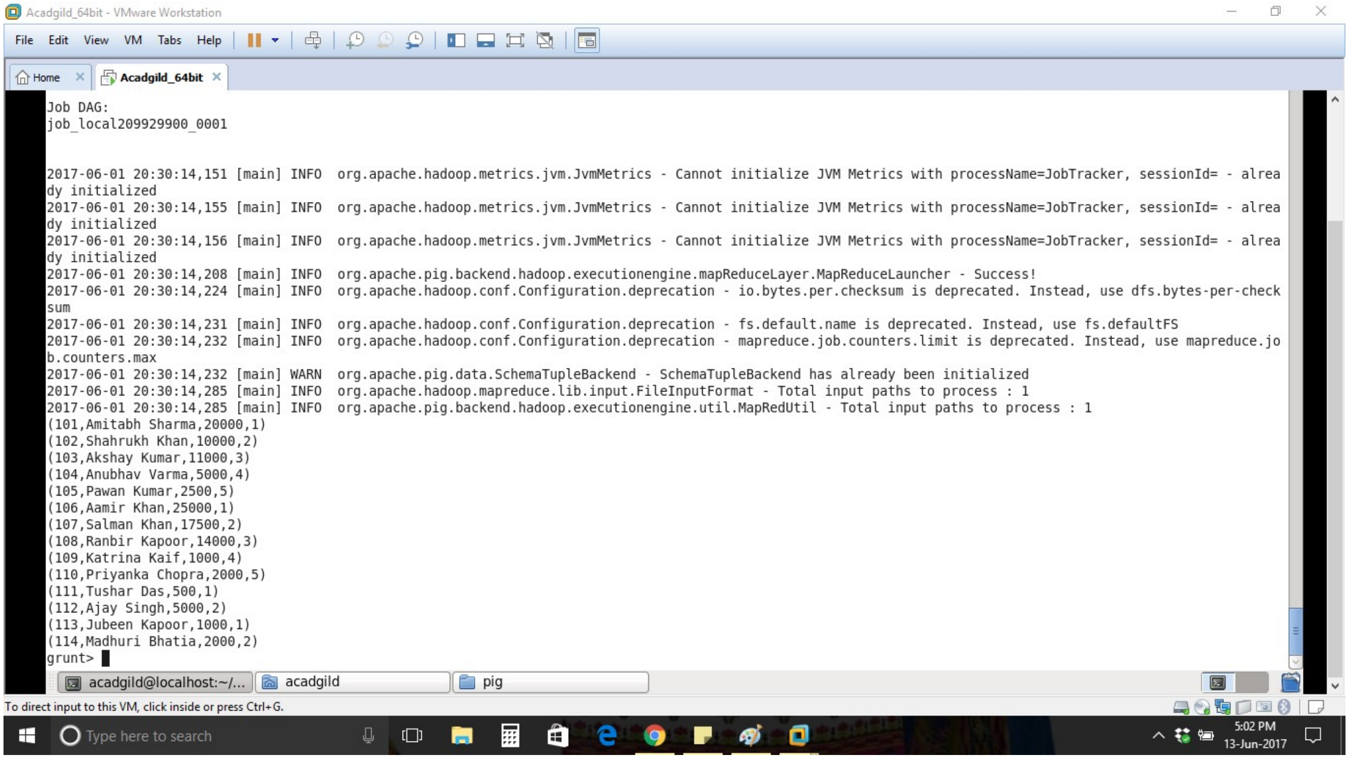
Assignment 8.2

Employee_details file getting loaded

emp = LOAD 'employee_details' USING PigStorage(',') AS (emp_id:int, emp_name:chararray, emp_salary:int);

dump emp;

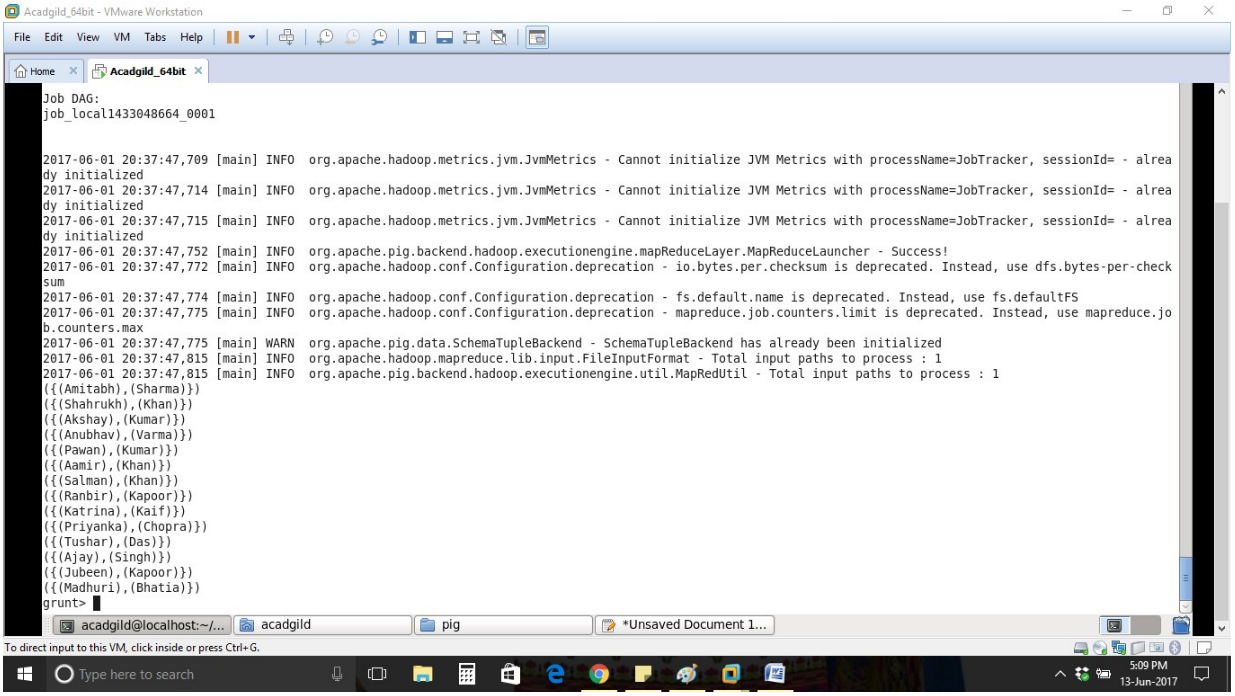


```
Job DAG:
job_local209929900_0001


2017-06-01 20:30:14,151 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:30:14,155 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:30:14,156 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:30:14,208 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 20:30:14,224 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 20:30:14,231 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 20:30:14,232 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 20:30:14,232 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:30:14,285 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:30:14,285 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh Sharma,20000,1)
(102,Shahrukh Khan,10000,2)
(103,Akshay Kumar,11000,3)
(104,Anubhav Varma,5000,4)
(105,Pawan Kumar,2500,5)
(106,Aamir Khan,25000,1)
(107,Salman Khan,17500,2)
(108,Ranbir Kapoor,14000,3)
(109,Katrina Kaif,1000,4)
(110,Priyanka Chopra,2000,5)
(111,Tushar Das,500,1)
(112,Ajay Singh,5000,2)
(113,Jubeen Kapoor,1000,1)
(114,Madhuri Bhatia,2000,2)
grunt>
```

8.2.1
TOKENIZE

emp_name_tokenize = foreach emp generate TOKENIZE(emp_name);

dump emp_name_tokenize;



Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit
acadgild@localhost:~/...
acadgild
pig
*Unsaved Document 1...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search
5:09 PM
13-Jun-2017

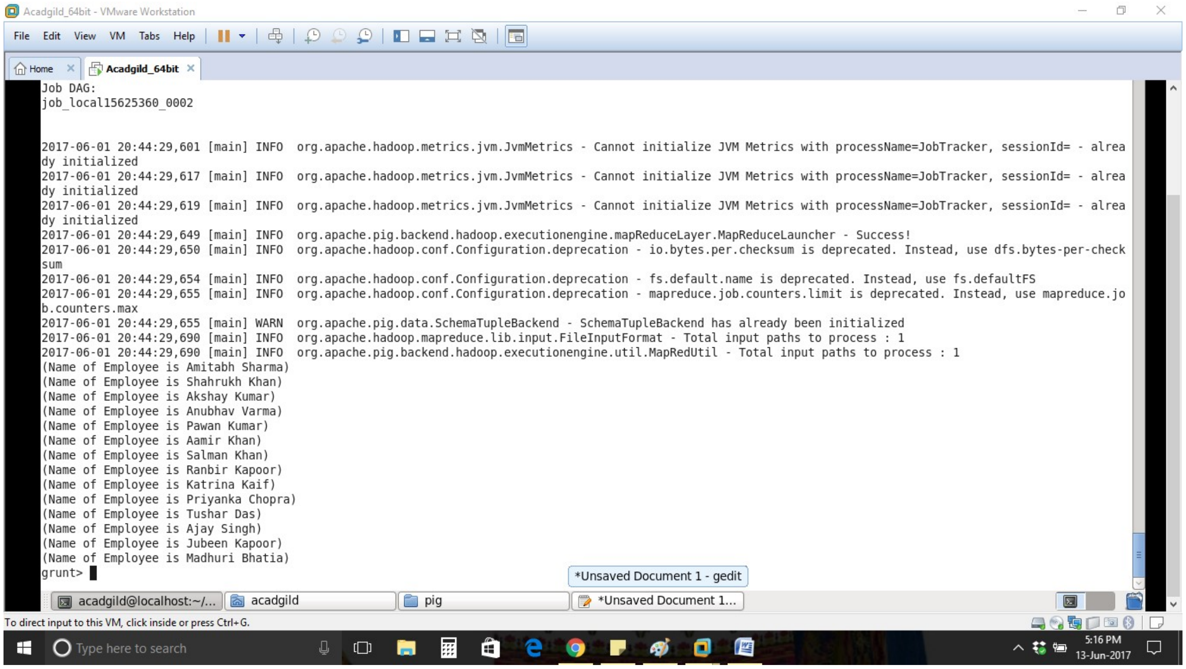
```
Job DAG:
job_local1433048664_0001


2017-06-01 20:37:47,709 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:37:47,714 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:37:47,715 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:37:47,752 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 20:37:47,772 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 20:37:47,774 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 20:37:47,775 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 20:37:47,775 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:37:47,815 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:37:47,815 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
({(Amitabh),(Sharma)})
({(Shahrukh),(Khan)})
({(Akshay),(Kumar)})
({(Anubhav),(Varma)})
({(Pawan),(Kumar)})
({(Aamir),(Khan)})
({(Salman),(Khan)})
({(Ranbir),(Kapoor)})
({(Katrina),(Kaif)})
({(Priyanka),(Chopra)})
({(Tushar),(Das)})
({(Ajay),(Singh)})
({(Jubeen),(Kapoor)})
({(Madhuri),(Bhatia)})
grunt> 
```

8.2.2

CONCAT

emp_name_detail = foreach emp generate CONCAT('Name of Employee is ' , emp_name);

dump emp_name_detail;



Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit
*Unsaved Document 1 - gedit
acadgild@localhost:~/...
acadgild
pig
*Unsaved Document 1...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search
5:16 PM
13-Jun-2017

```
Job DAG:
job_local15625360_0002


2017-06-01 20:44:29,601 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:44:29,617 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:44:29,619 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:44:29,649 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 20:44:29,650 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 20:44:29,654 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 20:44:29,655 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 20:44:29,655 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:44:29,690 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:44:29,690 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Name of Employee is Amitabh Sharma)
(Name of Employee is Shahrukh Khan)
(Name of Employee is Akshay Kumar)
(Name of Employee is Anubhav Varma)
(Name of Employee is Pawan Kumar)
(Name of Employee is Aamir Khan)
(Name of Employee is Salman Khan)
(Name of Employee is Ranbir Kapoor)
(Name of Employee is Katrina Kaif)
(Name of Employee is Priyanka Chopra)
(Name of Employee is Tushar Das)
(Name of Employee is Ajay Singh)
(Name of Employee is Jubeen Kapoor)
(Name of Employee is Madhuri Bhatia)
grunt>
```

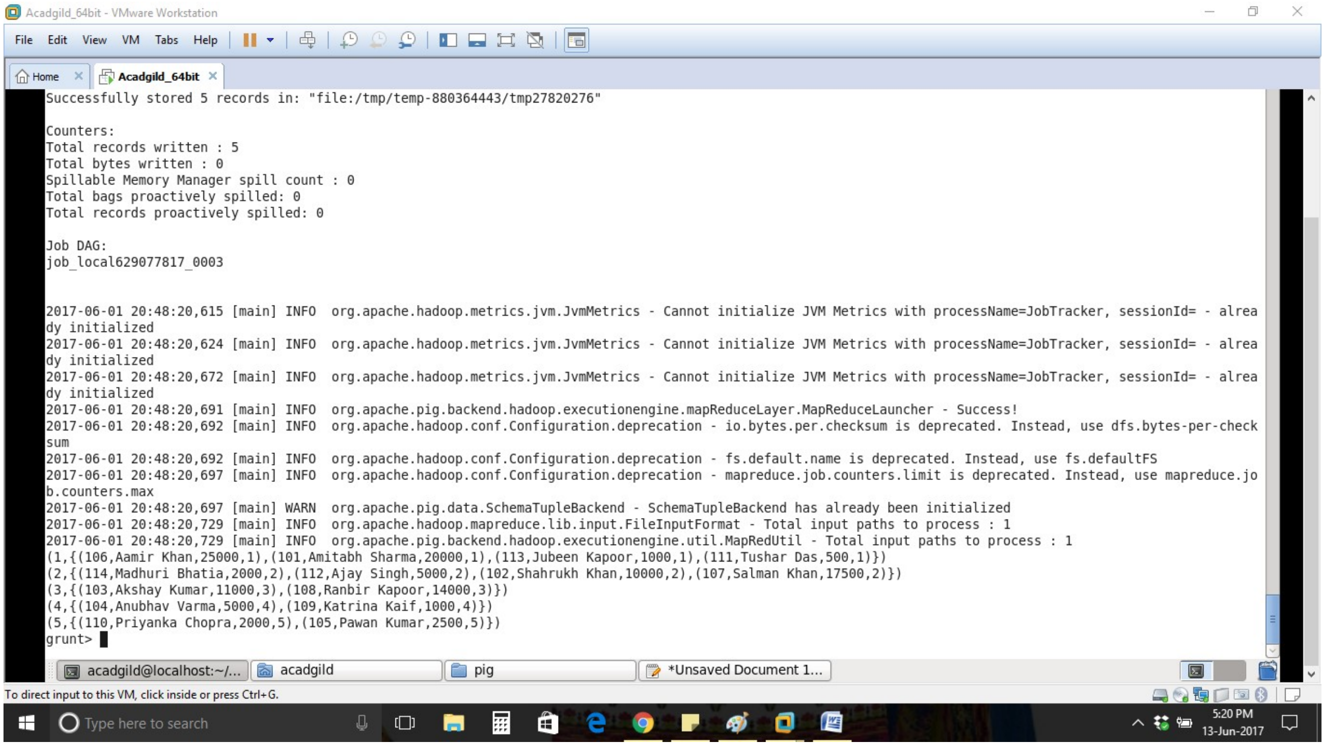
8.2.3

SUM

Employee details grouped by rating

emp_group_rating = group emp by emp_rating;

dump emp_group_rating;

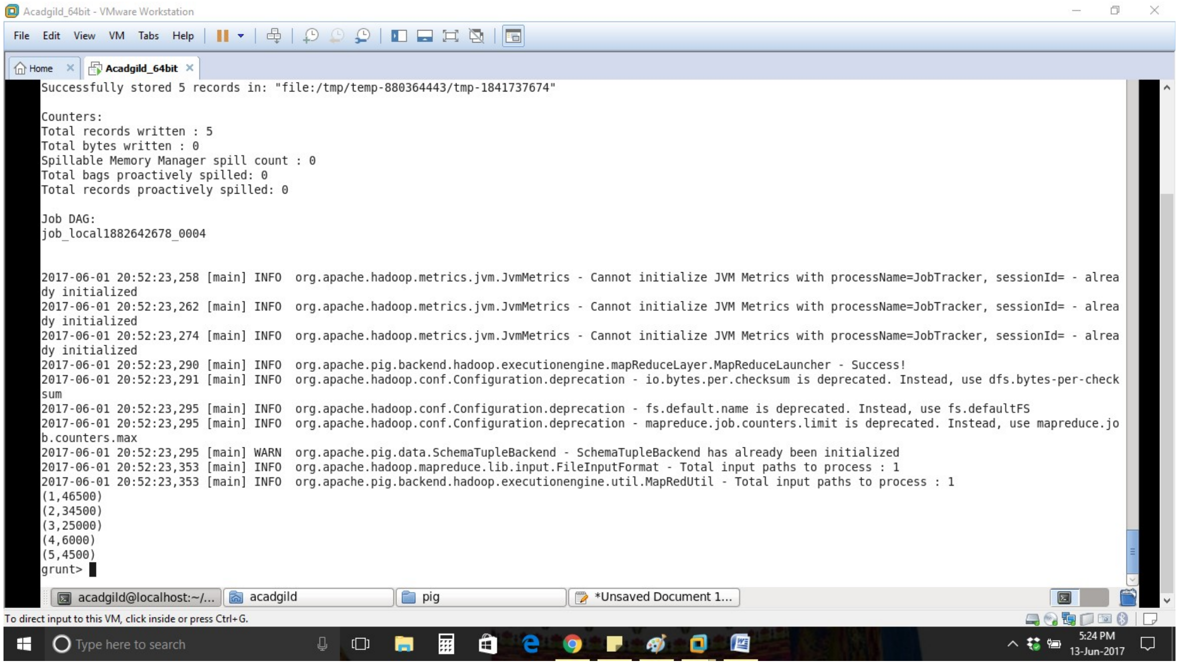

Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit

```
Successfully stored 5 records in: "file:/tmp/temp-880364443/tmp27820276"

Counters:
Total records written : 5
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local629077817_0003


2017-06-01 20:48:20,615 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:48:20,624 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:48:20,672 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:48:20,691 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 20:48:20,692 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 20:48:20,692 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 20:48:20,697 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 20:48:20,697 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:48:20,729 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:48:20,729 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,{(106,Aamir Khan,25000,1),(101,Amitabh Sharma,20000,1),(113,Jubeen Kapoor,1000,1),(111,Tushar Das,500,1)})
(2,{(114,Madhuri Bhatia,2000,2),(112,Ajay Singh,5000,2),(102,Shahrukh Khan,10000,2),(107,Salman Khan,17500,2)})
(3,{(103,Akshay Kumar,11000,3),(108,Ranbir Kapoor,14000,3)})
(4,{(104,Anubhav Varma,5000,4),(109,Katrina Kaif,1000,4)})
(5,{(110,Priyanka Chopra,2000,5),(105,Pawan Kumar,2500,5)})
grunt>
```

Sum of employee salaries grouped by rating

emp_sum_sal = foreach emp_group_rating generate group, SUM(emp.emp_salary);

dump emp_sum_sal;



Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit

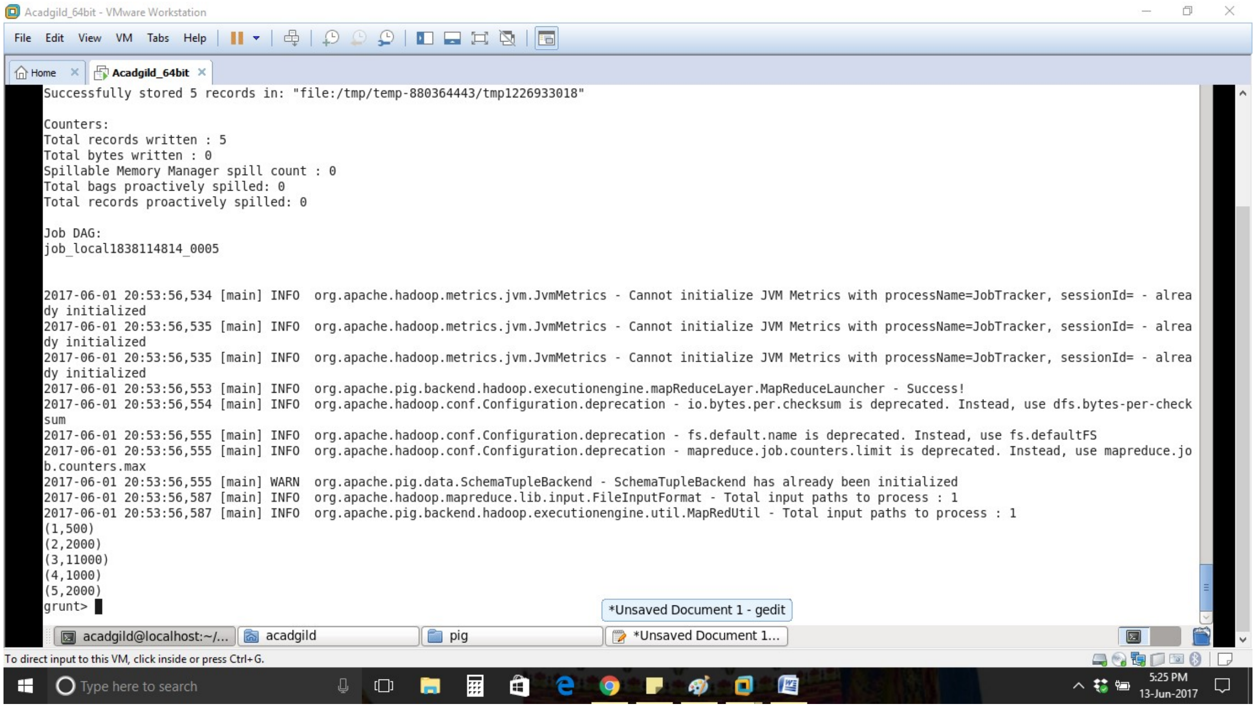
```
Successfully stored 5 records in: "file:/tmp/temp-880364443/tmp-1841737674"

Counters:
Total records written : 5
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1882642678_0004


2017-06-01 20:52:23,258 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:52:23,262 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:52:23,274 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:52:23,290 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 20:52:23,291 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 20:52:23,295 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 20:52:23,295 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 20:52:23,295 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:52:23,353 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:52:23,353 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,46500)
(2,34500)
(3,25000)
(4,6000)
(5,4500)
grunt> 
```

8.2.4

MIN

emp_min_sal = foreach emp_group_rating generate group, MIN(emp.emp_salary);

dump emp_min_sal;



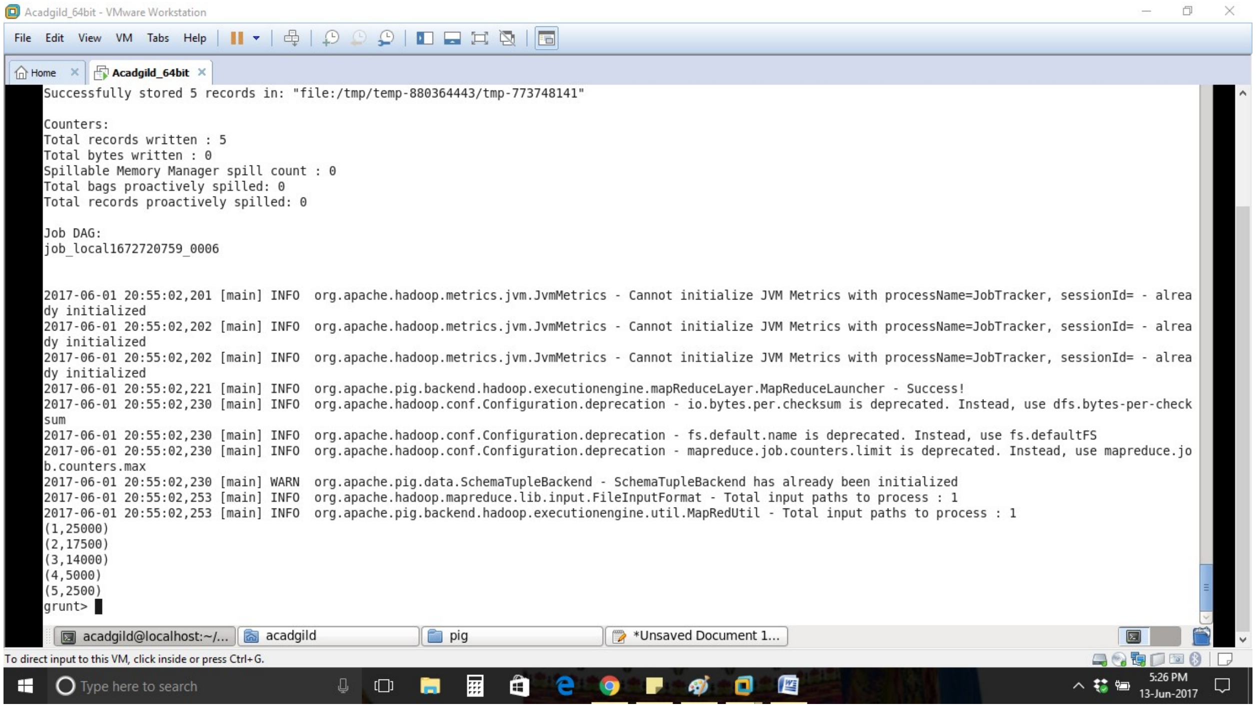
```
Successfully stored 5 records in: "file:/tmp/temp-880364443/tmp1226933018"

Counters:
Total records written : 5
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1838114814_0005


2017-06-01 20:53:56,534 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:53:56,535 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:53:56,535 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:53:56,553 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 20:53:56,554 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 20:53:56,555 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 20:53:56,555 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 20:53:56,555 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:53:56,587 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:53:56,587 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,500)
(2,2000)
(3,11000)
(4,1000)
(5,2000)
grunt> 
```

8.2.5

MAX

emp_max_sal = foreach emp_group_rating generate group, MAX(emp.emp_salary);

dump emp_max_sal;



Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit
acadgild@localhost:~/...
acadgild
pig
*Unsaved Document 1...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search
5:26 PM
13-Jun-2017

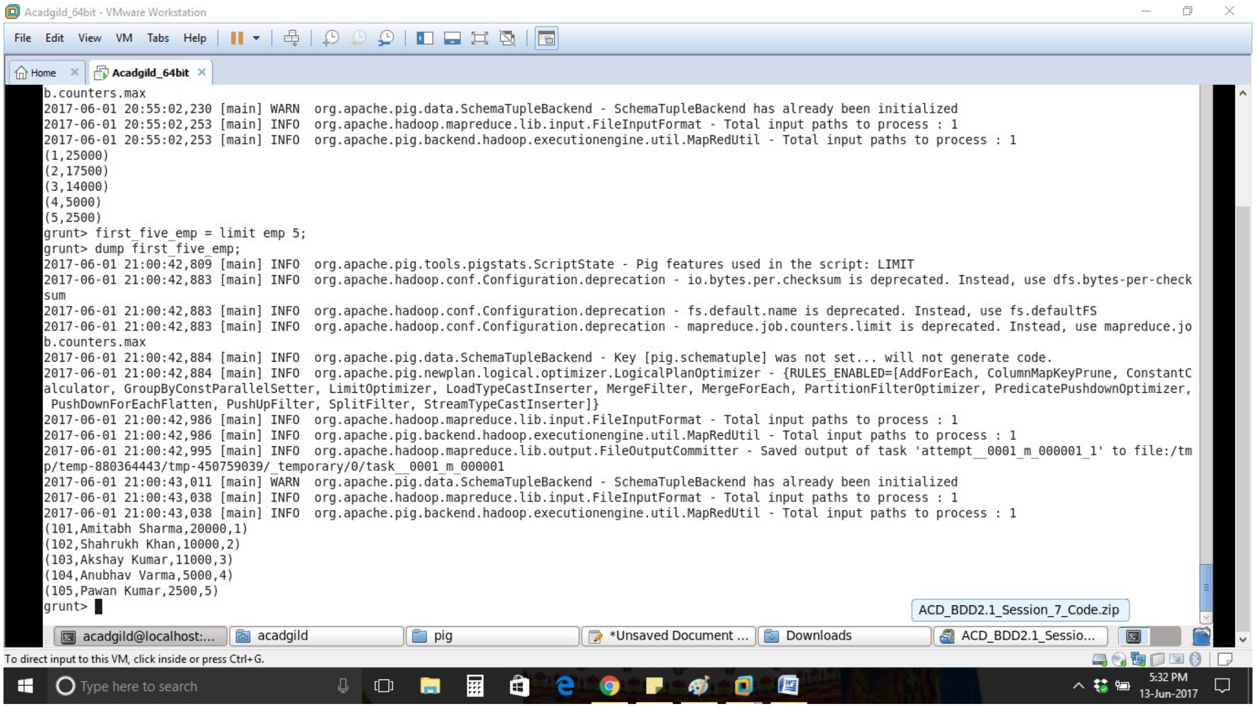
```
Successfully stored 5 records in: "file:/tmp/temp-880364443/tmp-773748141"

Counters:
Total records written : 5
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1672720759_0006


2017-06-01 20:55:02,201 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:55:02,202 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:55:02,202 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 20:55:02,221 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 20:55:02,230 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 20:55:02,230 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 20:55:02,230 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 20:55:02,230 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:55:02,253 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:55:02,253 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,25000)
(2,17500)
(3,14000)
(4,5000)
(5,2500)
grunt>
```

8.2.6

LIMIT

first_five_emp = limit emp 5;

dump first_five_emp;



```
b.counters.max
2017-06-01 20:55:02,230 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 20:55:02,253 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 20:55:02,253 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,25000)
(2,17500)
(3,14000)
(4,5000)
(5,2500)
grunt> first_five_emp = limit emp 5;
grunt> dump first_five_emp;
2017-06-01 21:00:42,809 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2017-06-01 21:00:42,883 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 21:00:42,883 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 21:00:42,883 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 21:00:42,884 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-06-01 21:00:42,884 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantC
alculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer,
 PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-06-01 21:00:42,986 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 21:00:42,986 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-06-01 21:00:42,995 [main] INFO  org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt__0001_m_000001_1' to file:/tm
p/temp-880364443/tmp-450759039/_temporary/0/task__0001_m_000001
2017-06-01 21:00:43,011 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 21:00:43,038 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 21:00:43,038 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh Sharma,20000,1)
(102,Shahrukh Khan,10000,2)
(103,Akshay Kumar,11000,3)
(104,Anubhav Varma,5000,4)
(105,Pawan Kumar,2500,5)
grunt> 
```

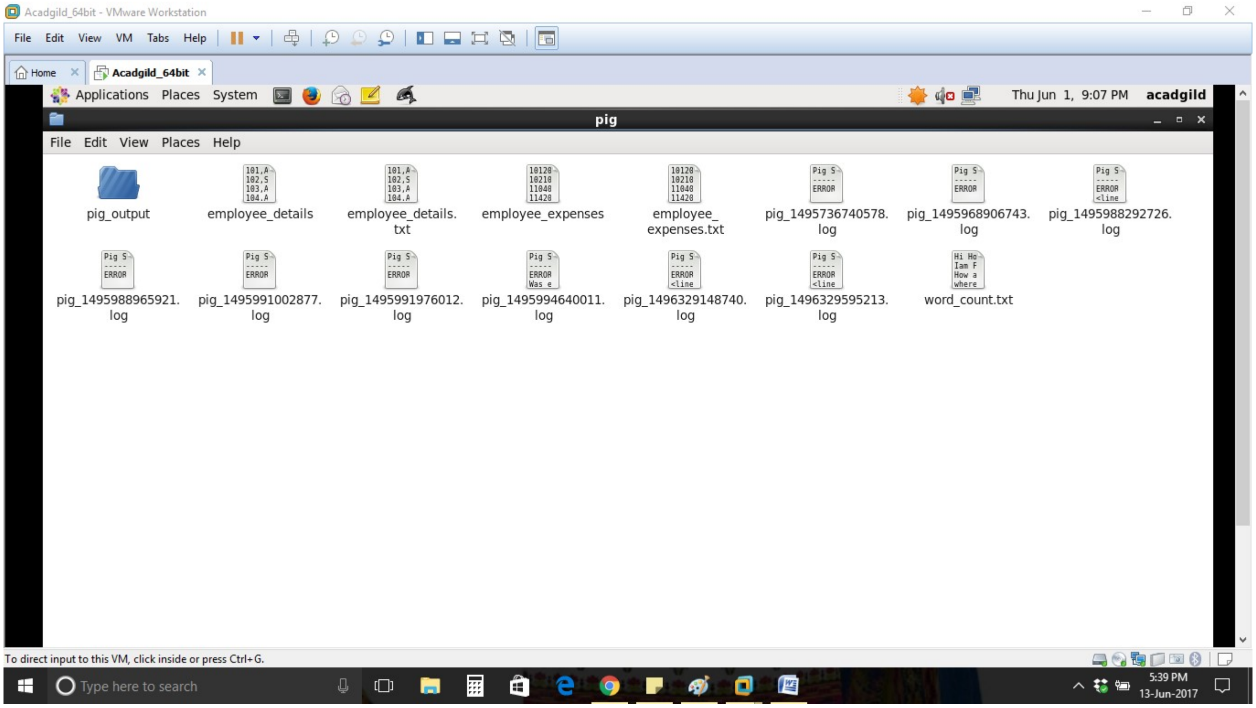
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit
ACD_BDD2.1_Session_7_Code.zip
acadgild@localhost:...
acadgild
pig
*Unsaved Document ...
Downloads
ACD_BDD2.1_Sessio...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search
5:32 PM
13-Jun-2017

8.2.7

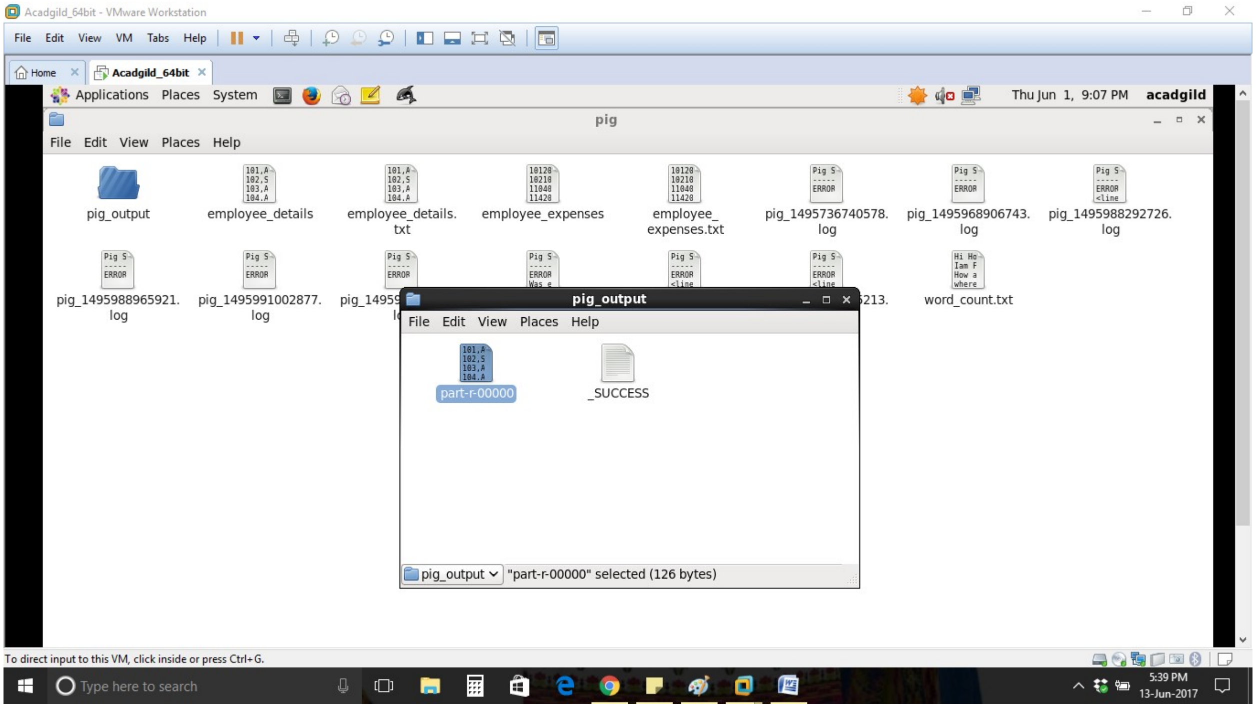
STORE

store first_five_emp INTO './pig_output/' USING PigStorage(',');

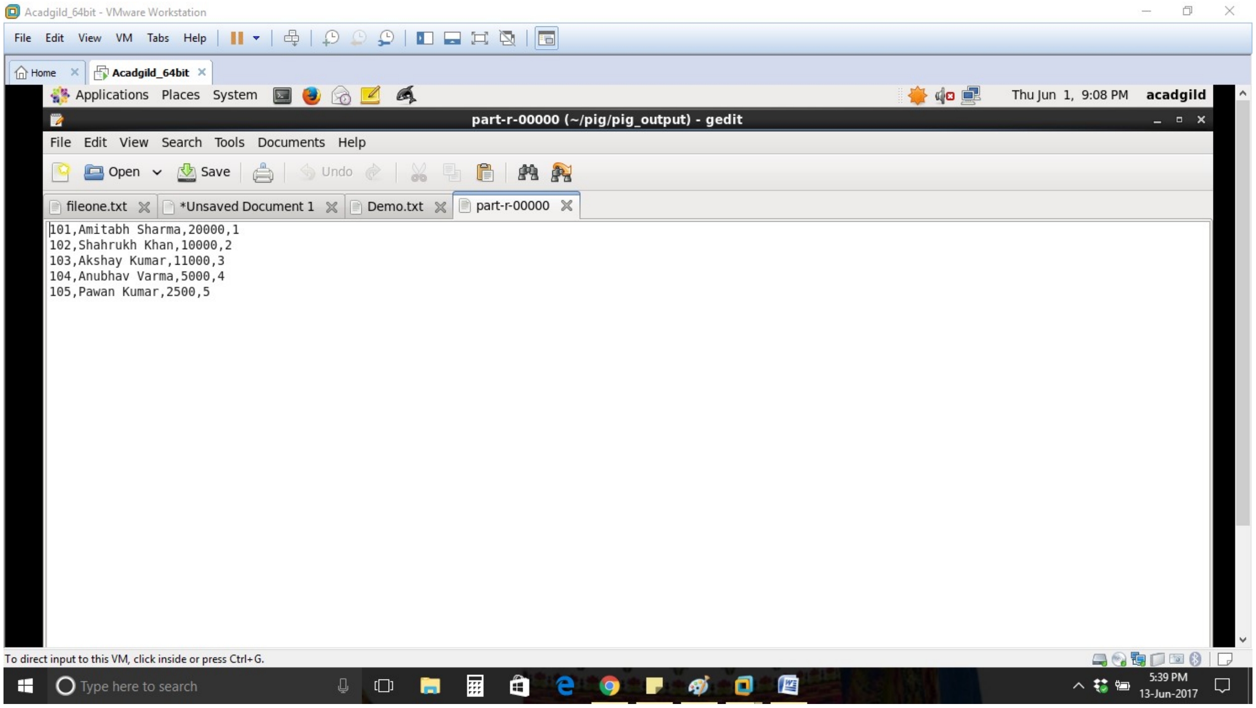
Directory pig_output created by STORE function



Two files created inside the pig_output folder



The file contains first five employee details


Acadgild_64bit - VMware Workstation
Applications Places System
Thu Jun 1, 9:08 PM acadgild
part-r-00000 (~/pig/pig_output) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
fileone.txt
*Unsaved Document 1
Demo.txt
part-r-00000
To direct input to this VM, click inside or press Ctrl+G.
Type here to search
5:39 PM 13-Jun-2017

```
101,Amitabh Sharma,20000,1
102,Shahrukh Khan,10000,2
103,Akshay Kumar,11000,3
104,Anubhav Varma,5000,4
105,Pawan Kumar,2500,5
```

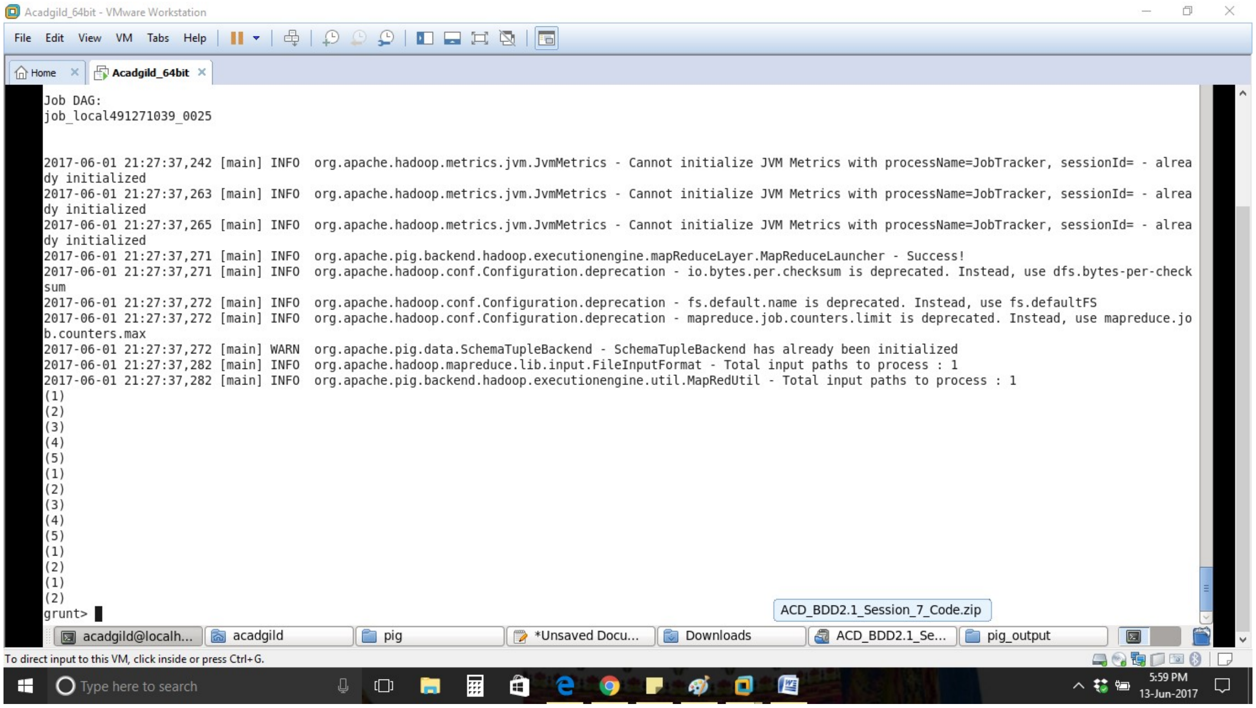
8.2.8

DISTINCT

emp_by_rating = foreach emp generate emp_rating as rating;

dump emp_by_rating;

Employee Ratings listed



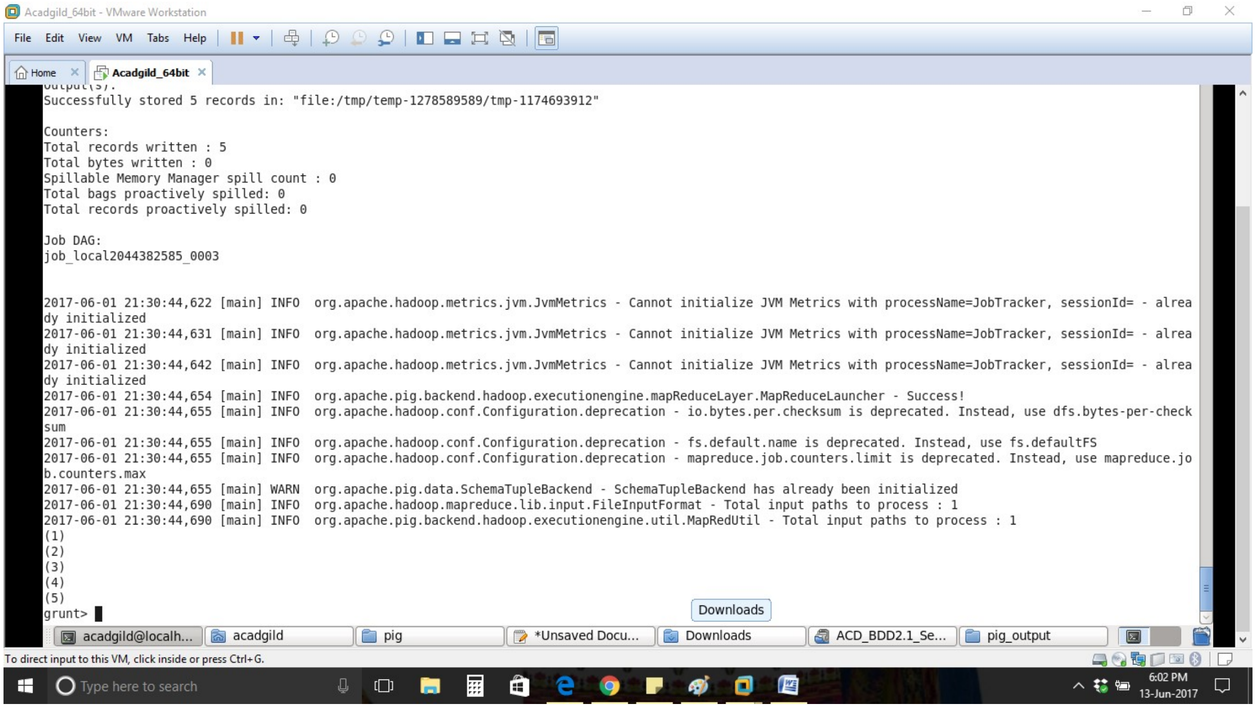
```
Job DAG:
job_local491271039_0025


2017-06-01 21:27:37,242 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:27:37,263 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:27:37,265 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:27:37,271 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 21:27:37,271 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 21:27:37,272 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 21:27:37,272 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 21:27:37,272 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 21:27:37,282 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 21:27:37,282 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1)
(2)
(3)
(4)
(5)
(1)
(2)
(3)
(4)
(5)
(1)
(2)
(1)
(2)
grunt> 
```

Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit
ACD_BDD2.1_Session_7_Code.zip
acadgild@localh...
acadgild
pig
*Unsaved Docu...
Downloads
ACD_BDD2.1_Se...
pig_output
To direct input to this VM, click inside or press Ctrl+G.
Type here to search
5:59 PM
13-Jun-2017

emp_distinct_rating = distinct emp_by_rating;

dump emp_distinct_rating

Only Distinct values in Rating Loaded avoiding duplicates



```
Successfully stored 5 records in: "file:/tmp/temp-1278589589/tmp-1174693912"

Counters:
Total records written : 5
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local2044382585_0003


2017-06-01 21:30:44,622 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:30:44,631 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:30:44,642 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:30:44,654 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 21:30:44,655 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 21:30:44,655 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 21:30:44,655 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 21:30:44,655 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 21:30:44,690 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 21:30:44,690 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1)
(2)
(3)
(4)
(5)
grunt>
```

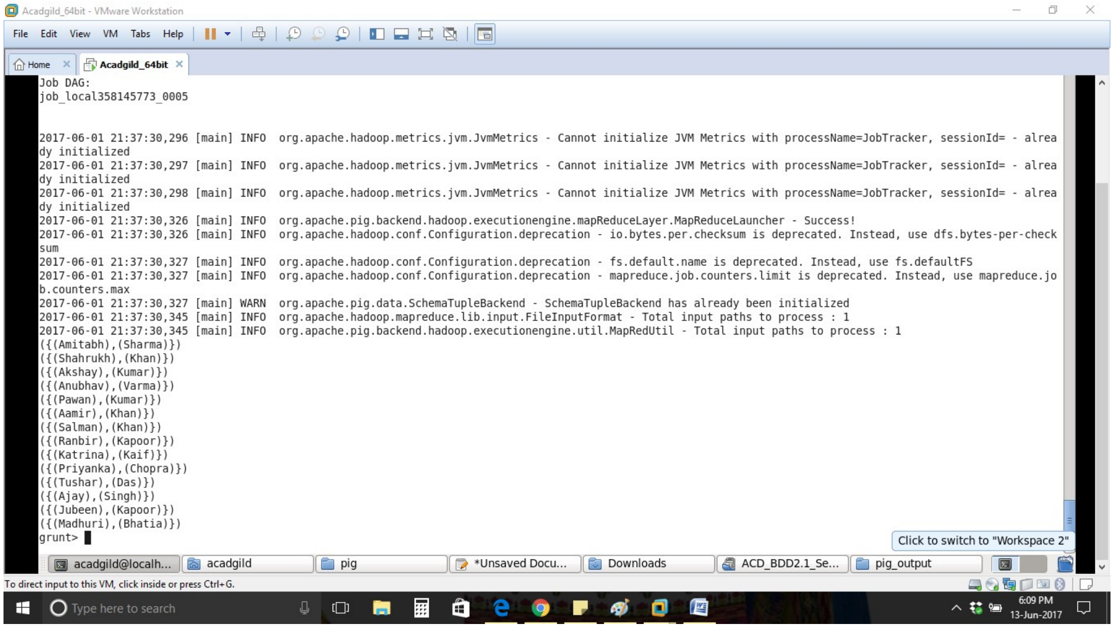
8.2.9

FLATTEN

emp_name_tokenize = foreach emp generate TOKENIZE(emp_name);

dump emp_name_tokenize;

employee name details tokenized

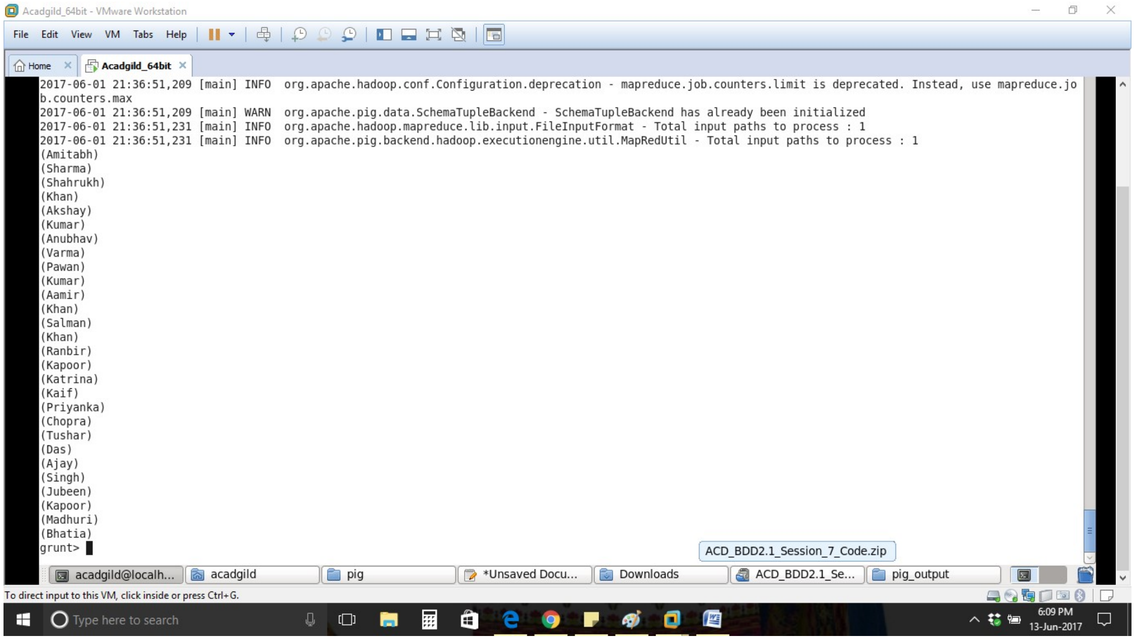

Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Home
Acadgild_64bit
acadgild@localh...
acadgild
pig
*Unsaved Docu...
Downloads
ACD_BDD2.1_Se...
pig_output
Click to switch to "Workspace 2"
To direct input to this VM, click inside or press Ctrl+G.
Type here to search
6:09 PM
13-Jun-2017

```
Job DAG:
job_local358145773_0005


2017-06-01 21:37:30,296 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:37:30,297 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:37:30,298 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-01 21:37:30,326 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-01 21:37:30,326 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-01 21:37:30,327 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-01 21:37:30,327 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 21:37:30,327 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 21:37:30,345 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 21:37:30,345 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
({(Amitabh),(Sharma)})
({(Shahrukh),(Khan)})
({(Akshay),(Kumar)})
({(Anubhav),(Varma)})
({(Pawan),(Kumar)})
({(Aamir),(Khan)})
({(Salman),(Khan)})
({(Ranbir),(Kapoor)})
({(Katrina),(Kaif)})
({(Priyanka),(Chopra)})
({(Tushar),(Das)})
({(Ajay),(Singh)})
({(Jubeen),(Kapoor)})
({(Madhuri),(Bhatia)})
grunt>
```

B = foreach emp_name_tokenize generate flatten($0);

Dump B;

FLATTEN function on Tokenized name



```
2017-06-01 21:36:51,209 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-01 21:36:51,209 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-01 21:36:51,231 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-01 21:36:51,231 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Amitabh)
(Sharma)
(Shahrukh)
(Khan)
(Akshay)
(Kumar)
(Anubhav)
(Varma)
(Pawan)
(Kumar)
(Aamir)
(Khan)
(Salman)
(Khan)
(Ranbir)
(Kapoor)
(Katrina)
(Kaif)
(Priyanka)
(Chopra)
(Tushar)
(Das)
(Ajay)
(Singh)
(Jubeen)
(Kapoor)
(Madhuri)
(Bhatia)
grunt>
```

8.2.10

IsEmpty()

```
emp_no_rating = filter emp by IsEmpty(emp_rating);
```

Giving error iterator could not be started.