

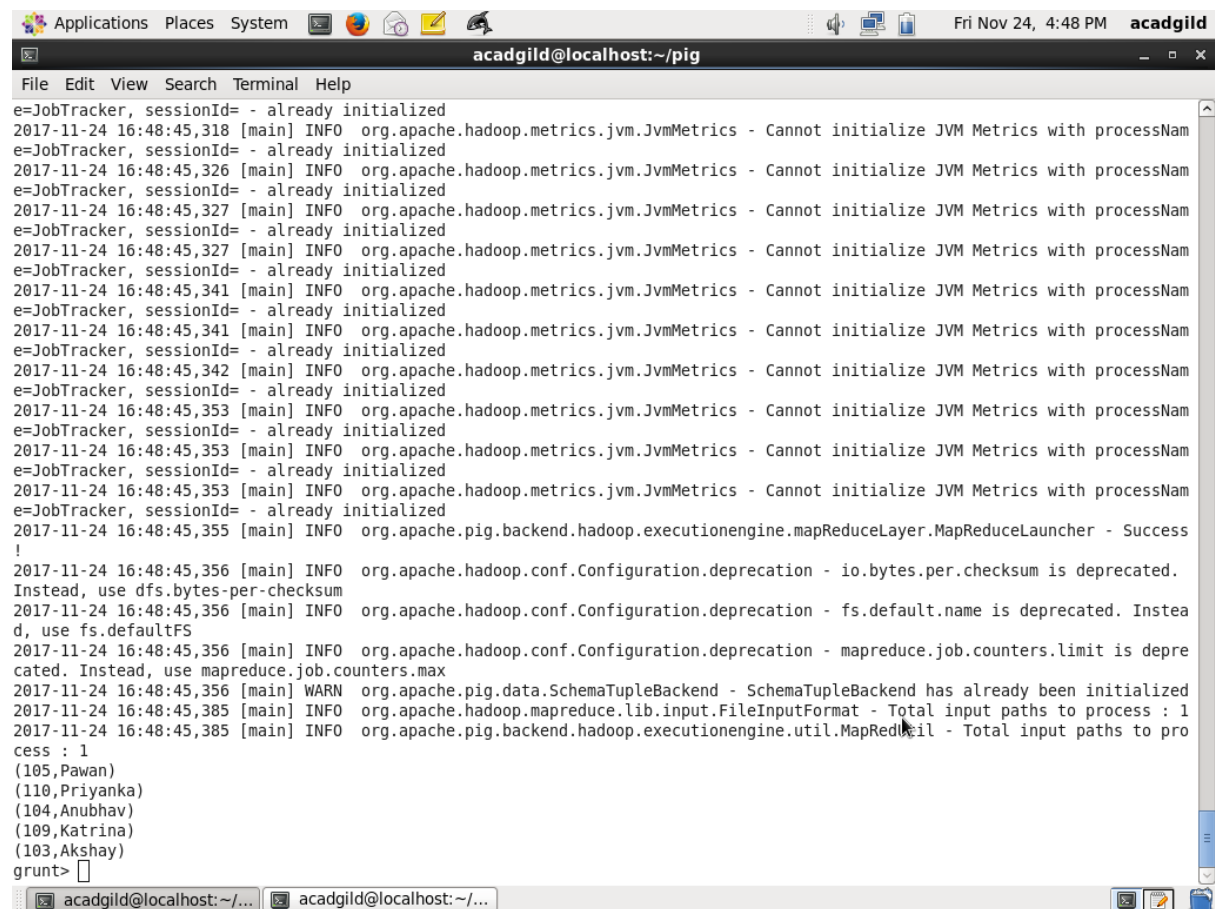
SESSION 5 ASSIGNMENT 1

Task_1:

```
grunt> emp_rating = ORDER emp_details BY rating DESC, name;

grunt> emp_rating_5 = LIMIT emp_rating 5;

grunt> emp = FOREACH emp_rating_5 GENERATE FLATTEN(id), FLATTEN(name);
```



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The window displays the output of a Pig script execution. The logs include several 'INFO' messages from 'org.apache.hadoop.metrics.jvm.JvmMetrics' stating 'Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized'. There are also 'WARN' messages from 'org.apache.hadoop.conf.Configuration.deprecation' about deprecated settings like 'io.bytes.per.checksum', 'fs.default.name', and 'mapreduce.job.counters.limit'. A 'Success' message from 'org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher' is followed by a list of output records: (105,Pawan), (110,Priyanka), (104,Anubhav), (109,Katrina), and (103,Akshay). The prompt 'grunt>' is visible at the bottom of the terminal.

```
File Edit View Search Terminal Help
e=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,318 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,326 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,327 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,327 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,341 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,341 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,342 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,353 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,353 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,353 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 16:48:45,355 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-24 16:48:45,356 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-24 16:48:45,356 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-24 16:48:45,356 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-24 16:48:45,356 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-24 16:48:45,385 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-24 16:48:45,385 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
(105,Pawan)
(110,Priyanka)
(104,Anubhav)
(109,Katrina)
(103,Akshay)
grunt>
```

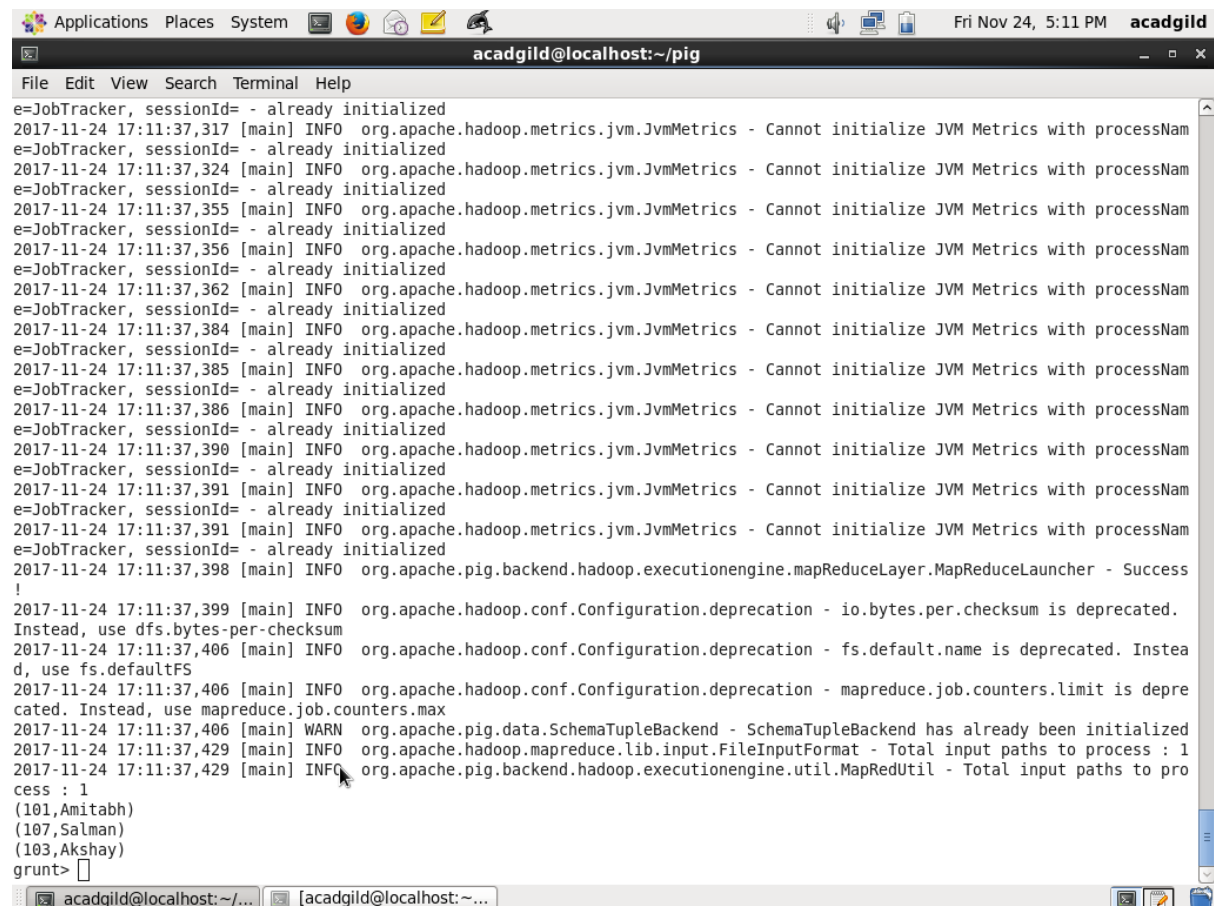
Task_2:

```
grunt> emp_odd = FILTER emp_details BY (id%2 != 0);

grunt> emp_odd_order = ORDER emp_odd BY salary desc, name;

grunt> emp = LIMIT emp_odd_order 3;

grunt> emp_result = FOREACH emp GENERATE id,name;
```



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The window contains a series of log messages from the Apache Pig execution engine. The logs indicate that several JVM Metrics could not be initialized for different processes. There are also deprecation warnings for Hadoop configuration properties like 'io.bytes.per.checksum', 'fs.default.name', and 'mapreduce.job.counters.limit'. A warning from the SchemaTupleBackend is also present. The logs conclude with a successful message from the MapReduceLauncher and a summary of the input paths processed. The terminal prompt 'grunt>' is visible at the bottom.

```
File Edit View Search Terminal Help
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,317 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,324 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,355 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,356 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,362 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,384 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,385 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,386 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,390 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,391 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,391 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-24 17:11:37,398 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2017-11-24 17:11:37,399 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
2017-11-24 17:11:37,406 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2017-11-24 17:11:37,406 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre
cated. Instead, use mapreduce.job.counters.max
2017-11-24 17:11:37,406 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-24 17:11:37,429 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-24 17:11:37,429 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(101,Amitabh)
(107,Salman)
(103,Akshay)
grunt>
```

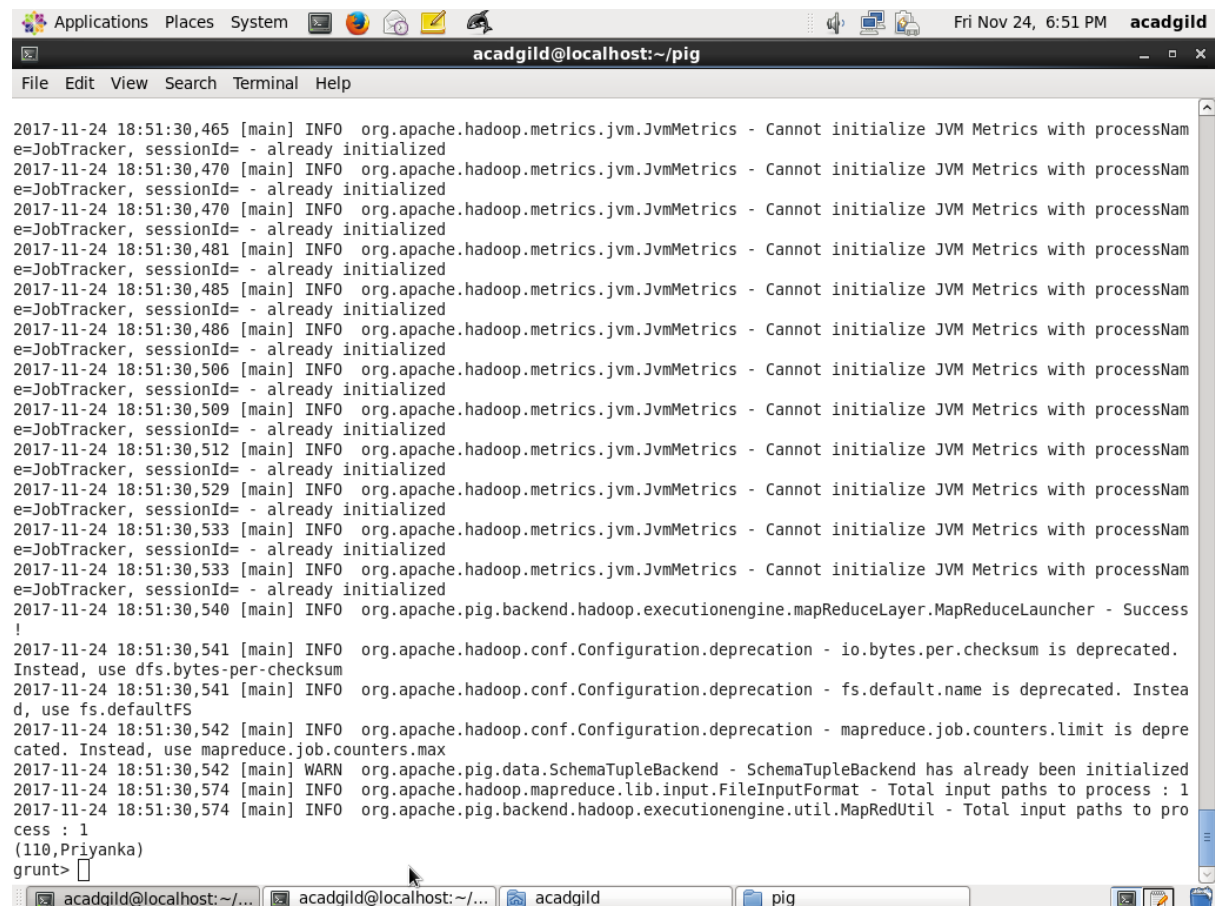
Task_3:

```
grunt> A = JOIN emp_details BY id, emp_expenses BY id;

grunt> B = ORDER A BY expense DESC,name;

grunt> C = LIMIT B 1;

grunt> D= FOREACH C GENERATE emp_details::id,name;
```



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The window displays a series of log messages from the Apache Pig execution engine. The logs indicate that several JVM metrics could not be initialized for the JobTracker process. There are also deprecation warnings for Hadoop configuration properties like 'io.bytes.per.checksum', 'fs.default.name', and 'mapreduce.job.counters.limit'. A warning from the SchemaTupleBackend is also present. The logs conclude with a success message for the MapReduce layer and a summary of the input paths. The terminal prompt 'grunt>' is visible at the bottom.

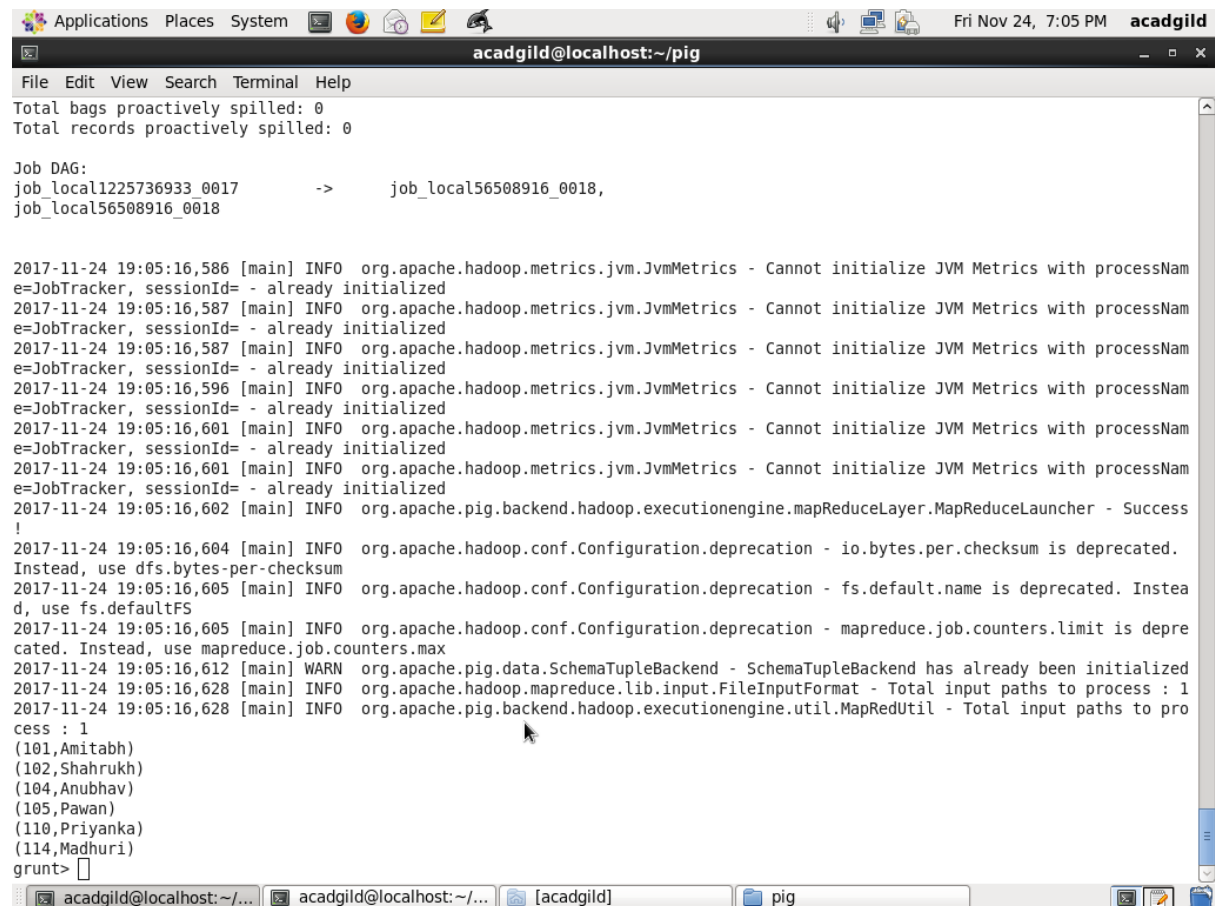
```
2017-11-24 18:51:30,465 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,470 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,470 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,481 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,485 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,486 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,506 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,509 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,512 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,529 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,533 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,533 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,533 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 18:51:30,540 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-24 18:51:30,541 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-24 18:51:30,541 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-24 18:51:30,542 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-24 18:51:30,542 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-24 18:51:30,574 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-24 18:51:30,574 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(110,Priyanka)
grunt>
```

Task_4:

```
grunt> joined = JOIN emp_details BY id, emp_expenses BY id;

grunt> emp = FOREACH joined GENERATE emp_details::id, name;

grunt> emp_data = DISTINCT emp;
```



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The window contains the following text:

```
File Edit View Search Terminal Help
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1225736933_0017    ->    job_local56508916_0018,
job_local56508916_0018

2017-11-24 19:05:16,586 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:05:16,587 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:05:16,587 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:05:16,596 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:05:16,601 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:05:16,601 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:05:16,602 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-24 19:05:16,604 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-24 19:05:16,605 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-24 19:05:16,605 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-24 19:05:16,612 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-24 19:05:16,628 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-24 19:05:16,628 [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)
grunt>
```

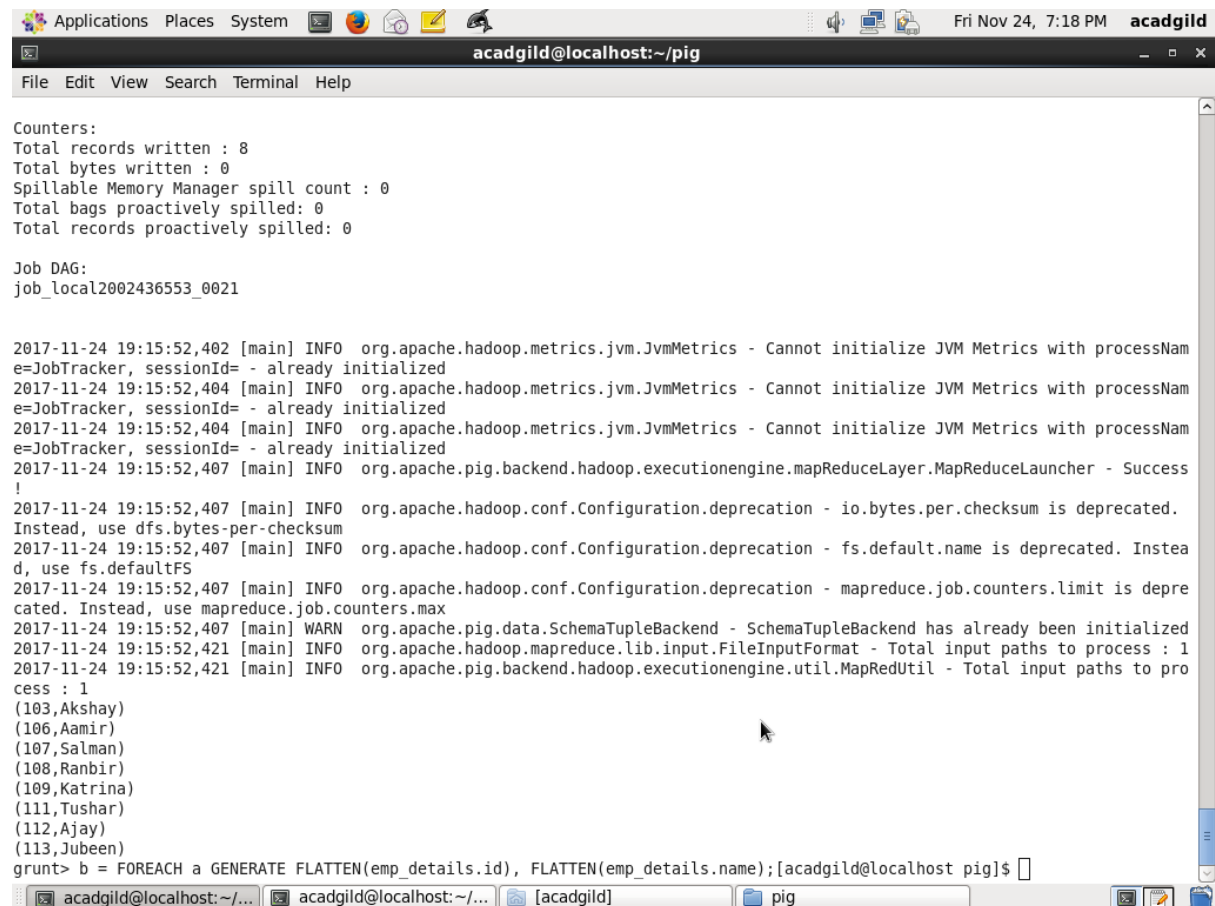
The terminal window also shows a taskbar at the bottom with several open windows: 'acadgild@localhost:~/...', '[acadgild]', and 'pig'.

Task_5:

```
grunt> joined = GROUP emp_details BY id, emp_expenses BY id;

grunt> a= FILTER joined BY IsEmpty(emp_expenses);

grunt> b = FOREACH a GENERATE FLATTEN(emp_details.id), FLATTEN(emp_details.name);
```



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig' with a menu bar (File, Edit, View, Search, Terminal, Help). The output of the Pig job is as follows:

```
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_local2002436553_0021

2017-11-24 19:15:52,402 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:15:52,404 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:15:52,404 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-24 19:15:52,407 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-24 19:15:52,407 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-24 19:15:52,407 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-24 19:15:52,407 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-24 19:15:52,407 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-24 19:15:52,421 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-24 19:15:52,421 [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)
grunt> b = FOREACH a GENERATE FLATTEN(emp_details.id), FLATTEN(emp_details.name);[acadgild@localhost pig]$
```

The terminal window also shows a taskbar at the bottom with several open windows, including 'acadgild@localhost:~/...', '[acadgild]', and 'pig'.