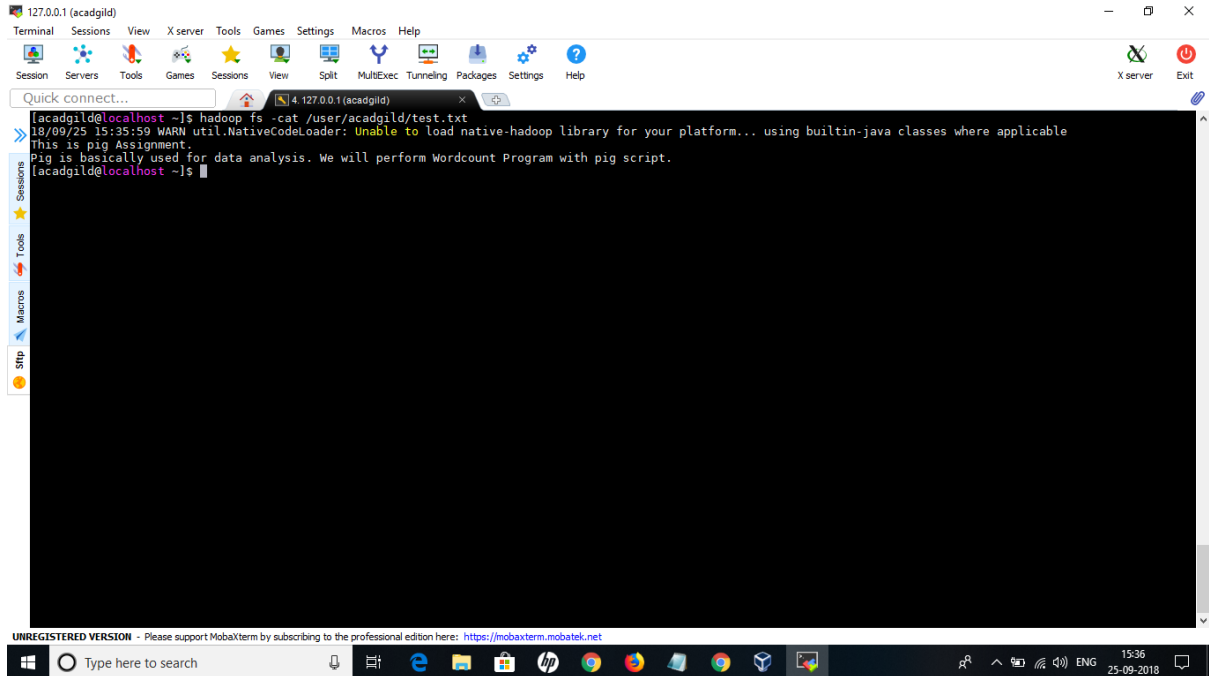


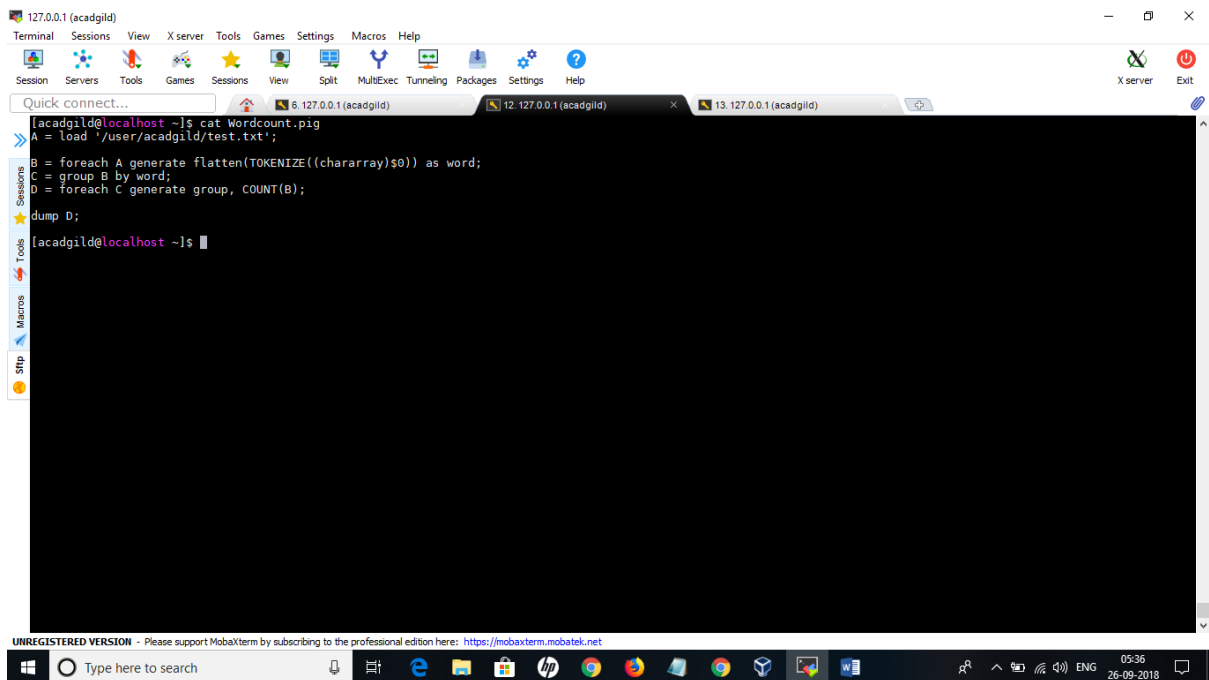
## Assignment No: 7(Exploring Apache Pig).

Task 1: Write a program to implement wordcount using Pig.

Solution: Please find the below attached snapshot:



```
127.0.0.1 (acagild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 4 127.0.0.1 (acagild)
[acagild@localhost ~]$ hadoop fs -cat /user/acagild/test.txt
18/09/25 15:35:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
This is pig Assignment.
Pig is basically used for data analysis. We will perform Wordcount Program with pig script.
[acagild@localhost ~]$
```



```
127.0.0.1 (acagild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 6 127.0.0.1 (acagild) 12 127.0.0.1 (acagild) 13 127.0.0.1 (acagild)
[acagild@localhost ~]$ cat Wordcount.pig
A = load '/user/acagild/test.txt';
B = foreach A generate flatten(TOKENIZE((chararray)$0)) as word;
C = group B by word;
D = foreach C generate group, COUNT(B);
dump D;
[acagild@localhost ~]$
```

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
[acadgild@localhost ~]$ pig Wordcount.pig
18/09/25 15:37:06 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/09/25 15:37:06 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/09/25 15:37:06 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-09-25 15:37:06,480 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-09-25 15:37:06,480 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/pig_1537870026475.log
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]
2018-09-25 15:37:07,412 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-09-25 15:37:07,889 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-09-25 15:37:08,257 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2018-09-25 15:37:08,258 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2018-09-25 15:37:09,344 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Wordcount.pig-d8738559-0ac0-4185-aed4-3428d2095f7b
2018-09-25 15:37:09,344 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline.service.enabled set to false
2018-09-25 15:37:10,421 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 15:37:10,579 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-09-25 15:37:10,926 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP BY
2018-09-25 15:37:11,019 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 15:37:11,029 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schema.tuple] was not set... will not generate code.
2018-09-25 15:37:11,102 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-09-25 15:37:11,253 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2018-09-25 15:37:11,295 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combiner
2018-09-25 15:37:11,302 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2018-09-25 15:37:11,392 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2018-09-25 15:37:11,464 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 15:37:11,542 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 15:37:12,108 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2018-09-25 15:37:12,125 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markrest.buffer.percent is deprecated. Instead, use mapred
```

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
Total records proactively spilled: 0
Job DAG:
job_1537869398305_0002
2018-09-25 15:37:59,448 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 15:37:59,457 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 15:37:59,556 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 15:37:59,569 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 15:37:59,683 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 15:37:59,694 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 15:37:59,805 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 15:37:59,817 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 15:37:59,920 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schema.tuple] was not set... will not generate code.
2018-09-25 15:37:59,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2018-09-25 15:37:59,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(We,1)
(is,2)
(pig,1)
(for,1)
(pig,2)
(This,1)
(data,1)
(used,1)
(will,1)
(with,1)
(Program,1)
(perform,1)
(script,1)
(wordcount,1)
(analysis,1)
(basically,1)
(assignment,1)
2018-09-25 15:38:00,128 [main] INFO org.apache.pig.Main - Pig script completed in 54 seconds and 297 milliseconds (54297 ms)
```

Task 2(a) 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).

Solution: Please see the below screenshots:

The screenshot shows a MobaXterm terminal window with the title '127.0.0.1 (acadgild)'. The terminal displays the execution of a Pig Latin script. The script defines a relation A from 'employee\_details.txt', orders it by rating, generates columns EmpID and EmpName, and limits the results to 5 rows. The output is stored in 'pig/employee/Task2a\_out'. Comments explain the steps: loading data, ordering by rating, generating columns, and limiting to 5 rows. The terminal prompt is [acadgild@localhost Desktop]\$. The window includes a menu bar (Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, Help) and a toolbar with icons for Session, Servers, Tools, Games, Sessions, View, Split, MultiExec, Tunneling, Packages, Settings, and Help. A sidebar on the left has buttons for Session, Servers, Tools, Games, Sessions, View, Split, MultiExec, Tunneling, Packages, Settings, and Help. The bottom status bar shows 'UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>' and the system tray with icons for search, network, volume, and date/time (01:43, 26-09-2018).

```
[acadgild@localhost Desktop]$ cat Task2_a.pig
/* Top 5 employees (employee id and employee name) with highest rating */
A = LOAD 'employee_details.txt' using PigStorage(',') AS (EmpID:int,EmpName:chararray,EmpSalary:double,Rating:int);
B = order A by Rating, EmpName ASC;
C = FOREACH B GENERATE EmpID,EmpName;
D = LIMIT C 5;
store D into 'pig/employee/Task2a_out';

/* In relation A we load the employee_details,
In relation B we arrange the empID and employee name in ascending order
In relation C we are generating two required columns empID and empName
In relation D we limit it till top 5 rows */
[acadgild@localhost Desktop]$
```

The screenshot shows the same MobaXterm terminal window after executing the Pig Latin script. The output of the 'cat Task2a\_out/part-r-00000' command is displayed, showing a list of employee IDs and names: 106 Amir, 101 Amitabh, 113 Jubeen, 111 Tushar, and 112 Ajay. A system message indicates new mail in the spool directory. The terminal prompt is [acadgild@localhost employee]\$. The window interface is identical to the previous screenshot.

```
[acadgild@localhost employee]$ cat Task2a_out/part-r-00000
106 Amir
101 Amitabh
113 Jubeen
111 Tushar
112 Ajay
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost employee]$
```

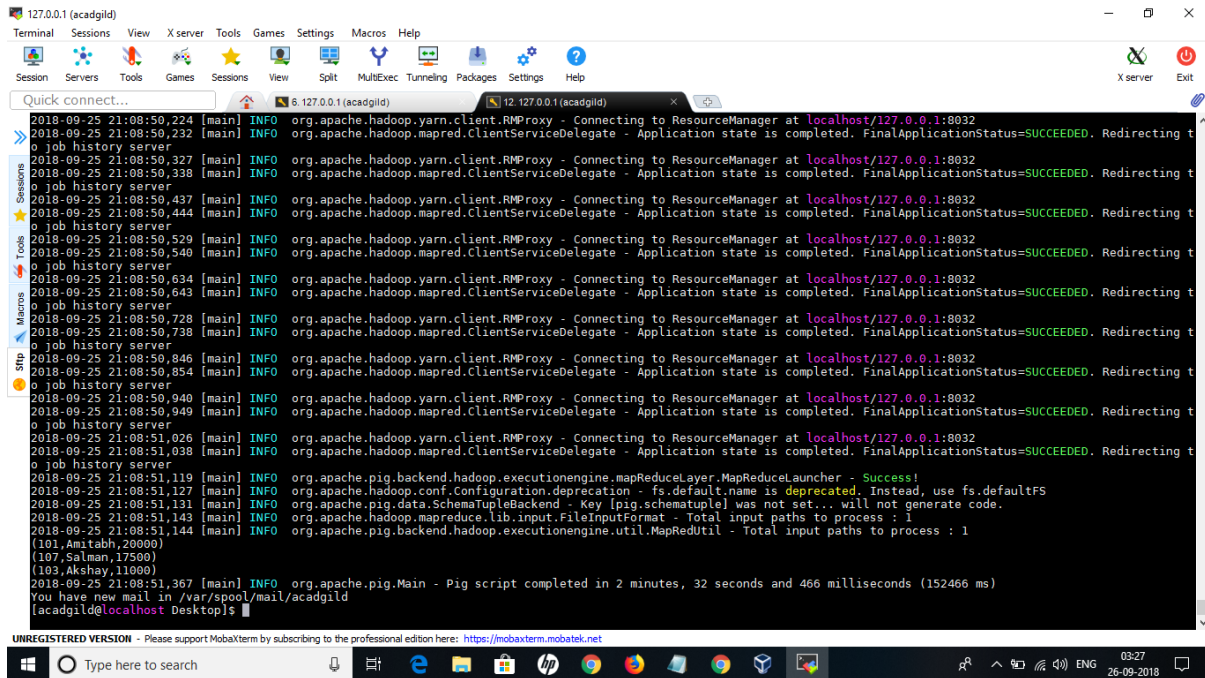
2(b) Top 3 employees (employee id and employee name) with highest salary, whose employee id is an odd number. (In case two employees have same salary, employee with name coming first in dictionary should get preference)

Solution: Please see the below attached screenshot :

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
[acadgild@localhost Desktop]$ cat Task2_b.pig
/* Top 3 employees (employee id and employee name) with highest salary, whose employee id is an odd number */
A = LOAD '/pig/employee/employee_details.txt' using PigStorage(',') AS (EmpID:int,EmpName:chararray,EmpSalary:long,Rating:int);
B = order A by EmpSalary DESC;
C = Filter B by EmpID%2==1;
D = FOREACH C GENERATE EmpID, EmpName, EmpSalary;
E = Limit D 3;

dump E;
/* In relation A we load the employee details.
In relation B we arrange the employee salary in descending order
In relation C we filter out empID with odd number
In relation D we are generating required columns empID and empName and empSalary
In relation E we limit it till top 3 rows */
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
2018-09-25 21:06:06,788 [main] INFO org.apache.pig.Main - Pig script completed in 6 seconds and 387 milliseconds (6387 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ pig Task2_b.pig
18/09/25 21:06:19 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/09/25 21:06:19 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/09/25 21:06:19 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-09-25 21:06:19,547 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-09-25 21:06:19,547 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/pig_1537889779544.log
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]
2018-09-25 21:06:20,458 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-09-25 21:06:20,955 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/acadgild/.pigbootup not found
2018-09-25 21:06:21,329 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2018-09-25 21:06:21,329 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:06:21,330 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2018-09-25 21:06:22,459 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task2_b.pig-50ccbdfa-e19a-449e-b470-51e49ecc8ba4
2018-09-25 21:06:22,459 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-09-25 21:06:23,641 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:06:24,061 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: ORDER_BY,FILTER,LIMIT
2018-09-25 21:06:24,204 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:06:24,219 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-25 21:06:24,302 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED:[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-09-25 21:06:24,358 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A: $3
2018-09-25 21:06:24,446 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-09-25 21:06:24,553 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2018-09-25 21:06:24,661 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope:30
2018-09-25 21:06:24,698 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 4
2018-09-25 21:06:24,699 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 4
2018-09-25 21:06:24,790 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

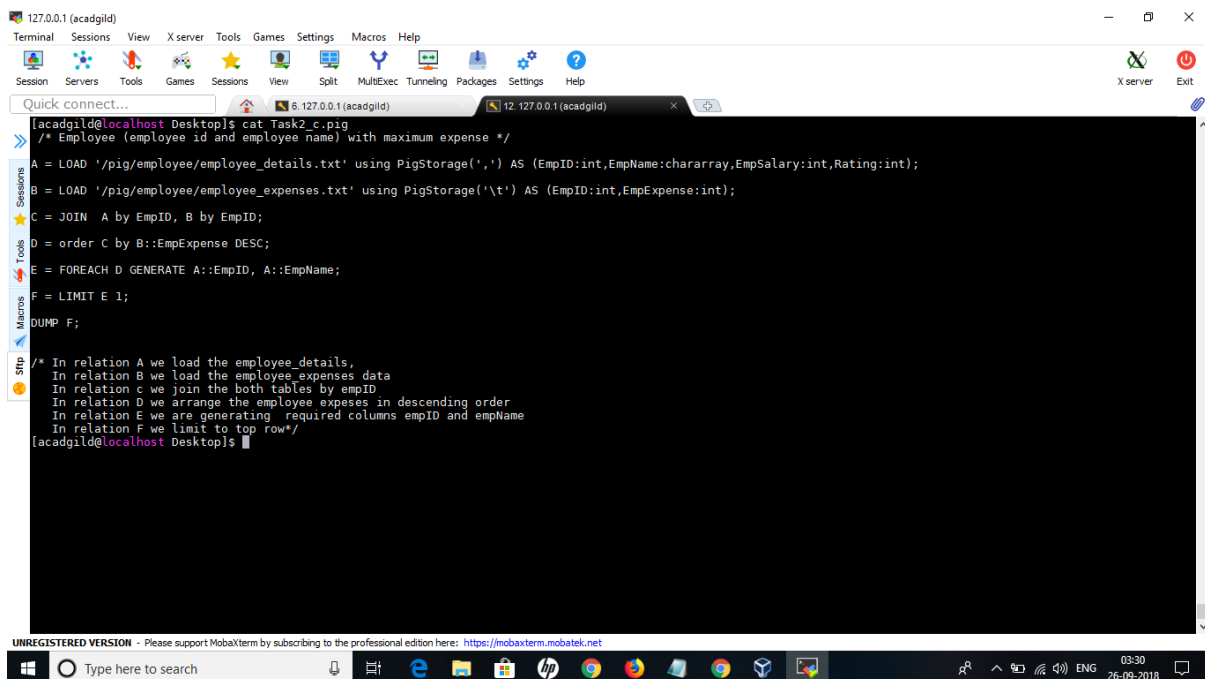


The screenshot shows a MobaXterm terminal window with a dark background. The title bar indicates the session is '127.0.0.1 (acadgild)'. The terminal displays a series of log messages from Hadoop and Pig. The logs show the connection of the RMPProxy to the ResourceManager at localhost/127.0.0.1:8032, the completion of the Application state, and the redirection of the job history server. The logs also show the execution of the Pig script, which includes a deprecation warning for 'fs.default.name' and a successful execution of the MapReduce job. The final output of the Pig script is displayed at the bottom of the terminal.

```
2018-09-25 21:08:50,224 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,232 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:50,327 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,338 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:50,437 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,444 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:50,529 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,540 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:50,634 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,643 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:50,728 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,738 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:50,846 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,854 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:50,940 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:50,949 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:51,026 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:08:51,038 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:08:51,119 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 21:08:51,127 [main] INFO org.apache.pig.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:08:51,131 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key (pig.schematuple) was not set... Will not generate code.
2018-09-25 21:08:51,143 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 21:08:51,144 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh,20000)
(107,Salman,17500)
(103,Akshay,11000)
2018-09-25 21:08:51,367 [main] INFO org.apache.pig.Main - Pig script completed in 2 minutes, 32 seconds and 466 milliseconds (152466 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

2(c) Employee (employee id and employee name) with maximum expense (In case two employees have same expense, employee with name coming first in dictionary should get preference).

Solution: Please find the attached snapshot:



The screenshot shows a MobaXterm terminal window with a dark background. The title bar indicates the session is '127.0.0.1 (acadgild)'. The terminal displays a Pig script that loads two tables, 'employee\_details.txt' and 'employee\_expenses.txt', and joins them by employee ID. The script then orders the results by employee expense in descending order and limits the output to the top row. The final output of the script is displayed at the bottom of the terminal.

```
[acadgild@localhost Desktop]$ cat Task2_c.pig
/* Employee (employee id and employee name) with maximum expense */
A = LOAD '/pig/employee/employee_details.txt' using PigStorage(',') AS (EmpID:int,EmpName:chararray,EmpSalary:int,Rating:int);
B = LOAD '/pig/employee/employee_expenses.txt' using PigStorage('\t') AS (EmpID:int,EmpExpense:int);
C = JOIN A by EmpID, B by EmpID;
D = order C by B::EmpExpense DESC;
E = FOREACH D GENERATE A::EmpID, A::EmpName;
F = LIMIT E 1;
DUMP F;

/* In relation A we load the employee_details,
In relation B we load the employee_expenses data
In relation C we join the both tables by empID
In relation D we arrange the employee expenses in descending order
In relation E we are generating required columns empID and empName
In relation F we limit to top row*/
[acadgild@localhost Desktop]$
```



```
127.0.0.1 (acadgild) Desktop$
acadgild@localhost Desktop$ pig Task2_c.pig
18/09/25 21:13:16 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/09/25 21:13:16 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/09/25 21:13:16 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-09-25 21:13:16,266 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-09-25 21:13:16,267 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/pig_1537890196264.log
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]
2018-09-25 21:13:17,125 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-09-25 21:13:17,617 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-09-25 21:13:17,995 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2018-09-25 21:13:17,995 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:13:17,995 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2018-09-25 21:13:19,103 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task2_c.pig-7ef8ba84-59de-4b8f-b980-d4ebf018d2cb
2018-09-25 21:13:19,107 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline.service.enabled set to false
2018-09-25 21:13:20,281 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:13:20,496 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:13:20,934 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-09-25 21:13:20,987 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN,ORDER_BY,LIMIT
2018-09-25 21:13:21,135 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:13:21,159 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schemaTuple] was not set... will not generate code.
2018-09-25 21:13:21,328 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConsParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-09-25 21:13:21,426 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A: $2, $3
2018-09-25 21:13:21,606 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2018-09-25 21:13:21,697 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope-48
2018-09-25 21:13:21,765 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage$POForEach to POPackage$JoinPackage
2018-09-25 21:13:21,729 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 4
```

```
127.0.0.1 (acadgild) Desktop$
acadgild@localhost Desktop$ pig Task2_c.pig
2018-09-25 21:16:07,118 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,233 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,239 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,356 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,378 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,475 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,481 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,581 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,588 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,676 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,689 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,787 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,796 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,885 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,891 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:07,966 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:07,983 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:08,091 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:16:08,097 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:16:08,184 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 21:16:08,197 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:16:08,198 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schemaTuple] was not set... will not generate code.
2018-09-25 21:16:08,213 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 21:16:08,213 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
2018-09-25 21:16:08,435 [main] INFO org.apache.pig.Main - Pig script completed in 2 minutes, 52 seconds and 872 milliseconds (172872 ms)
You have new mail in /var/spool/mail/acadgild
acadgild@localhost Desktop$
```

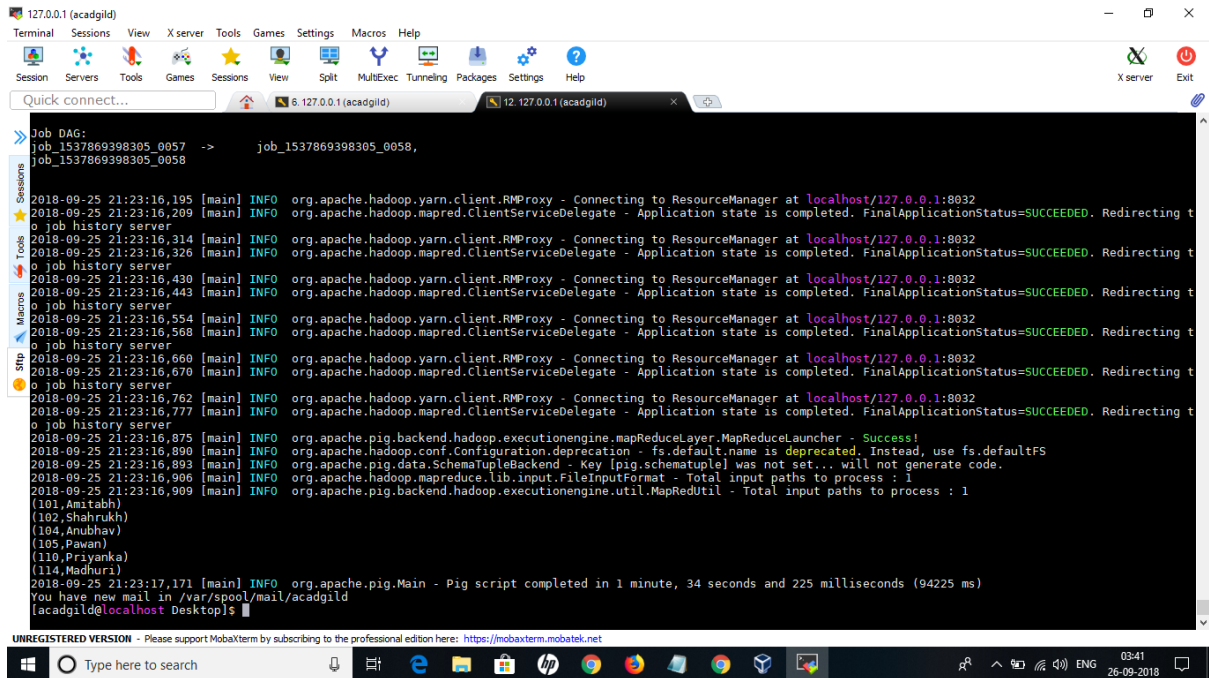
2(d) List of employees (employee id and employee name) having entries in employee\_expenses file.

Solution: Please find the below attached screenshot:

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
[acadgild@localhost Desktop]$ nano Task2_d.pig
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ cat Task2_d.pig
/*List of employees (employee id and employee name) having entries in employee_expenses file */
A = LOAD '/pig/employee/employee_details.txt' using PigStorage(',') AS (EmpID:int,EmpName:chararray,EmpSalary:int,Rating:int);
B = LOAD '/pig/employee/employee_expenses.txt' using PigStorage('\t') AS (EmpID:int,EmpExpense:int);
C = JOIN A by EmpID, B by EmpID;
D = FOREACH C GENERATE A::EmpID, A::EmpName;
E = DISTINCT D;
DUMP E;

/* In relation A we load the employee_details,
In relation B we load the employee_expenses data
In relation C we join the both tables by empID
In relation D we are generating required columns empID and empName
In relation E we generating empID and empName in relation D */
[acadgild@localhost Desktop]$
```

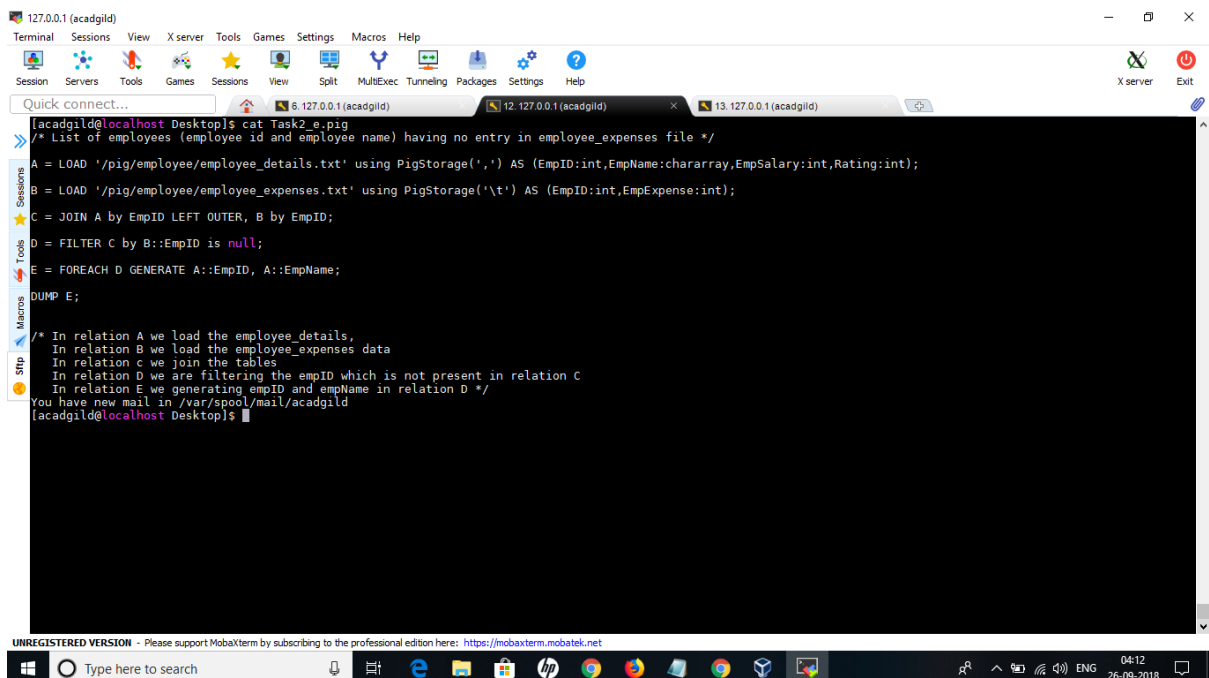
```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
[acadgild@localhost Desktop]$ pig Task2_d.pig
18/09/25 21:21:43 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/09/25 21:21:43 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/09/25 21:21:43 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-09-25 21:21:43,614 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-09-25 21:21:43,614 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/pig_1537890703608.log
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]
2018-09-25 21:21:44,520 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-09-25 21:21:45,054 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/acadgild/pigbootstrap not found
2018-09-25 21:21:45,423 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use fs.defaultFS
2018-09-25 21:21:45,423 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:21:45,423 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2018-09-25 21:21:46,512 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task2_d.pig-a967cbc6-0e2f-4f75-a251-6d1f55f7852f
2018-09-25 21:21:46,512 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-09-25 21:21:47,707 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:21:47,935 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:21:48,293 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-09-25 21:21:48,327 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN,DISTINCT
2018-09-25 21:21:48,436 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:21:48,453 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pigsschematuple] was not set... will not generate code.
2018-09-25 21:21:48,572 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-09-25 21:21:48,632 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A: $2, $3
2018-09-25 21:21:48,639 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for B: $1
2018-09-25 21:21:48,805 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2018-09-25 21:21:48,866 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEach to POPackage(JoinPackager)
2018-09-25 21:21:48,883 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2
2018-09-25 21:21:48,884 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2
```



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
Job DAG:
job_1537869398305_0057 -> job_1537869398305_0058,
job_1537869398305_0058
2018-09-25 21:23:16,185 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:23:16,209 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:23:16,314 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:23:16,326 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:23:16,430 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:23:16,443 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:23:16,554 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:23:16,568 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:23:16,660 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:23:16,670 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:23:16,762 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:23:16,777 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-09-25 21:23:16,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 21:23:16,890 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:23:16,893 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-25 21:23:16,906 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 21:23:16,909 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh)
(102,Shahrukh)
(104,Anubhav)
(105,Pawan)
(110,Priyanka)
(114,Madhuri)
2018-09-25 21:23:17,171 [main] INFO org.apache.pig.Main - Pig script completed in 1 minute, 34 seconds and 225 milliseconds (94225 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

2(e) List of employees (employee id and employee name) having no entry in employee\_expenses file.

Solution: Please find the attached snapshot:



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
[acadgild@localhost Desktop]$ cat Task2.e.pig
/* List of employees (employee id and employee name) having no entry in employee_expenses file */
A = LOAD '/pig/employee/employee_details.txt' using PigStorage(',') AS (EmpID:int,EmpName:chararray,EmpSalary:int,Rating:int);
B = LOAD '/pig/employee/employee_expenses.txt' using PigStorage('\t') AS (EmpID:int,EmpExpense:int);
C = JOIN A by EmpID LEFT OUTER, B by EmpID;
D = FILTER C by B::EmpID is null;
E = FOREACH D GENERATE A::EmpID, A::EmpName;
DUMP E;

/* In relation A we load the employee_details,
In relation B we load the employee_expenses data
In relation C we join the tables
In relation D we are filtering the empID which is not present in relation C
In relation E we generating empID and empName in relation D */
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ pig Task2_e.pig
18/09/25 21:55:15 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/09/25 21:55:15 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/09/25 21:55:15 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-09-25 21:55:15,324 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:16:49
2018-09-25 21:55:15,325 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/pig_1537892715321.log
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]
2018-09-25 21:55:16,209 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-09-25 21:55:16,687 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-09-25 21:55:17,059 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2018-09-25 21:55:17,059 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:55:17,063 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2018-09-25 21:55:18,132 [main] INFO org.apache.pig.PigServer - Pig script ID for the session: PIG-Task2_e.pig-c806996a-1ae6-4958-8348-addcbfa0c020
2018-09-25 21:55:18,132 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-09-25 21:55:19,297 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:55:19,520 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:55:19,861 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThresh
hold = 489350752, usageThreshold = 489350752
2018-09-25 21:55:19,897 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH JOIN, FILTER
2018-09-25 21:55:20,002 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:55:20,018 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-25 21:55:20,130 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculato
r, GroupByConsistentSorter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEach
Flatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-09-25 21:55:20,218 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A: $2, $3
2018-09-25 21:55:20,224 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for B: $1
2018-09-25 21:55:20,417 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2018-09-25 21:55:20,482 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2018-09-25 21:55:20,482 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2018-09-25 21:55:20,564 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:55:20,582 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
Successfully stored 8 records (118 bytes) in: "hdfs://localhost:8020/tmp/temp671180332/tmp-890713397"
Counters:
Total records written : 8
Total bytes written : 118
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_153789398305_0064
2018-09-25 21:56:10,953 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:56:10,982 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:56:11,113 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:56:11,130 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:56:11,239 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 21:56:11,254 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 21:56:11,362 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 21:56:11,373 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 21:56:11,377 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-25 21:56:11,429 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 21:56:11,429 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(108, Akshay)
(106, Aamir)
(107, Salman)
(108, Ranbir)
(109, Katrina)
(111, Tushar)
(112, Ajay)
(113, Jubeen)
2018-09-25 21:56:11,688 [main] INFO org.apache.pig.Main - Pig script completed in 57 seconds and 30 milliseconds (57030 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

3(a) Find out the top 5 most visited destinations.

Solution: CMD:

REGISTER '/home/acadgild/Desktop/piggybank.jar';

A = load '/airline\_usecase/DelayedFlights.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage('','NO\_MULTILINE','UNIX','SKIP\_INPUT\_  
HEADER');

B = foreach A generate (int)\$1 as year, (int)\$10 as flight\_num, (chararray)\$17 as origin, (chararray) \$18 as dest;

C = filter B by dest is not null;

D = group C by dest;

E = foreach D generate group, COUNT(C.dest);

F = order E by \$1 DESC;

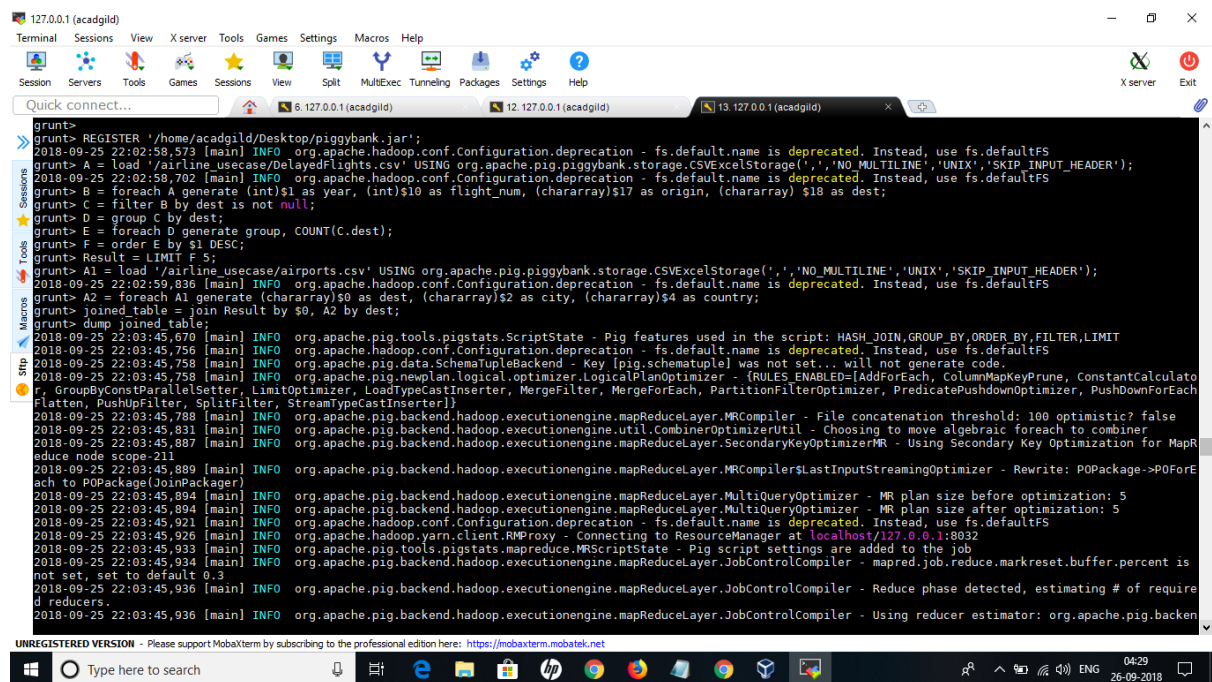
Result = LIMIT F 5;

A1 = load '/home/acadgild/airline\_usecase/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage('','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

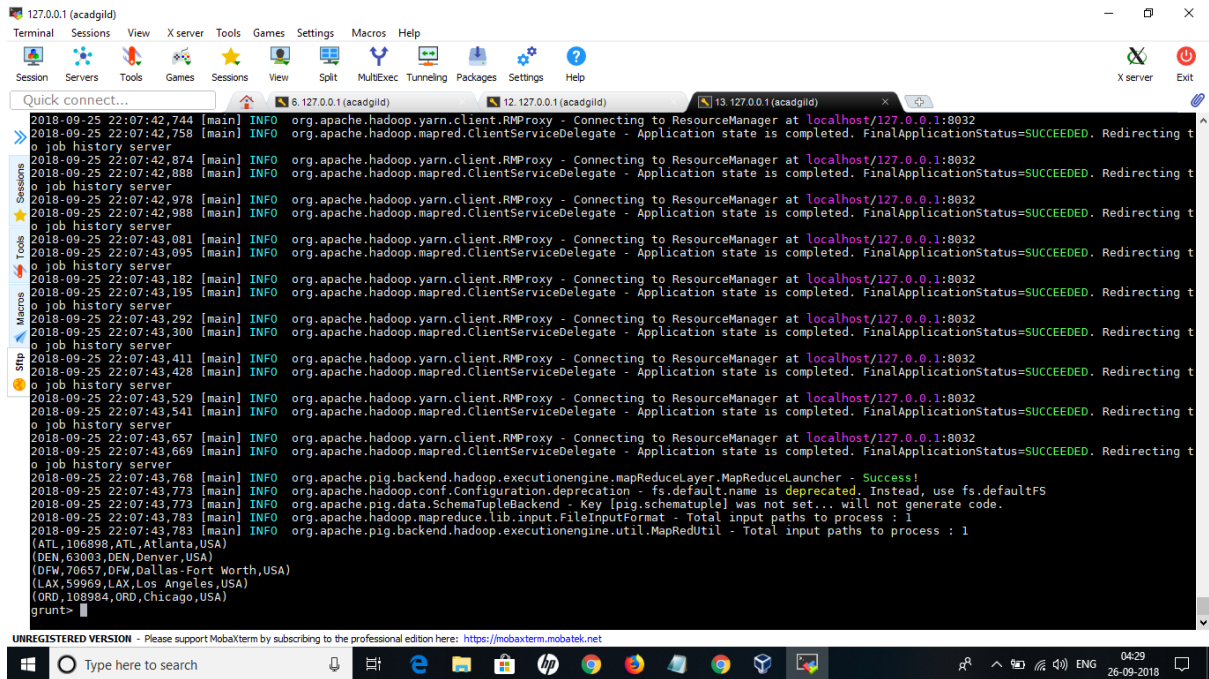
A2 = foreach A1 generate (chararray)\$0 as dest, (chararray)\$2 as city, (chararray)\$4 as country;

joined\_table = join Result by \$0, A2 by dest;

dump joined\_table;



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
grunt> REGISTER '/home/acadgild/Desktop/piggybank.jar';
2018-09-25 22:02:58,573 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A = load '/airline_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage('','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2018-09-25 22:02:58,702 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = foreach A generate (int)$1 as year, (int)$10 as flight_num, (chararray)$17 as origin, (chararray) $18 as dest;
grunt> C = filter B by dest is not null;
grunt> D = group C by dest;
grunt> E = foreach D generate group, COUNT(C.dest);
grunt> F = order E by $1 DESC;
grunt> Result = LIMIT F 5;
grunt> A1 = load '/airline_usecase/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage('','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2018-09-25 22:02:59,836 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A2 = foreach A1 generate (chararray)$0 as dest, (chararray)$2 as city, (chararray)$4 as country;
grunt> joined_table = join Result by $0, A2 by dest;
grunt> dump joined_table;
2018-09-25 22:03:45,670 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN, GROUP_BY, ORDER_BY, FILTER, LIMIT
2018-09-25 22:03:45,756 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 22:03:45,758 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code
2018-09-25 22:03:45,758 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[addForEach, columnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-09-25 22:03:45,788 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2018-09-25 22:03:45,831 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combiner
2018-09-25 22:03:45,887 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope=211
2018-09-25 22:03:45,889 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEach to POPackage(JoinPackager)
2018-09-25 22:03:45,894 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 5
2018-09-25 22:03:45,894 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 5
2018-09-25 22:03:45,921 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 22:03:45,926 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:03:45,933 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2018-09-25 22:03:45,934 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2018-09-25 22:03:45,936 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers
2018-09-25 22:03:45,936 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backen
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
Type here to search
```



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
127.0.0.1 (acadgild) 12 127.0.0.1 (acadgild) 13 127.0.0.1 (acadgild)
2018-09-25 22:07:42,744 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:42,758 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:42,874 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:42,888 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:42,978 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:42,988 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:43,081 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:43,095 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:43,182 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:43,195 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:43,292 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:43,300 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:43,411 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:43,428 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:43,529 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:43,541 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:43,657 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:07:43,669 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:07:43,768 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 22:07:43,773 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 22:07:43,773 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key (pig.schematuple) was not set... Will not generate code.
2018-09-25 22:07:43,783 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 22:07:43,783 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - 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,106984,ORD,Chicago,USA)
grunt>
```

3(b) Which month has seen the most number of cancellations due to bad weather?

Solution: CMD :

REGISTER '/home/acadgild/Desktop/piggybank.jar';

A = load '/airline\_usecase/DelayedFlights.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage('','NO\_MULTILINE','UNIX','SKIP\_INPUT\_  
HEADER');

B = foreach A generate (int)\$2 as month,(int)\$10 as flight\_num,(int)\$22 as  
cancelled,(chararray)\$23 as cancel\_code;

C = filter B by cancelled == 1 AND cancel\_code == 'B';

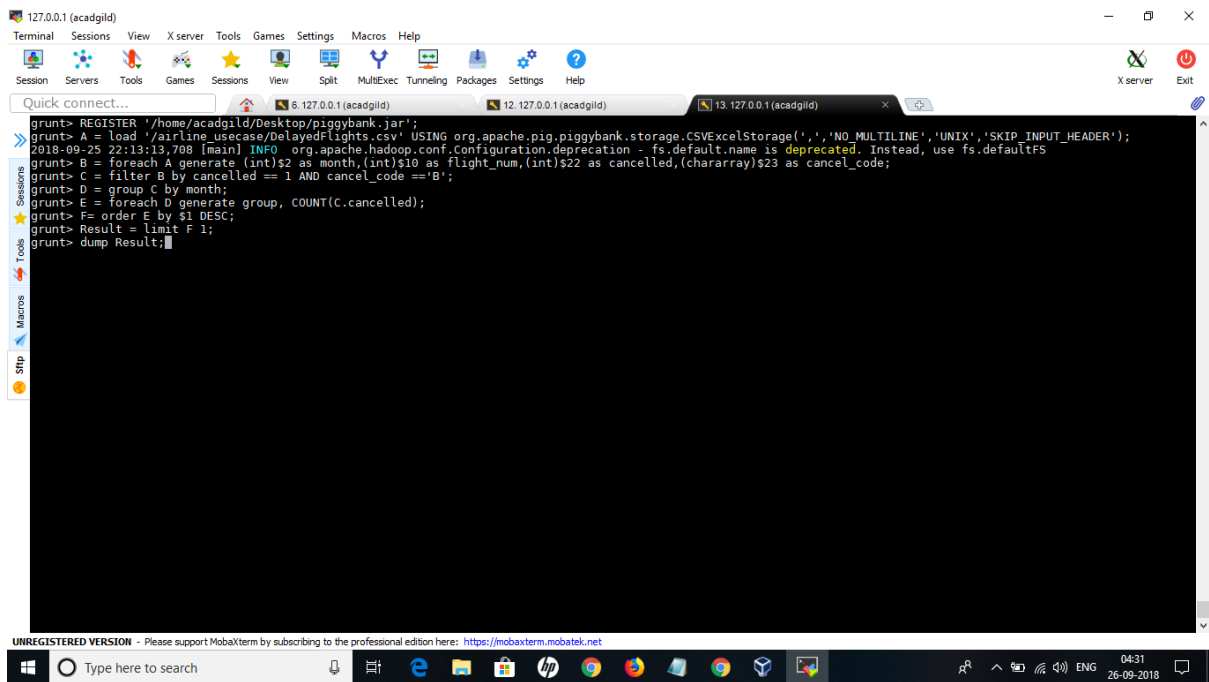
D = group C by month;

E = foreach D generate group, COUNT(C.cancelled);

F= order E by \$1 DESC;

Result = limit F 1;

dump Result;



127.0.0.1 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

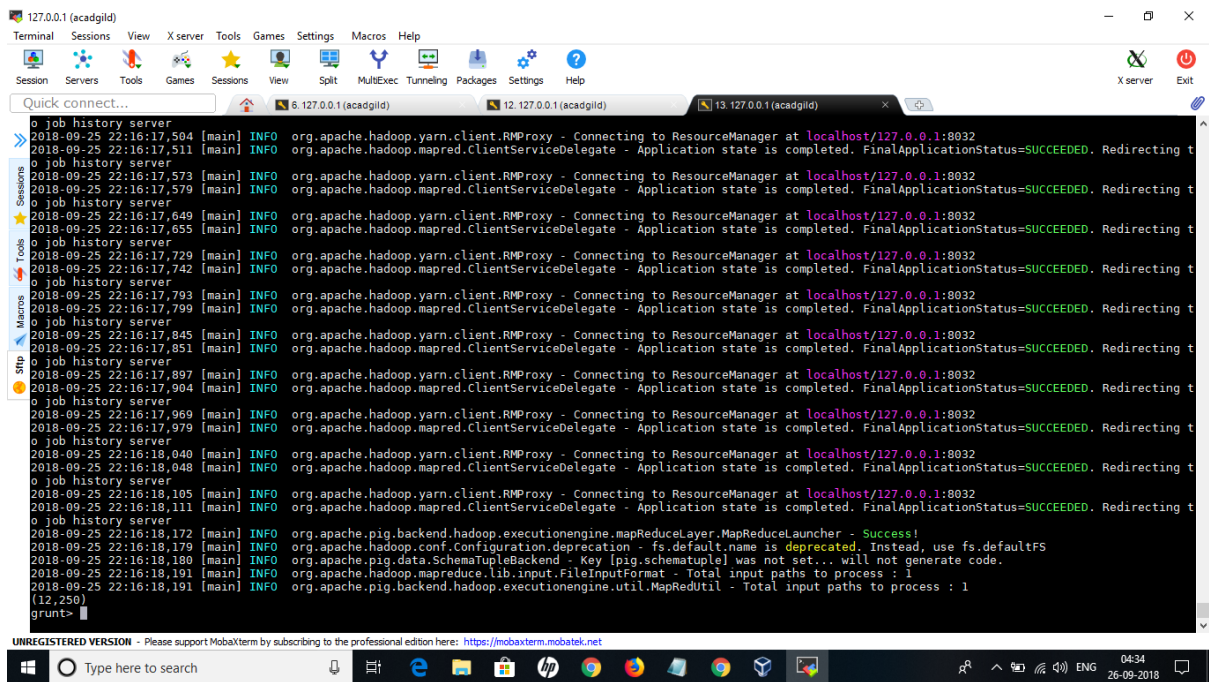
Quick connect...

8 127.0.0.1 (acadgild) 12 127.0.0.1 (acadgild) 13 127.0.0.1 (acadgild)

```
grunt> REGISTER '/home/acadgild/Desktop/piggybank.jar';
grunt> A = load '/airline_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage('','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2018-09-25 22:13:13,788 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = foreach A generate (int)$2 as month,(int)$10 as flight_num,(int)$22 as cancelled,(chararray)$23 as cancel_code;
grunt> C = filter B by cancelled == 1 AND cancel_code == 'B';
grunt> D = group C by month;
grunt> E = foreach D generate group, COUNT(C.cancelled);
grunt> F = order E by $1 DESC;
grunt> Result = limit F 1;
grunt> dump Result;
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Type here to search



127.0.0.1 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

6 127.0.0.1 (acadgild) 12 127.0.0.1 (acadgild) 13 127.0.0.1 (acadgild)

```
o job history server
2018-09-25 22:16:17,504 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,511 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:17,573 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,579 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:17,649 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,655 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:17,729 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,742 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:17,793 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,799 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:17,845 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,851 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:17,897 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,904 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:17,969 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:17,979 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:18,040 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:18,048 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:18,105 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:16:18,111 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:16:18,172 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 22:16:18,179 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 22:16:18,180 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-25 22:16:18,191 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 22:16:18,191 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(12,250)
grunt>
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Type here to search

3(c) Top ten origins with the highest AVG departure delay

Solution: CMD:

REGISTER '/home/acadgild/Desktop/piggybank.jar';

A = load '/airline\_usecase/DelayedFlights.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage('','NO\_MULTILINE','UNIX','SKIP\_INPUT\_  
HEADER');

B1 = foreach A generate (int)\$16 as dep\_delay, (chararray)\$17 as origin;

C1 = filter B1 by (dep\_delay is not null) AND (origin is not null);

D1 = group C1 by origin;

E1 = foreach D1 generate group, AVG(C1.dep\_delay);

Result = order E1 by \$1 DESC;

Top\_ten = limit Result 10;

Lookup = load '/airline\_usecase/airports.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO\_MULTILINE', 'UNIX', 'SKIP\_INPUT\_HEADER');

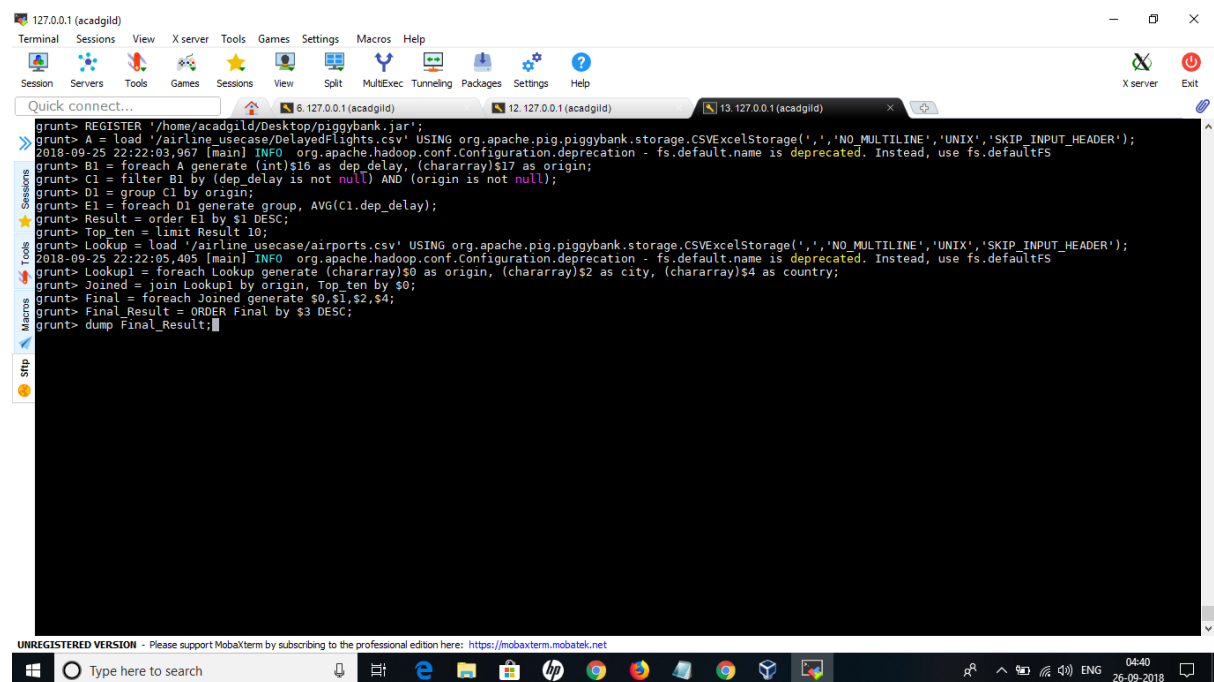
Lookup1 = foreach Lookup generate (chararray)\$0 as origin, (chararray)\$2 as city,  
(chararray)\$4 as country;

Joined = join Lookup1 by origin, Top\_ten by \$0;

Final = foreach Joined generate \$0,\$1,\$2,\$4;

Final\_Result = ORDER Final by \$3 DESC;

dump Final\_Result;



The screenshot shows a MobaXterm terminal window with a dark background. The terminal displays a series of Pig Latin commands and their corresponding output. The commands are: REGISTER, load, foreach, filter, group, foreach, order, limit, load, foreach, join, foreach, ORDER, and dump. The output includes timestamps, log levels (INFO), and deprecation warnings from the Apache Hadoop configuration. The terminal window has a standard Linux-style menu bar at the top and a Windows taskbar at the bottom.

```
grunt> REGISTER '/home/acadgild/Desktop/piggybank.jar';
grunt> A = load '/airline_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER');
2018-09-25 22:22:03.967 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B1 = foreach A generate (int)$16 as dep_delay, (chararray)$17 as origin;
grunt> C1 = filter B1 by (dep_delay is not null) AND (origin is not null);
grunt> D1 = group C1 by origin;
grunt> E1 = foreach D1 generate group, AVG(C1.dep_delay);
grunt> Result = order E1 by $1 DESC;
grunt> Top_ten = limit Result 10;
grunt> Lookup = load '/airline_usecase/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER');
2018-09-25 22:22:05.405 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> Lookup1 = foreach Lookup generate (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as country;
grunt> Joined = join Lookup1 by origin, Top_ten by $0;
grunt> Final = foreach Joined generate $0,$1,$2,$4;
grunt> Final_Result = ORDER Final by $3 DESC;
grunt> dump Final_Result;
```



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
127.0.0.1 (acadgild) 12 127.0.0.1 (acadgild) 13 127.0.0.1 (acadgild)
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:32,896 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:32,961 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,004 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,040 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,043 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,072 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,080 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,152 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,155 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,190 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:27:33,199 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 22:27:33,232 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 22:27:33,233 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-25 22:27:33,234 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 22:27:33,241 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2018-09-25 22:27:33,241 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - 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, 79.59665024638542)
(ACY, Atlantic City, USA, 79.3103448275862)
(MOT, Minot, USA, 78.66165413533835)
(HHH, NA, USA, 76.53005464480874)
(EGE, Eagle, USA, 74.12891986062718)
(BGM, Binghamton, USA, 73.15333980582525)
grunt>
```

3(d) Which route (origin & destination) has seen the maximum diversion?

Solution: CMD :

REGISTER '/home/acadgild/Desktop/piggybank.jar';

A = load '/airline\_usecase/DelayedFlights.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage('','NO\_MULTILINE','UNIX','SKIP\_INPUT\_  
HEADER');

B = FOREACH A GENERATE (chararray)\$17 as origin, (chararray)\$18 as dest, (int)\$24 as  
diversion;

C = FILTER B BY (origin is not null) AND (dest is not null) AND (diversion == 1);

D = GROUP C by (origin,dest);

E = FOREACH D generate group, COUNT(C.diversion);

F = ORDER E BY \$1 DESC;

Result = limit F 10;

dump Result;

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
grunt> REGISTER '/home/acadgild/Desktop/piggybank.jar';
grunt> A = load '/airline/uscases/delayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage('','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2018-09-25 22:34:51,099 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = FOREACH A GENERATE (chararray)$17 as origin, (chararray)$18 as dest, (int)$24 as diversion;
grunt> C = FILTER B BY (origin is not null) AND (dest is not null) AND (diversion == 1);
grunt> D = GROUP C by (origin,dest);
grunt> E = FOREACH D generate group, COUNT(C.diversion);
grunt> F = ORDER E BY $1 DESC;
grunt> Result = limit F 10;
grunt> dump Result;
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
o job history server
2018-09-25 22:37:59,753 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:37:59,757 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:37:59,791 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:37:59,794 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:37:59,848 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:37:59,853 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:37:59,925 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:37:59,933 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:37:59,995 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:37:59,998 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:38:00,045 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:38:00,048 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:38:00,082 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-25 22:38:00,090 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-25 22:38:00,125 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-25 22:38:00,126 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-25 22:38:00,127 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-25 22:38:00,131 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-25 22:38:00,131 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((ORD,LGA),36)
((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)
grunt>
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>