1. Write a MapReduce/Pig program to calculate the number of cases investigated

under each

FBI code

REGISTER '/home/acadgild/Downloads/jarfiles/piggybank.jar'

A = load '/home/acadgild/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;

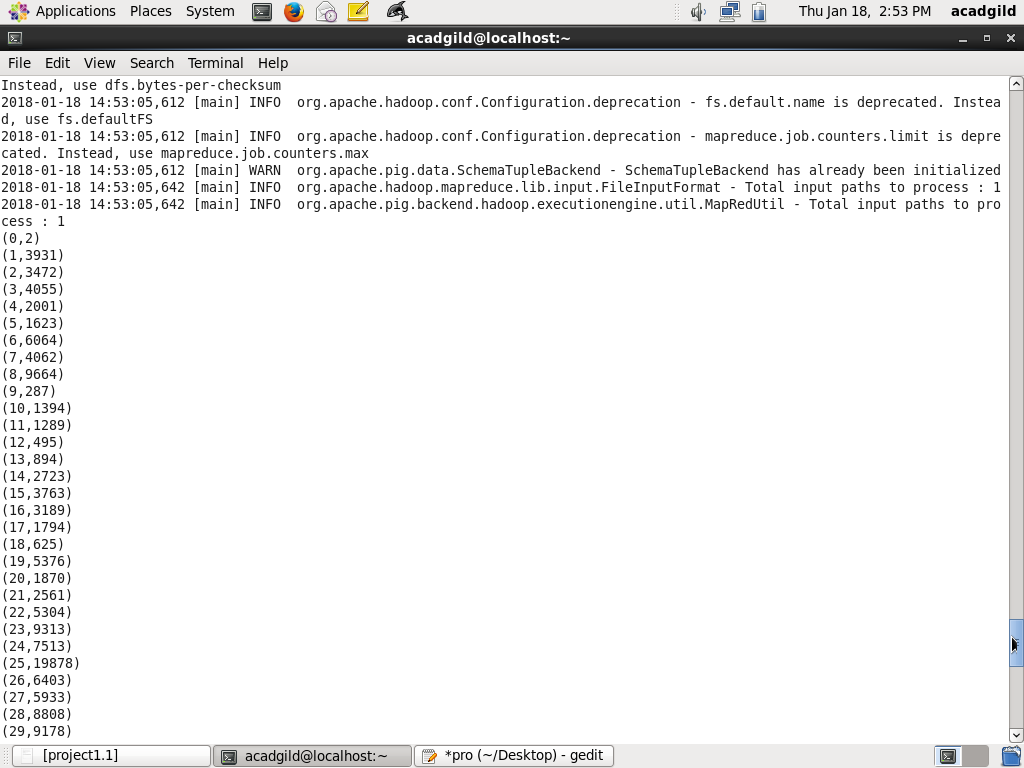
D = group C by FBICode;

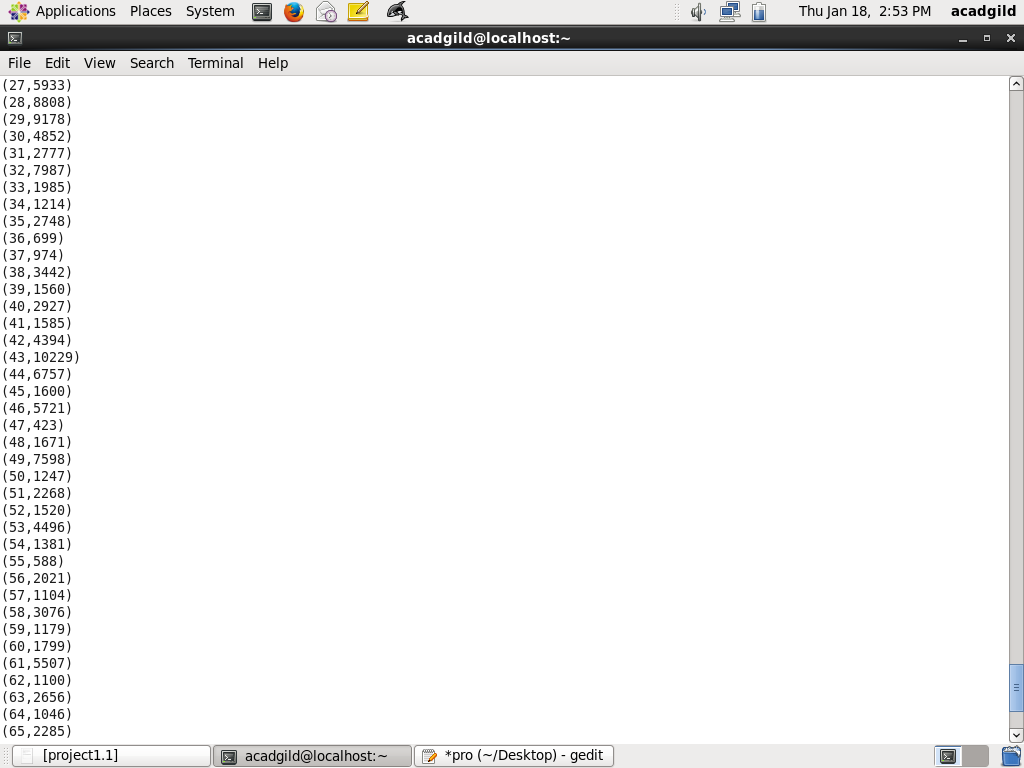
E = foreach D generate group, COUNT(C.FBICode);

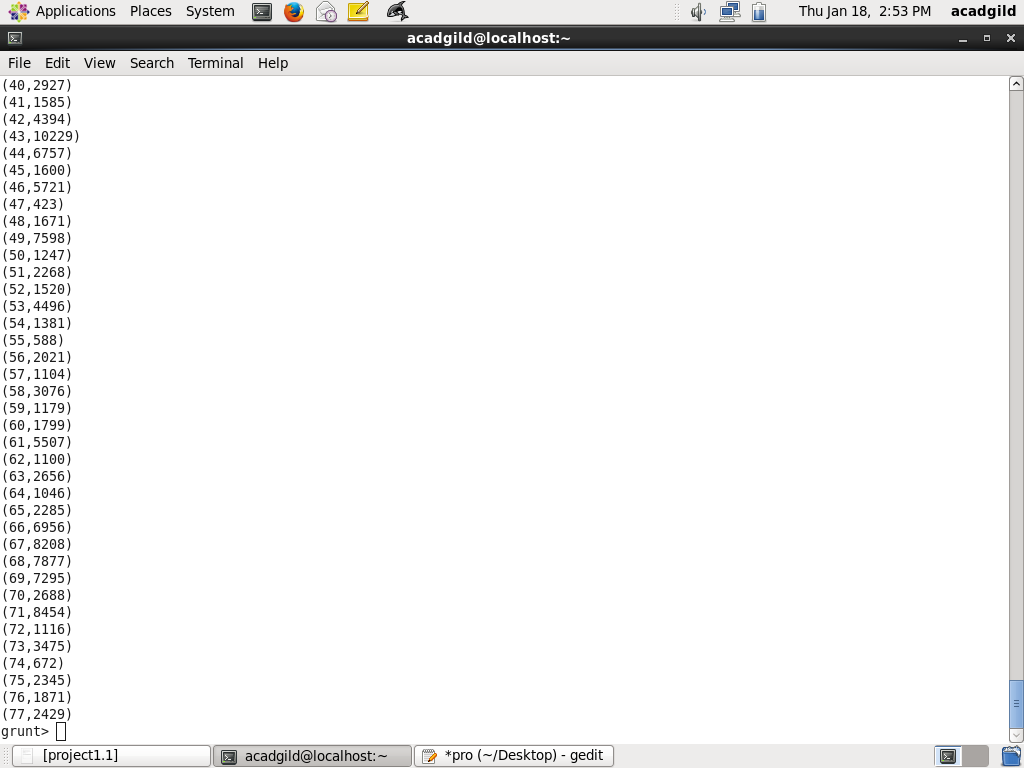
Dump E;

Execution: pig -x local <file\_name>

Output:







2. Write a MapReduce/Pig program to calculate the number of cases investigated

under FBI

code 32.

REGISTER '/home/acadgild/Downloads/jarfiles/piggybank.jar'

A = load '/home/acadgild/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) $13 as FBICode;

C = filter B by FBICode is not null and FBICode == '32';

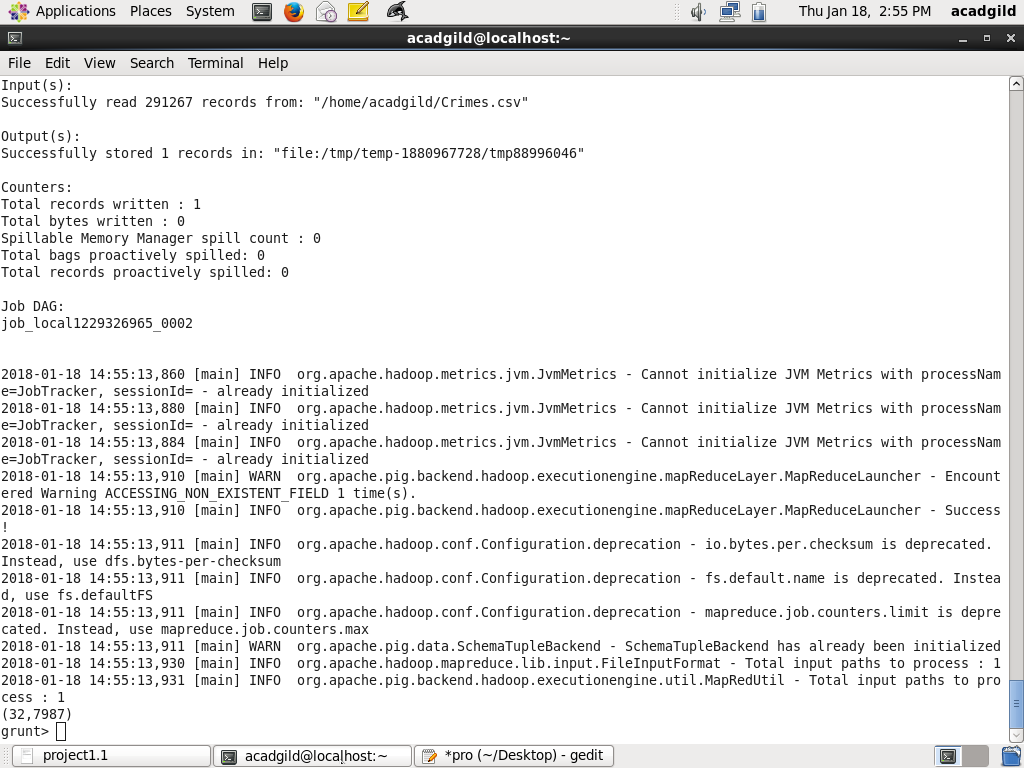
D = group C by FBICode;

E = foreach D generate group, COUNT(C.FBICode);

Dump E;

Execution: pig -x local <file\_name>

Output:



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

wise.

REGISTER '/home/acadgild/Downloads/jarfiles/piggybank.jar'

A = load '/home/acadgild/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) $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;

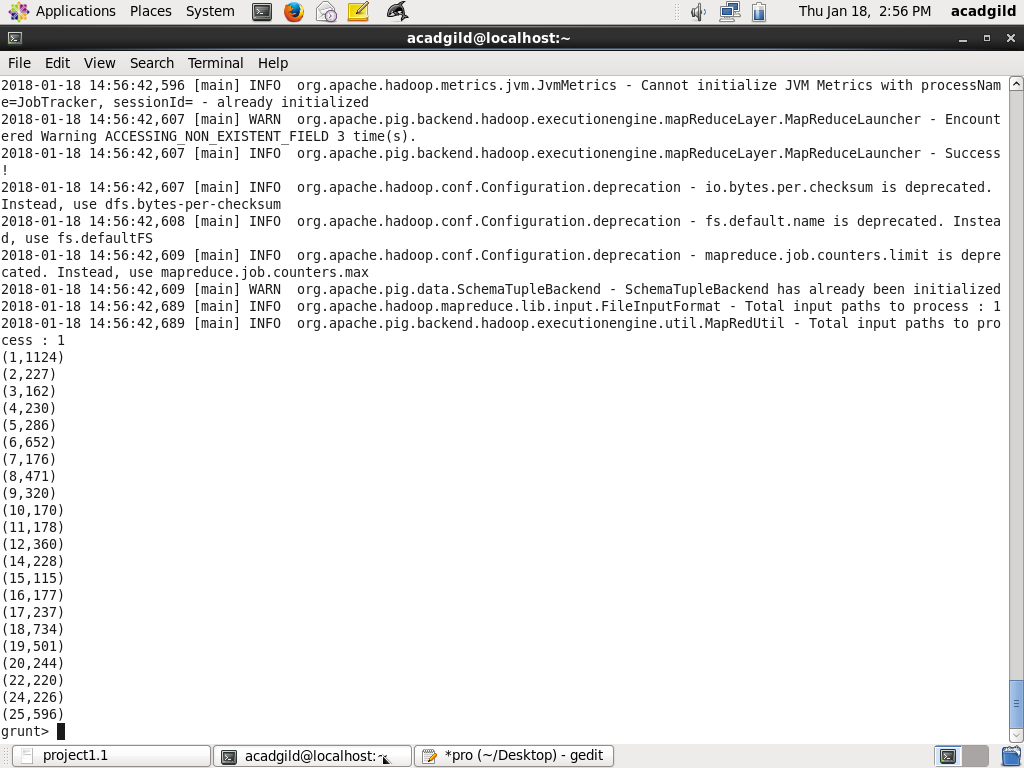
D = group C by district;

E = foreach D generate group, COUNT(C.district);

dump E;

Execution: pig -x local <file\_name>

Output:



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

October

2014 and October 2015.

Pig Script:

REGISTER '/home/acadgild/Downloads/jarfiles/piggybank.jar'

A = load '/home/acadgild/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) $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:

