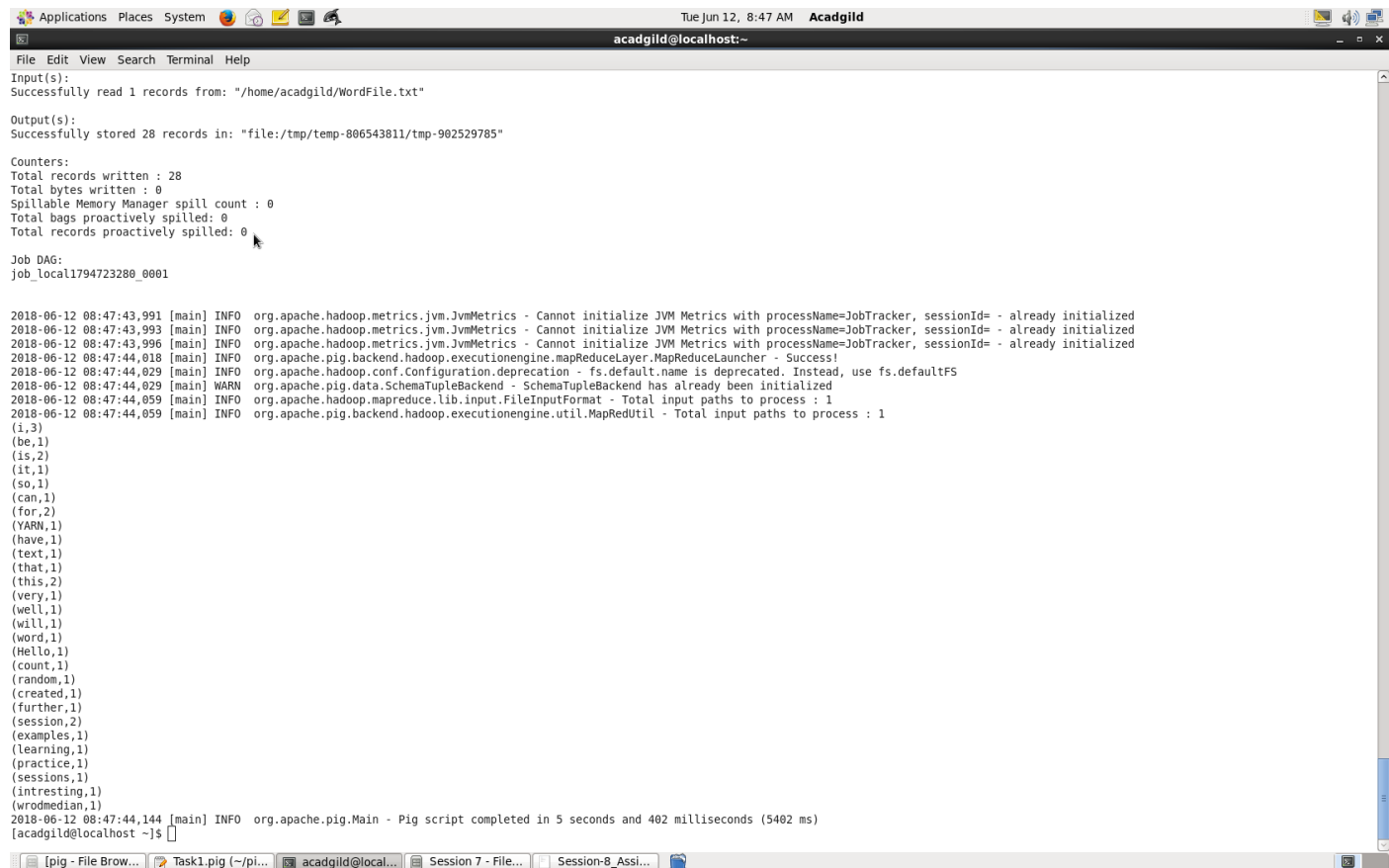


## Task 1

Write a program to implement wordcount using Pig.

I have created WordFile in /home/acadgildWordFile.txt then executed

Executed pig script as **pig -x local /home/acadgild/Task1.pig**



```
Applications Places System Tue Jun 12, 8:47 AM Acadgild
acadgild@localhost:~$
File Edit View Search Terminal Help
Input(s):
Successfully read 1 records from: "/home/acadgild/WordFile.txt"

Output(s):
Successfully stored 28 records in: "file:/tmp/temp-806543811/tmp-902529785"

Counters:
Total records written : 28
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1794723280_0001

2018-06-12 08:47:43,991 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:47:43,993 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:47:43,996 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:47:44,018 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Success!
2018-06-12 08:47:44,029 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:47:44,029 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:47:44,059 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:47:44,059 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(i,3)
(be,1)
(is,2)
(it,1)
(so,1)
(can,1)
(for,2)
(YARN,1)
(have,1)
(text,1)
(that,1)
(this,2)
(very,1)
(well,1)
(will,1)
(word,1)
>Hello,1)
(count,1)
(random,1)
(created,1)
(further,1)
(session,2)
(examples,1)
(learning,1)
(practice,1)
(sessions,1)
(intresting,1)
>wrodmedian,1)
2018-06-12 08:47:44,144 [main] INFO org.apache.pig.Main - Pig script completed in 5 seconds and 402 milliseconds (5402 ms)
[acadgild@localhost ~]$
```

## 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)

Executed pig script as **pig -x local /home/acadgild/Task2 a.pig**

Script is also attached in Github

(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)

Executed pig script as **pig -x local /home/acadgild/Task2 b.pig**

Script is also attached in Github

```
Applications Places System Tue Jun 12, 8:49 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.5 0.16.0 acadgild 2018-06-12 08:48:59 2018-06-12 08:49:04 ORDER_BY,FILTER,LIMIT

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job_local1068251798_0002 1 1 n/a n/a n/a n/a n/a n/a n/a 8 SAMPLER
job_local1176031040_0003 1 1 n/a n/a n/a n/a n/a n/a n/a 8 ORDER_BY,COMBINER
job_local1382529486_0004 1 1 n/a n/a n/a n/a n/a n/a n/a 8,B,E file:/tmp/temp1477455684/tmp-1833596166,
job_local694880355_0001 1 0 n/a n/a n/a n/a 0 0 0 A,C MAP_ONLY

Input(s):
Successfully read 14 records from: "/home/acadgild/employee_details.txt"

Output(s):
Successfully stored 3 records in: "file:/tmp/temp1477455684/tmp-1833596166"

Counters:
Total records written : 3
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local694880355_0001 -> job_local1068251798_0002,
job_local1068251798_0002 -> job_local1176031040_0003,
job_local1176031040_0003 -> job_local1382529486_0004,
job_local1382529486_0004

2018-06-12 08:49:04,874 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,875 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,875 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,898 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,899 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,912 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,913 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,913 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,917 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,918 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,931 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:04,941 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2018-06-12 08:49:04,949 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:49:04,949 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:49:04,976 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:49:04,977 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh)
(107,Salman)
(103,Akshay)
2018-06-12 08:49:05,038 [main] INFO org.apache.pig.Main - Pig script completed in 8 seconds and 28 milliseconds (8028 ms)
[acadgild@localhost ~]$
```

(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)

Executed pig script as **pig -x local /home/acadgild/Task2 c.pig**

Script is also attached in Github

```
Applications Places System Tue Jun 12, 8:49 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job_local1131692176_0003 1 1 n/a n/a n/a n/a n/a n/a F SAMPLER
job_local1149572704_0005 1 1 n/a n/a n/a n/a n/a n/a F,H file:/tmp/temp-195849170/tmp838889309,
job_local1207289645_0002 2 1 n/a n/a n/a n/a n/a n/a A,E HASH JOIN
job_local433559458_0004 1 1 n/a n/a n/a n/a n/a n/a F ORDER_BY_COMBINER
job_local454947053_0001 1 1 n/a n/a n/a n/a n/a n/a B,C,D GROUP_BY_COMBINER

Input(s):
Successfully read 9 records from: "/home/acadgild/employee_expenses.txt"
Successfully read 14 records from: "/home/acadgild/employee_details.txt"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp-195849170/tmp838889309"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

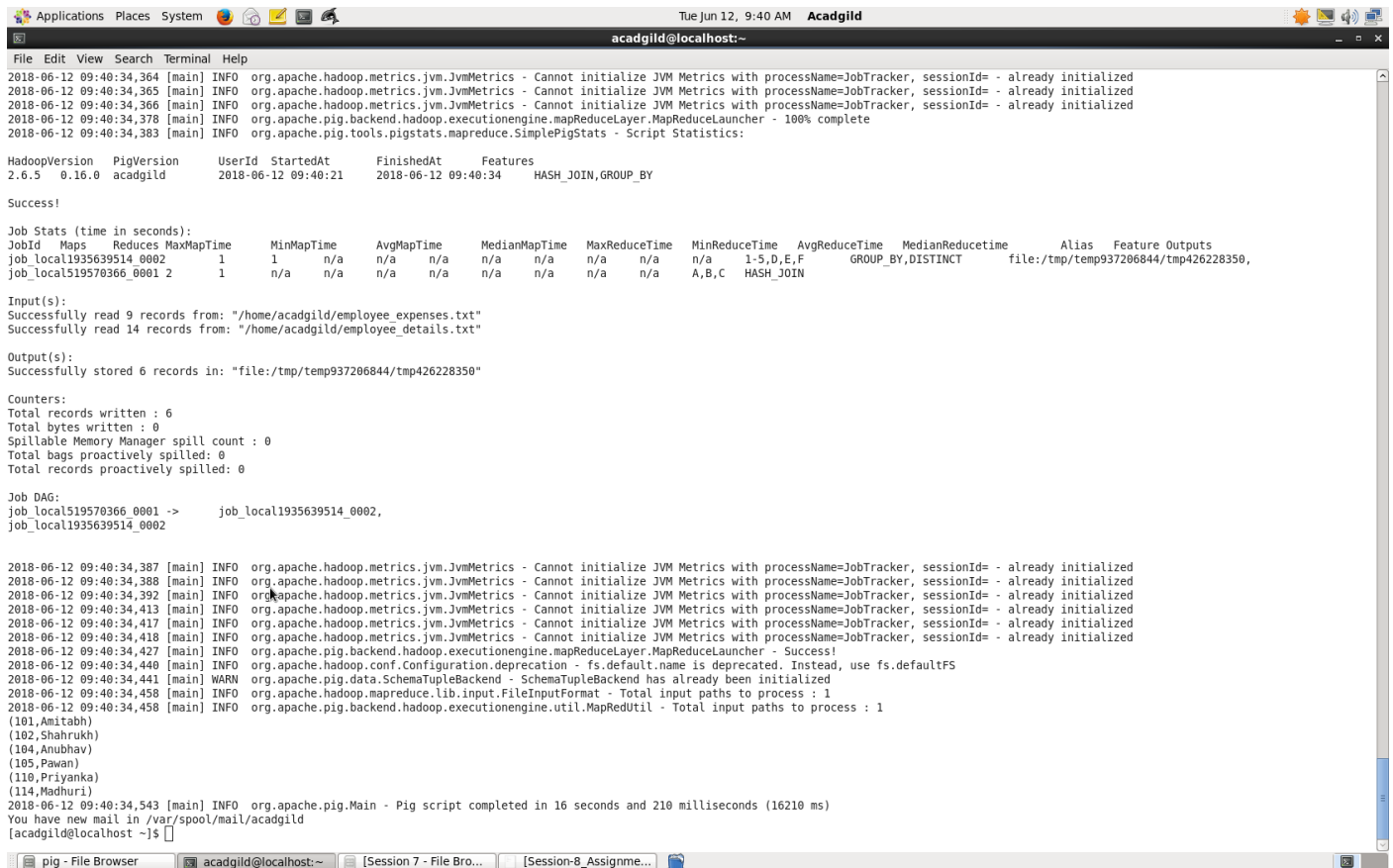
Job DAG:
job_local454947053_0001 -> job_local1207289645_0002,
job_local1207289645_0002 -> job_local1131692176_0003,
job_local1131692176_0003 -> job_local433559458_0004,
job_local433559458_0004 -> job_local1149572704_0005,
job_local1149572704_0005

2018-06-12 08:49:23,964 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:23,970 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:23,972 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:23,990 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:23,996 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:23,997 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,021 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,022 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,022 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,038 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,038 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,039 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,048 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,049 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,056 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:24,059 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2018-06-12 08:49:24,069 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:49:24,069 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:49:24,100 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:49:24,100 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(102,Shahrukh)
2018-06-12 08:49:24,181 [main] INFO org.apache.pig.Main - Pig script completed in 10 seconds and 306 milliseconds (10306 ms)
[acadgild@localhost ~]$
```

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

Executed pig script as **pig -x local /home/acadgild/Task2 d.pig**

Script is also attached in Github



```
Applications Places System Tue Jun 12, 9:40 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
2018-06-12 09:40:34,364 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,365 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,366 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,378 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2018-06-12 09:40:34,383 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.5 0.16.0 acadgild 2018-06-12 09:40:21 2018-06-12 09:40:34 HASH_JOIN, GROUP_BY

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job_local1935639514_0002 1 1 n/a n/a n/a n/a n/a n/a 1-5,D,E,F GROUP_BY,DISTINCT file:/tmp/temp937206844/tmp426228350,
job_local19570366_0001 2 1 n/a n/a n/a n/a n/a n/a A,B,C HASH_JOIN

Input(s):
Successfully read 9 records from: "/home/acadgild/employee_expenses.txt"
Successfully read 14 records from: "/home/acadgild/employee_details.txt"

Output(s):
Successfully stored 6 records in: "file:/tmp/temp937206844/tmp426228350"

Counters:
Total records written : 6
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

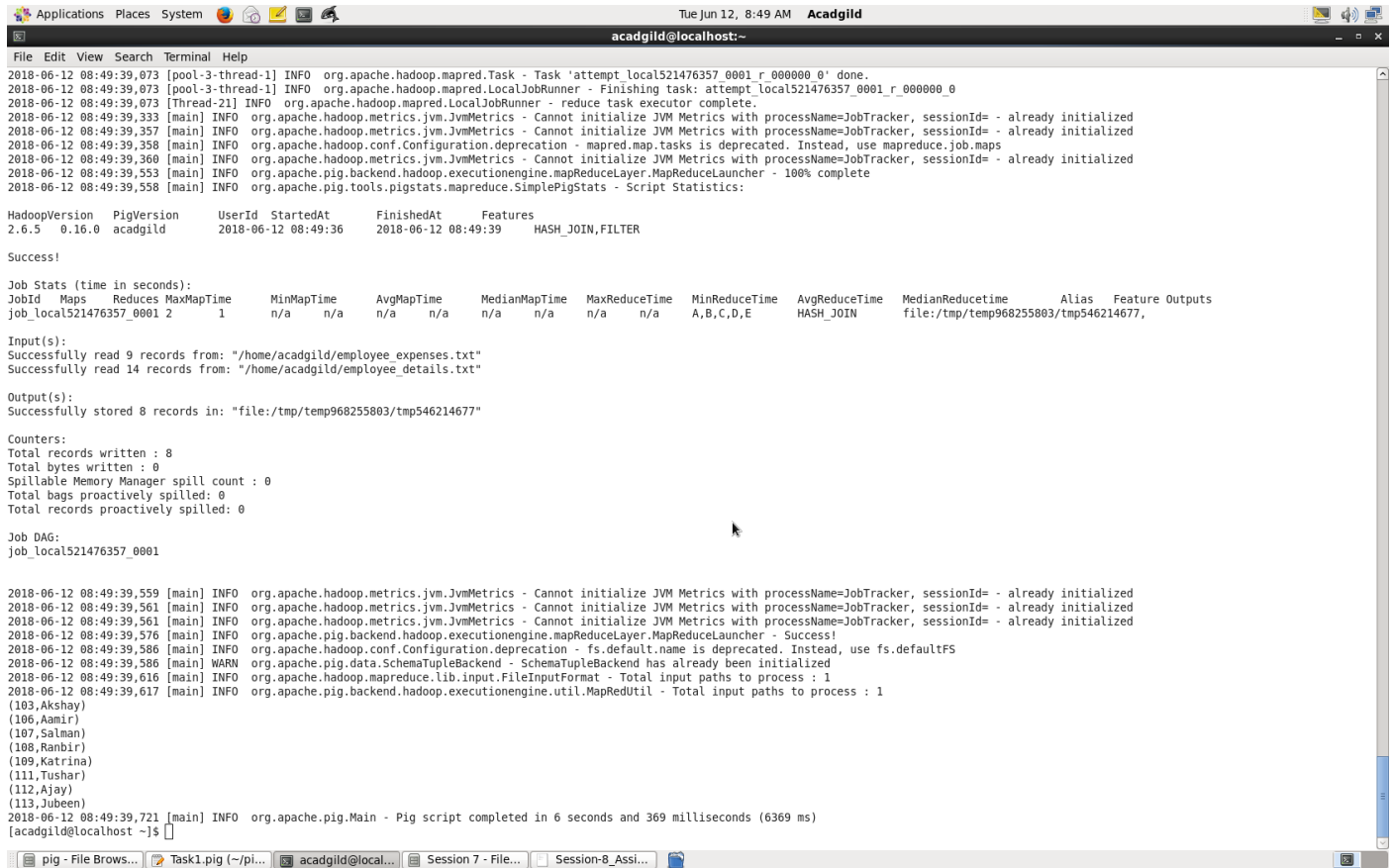
Job DAG:
job_local19570366_0001 -> job_local1935639514_0002,
job_local1935639514_0002

2018-06-12 09:40:34,387 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,388 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,392 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,413 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,417 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,418 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 09:40:34,427 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-06-12 09:40:34,440 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 09:40:34,441 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 09:40:34,458 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 09:40:34,458 [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)
2018-06-12 09:40:34,543 [main] INFO org.apache.pig.Main - Pig script completed in 16 seconds and 210 milliseconds (16210 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

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

Executed pig script as **pig -x local /home/acadgild/Task2 e.pig**

Script is also attached in Github



```
Applications Places System Tue Jun 12, 8:49 AM Acadgild
acadgild@localhost:~$ pig -x local /home/acadgild/Task2 e.pig
2018-06-12 08:49:39,073 [pool-3-thread-1] INFO org.apache.hadoop.mapred.Task - Task 'attempt local521476357_0001_r_000000_0' done.
2018-06-12 08:49:39,073 [pool-3-thread-1] INFO org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt local521476357_0001_r_000000_0
2018-06-12 08:49:39,073 [Thread-21] INFO org.apache.hadoop.mapred.LocalJobRunner - reduce task executor complete.
2018-06-12 08:49:39,333 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:39,357 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:39,358 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2018-06-12 08:49:39,360 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:39,553 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2018-06-12 08:49:39,558 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.5 0.16.0 acadgild 2018-06-12 08:49:36 2018-06-12 08:49:39 HASH_JOIN,FILTER

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job_local521476357_0001 2 1 n/a n/a n/a n/a n/a n/a A,B,C,D,E HASH_JOIN file:/tmp/temp968255803/tmp546214677,

Input(s):
Successfully read 9 records from: "/home/acadgild/employee_expenses.txt"
Successfully read 14 records from: "/home/acadgild/employee_details.txt"

Output(s):
Successfully stored 8 records in: "file:/tmp/temp968255803/tmp546214677"

Counters:
Total records written : 8
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local521476357_0001

2018-06-12 08:49:39,559 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:39,561 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:39,561 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:49:39,576 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-06-12 08:49:39,586 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:49:39,586 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:49:39,616 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:49:39,617 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(103,Akshay)
(106,Aamir)
(107,Salman)
(108,Ranbir)
(109,Katrina)
(111,Tushar)
(112,Ajay)
(113,Jubeen)
2018-06-12 08:49:39,721 [main] INFO org.apache.pig.Main - Pig script completed in 6 seconds and 369 milliseconds (6369 ms)
acadgild@localhost ~]$
```

## Task 3

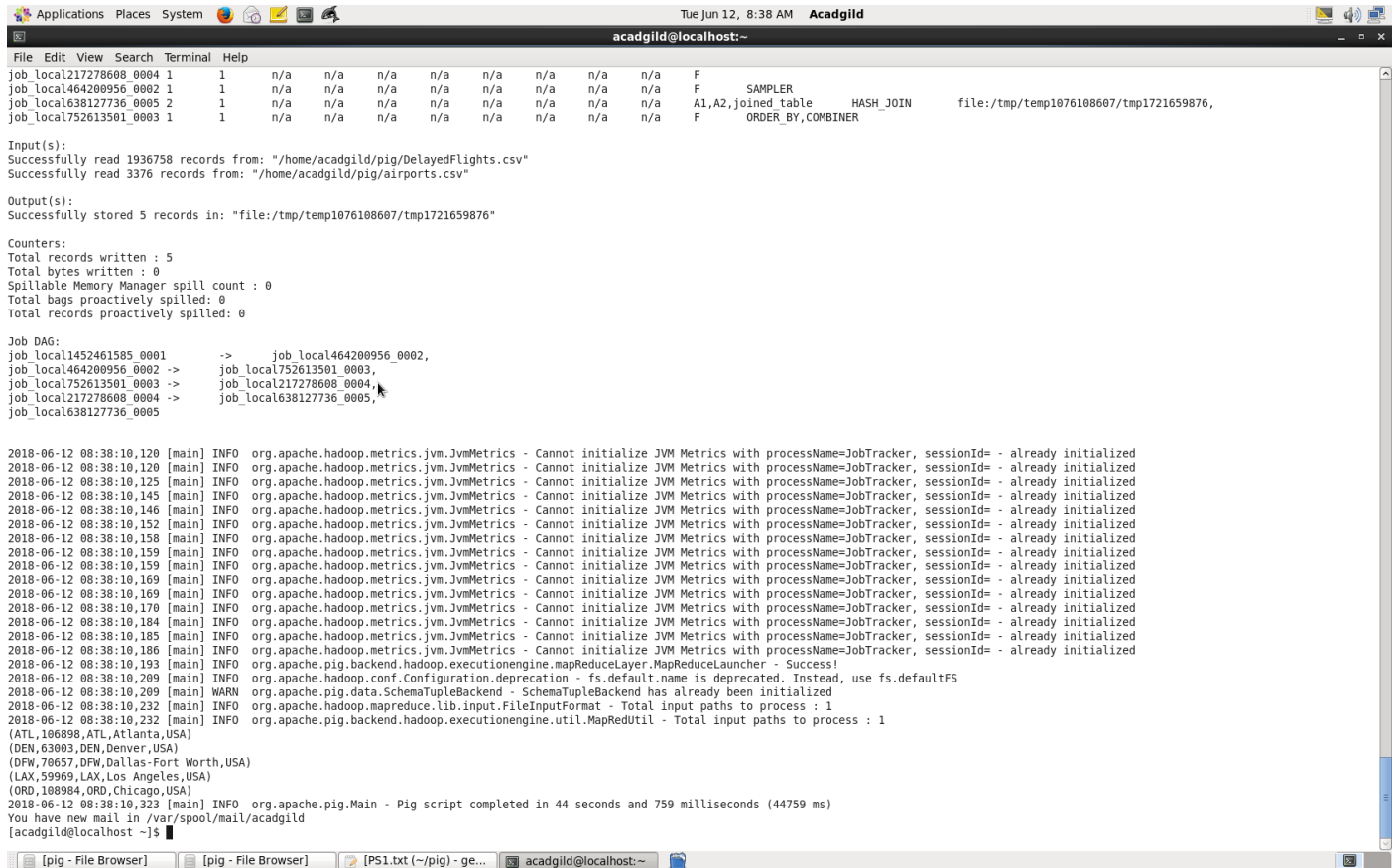
Implement the use case present in below blog link and share the complete steps along with screenshot(s) from your end.

<https://acadgild.com/blog/aviation-data-analysis-using-apache-pig/>

Executed pig script as **pig -x local /home/acadgild/PS1.pig**

Script is also attached in Github

## Problem Statement 1:



```
Applications Places System Tue Jun 12, 8:38 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
job local217278608_0004 1 1 n/a n/a n/a n/a n/a n/a n/a n/a F SAMPLER
job local464200956_0002 1 1 n/a n/a n/a n/a n/a n/a n/a n/a F A1,A2,joined_table HASH_JOIN file:/tmp/temp1076108607/tmp1721659876,
job local638127736_0005 2 1 n/a n/a n/a n/a n/a n/a n/a n/a F ORDER_BY,COMBINER
Input(s):
Successfully read 1936758 records from: "/home/acadgild/pig/DelayedFlights.csv"
Successfully read 3376 records from: "/home/acadgild/pig/airports.csv"
Output(s):
Successfully stored 5 records in: "file:/tmp/temp1076108607/tmp1721659876"
Counters:
Total records written : 5
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job local1452461585_0001 -> job local464200956_0002,
job local464200956_0002 -> job local752613501_0003,
job local752613501_0003 -> job local217278608_0004,
job local217278608_0004 -> job local638127736_0005,
job local638127736_0005
2018-06-12 08:38:10,120 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,120 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,125 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,145 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,146 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,152 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,158 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,159 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,159 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,169 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,169 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,170 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,184 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,185 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,186 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:38:10,193 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2018-06-12 08:38:10,209 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:38:10,209 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:38:10,232 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:38:10,232 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ATL,106698,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)
2018-06-12 08:38:10,323 [main] INFO org.apache.pig.Main - Pig script completed in 44 seconds and 759 milliseconds (44759 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

## Problem Statement 2:

Executed pig script as **pig -x local /home/acadgild/PS2.pig**

Script is also attached in Github

```
Applications Places System Tue Jun 12, 8:39 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
2018-06-12 08:39:32,714 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.5 0.16.0 acadgild 2018-06-12 08:39:17 2018-06-12 08:39:32 GROUP_BY,ORDER_BY,FILTER,LIMIT

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job_local398793781_0003 1 1 n/a n/a n/a n/a n/a n/a n/a F ORDER_BY,COMBINER
job_local615328307_0001 8 1 n/a n/a n/a n/a n/a n/a n/a A,B,C,D,E GROUP_BY,COMBINER
job_local636160859_0004 1 1 n/a n/a n/a n/a n/a n/a n/a F file:/tmp/temp578350064/tmp1477596117,
job_local972305287_0002 1 1 n/a n/a n/a n/a n/a n/a n/a F SAMPLER

Input(s):
Successfully read 1936758 records from: "/home/acadgild/pig/DelayedFlights.csv"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp578350064/tmp1477596117"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local615328307_0001 -> job_local972305287_0002,
job_local972305287_0002 -> job_local398793781_0003,
job_local398793781_0003 -> job_local636160859_0004,
job_local636160859_0004

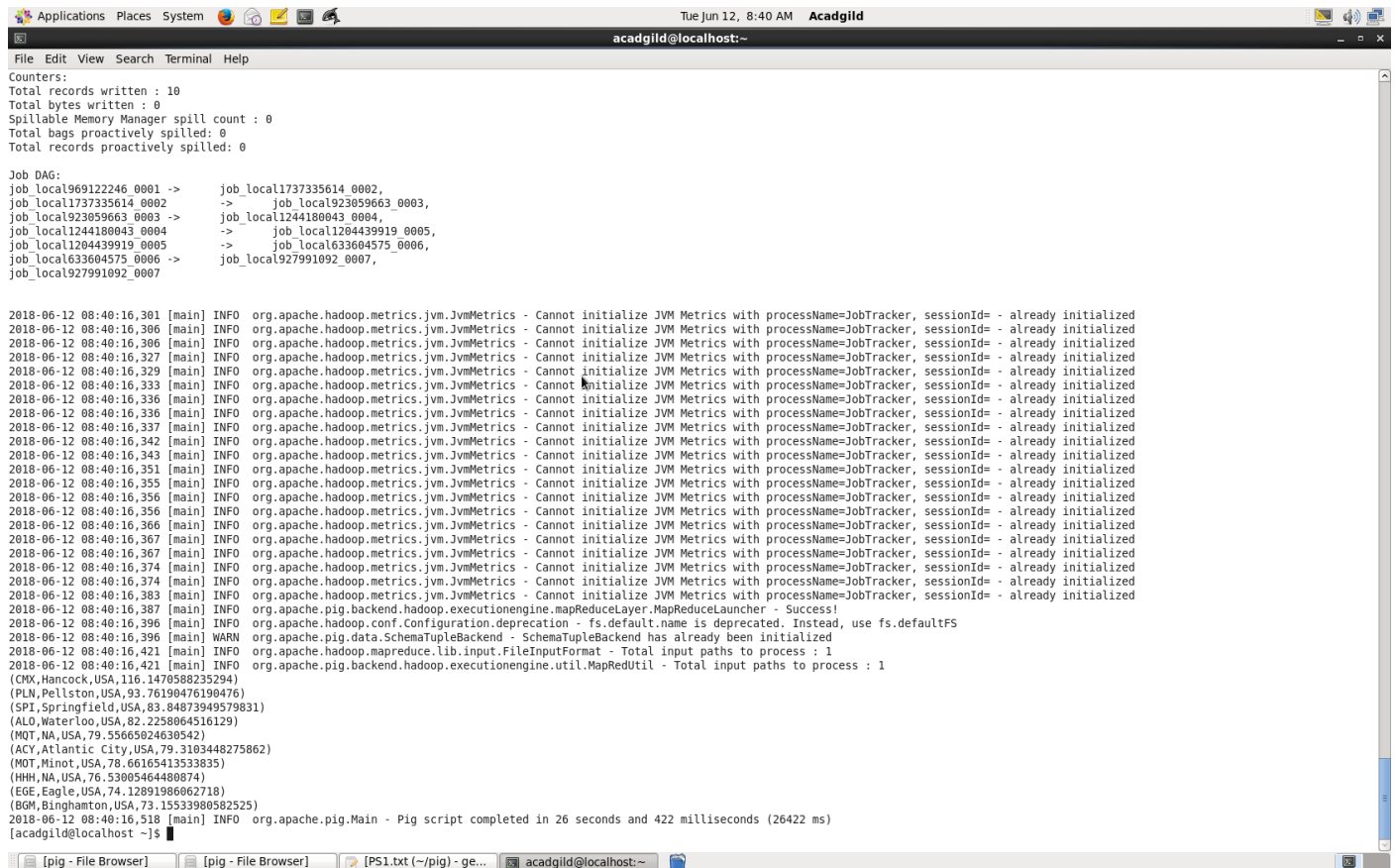
2018-06-12 08:39:32,720 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,721 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,721 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,744 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,744 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,746 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,755 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,755 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,755 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,767 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,768 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,768 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-06-12 08:39:32,780 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Success!
2018-06-12 08:39:32,786 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:39:32,786 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:39:32,805 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:39:32,805 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(12,250)
2018-06-12 08:39:32,886 [main] INFO org.apache.pig.Main - Pig script completed in 17 seconds and 828 milliseconds (17828 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```



## Problem Statement 3:

Executed pig script as **pig -x local /home/acadgild/PS3.pig**

Script is also attached in Github



```
Applications Places System Tue Jun 12, 8:40 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
Counters:
Total records written : 10
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local969122246_0001 -> job_local1737335614_0002,
job_local1737335614_0002 -> job_local923059663_0003,
job_local923059663_0003 -> job_local1244180043_0004,
job_local1244180043_0004 -> job_local1204439919_0005,
job_local1204439919_0005 -> job_local633604575_0006,
job_local633604575_0006 -> job_local927991092_0007,
job_local927991092_0007

2018-06-12 08:40:16,301 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,306 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,306 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,327 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,329 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,333 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,336 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,336 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,337 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,342 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,343 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,351 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,355 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,356 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,356 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,366 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,367 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,367 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,374 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,374 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,383 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:16,387 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-06-12 08:40:16,396 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:40:16,396 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:40:16,421 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:40:16,421 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(CMX,Hancock,USA,116.1470580235294)
(PLN,Pellston,USA,93.76190476190476)
(SPI,Springfield,USA,83.84873949579831)
(ALO,Waterloo,USA,82.2258064516129)
(MQT,NA,USA,79.55665024630542)
(ACY,Atlantic City,USA,79.3103448275862)
(MDT,Minot,USA,78.66165413533835)
(HHH,NA,USA,76.53005464480874)
(EGE,Eagle,USA,74.1289190602718)
(BGM,Binghanton,USA,73.15533980582525)
2018-06-12 08:40:16,518 [main] INFO org.apache.pig.Main - Pig script completed in 26 seconds and 422 milliseconds (26422 ms)
[acadgild@localhost ~]$
```

## Problem Statement 4:

Executed pig script as **pig -x local /home/acadgild/PS4.pig**

Script is also attached in Github

```
Applications Places System Tue Jun 12, 8:40 AM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
job_local1206362264_0002 1 1 n/a n/a n/a n/a n/a n/a n/a F SAMPLER
job_local1758108660_0003 1 1 n/a n/a n/a n/a n/a n/a n/a F ORDER_BY_COMBINER
job_local1920979279_0001 8 1 n/a n/a n/a n/a n/a n/a n/a A,B,C,D,E GROUP_BY_COMBINER
job_local431886932_0004 1 1 n/a n/a n/a n/a n/a n/a n/a F file:/tmp/temp1239810196/tmp-97497391,

Input(s):
Successfully read 1936758 records from: "/home/acadgild/pig/DelayedFlights.csv"

Output(s):
Successfully stored 10 records in: "file:/tmp/temp1239810196/tmp-97497391"

Counters:
Total records written : 10
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1920979279_0001 -> job_local1206362264_0002,
job_local1206362264_0002 -> job_local1758108660_0003,
job_local1758108660_0003 -> job_local431886932_0004,
job_local431886932_0004

2018-06-12 08:40:46,493 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,494 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,494 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,517 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,518 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,526 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,530 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,530 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,531 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,542 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,550 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,550 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-06-12 08:40:46,554 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-06-12 08:40:46,564 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-12 08:40:46,564 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-06-12 08:40:46,587 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-06-12 08:40:46,587 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - 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)
2018-06-12 08:40:46,665 [main] INFO org.apache.pig.Main - Pig script completed in 19 seconds and 293 milliseconds (19293 ms)
You have new mail in /var/spool/mail/acadgild
acadgild@localhost ~$
```