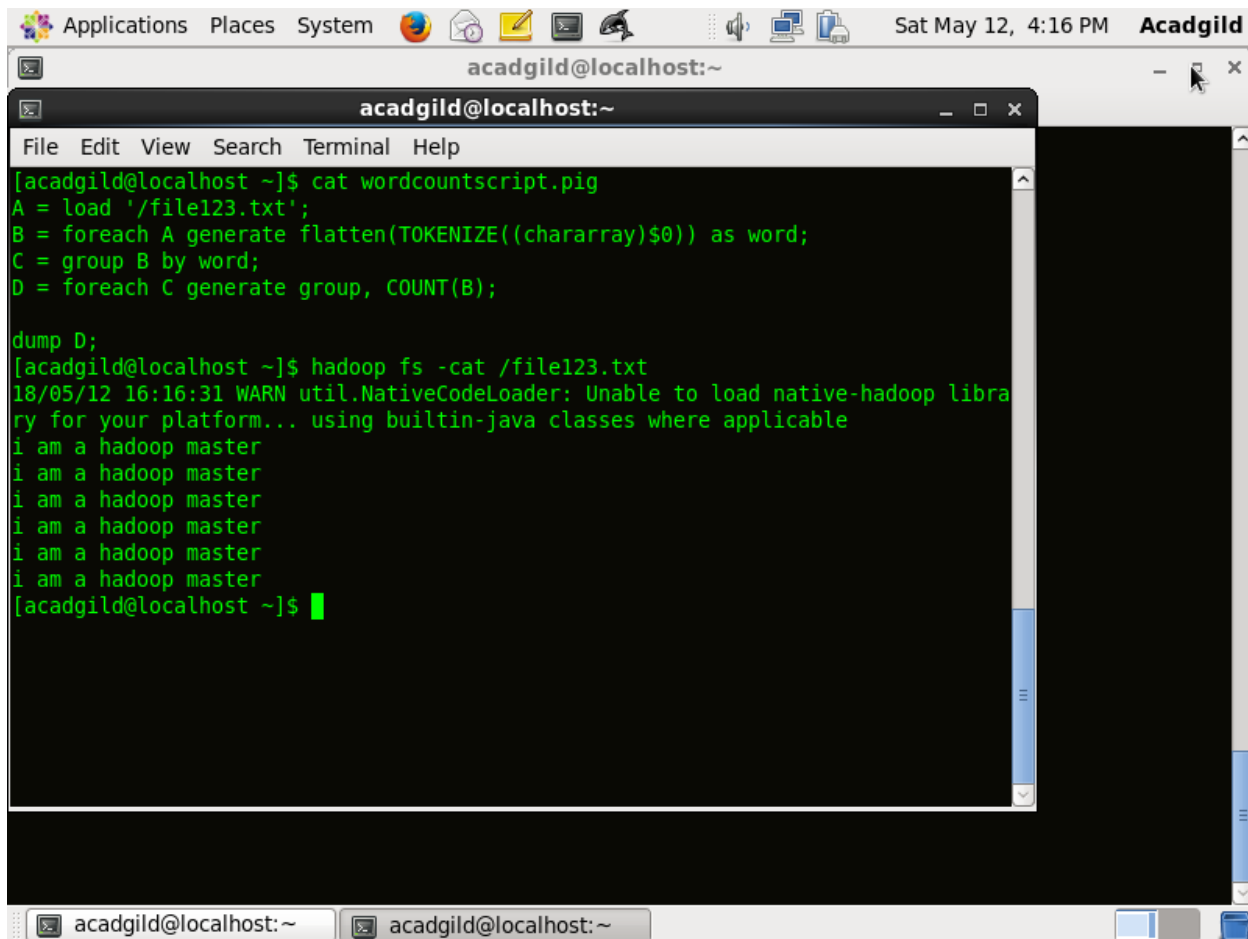


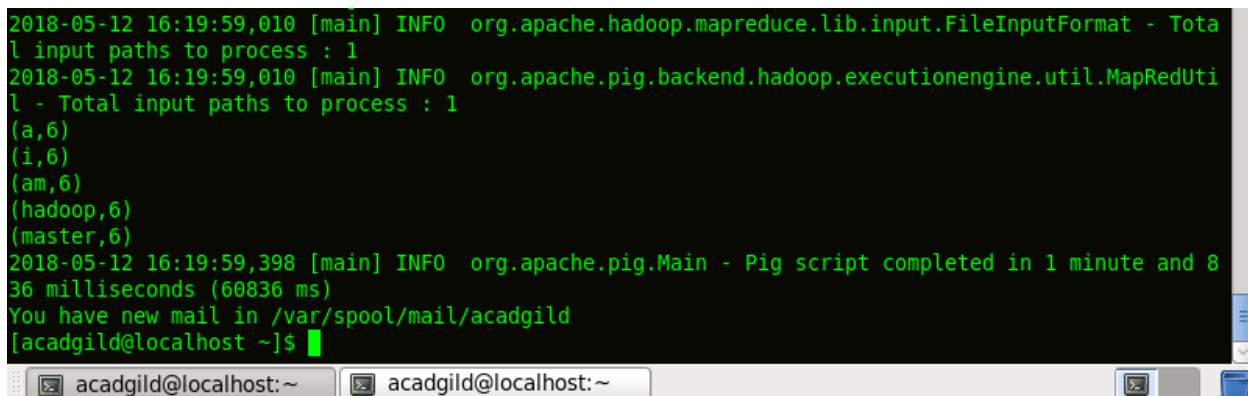
The pig script for word count has been placed in file named wordcountscript.pig.
Use cat wordcountscript.pig to view the contents of the file.
Wordcount is performed on a file named file123.txt stored on hdfs.
use hadoop fs -cat to see the contents of file123.txt



```
acadmild@localhost:~  
[acadmild@localhost ~]$ cat wordcountscript.pig  
A = load '/file123.txt';  
B = foreach A generate flatten(TOKENIZE((chararray)$0)) as word;  
C = group B by word;  
D = foreach C generate group, COUNT(B);  
  
dump D;  
[acadmild@localhost ~]$ hadoop fs -cat /file123.txt  
18/05/12 16:16:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
i am a hadoop master  
i am a hadoop master  
i am a hadoop master  
i am a hadoop master  
i am a hadoop master  
i am a hadoop master  
[acadmild@localhost ~]$
```

Run the command : pig <pig-script-file-name>

The output returns the number of occurrences of each word as shown in the screenshot



```
2018-05-12 16:19:59,010 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2018-05-12 16:19:59,010 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(a,6)  
(i,6)  
(am,6)  
(hadoop,6)  
(master,6)  
2018-05-12 16:19:59,398 [main] INFO org.apache.pig.Main - Pig script completed in 1 minute and 836 milliseconds (60836 ms)  
You have new mail in /var/spool/mail/acadmild  
[acadmild@localhost ~]$
```

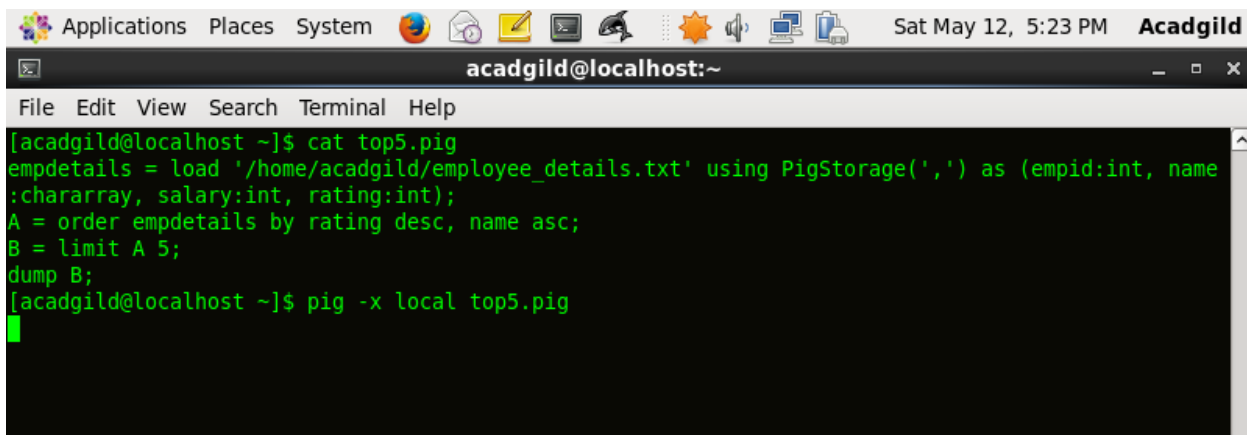
Top 5 employees (employee id and employee name) with highest rating. (In case two employees have same rating, employee with name coming first in dictionary should get preference)

I have used order by in my script top5.pig

I have used cat top5.pig to show the content of my pig script file. Since the data is on local and the path provided in the script file also points to the local mode.

So I am running the pig script in local mode. To do that use the command :

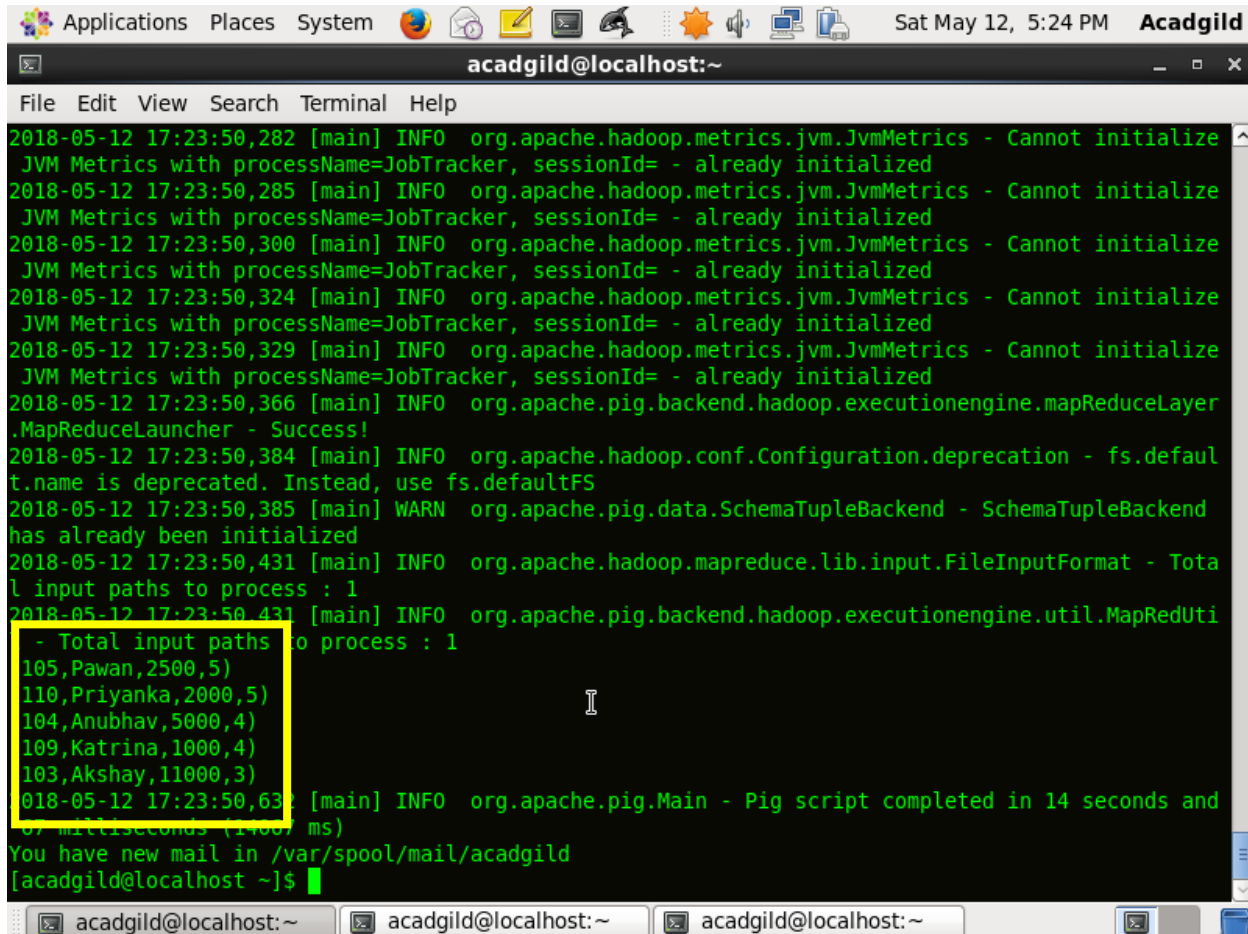
`pig -x local top5.pig`

A screenshot of a Linux terminal window. The window title bar shows 'Applications Places System' on the left, a system status area with icons for network, volume, and battery in the center, and 'Sat May 12, 5:23 PM Acadgild' on the right. The terminal itself has a title bar 'acadgild@localhost:~' and a menu bar 'File Edit View Search Terminal Help'. The terminal content shows a user running 'cat top5.pig' which displays a Pig script. The script loads data from '/home/acadgild/employee_details.txt' using 'PigStorage' with schema '(empid:int, name:chararray, salary:int, rating:int)', orders the results by 'rating desc, name asc', limits the output to 5 rows, and dumps the result. Finally, the user runs 'pig -x local top5.pig' to execute the script in local mode. The terminal text is as follows:

```
[acadgild@localhost ~]$ cat top5.pig
empdetails = load '/home/acadgild/employee_details.txt' using PigStorage(',') as (empid:int, name:chararray, salary:int, rating:int);
A = order empdetails by rating desc, name asc;
B = limit A 5;
dump B;
[acadgild@localhost ~]$ pig -x local top5.pig
```

Screenshot of output of top5.pig:

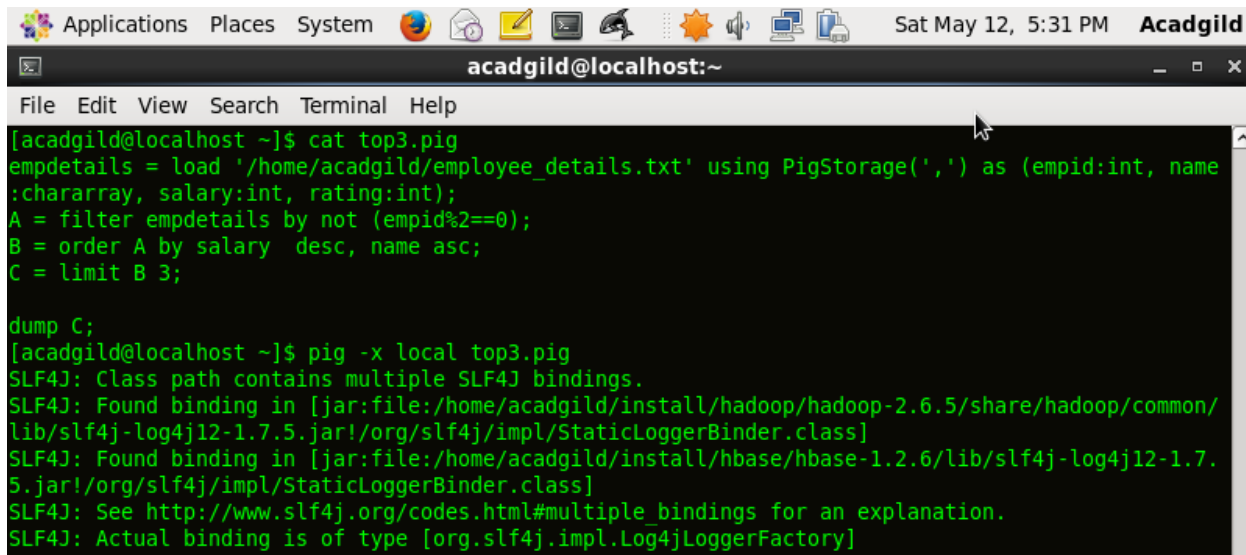
As one can see the records with highest ratings have been fetched and in case of same rating, it has been ordered by the name field.



```
2018-05-12 17:23:50,282 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:23:50,285 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:23:50,300 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:23:50,324 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:23:50,329 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:23:50,366 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Success!
2018-05-12 17:23:50,384 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
t.name is deprecated. Instead, use fs.defaultFS
2018-05-12 17:23:50,385 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
has already been initialized
2018-05-12 17:23:50,431 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tota
l input paths to process : 1
2018-05-12 17:23:50,431 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
l - Total input paths to process : 1
105,Pawan,2500,5)
110,Priyanka,2000,5)
104,Anubhav,5000,4)
109,Katrina,1000,4)
103,Akshay,11000,3)
2018-05-12 17:23:50,631 [main] INFO org.apache.pig.Main - Pig script completed in 14 seconds and
07 milliseconds (14007 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

To find top 3 employees with highest salary and whose ids are odd, I use filter to remove the records with even employee ids and then use orderby to arrange it according to salary and name and then limit the output to three. I have use cat command to show my script code in the screenshot.

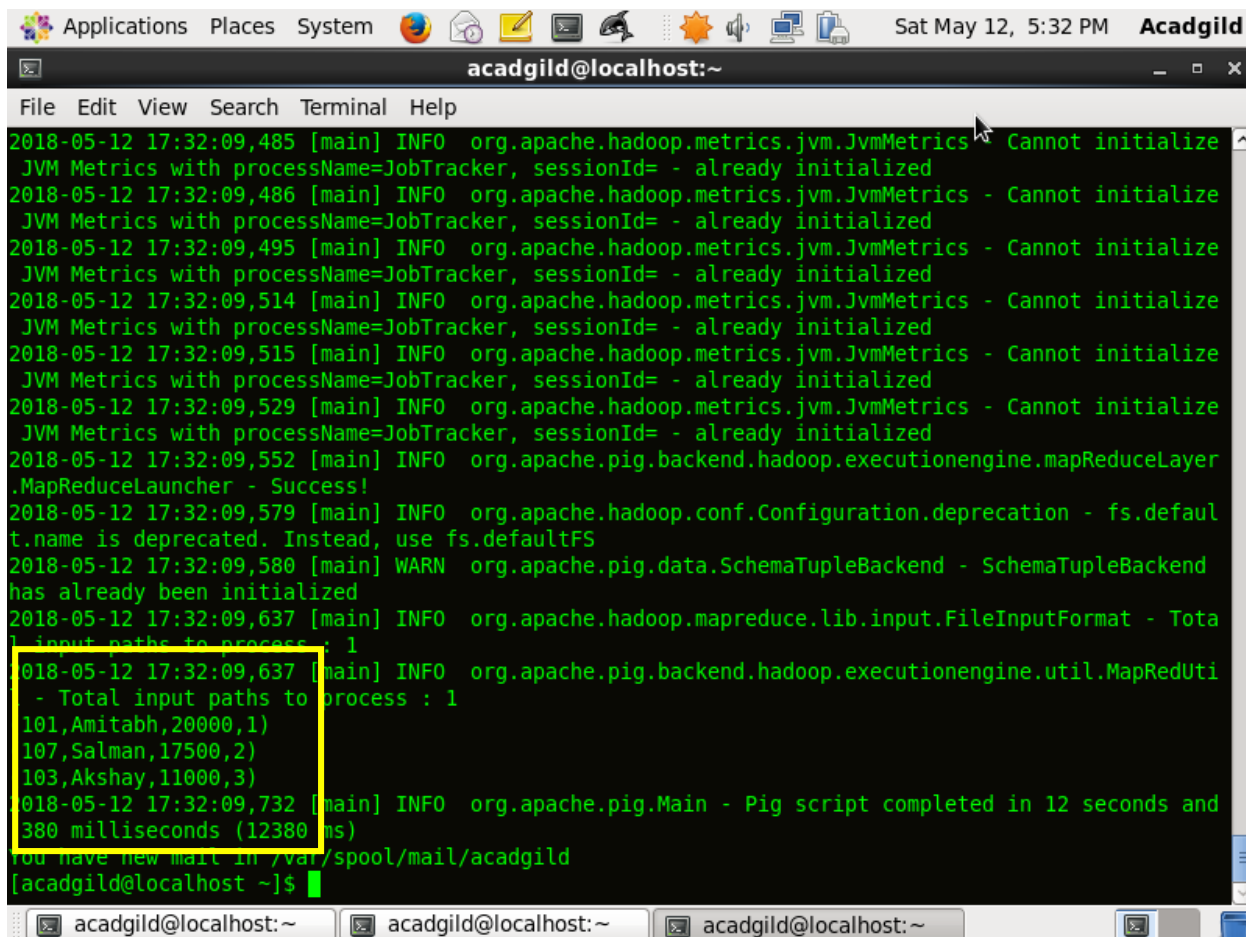
Then run the script in local mode to get the output



```
Applications Places System Sat May 12, 5:31 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ cat top3.pig
empdetails = load '/home/acadgild/employee_details.txt' using PigStorage(',') as (empid:int, name:chararray, salary:int, rating:int);
A = filter empdetails by not (empid%2==0);
B = order A by salary desc, name asc;
C = limit B 3;

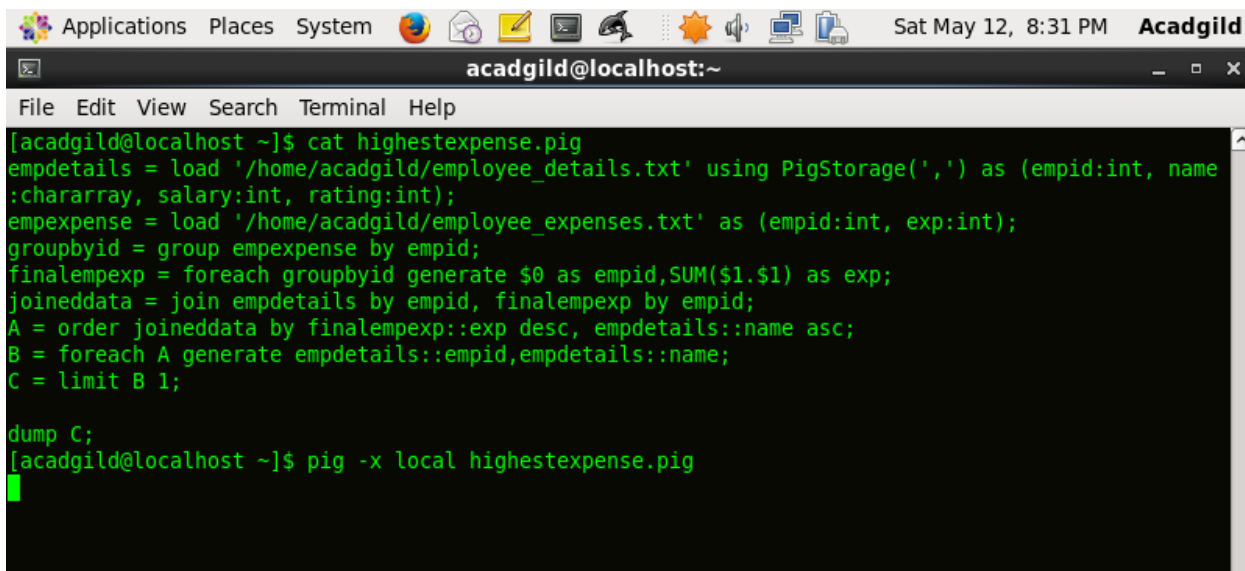
dump C;
[acadgild@localhost ~]$ pig -x local top3.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
```

As you can see, the employee ids are odd in the screenshot of the output:



```
Applications Places System Sat May 12, 5:32 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
2018-05-12 17:32:09,485 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:32:09,486 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:32:09,495 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:32:09,514 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:32:09,515 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:32:09,529 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 17:32:09,552 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-05-12 17:32:09,579 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-05-12 17:32:09,580 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-05-12 17:32:09,637 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-05-12 17:32:09,637 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
101,Amitabh,20000,1)
107,Salman,17500,2)
103,Akshay,11000,3)
2018-05-12 17:32:09,732 [main] INFO org.apache.pig.Main - Pig script completed in 12 seconds and 380 milliseconds (12380 ms)
you have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

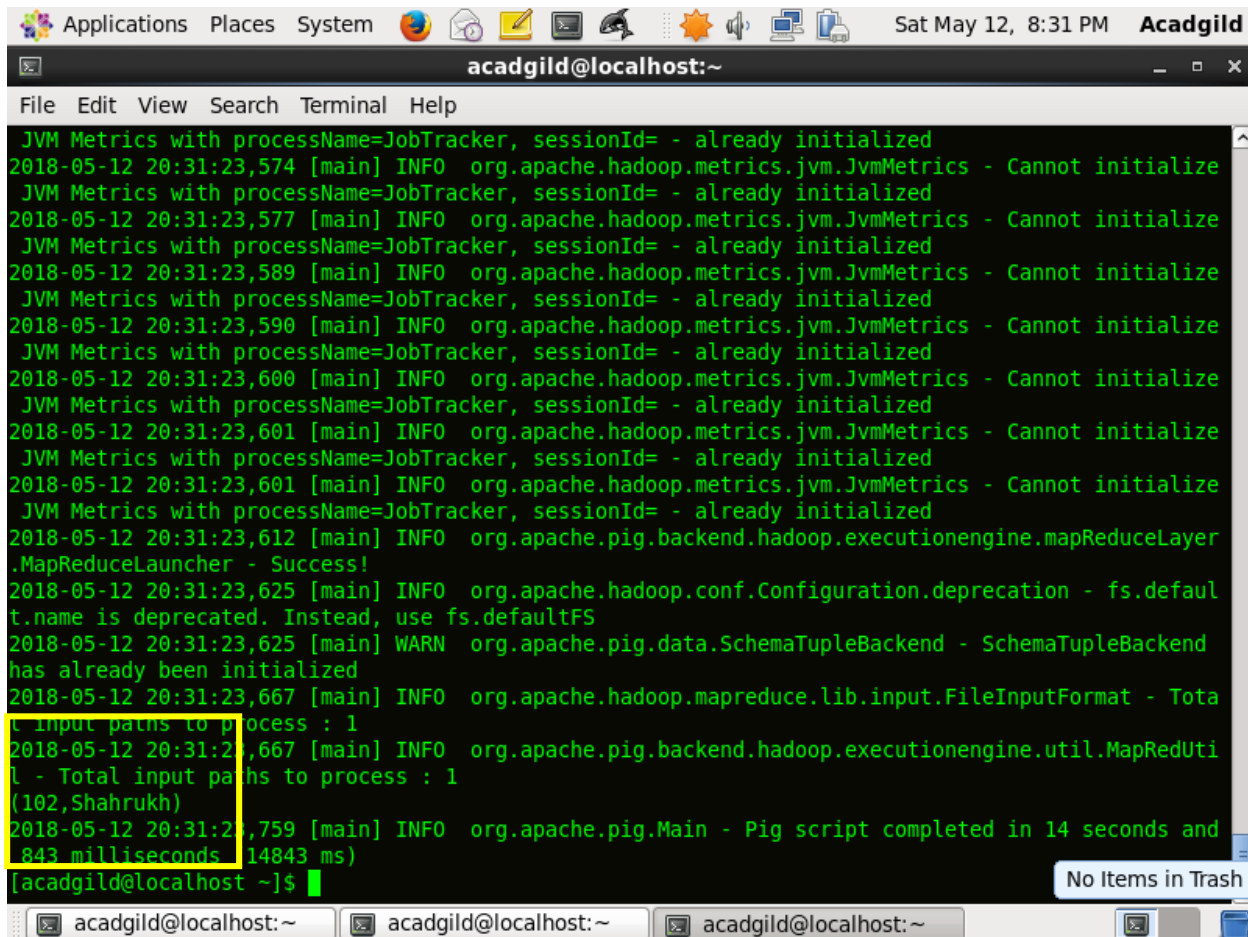
To show the employee id and name with highest expense one has to join the details and expense table. But the expense table has multiple entries of an employee id and its expenses. Thus these expenses has to be added to find the total expense of an employee. employee expenses are grouped by ids and then expenses added for each employee id. And Then joined with details table which is then ordered by expense and name. By limiting the output to 1 we get the employee with highest expense. The screenshot below the shows the pig script used to achieve the objective

A screenshot of a Linux terminal window. The window title is 'Acadgild' and the system clock shows 'Sat May 12, 8:31 PM'. The terminal prompt is '[acadgild@localhost ~]'. The user has entered a Pig script to find the employee with the highest expense. The script loads two tables: 'employee_details.txt' and 'employee_expenses.txt'. It groups the expenses by employee ID, calculates the sum for each, joins this with the employee details, orders the results by total expense in descending order and then by name in ascending order, and finally limits the output to one row. The script is saved as 'highestexpense.pig' and is executed with 'pig -x local highestexpense.pig'. The output is not visible in the screenshot.

```
[acadgild@localhost ~]$ cat highestexpense.pig
empdetails = load '/home/acadgild/employee_details.txt' using PigStorage(',') as (empid:int, name:chararray, salary:int, rating:int);
empexpense = load '/home/acadgild/employee_expenses.txt' as (empid:int, exp:int);
groupbyid = group empexpense by empid;
finalempexp = foreach groupbyid generate $0 as empid,SUM($1.$1) as exp;
joineddata = join empdetails by empid, finalempexp by empid;
A = order joineddata by finalempexp::exp desc, empdetails::name asc;
B = foreach A generate empdetails::empid,empdetails::name;
C = limit B 1;

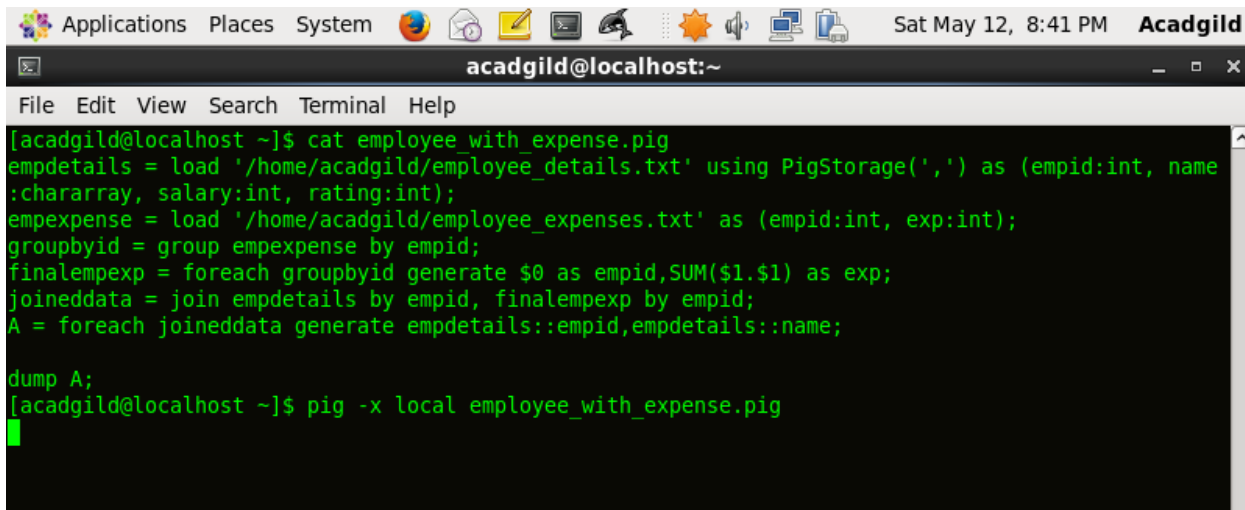
dump C;
[acadgild@localhost ~]$ pig -x local highestexpense.pig
```

Output:



```
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:31:23,574 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:31:23,577 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:31:23,589 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:31:23,590 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:31:23,600 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:31:23,601 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:31:23,612 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Success!
2018-05-12 20:31:23,625 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
t.name is deprecated. Instead, use fs.defaultFS
2018-05-12 20:31:23,625 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
has already been initialized
2018-05-12 20:31:23,667 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tota
l input paths to process : 1
2018-05-12 20:31:23,667 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
l - Total input paths to process : 1
(102,Shahrukh)
2018-05-12 20:31:23,759 [main] INFO org.apache.pig.Main - Pig script completed in 14 seconds and
843 milliseconds (14843 ms)
[acadgild@localhost ~]$
```

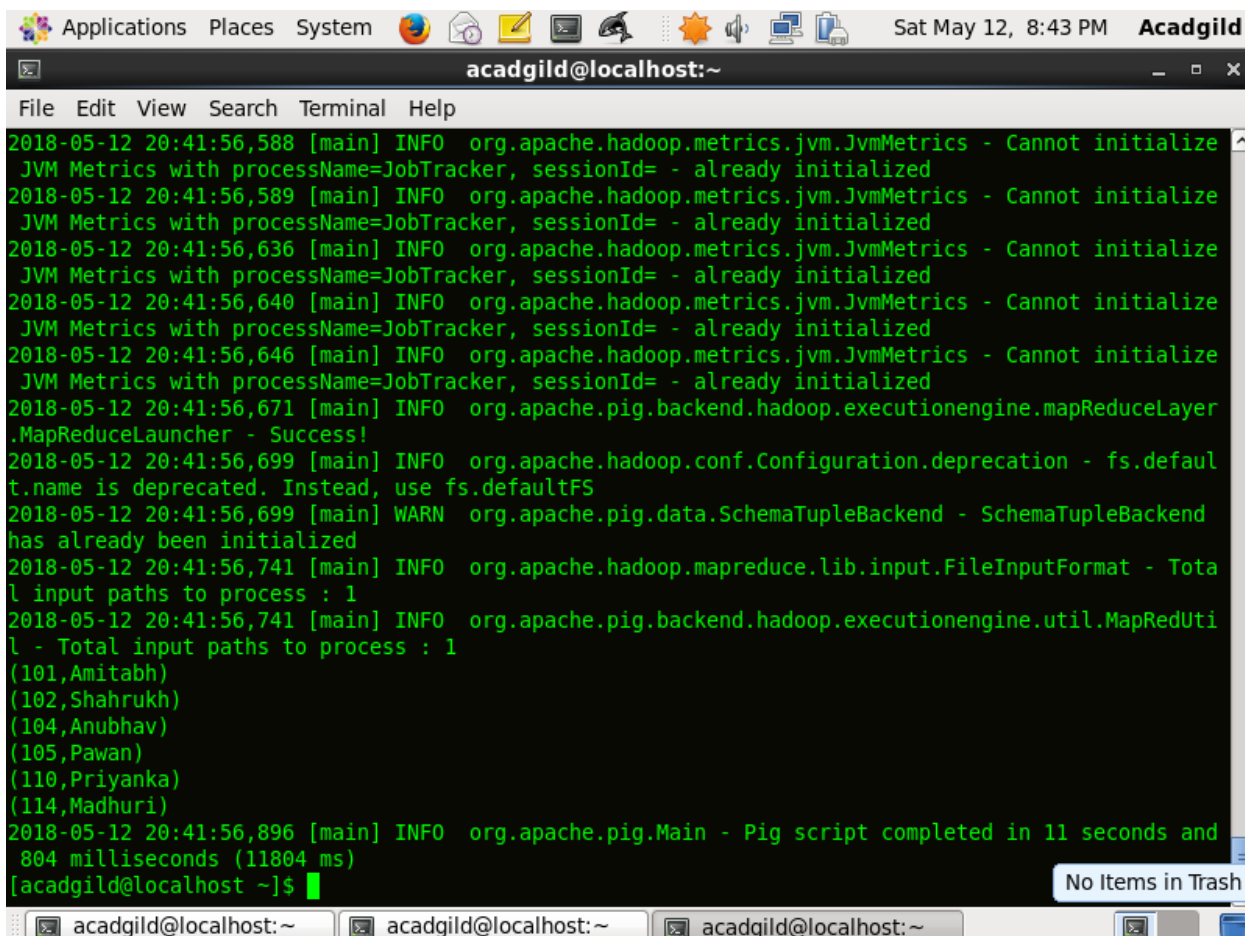
To find all the employees with entries in the expense file, we take a similar approach like the last problem. We do a simple join of the details table by id and the expense table by id. The joineddata variable will contain all the records wherever the ids from both tables have matched. No need to apply any order in the output. In the screenshot below one can see the code of the pig script employee_with_expense.pig



```
Applications Places System Sat May 12, 8:41 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ cat employee_with_expense.pig
empdetails = load '/home/acadgild/employee_details.txt' using PigStorage(',') as (empid:int, name:chararray, salary:int, rating:int);
empexpense = load '/home/acadgild/employee_expenses.txt' as (empid:int, exp:int);
groupbyid = group empexpense by empid;
finalempexp = foreach groupbyid generate $0 as empid,SUM($1.$1) as exp;
joineddata = join empdetails by empid, finalempexp by empid;
A = foreach joineddata generate empdetails::empid,empdetails::name;

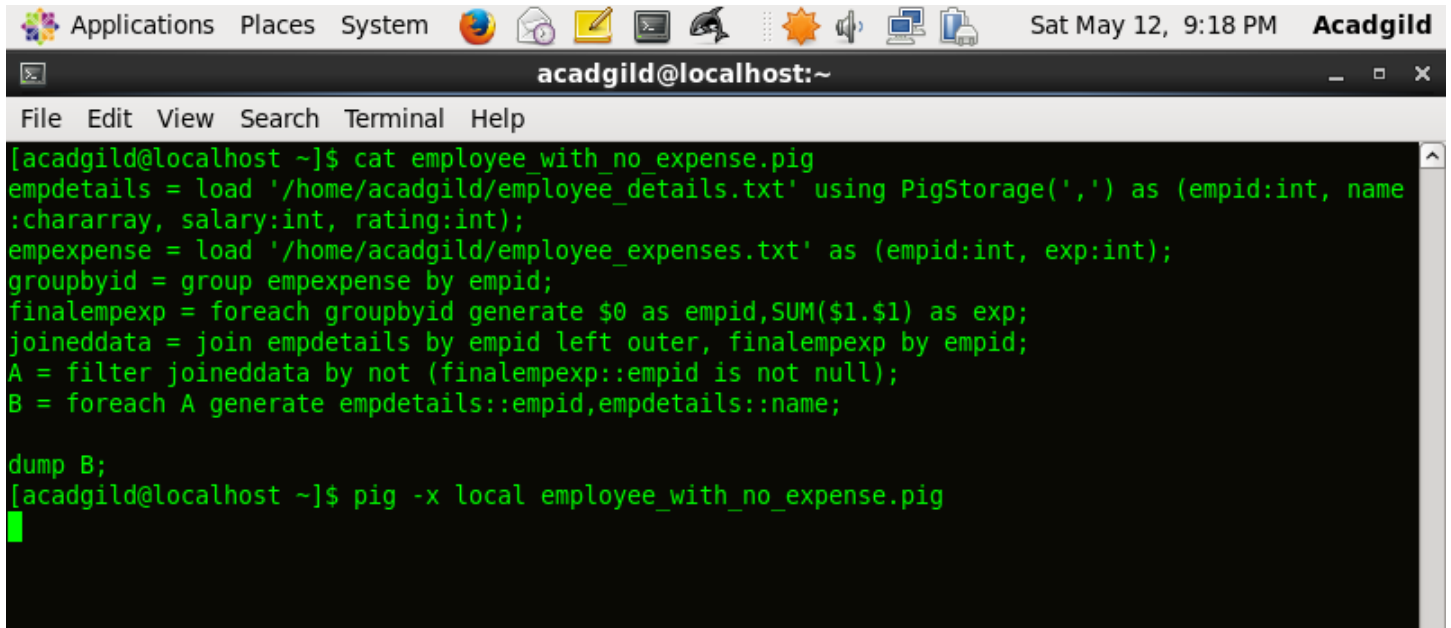
dump A;
[acadgild@localhost ~]$ pig -x local employee_with_expense.pig
```

Output:



```
Applications Places System Sat May 12, 8:43 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
2018-05-12 20:41:56,588 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:41:56,589 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:41:56,636 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:41:56,640 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:41:56,646 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 20:41:56,671 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Success!
2018-05-12 20:41:56,699 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
t.name is deprecated. Instead, use fs.defaultFS
2018-05-12 20:41:56,699 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
has already been initialized
2018-05-12 20:41:56,741 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tota
l input paths to process : 1
2018-05-12 20:41:56,741 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
l - Total input paths to process : 1
(101,Amitabh)
(102,Shahrukh)
(104,Anubhav)
(105,Pawan)
(110,Priyanka)
(114,Madhuri)
2018-05-12 20:41:56,896 [main] INFO org.apache.pig.Main - Pig script completed in 11 seconds and
804 milliseconds (11804 ms)
[acadgild@localhost ~]$
```

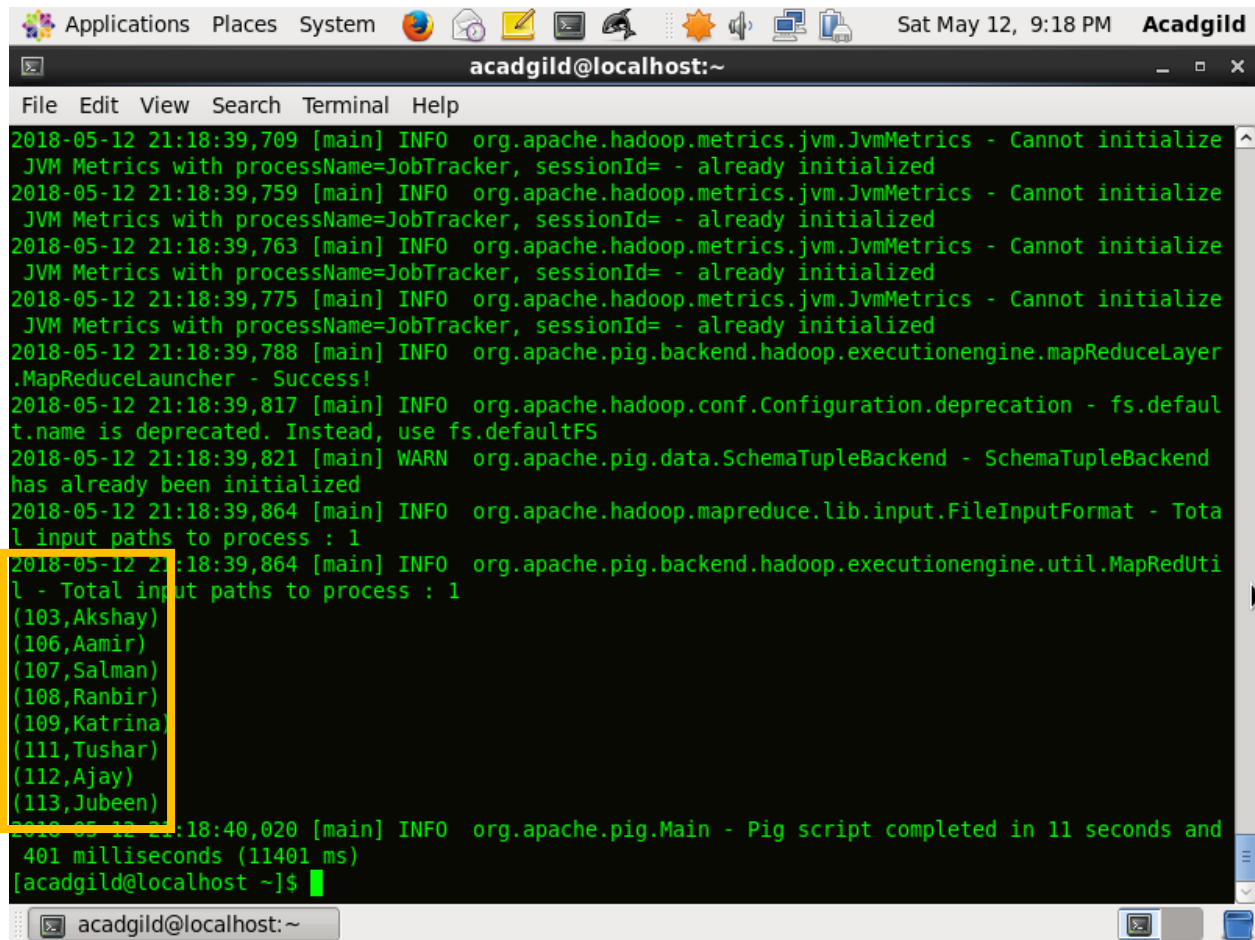
To find out the employees with no entries in the expense table, we do a left outer join where the left table data are retained in the joined entity irrespective of the fact that a match has found or no. So to find out are required data we just check whether a record in joineddata has data for expenses. All the records with no records of expenses are filtered and displayed. The code is given in the screenshot.

A screenshot of a Linux terminal window. The window title bar shows 'Applications Places System' on the left, a clock and 'Acadgild' on the right, and 'acadgild@localhost:~' in the center. The terminal has a menu bar with 'File Edit View Search Terminal Help'. The command prompt is '[acadgild@localhost ~]\$'. The user has entered a series of Pig Latin commands to load employee details and expenses, perform a left outer join, filter for employees with no expenses, and dump the results. The final command is 'pig -x local employee_with_no_expense.pig', and a green cursor is visible on the line following it.

```
[acadgild@localhost ~]$ cat employee_with_no_expense.pig
empdetails = load '/home/acadgild/employee_details.txt' using PigStorage(',') as (empid:int, name:chararray, salary:int, rating:int);
empexpense = load '/home/acadgild/employee_expenses.txt' as (empid:int, exp:int);
groupbyid = group empexpense by empid;
finalempexp = foreach groupbyid generate $0 as empid, SUM($1.$1) as exp;
joineddata = join empdetails by empid left outer, finalempexp by empid;
A = filter joineddata by not (finalempexp::empid is not null);
B = foreach A generate empdetails::empid, empdetails::name;

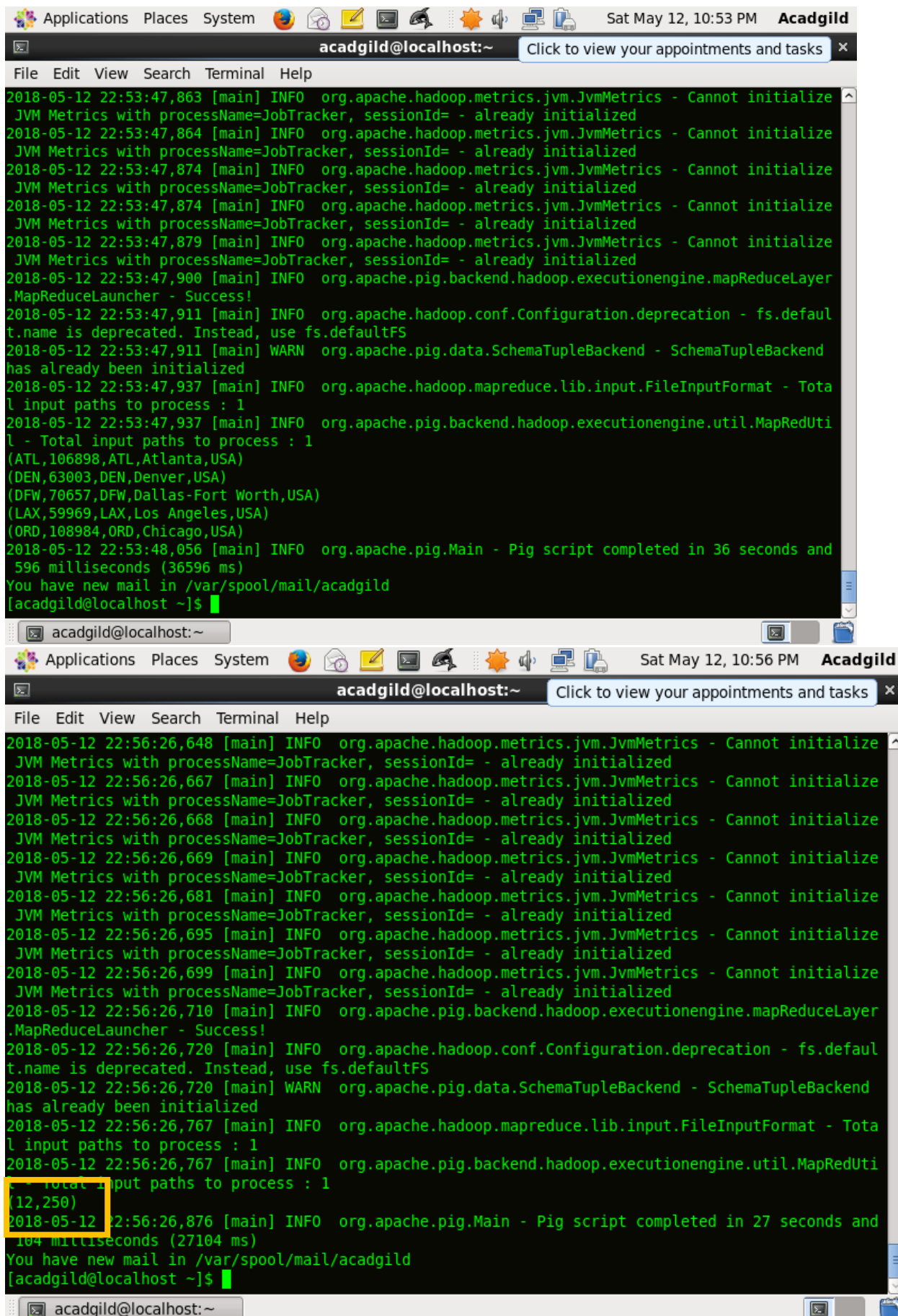
dump B;
[acadgild@localhost ~]$ pig -x local employee_with_no_expense.pig
```


Ouput:



```
Applications Places System Sat May 12, 9:18 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
2018-05-12 21:18:39,709 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 21:18:39,759 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 21:18:39,763 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 21:18:39,775 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 21:18:39,788 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Success!
2018-05-12 21:18:39,817 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
t.name is deprecated. Instead, use fs.defaultFS
2018-05-12 21:18:39,821 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
has already been initialized
2018-05-12 21:18:39,864 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tota
l input paths to process : 1
2018-05-12 21:18:39,864 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
l - Total input paths to process : 1
(103,Akshay)
(106,Aamir)
(107,Salman)
(108,Ranbir)
(109,Katrina)
(111,Tushar)
(112,Ajay)
(113,Jubeen)
2018-05-12 21:18:40,020 [main] INFO org.apache.pig.Main - Pig script completed in 11 seconds and
401 milliseconds (11401 ms)
[acadgild@localhost ~]$
```

Output of aviation data analysis case study:



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
2018-05-12 22:53:47,863 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:53:47,864 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:53:47,874 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:53:47,874 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:53:47,879 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:53:47,900 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer  
.MapReduceLauncher - Success!  
2018-05-12 22:53:47,911 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default  
t.name is deprecated. Instead, use fs.defaultFS  
2018-05-12 22:53:47,911 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend  
has already been initialized  
2018-05-12 22:53:47,937 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total  
input paths to process : 1  
2018-05-12 22:53:47,937 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti  
l - Total input paths to process : 1  
(ATL,106898,ATL,Atlanta,USA)  
(DEN,63003,DEN,Denver,USA)  
(DFW,70657,DFW,Dallas-Fort Worth,USA)  
(LAX,59969,LAX,Los Angeles,USA)  
(ORD,108984,ORD,Chicago,USA)  
2018-05-12 22:53:48,056 [main] INFO org.apache.pig.Main - Pig script completed in 36 seconds and  
596 milliseconds (36596 ms)  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$  
  
acadgild@localhost:~  
File Edit View Search Terminal Help  
2018-05-12 22:56:26,648 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:56:26,667 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:56:26,668 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:56:26,669 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:56:26,681 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:56:26,695 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:56:26,699 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize  
JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2018-05-12 22:56:26,710 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer  
.MapReduceLauncher - Success!  
2018-05-12 22:56:26,720 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default  
t.name is deprecated. Instead, use fs.defaultFS  
2018-05-12 22:56:26,720 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend  
has already been initialized  
2018-05-12 22:56:26,767 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total  
input paths to process : 1  
2018-05-12 22:56:26,767 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti  
l - Total input paths to process : 1  
(12,250)  
2018-05-12 22:56:26,876 [main] INFO org.apache.pig.Main - Pig script completed in 27 seconds and  
104 milliseconds (27104 ms)  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$
```

```
acadgild@localhost:~ Click to view your appointments and tasks
File Edit View Search Terminal Help
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 22:58:38,180 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 22:58:38,188 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 22:58:38,190 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Success!
2018-05-12 22:58:38,202 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
t.name is deprecated. Instead, use fs.defaultFS
2018-05-12 22:58:38,202 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
has already been initialized
2018-05-12 22:58:38,235 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total
input paths to process : 1
2018-05-12 22:58:38,235 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
l - Total input paths to process : 1
(CMX,Hancock,USA,116.1470588235294)
(PLN,Pellston,USA,93.76190476190476)
(SPI,Springfield,USA,83.84873949579831)
(ALO,Waterloo,USA,82.2258064516129)
(MOT,NA,USA,79.55665024630542)
(ACY,Atlantic City,USA,79.3103448275862)
(MOT,Minot,USA,78.66165413533835)
(HHH,NA,USA,76.53005464480874)
(EGE,Eagle,USA,74.12891986062718)
(BGM,Binghamton,USA,73.15533980582525)
2018-05-12 22:58:38,336 [main] INFO org.apache.pig.Main - Pig script completed in 37 seconds and
269 milliseconds (37269 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

```
acadgild@localhost:~ Click to view your appointments and tasks
File Edit View Search Terminal Help
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 22:59:50,193 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 22:59:50,194 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-05-12 22:59:50,197 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Success!
2018-05-12 22:59:50,206 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
t.name is deprecated. Instead, use fs.defaultFS
2018-05-12 22:59:50,206 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
has already been initialized
2018-05-12 22:59:50,230 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total
input paths to process : 1
2018-05-12 22:59:50,230 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
l - Total input paths to process : 1
((ORD,LGA),39)
((DAL,HOU),35)
((DFW,LGA),33)
((ATL,LGA),32)
((ORD,SNA),31)
((SLC,SUN),31)
((MIA,LGA),31)
((BUR,JFK),29)
((HRL,HOU),28)
((BUR,DFW),25)
2018-05-12 22:59:50,348 [main] INFO org.apache.pig.Main - Pig script completed in 27 seconds and
400 milliseconds (27400 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```