

Assignment 9.2

```
student_data = LOAD 'studentDS' USING PigStorage(',') AS (stu_name:chararray,  
college:chararray, DOB:chararray,stream:chararray,grade:float,state:chararray,city:chararray);  
dump student_data;
```

1.Problem Statement

Find the number of students who scored less than 5(50%).

```
student_less_than_50 = filter student_data by grade<5.0;
```

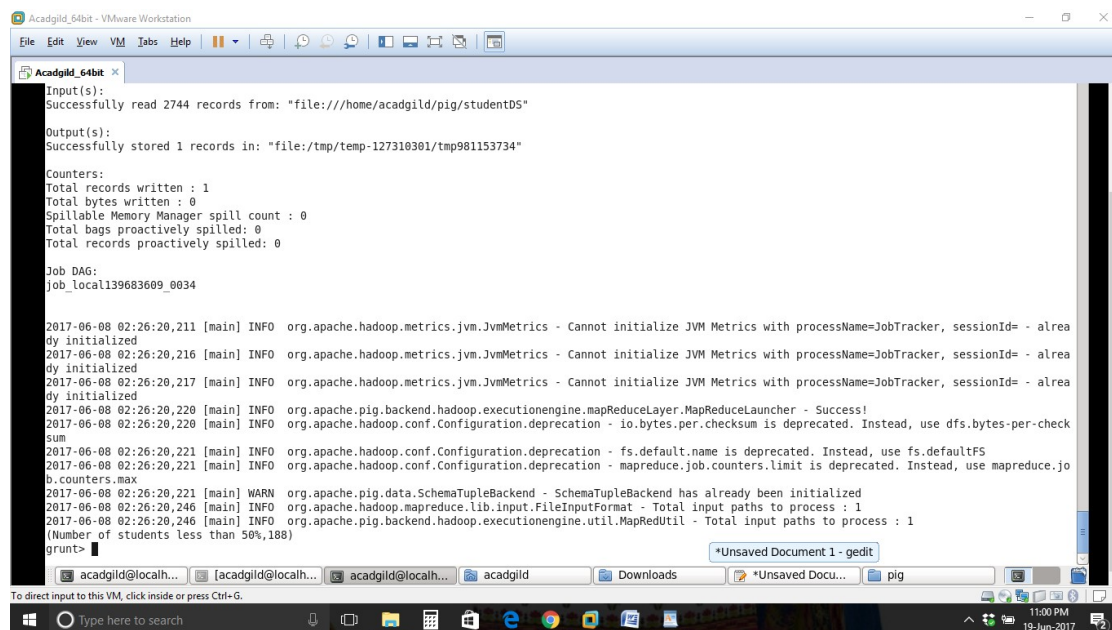
```
student_less_than_50_group = group student_less_than_50 all;
```

```
student_less_than_50_count = foreach student_less_than_50_group generate 'Number of  
students less than 50%' , COUNT($0);
```

```
student_less_than_50_count = foreach student_less_than_50_group generate 'Number of  
students less than 50%' , COUNT(student_less_than_50.grade);
```

```
dump student_less_than_50_count;
```

Number of Students who scored less than 50%



```
Acadgild_64bit x
File Edit View VM Tabs Help
Input(s):
Successfully read 2744 records from: "file:///home/acadgild/pig/studentDS"
Output(s):
Successfully stored 1 records in: "file:///tmp/temp-127310301/tmp981153734"
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_local139683609_0034
2017-06-08 02:26:20,211 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-06-08 02:26:20,216 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-06-08 02:26:20,217 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-06-08 02:26:20,220 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-08 02:26:20,220 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-06-08 02:26:20,221 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-08 02:26:20,221 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-06-08 02:26:20,221 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-08 02:26:20,246 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-08 02:26:20,246 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
(Number of students less than 50%,188)
ggrunt>
```

2. Problem Statement

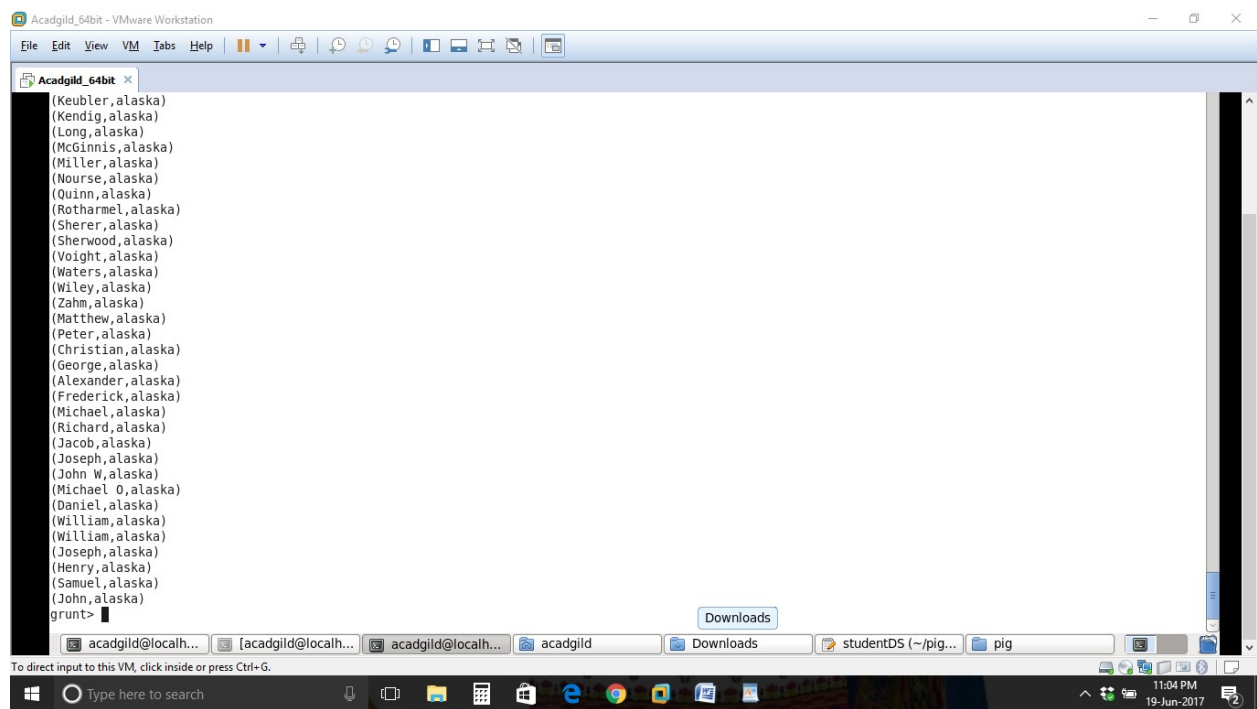
List the name of students who are from Alaska.

```
student_state = foreach student_data generate stu_name,state;
```

```
student_Alaska = filter student_state by state='alaska';
```

```
dump student_Alaska
```

Name of Students in Alaska



The screenshot shows a VMware Workstation window titled "Acadgild_64bit - VMware Workstation". Inside the window, a terminal window titled "Acadgild_64bit" displays a list of student names followed by the state "alaska" in parentheses. The list includes: (Keubler,alaska), (Kendig,alaska), (Long,alaska), (McGinnis,alaska), (Miller,alaska), (Nourse,alaska), (Quinn,alaska), (Rotharmel,alaska), (Sherer,alaska), (Sherwood,alaska), (Voight,alaska), (Waters,alaska), (Wiley,alaska), (Zahn,alaska), (Matthew,alaska), (Peter,alaska), (Christian,alaska), (George,alaska), (Alexander,alaska), (Frederick,alaska), (Michael,alaska), (Richard,alaska), (Jacob,alaska), (Joseph,alaska), (John W,alaska), (Michael O,alaska), (Daniel,alaska), (William,alaska), (William,alaska), (Joseph,alaska), (Henry,alaska), (Samuel,alaska), and (John,alaska). The terminal prompt "grunt>" is visible at the bottom. The VMware interface includes a menu bar (File, Edit, View, VM, Tabs, Help) and a toolbar. The host operating system's taskbar is visible at the bottom, showing the time as 11:04 PM on 19-Jun-2017.

```
(Keubler,alaska)
(Kendig,alaska)
(Long,alaska)
(McGinnis,alaska)
(Miller,alaska)
(Nourse,alaska)
(Quinn,alaska)
(Rotharmel,alaska)
(Sherer,alaska)
(Sherwood,alaska)
(Voight,alaska)
(Waters,alaska)
(Wiley,alaska)
(Zahn,alaska)
(Matthew,alaska)
(Peter,alaska)
(Christian,alaska)
(George,alaska)
(Alexander,alaska)
(Frederick,alaska)
(Michael,alaska)
(Richard,alaska)
(Jacob,alaska)
(Joseph,alaska)
(John W,alaska)
(Michael O,alaska)
(Daniel,alaska)
(William,alaska)
(William,alaska)
(Joseph,alaska)
(Henry,alaska)
(Samuel,alaska)
(John,alaska)
grunt>
```

3. Problem Statement

How many government collages are there in Alabama?

```
college_state_data = foreach student_data generate college,state;
```

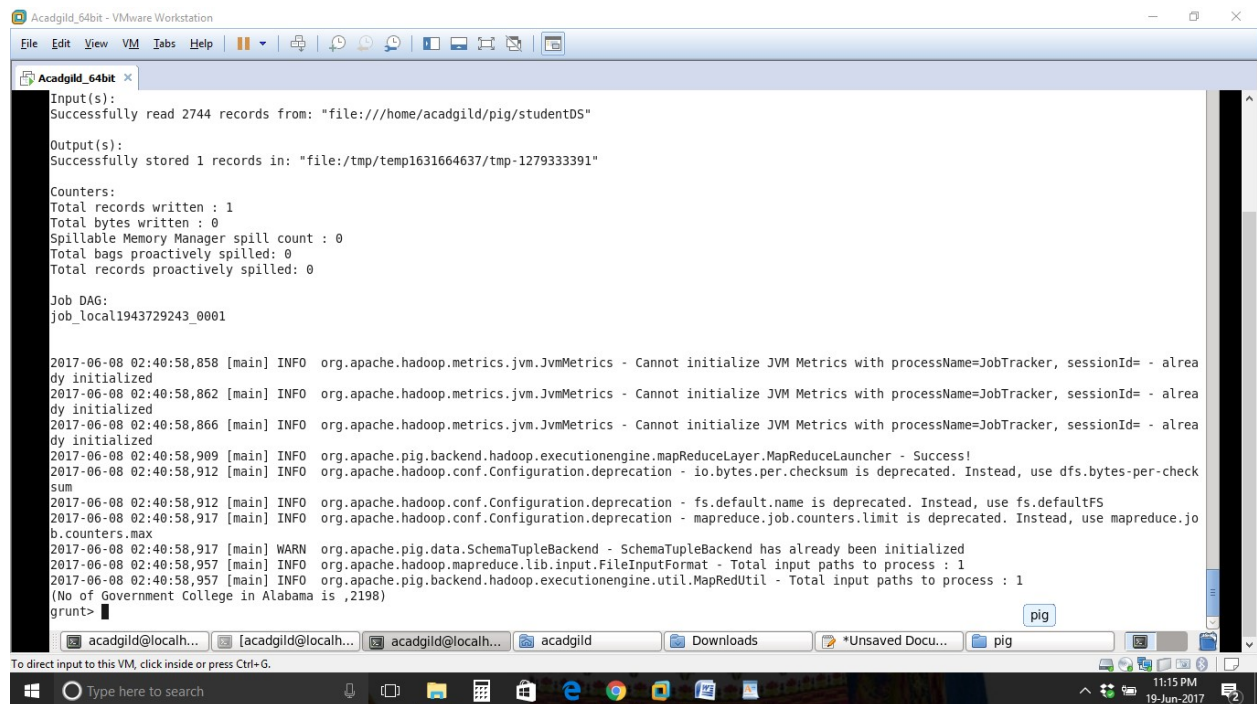
```
college_in_alabama = filter college_state_data by college matches'goverenment' and state matches'alabama';
```

```
government_college_alabama_group = group college_in_alabama all;
```

```
government_college_alabama_count = foreach government_college_alabama_group generate 'No of Government College in Alabama is ', COUNT(college_in_alabama.college);
```

```
dump government_college_alabama_count;
```

No. of Government Colleges in Alabama



```
Acadgild_64bit - VMWare Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Input(s):
Successfully read 2744 records from: "file:///home/acadgild/pig/studentDS"
Output(s):
Successfully stored 1 records in: "file:/tmp/temp1631664637/tmp-1279333391"
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_local1943729243_0001
2017-06-08 02:40:58,858 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-06-08 02:40:58,862 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-06-08 02:40:58,866 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-06-08 02:40:58,909 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-08 02:40:58,912 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-06-08 02:40:58,912 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-08 02:40:58,917 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-06-08 02:40:58,917 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-08 02:40:58,957 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-08 02:40:58,957 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(No of Government College in Alabama is ,2198)
grunt>
```

4.Problem Statement

List the name of students who are from Oregon and persuing BE.

```
student_name_details = foreach student_data generate stu_name,state,stream1;
```

```
student_oregon_BE = filter student_name_details by state matches'oregon' and stream1  
matches'BE';
```

```
student_name_oregon_BE = foreach student_oregon_BE generate stu_name;
```

```
dump student_name_oregon_BE;
```

