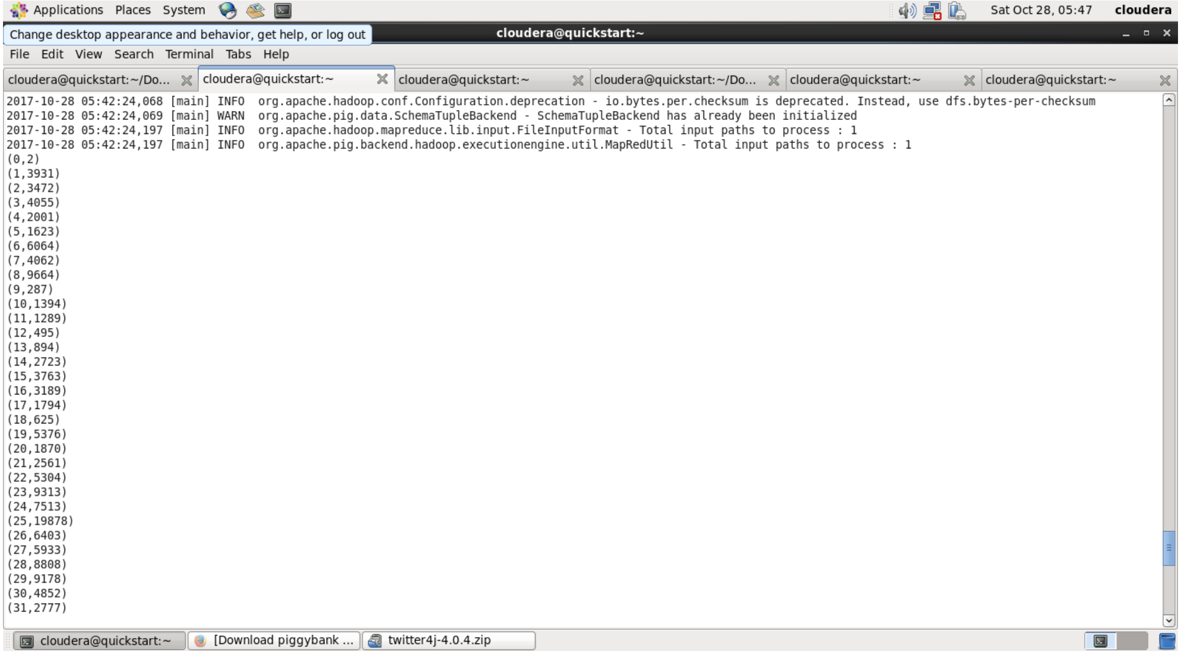
1. Write a MapReduce/Pig program to calculate the number of cases investigated under each FBI code

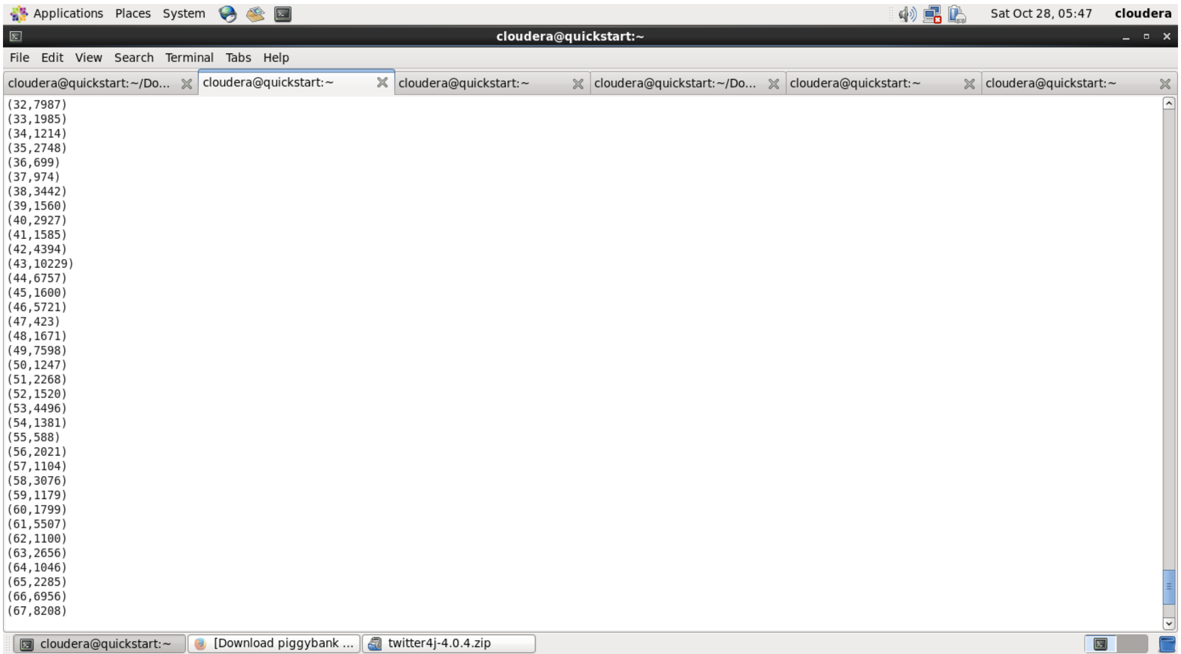
```
REGISTER '/home/cloudera/Downloads/jarfiles/piggybank-0.17.0.jar'
A = load '/home/cloudera/Downloads/Crimes_2001_to_present.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKI
P_INPUT_HEADER');
B = foreach A generate (chararray) $1 as case_number, (chararray) $9 as Arrest,
(chararray) $11 as District, (chararray) $13 as FBICode, (int)$17 as year;
C = filter B by FBICode is not null;
D = group C by FBICode;
E = foreach D generate group, COUNT(C.FBICode);
Dump E;
```

Execution: pig -x local <file_name>

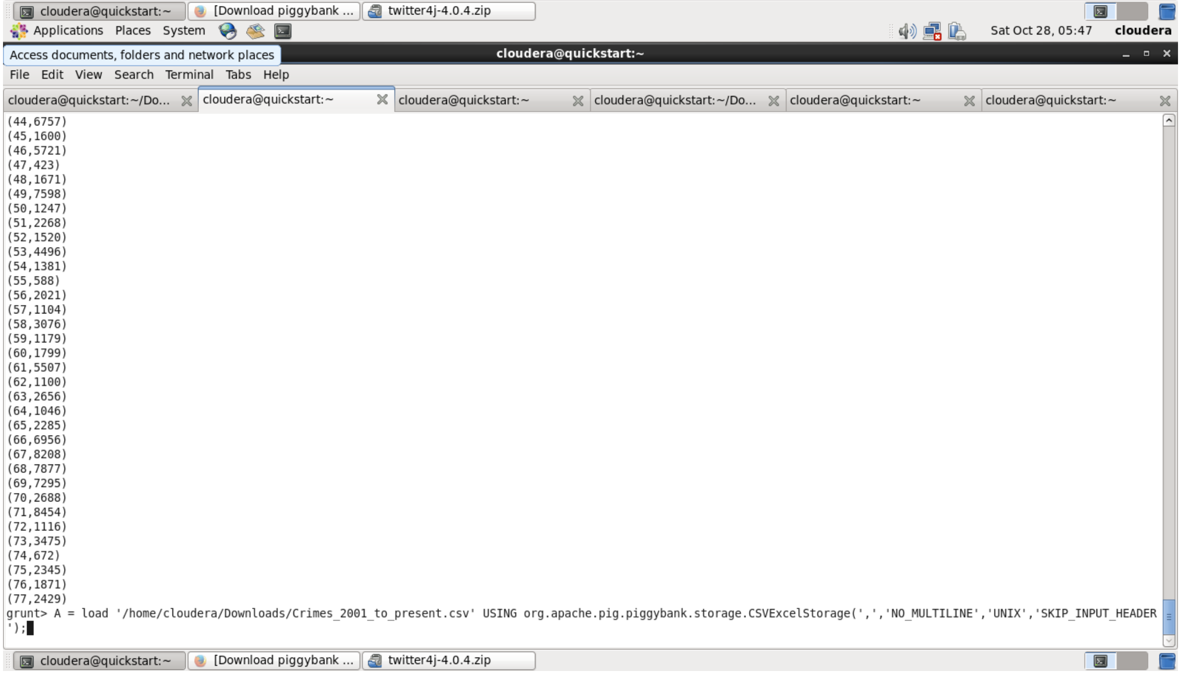
Output:



```
2017-10-28 05:42:24,068 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-28 05:42:24,069 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-28 05:42:24,197 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-28 05:42:24,197 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(0,2)
(1,3931)
(2,3472)
(3,4055)
(4,2001)
(5,1623)
(6,6064)
(7,4062)
(8,9664)
(9,287)
(10,1394)
(11,1289)
(12,495)
(13,894)
(14,2723)
(15,3763)
(16,3189)
(17,1794)
(18,625)
(19,5376)
(20,1870)
(21,2561)
(22,5304)
(23,9313)
(24,7513)
(25,19878)
(26,6403)
(27,5933)
(28,8808)
(29,9178)
(30,4852)
(31,2777)
```



```
(32,7987)
(33,1985)
(34,1214)
(35,2748)
(36,699)
(37,974)
(38,3442)
(39,1560)
(40,2927)
(41,1585)
(42,4394)
(43,10229)
(44,6757)
(45,1600)
(46,5721)
(47,423)
(48,1671)
(49,7598)
(50,1247)
(51,2268)
(52,1520)
(53,4496)
(54,1381)
(55,588)
(56,2021)
(57,1104)
(58,3076)
(59,1179)
(60,1799)
(61,5507)
(62,1100)
(63,2656)
(64,1046)
(65,2285)
(66,6956)
(67,8208)
```


Applications Places System
Sat Oct 28, 05:47 cloudera
Access documents, folders and network places
cloudera@quickstart:~
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~/Do...
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~/Do...
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~
[Download piggybank ...
twitter4j-4.0.4.zip

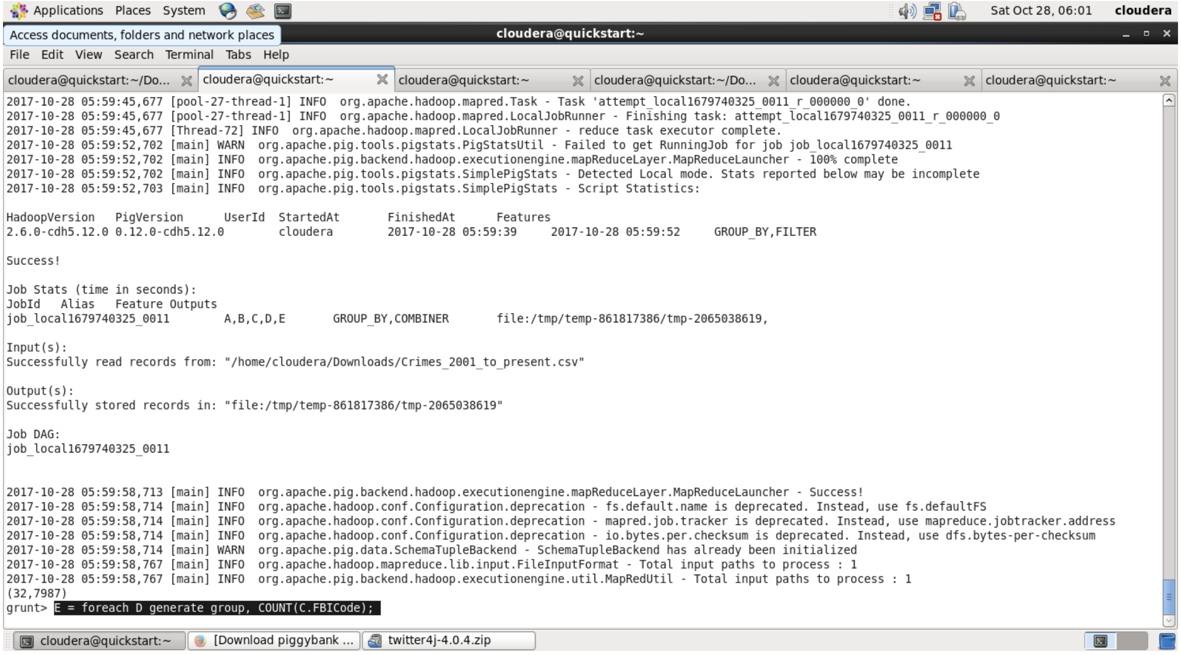
```
(44,6757)
(45,1600)
(46,5721)
(47,423)
(48,1671)
(49,7598)
(50,1247)
(51,2268)
(52,1520)
(53,4496)
(54,1381)
(55,588)
(56,2021)
(57,1104)
(58,3076)
(59,1179)
(60,1799)
(61,5507)
(62,1100)
(63,2656)
(64,1046)
(65,2285)
(66,6956)
(67,8208)
(68,7877)
(69,7295)
(70,2688)
(71,8454)
(72,1116)
(73,3475)
(74,672)
(75,2345)
(76,1871)
(77,2429)
grunt> A = load '/home/cloudera/Downloads/Crimes_2001_to_present.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER
');
```

2. Write a MapReduce/Pig program to calculate the number of cases investigated under FBI
code 32.

```
REGISTER '/home/cloudera/Downloads/jarfiles/piggybank-0.17.0.jar'
A = load '/home/cloudera/Downloads/Crimes_2001_to_present.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
B = foreach A generate (chararray) $13 as FBICode;
C = filter B by FBICode is not null and FBICode == '32';
D = group C by FBICode;
E = foreach D generate group, COUNT(C.FBICode);
Dump E;
```

Execution: pig -x local <file_name>

Output:



Applications Places System
Sat Oct 28, 06:01 cloudera
Access documents, folders and network places
cloudera@quickstart:~
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~/Do...
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~/Do...
cloudera@quickstart:~
cloudera@quickstart:~

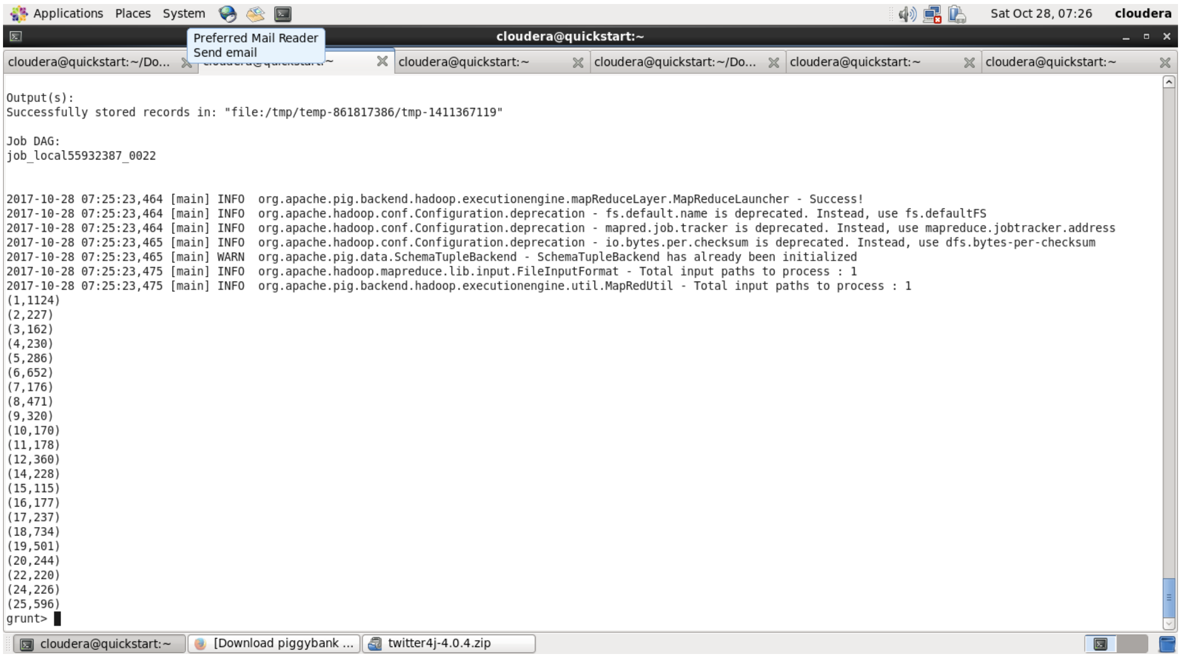
```
2017-10-28 05:59:45,677 [pool-27-thread-1] INFO  org.apache.hadoop.mapred.Task - Task 'attempt_local1679740325_0011_r_000000_0' done.
2017-10-28 05:59:45,677 [pool-27-thread-1] INFO  org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local1679740325_0011_r_000000_0
2017-10-28 05:59:45,677 [Thread-72] INFO  org.apache.hadoop.mapred.LocalJobRunner - reduce task executor complete.
2017-10-28 05:59:52,702 [main] WARN  org.apache.pig.tools.pigstats.PigStatsUtil - Failed to get RunningJob for job job_local1679740325_0011
2017-10-28 05:59:52,702 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2017-10-28 05:59:52,702 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Detected Local mode. Stats reported below may be incomplete
2017-10-28 05:59:52,703 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2017-10-28 05:59:39     2017-10-28 05:59:52     GROUP_BY,FILTER

Success!

Job Stats (time in seconds):
JobId   Alias   Feature Outputs
job_local1679740325_0011        A,B,C,D,E       GROUP_BY,COMBINER       file:/tmp/temp-861817386/tmp-2065038619,

Input(s):
Successfully read records from: "/home/cloudera/Downloads/Crimes_2001_to_present.csv"

Output(s):
Successfully stored records in: "file:/tmp/temp-861817386/tmp-2065038619"

Job DAG:
job_local1679740325_0011


2017-10-28 05:59:58,713 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-10-28 05:59:58,714 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-28 05:59:58,714 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-10-28 05:59:58,714 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-28 05:59:58,714 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-28 05:59:58,767 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-28 05:59:58,767 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(32,7987)
grunt> E = foreach D generate group, COUNT(C.FBICode);
```

3. Write a MapReduce/Pig program to calculate the number of arrests in theft district wise.

```
REGISTER '/home/cloudera/Downloads/jarfiles/piggybank-0.17.0.jar'
A = load '/home/cloudera/Downloads/Crimes_2001_to_present.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
B = foreach A generate (chararray) $8 as Arrest, (chararray) $5 as type, (int)$11 as district;
C = filter B by type == 'THEFT' and Arrest == 'true' and district is not null;
D = group C by district;
E = foreach D generate group, COUNT(C.district);
dump E;
```

Execution: pig -x local <file_name>

Output:



```
Output(s):
Successfully stored records in: "file:/tmp/temp-861817386/tmp-1411367119"

Job DAG:
job_local55932387_0022


2017-10-28 07:25:23,464 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-10-28 07:25:23,464 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-28 07:25:23,464 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-10-28 07:25:23,465 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-28 07:25:23,465 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-28 07:25:23,475 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-28 07:25:23,475 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,1124)
(2,227)
(3,162)
(4,230)
(5,286)
(6,652)
(7,176)
(8,471)
(9,320)
(10,170)
(11,178)
(12,360)
(14,228)
(15,115)
(16,177)
(17,237)
(18,734)
(19,501)
(20,244)
(22,220)
(24,226)
(25,596)
grunt> 
```

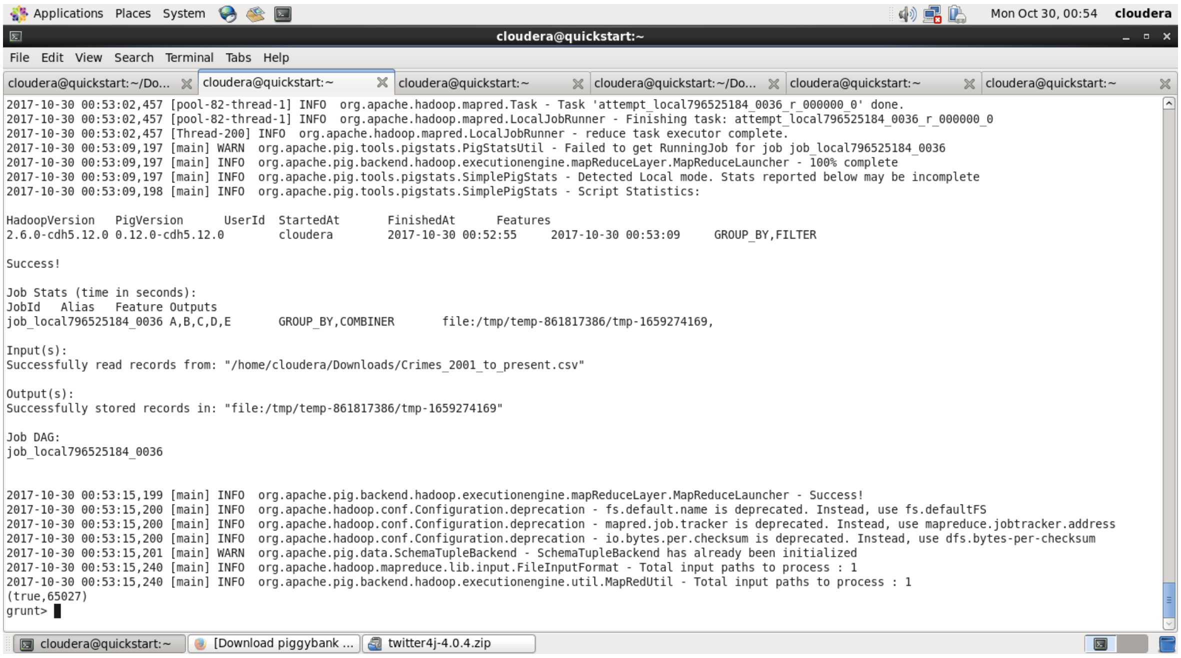
4. Write a MapReduce/Pig program to calculate the number of arrests done between October
2014 and October 2015.

Pig Script:

```
REGISTER '/home/cloudera/Downloads/jarfiles/piggybank-0.17.0.jar'
A = load '/home/cloudera/Downloads/Crimes_2001_to_present.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
B = foreach A generate (chararray) $8 as Arrest, ToDate($2,'MM/dd/yyyy HH:mm:ss aaa','America/Los_Angeles') as date;
C = filter B by Arrest == 'true' and date>=ToDate('2014-10-01') and date<=ToDate('2015-10-30’);
D = group C by Arrest;
E = foreach D generate group, COUNT(C.Arrest);
dump E;
```

Execution: pig -x local <file_name>

Output:



Applications Places System
Mon Oct 30, 00:54 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~/Do...
cloudera@quickstart:~
cloudera@quickstart:~
cloudera@quickstart:~/Do...
cloudera@quickstart:~
cloudera@quickstart:~

```
2017-10-30 00:53:02,457 [pool-82-thread-1] INFO  org.apache.hadoop.mapred.Task - Task 'attempt_local796525184_0036_r_000000_0' done.
2017-10-30 00:53:02,457 [pool-82-thread-1] INFO  org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local796525184_0036_r_000000_0
2017-10-30 00:53:02,457 [Thread-200] INFO  org.apache.hadoop.mapred.LocalJobRunner - reduce task executor complete.
2017-10-30 00:53:09,197 [main] WARN  org.apache.pig.tools.pigstats.PigStatsUtil - Failed to get RunningJob for job job_local796525184_0036
2017-10-30 00:53:09,197 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2017-10-30 00:53:09,197 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Detected Local mode. Stats reported below may be incomplete
2017-10-30 00:53:09,198 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2017-10-30 00:52:55     2017-10-30 00:53:09     GROUP_BY,FILTER

Success!

Job Stats (time in seconds):
JobId   Alias   Feature Outputs
job_local796525184_0036 A,B,C,D,E       GROUP_BY,COMBINER       file:/tmp/temp-861817386/tmp-1659274169,

Input(s):
Successfully read records from: "/home/cloudera/Downloads/Crimes_2001_to_present.csv"

Output(s):
Successfully stored records in: "file:/tmp/temp-861817386/tmp-1659274169"

Job DAG:
job_local796525184_0036


2017-10-30 00:53:15,199 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-10-30 00:53:15,200 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-30 00:53:15,200 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-10-30 00:53:15,200 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-30 00:53:15,201 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-30 00:53:15,240 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-30 00:53:15,240 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(true,65027)
grunt> 
```