

SESSION 4 ASSIGNMENT 3

```
grunt> lines = LOAD 'Input_File.txt' AS (line:chararray);
```

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

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

Job DAG:
job_local870016559_0012

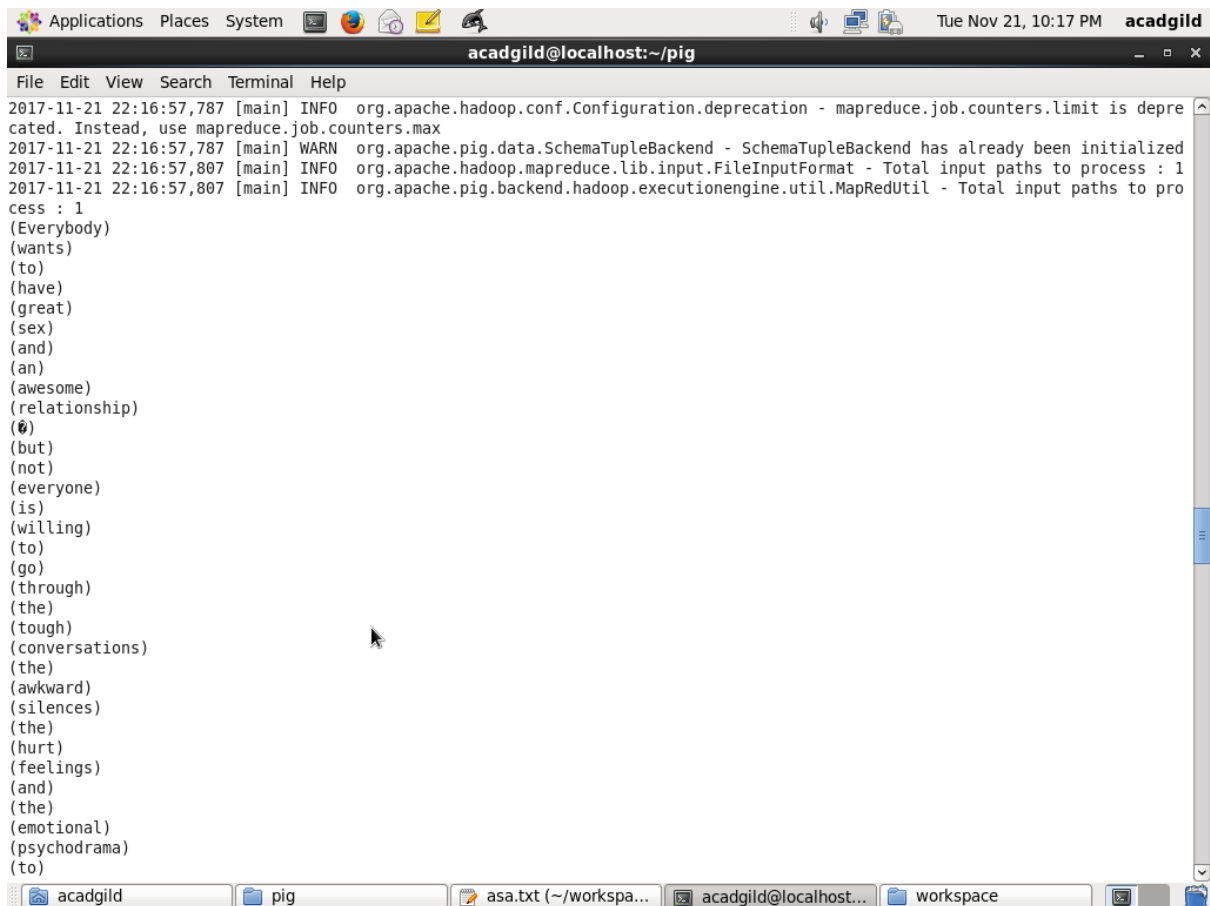
2017-11-21 22:14:48,961 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 22:14:48,965 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 22:14:48,966 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 22:14:48,967 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-21 22:14:48,967 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-21 22:14:48,968 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-21 22:14:48,968 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-21 22:14:48,968 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 22:14:48,993 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 22:14:48,993 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

(Everybody wants to have great sex and an awesome relationship but not everyone is willing to go through the tough conversations, the awkward silences, the hurt feelings and the emotional psychodrama to get there. And so they settle. They settle and wonder What if? for years and years until the question morphs from What if? into Was that it? And when the lawyers go home and the alimony check is in the mail they say, What was that for? if not for their lowered standards and expectations 20 years prior, then what for? Because happiness requires struggle. The positive is the side effect of handling the negative. You can only avoid negative experiences for so long before they come roaring back to life. At the core of all human behavior, our needs are more or less similar. Positive experience is easy to handle. It's negative experience that we all, by definition, struggle with. Therefore, what we get out of life is not determined by the good feelings we desire but by what bad feelings we're willing and able to sustain to get us to those good feelings. People want an amazing physique. But you don't end up with one unless you legitimately appreciate the pain and physical stress that comes with living inside a gym for hour upon hour, unless you love calculating and calibrating the food you eat, planning your life out in tiny plate-sized portions.)

grunt>
```

The terminal window has a taskbar at the bottom with several open tabs: 'acadgild', 'pig', 'asa.txt (~/workspa...', 'acadgild@localhost...', and 'workspace'.

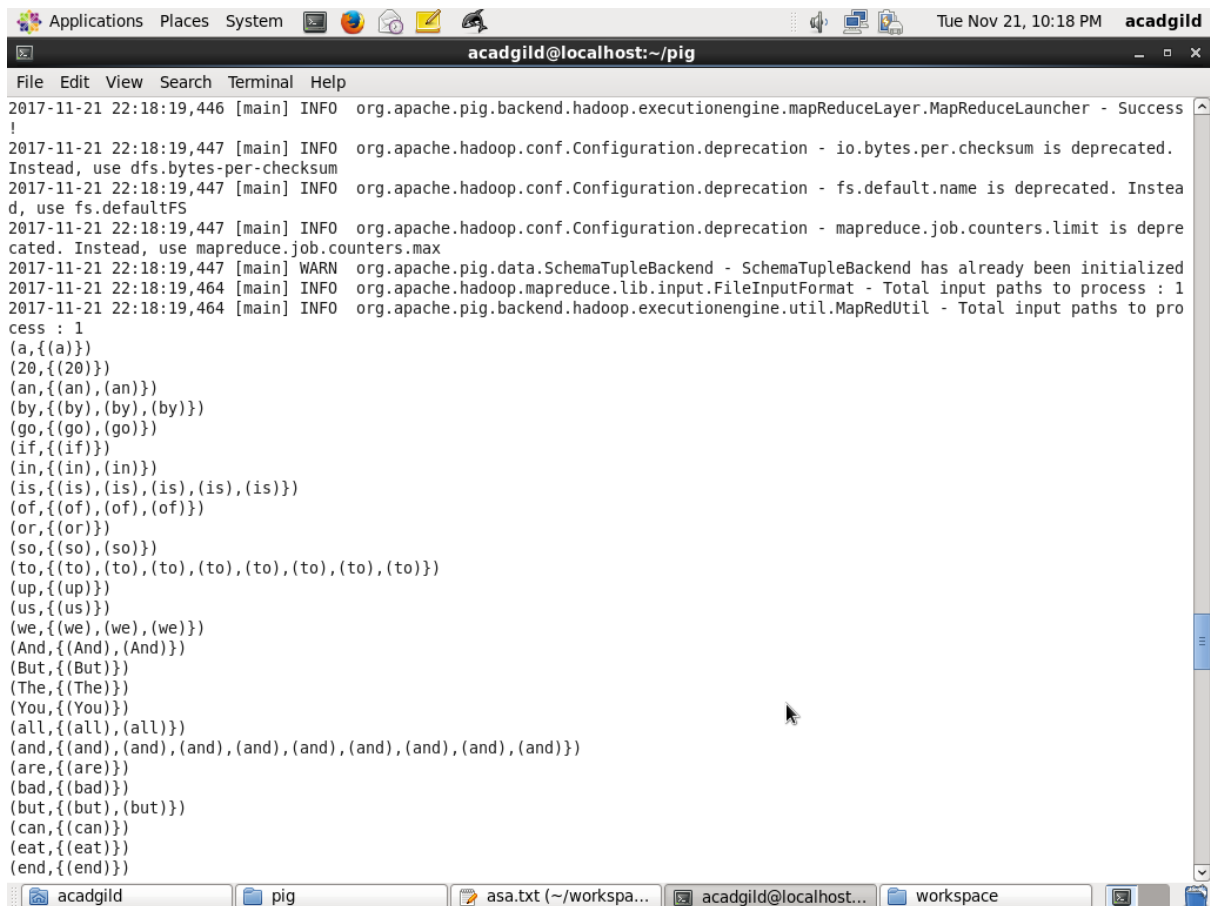
```
grunt> words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;
```



The screenshot shows a terminal window titled "acadgild@localhost:~/pig". The window contains log output from a Pig job execution. The logs show several messages from the main process, including warnings and information about the configuration and input paths. Below the logs, a list of words is displayed, each enclosed in parentheses. The words are: (Everybody), (wants), (to), (have), (great), (sex), (and), (an), (awesome), (relationship), (0), (but), (not), (everyone), (is), (willing), (to), (go), (through), (the), (tough), (conversations), (the), (awkward), (silences), (the), (hurt), (feelings), (and), (the), (emotional), (psychodrama), and (to). The terminal window also shows a taskbar at the bottom with several open applications, including "acadgild", "pig", "asa.txt (~/workspa...", "acadgild@localhost...", and "workspace".

```
File Edit View Search Terminal Help
2017-11-21 22:16:57,787 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre
cated. Instead, use mapreduce.job.counters.max
2017-11-21 22:16:57,787 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 22:16:57,807 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 22:16:57,807 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(Everybody)
(wants)
(to)
(have)
(great)
(sex)
(and)
(an)
(awesome)
(relationship)
(0)
(but)
(not)
(everyone)
(is)
(willing)
(to)
(go)
(through)
(the)
(tough)
(conversations)
(the)
(awkward)
(silences)
(the)
(hurt)
(feelings)
(and)
(the)
(emotional)
(psychodrama)
(to)
```

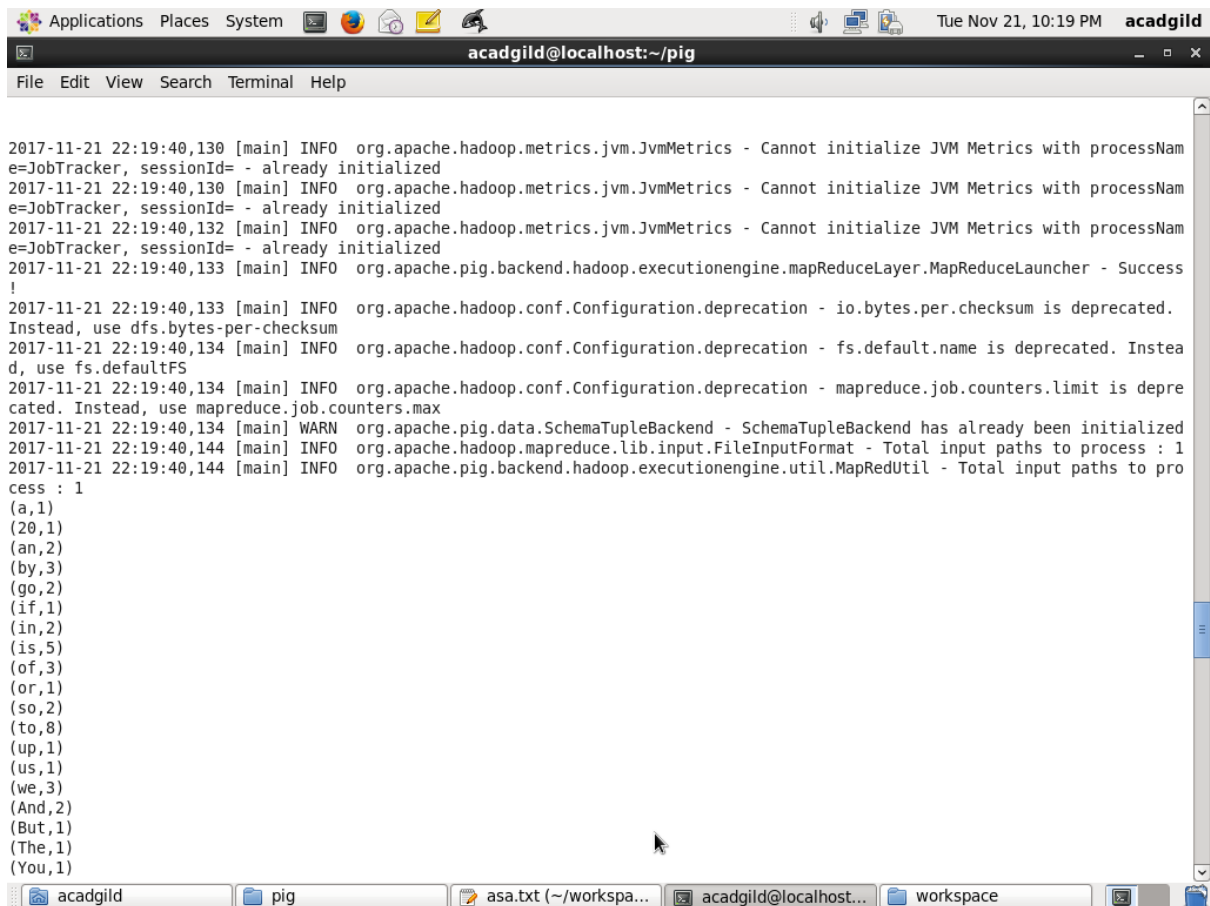
Grunt> grouped = GROUP words BY word;



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The window displays log output from a Pig job execution. The logs include several INFO and WARN messages from the org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher, org.apache.hadoop.conf.Configuration.deprecation, org.apache.pig.data.SchemaTupleBackend, org.apache.hadoop.mapreduce.lib.input.FileInputFormat, and org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil. The output of the 'grouped' command is a list of words and their corresponding groups, such as (a, {(a)}), (20, {(20)}), (an, {(an), (an)}), (by, {(by), (by), (by)}), (go, {(go), (go)}), (if, {(if)}), (in, {(in), (in)}), (is, {(is), (is), (is), (is), (is), (is)}), (of, {(of), (of), (of)}), (or, {(or)}), (so, {(so), (so)}), (to, {(to), (to), (to), (to), (to), (to), (to), (to)}), (up, {(up)}), (us, {(us)}), (we, {(we), (we), (we)}), (And, {(And), (And)}), (But, {(But)}), (The, {(The)}), (You, {(You)}), (all, {(all), (all)}), (and, {(and), (and), (and), (and), (and), (and), (and), (and)}), (are, {(are)}), (bad, {(bad)}), (but, {(but), (but)}), (can, {(can)}), (eat, {(eat)}), and (end, {(end)}).

```
2017-11-21 22:18:19,446 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2017-11-21 22:18:19,447 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
2017-11-21 22:18:19,447 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead,
use fs.defaultFS
2017-11-21 22:18:19,447 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated.
Instead, use mapreduce.job.counters.max
2017-11-21 22:18:19,447 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 22:18:19,464 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 22:18:19,464 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(a, {(a)})
(20, {(20)})
(an, {(an), (an)})
(by, {(by), (by), (by)})
(go, {(go), (go)})
(if, {(if)})
(in, {(in), (in)})
(is, {(is), (is), (is), (is), (is), (is)})
(of, {(of), (of), (of)})
(or, {(or)})
(so, {(so), (so)})
(to, {(to), (to), (to), (to), (to), (to), (to), (to)})
(up, {(up)})
(us, {(us)})
(we, {(we), (we), (we)})
(And, {(And), (And)})
(But, {(But)})
(The, {(The)})
(You, {(You)})
(all, {(all), (all)})
(and, {(and), (and), (and), (and), (and), (and), (and), (and)})
(are, {(are)})
(bad, {(bad)})
(but, {(but), (but)})
(can, {(can)})
(eat, {(eat)})
(end, {(end)})
```

grunt> wordcount = FOREACH grouped GENERATE group, COUNT(words);



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The terminal output consists of several log lines from Hadoop and Pig. The first three lines are warnings about JVM metrics initialization. The fourth line is an info message about the MapReduce launcher. The next three lines are deprecation warnings for Hadoop configuration properties: 'io.bytes.per.checksum', 'fs.default.name', and 'mapreduce.job.counters.limit'. The seventh line is a warning that 'SchemaTupleBackend' has already been initialized. The eighth and ninth lines are info messages about input paths. The final line of the log is a list of words and their counts: (a,1), (20,1), (an,2), (by,3), (go,2), (if,1), (in,2), (is,5), (of,3), (or,1), (so,2), (to,8), (up,1), (us,1), (we,3), (And,2), (But,1), (The,1), and (You,1). The taskbar at the bottom shows several open windows: 'acadgild', 'pig', 'asa.txt (~/workspa...', 'acadgild@localhost...', and 'workspace'.

```
2017-11-21 22:19:40,130 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 22:19:40,130 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 22:19:40,132 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 22:19:40,133 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-21 22:19:40,133 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-21 22:19:40,134 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-21 22:19:40,134 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-21 22:19:40,134 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 22:19:40,144 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 22:19:40,144 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(a,1)
(20,1)
(an,2)
(by,3)
(go,2)
(if,1)
(in,2)
(is,5)
(of,3)
(or,1)
(so,2)
(to,8)
(up,1)
(us,1)
(we,3)
(And,2)
(But,1)
(The,1)
(You,1)
```