**USA Crime Analysis**

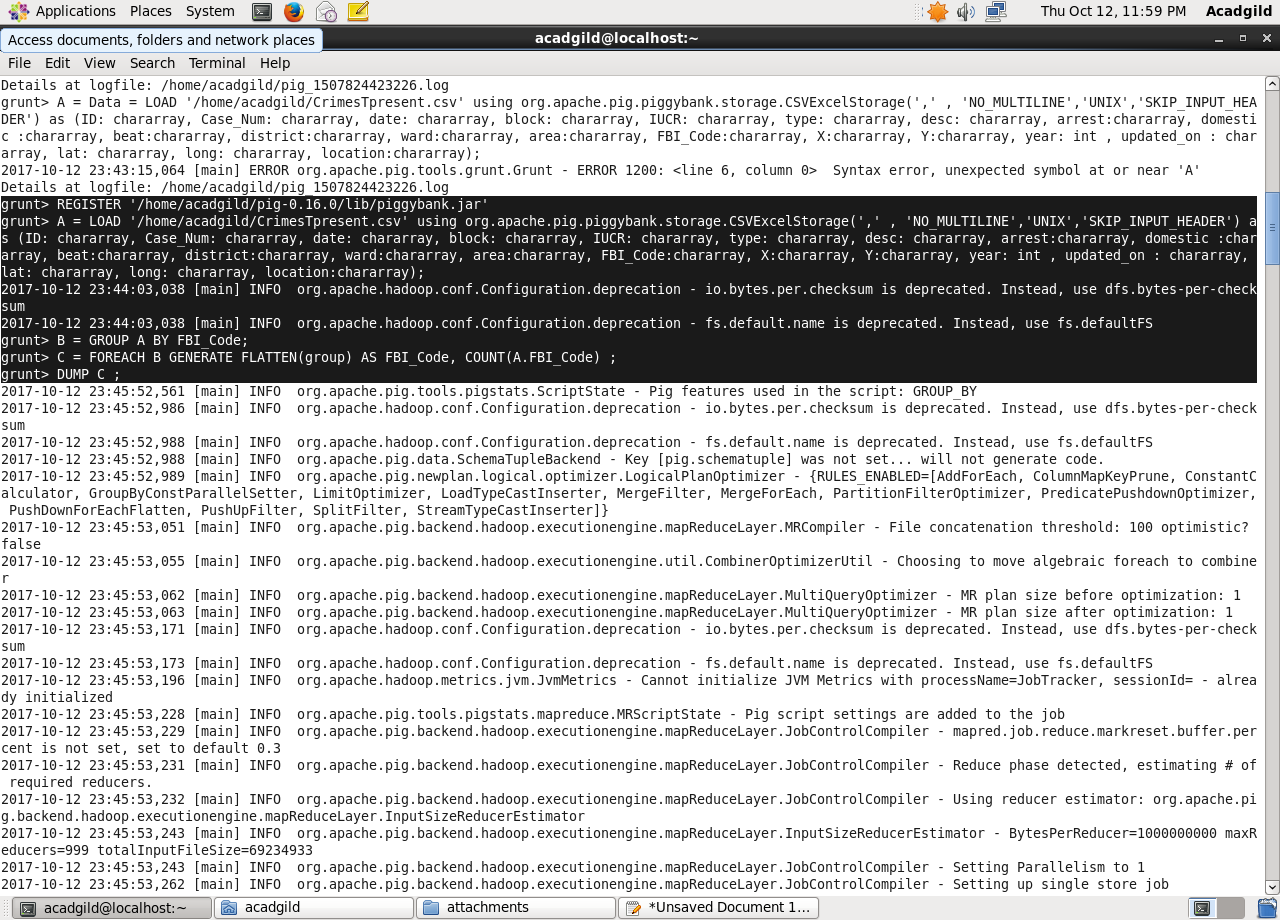
1. Write a MapReduce/pig program to calculate the number of cases investigated under each FBI code

Sol:

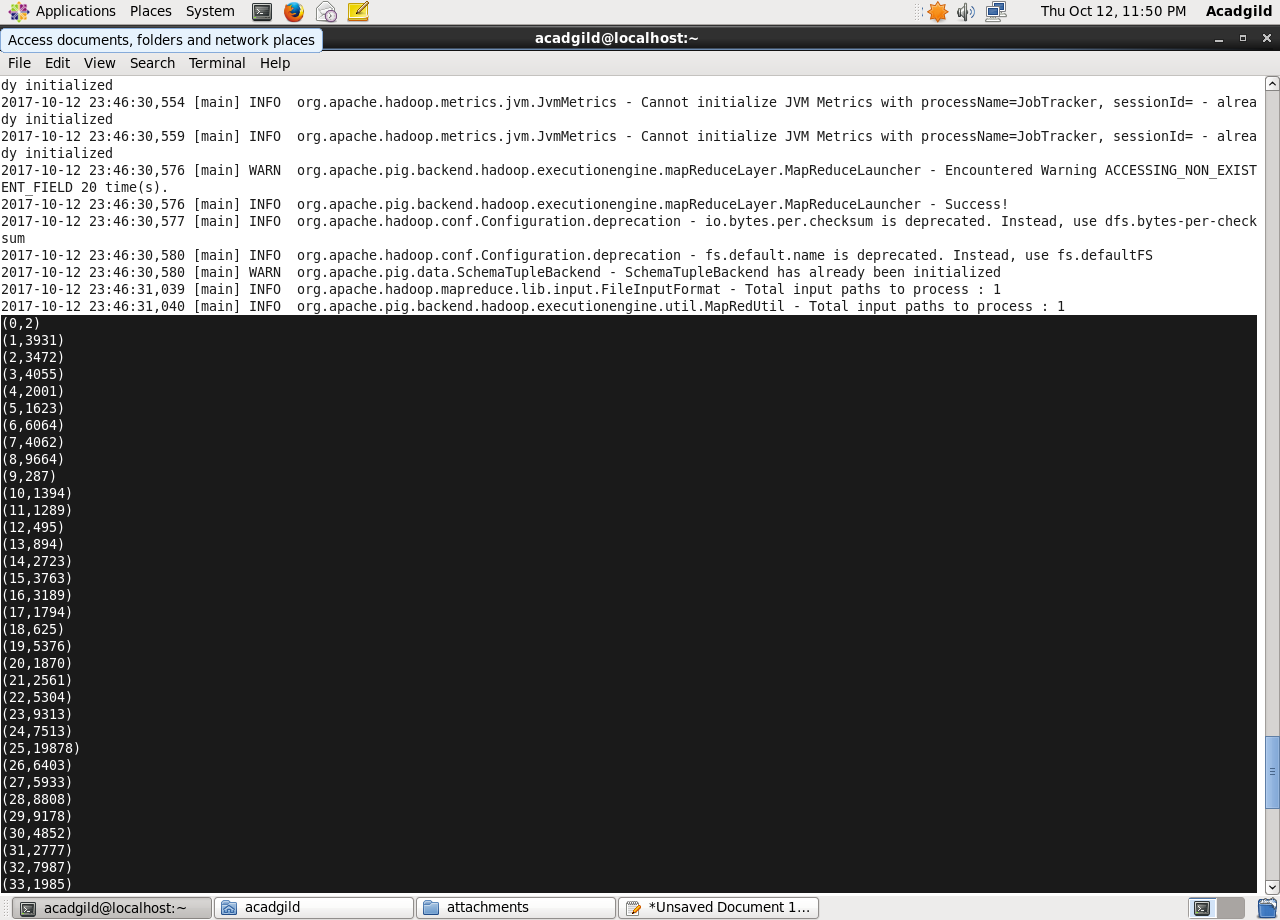
* Register piggybank.jar File.
* Download csvfile(dataset) and load csv using CsvExcelStorage command and perform the operations.

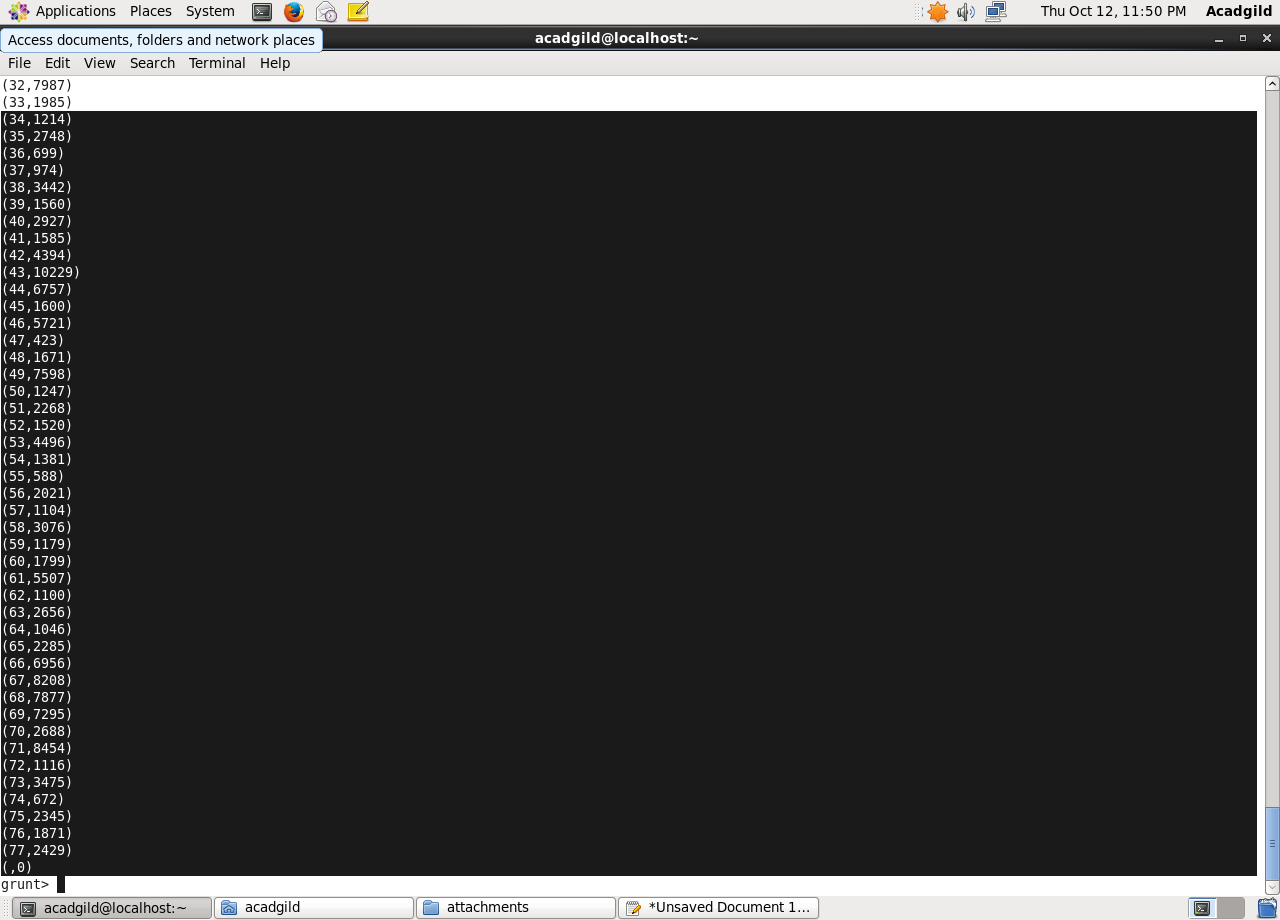
Commands:

* REGISTER '/home/acadgild/pig-0.16.0/lib/piggybank.jar';
* A = LOAD '/home/acadgild/CrimesTpresent.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',' , 'NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER') as (ID: chararray, Case\_Num: chararray, date: chararray, block: chararray, IUCR: chararray, type: chararray, desc: chararray, arrest:chararray, domestic :chararray, beat:chararray, district:chararray, ward:chararray, area:chararray, FBI\_Code:chararray, X:chararray, Y:chararray, year: int , updated\_on : chararray, lat: chararray, long: chararray, location:chararray);
* B = GROUP A BY FBI\_Code;
* C = FOREACH B GENERATE FLATTEN(group) AS FBI\_Code, COUNT(A.FBI\_Code) ;
* DUMP C;



Output:



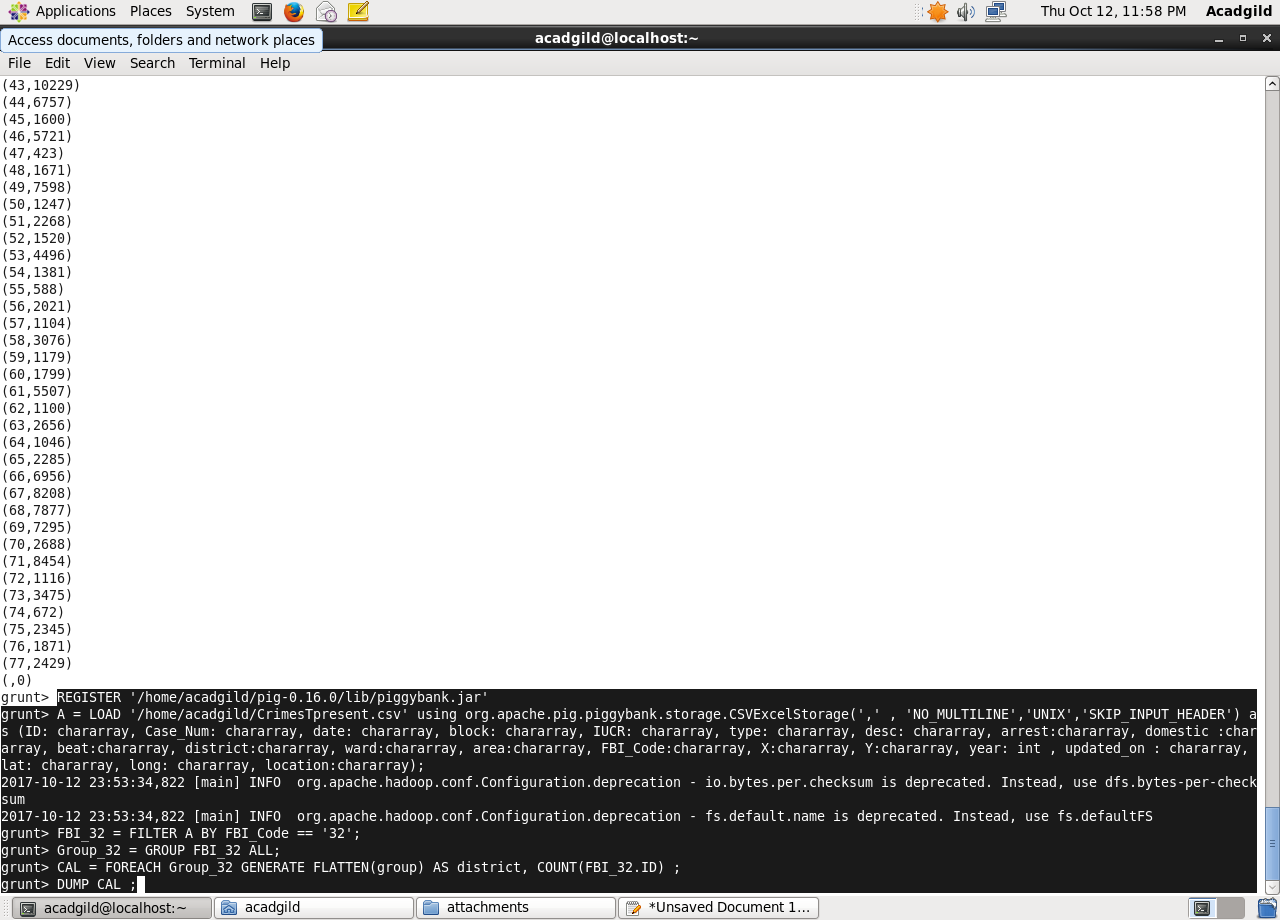


1. Write a mapreduce/Pig program to calculate the number of cases investigated under FBI code 32.

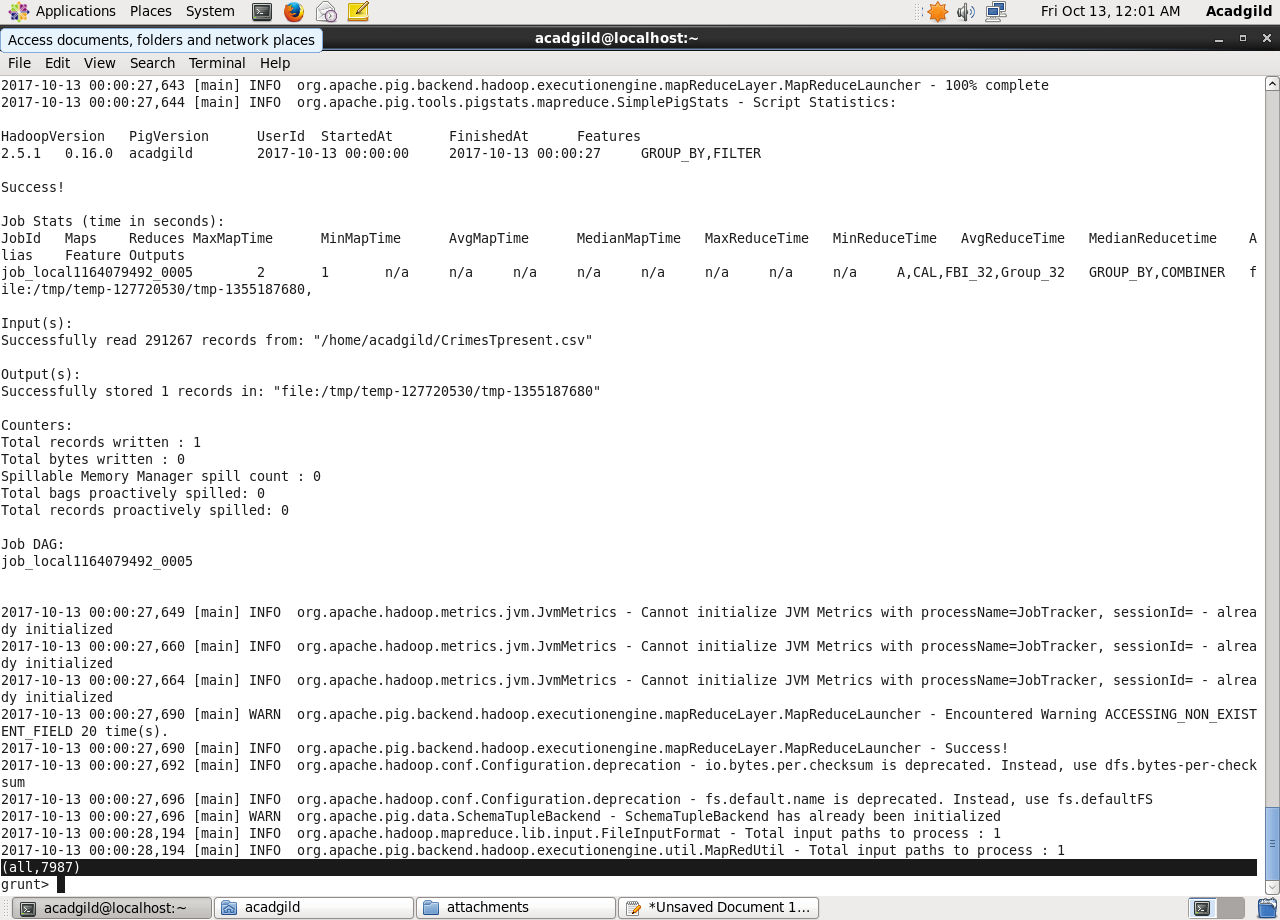
Sol:

* REGISTER ‘/home/acadgild/pig-0.16.0/lib/piggybank.jar';
* A = LOAD '/home/acadgild/CrimesTpresent.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',' , 'NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER') as (ID: chararray, Case\_Num: chararray, date: chararray, block: chararray, IUCR: chararray, type: chararray, desc: chararray, arrest:chararray, domestic :chararray, beat:chararray, district:chararray, ward:chararray, area:chararray, FBI\_Code:chararray, X:chararray, Y:chararray, year: int , updated\_on : chararray, lat: chararray, long: chararray, location:chararray);
* FBI\_32 = FILTER A BY FBI\_Code == '32';
* Group\_32 = GROUP FilterData ALL;
* CAL = FOREACH Group\_32 GENERATE FLATTEN(group) AS district, COUNT(FBI\_32.ID) ;
* DUMP CAL;

Commands:



Output:

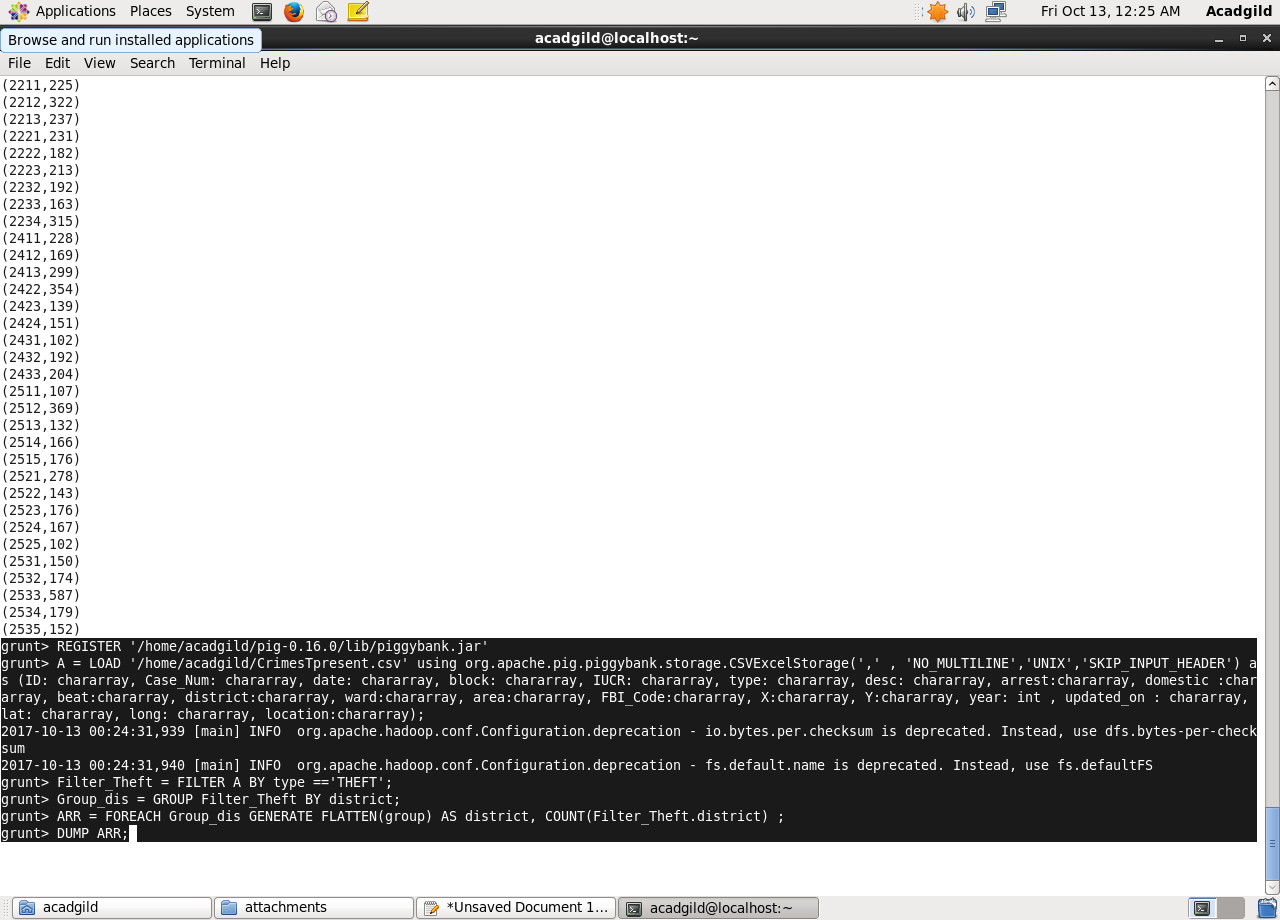


3.Write a Mapreduce/pig program to calculate the number of arrests in theft district wise.

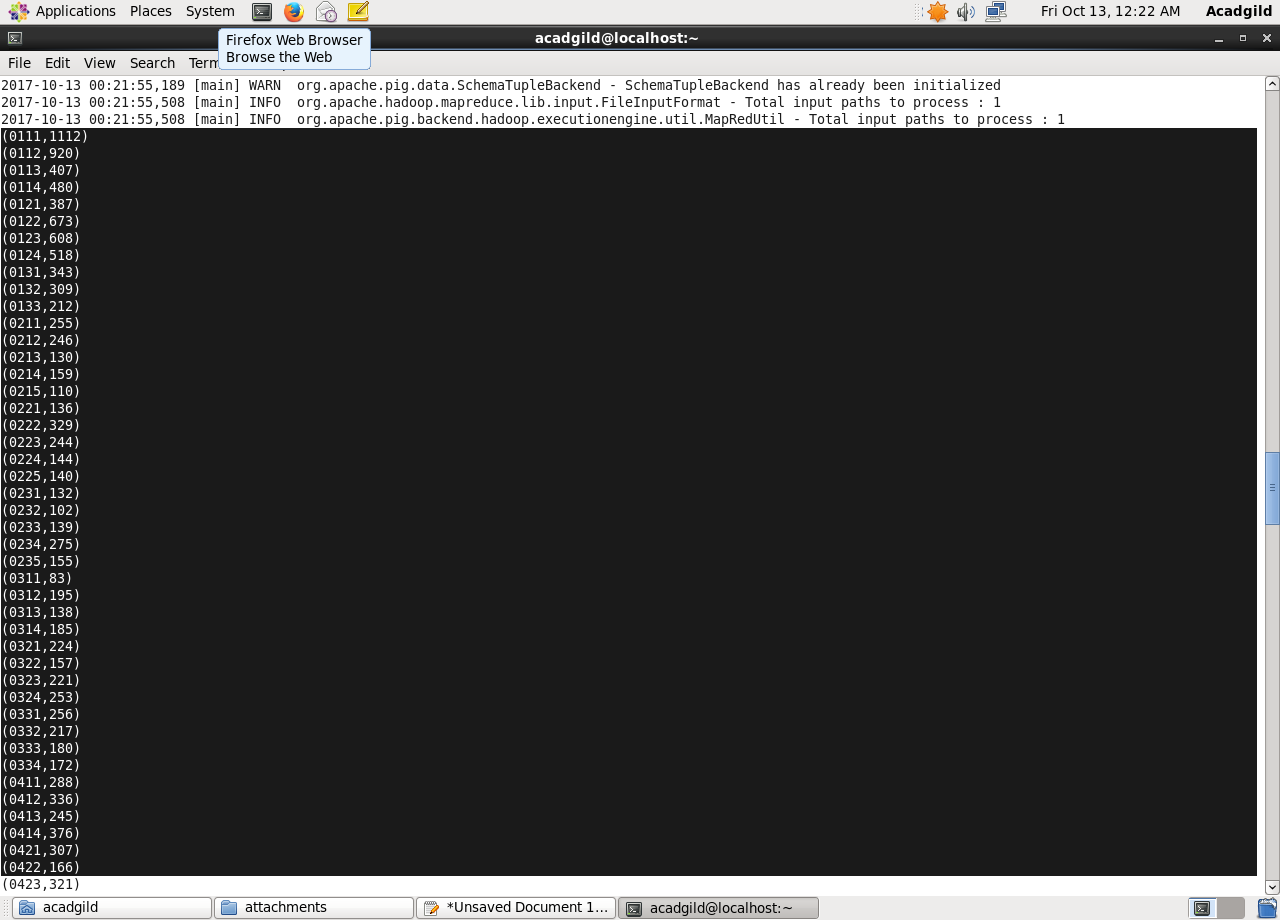
Sol:

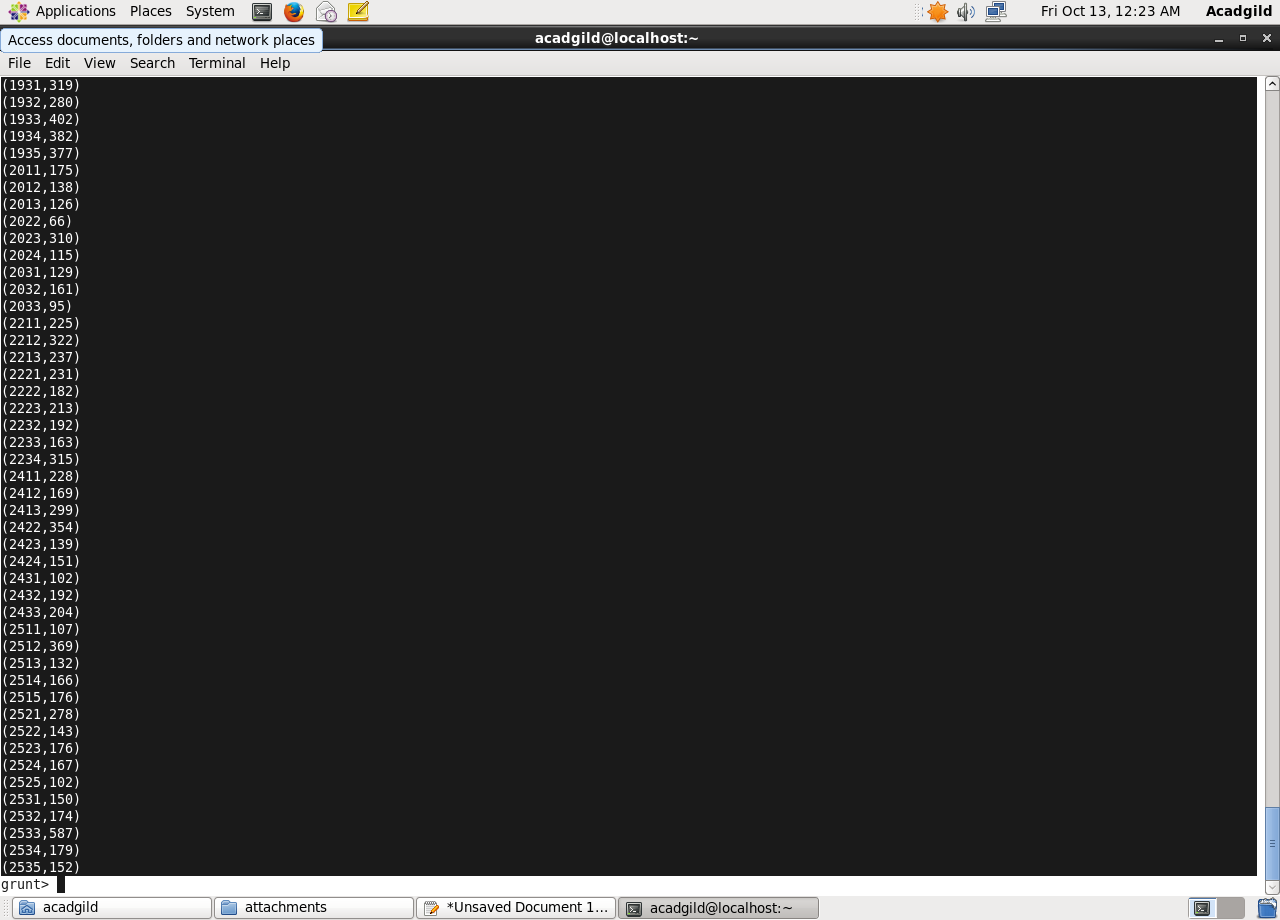
* REGISTER ‘/home/acadgild/pig-0.16.0/lib/piggybank.jar';
* A = LOAD '/home/acadgild/CrimesTpresent.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',' , 'NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER') as (ID: chararray, Case\_Num: chararray, date: chararray, block: chararray, IUCR: chararray, type: chararray, desc: chararray, arrest:chararray, domestic :chararray, beat:chararray, district:chararray, ward:chararray, area:chararray, FBI\_Code:chararray, X:chararray, Y:chararray, year: int , updated\_on : chararray, lat: chararray, long: chararray, location:chararray);
* Filter\_Theft = FILTER A BY type =='THEFT';
* Group\_dis = GROUP Filter\_Theft BY district;
* ARR = FOREACH Group\_dis GENERATE FLATTEN(group) AS district, COUNT(Filter\_Theft.district) ;
* DUMP ARR;

Commands:



Output:





4.Write a MapReduce/PIG script to calculate the number of arrests done between october 2014 and october 2015

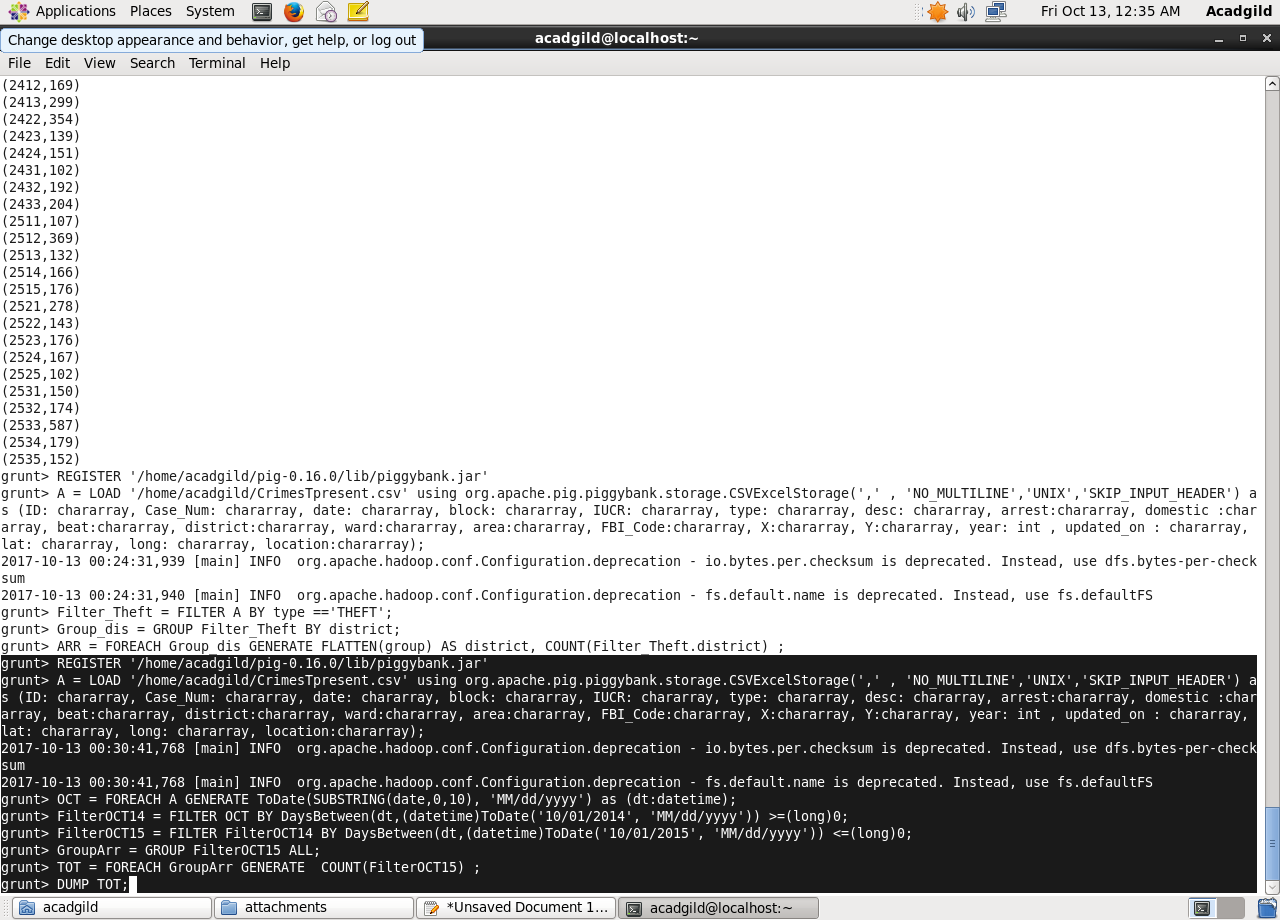
Sol:

* REGISTER ‘/home/acadgild/pig-0.16.0/lib/piggybank.jar';
* A = LOAD '/home/acadgild/CrimesTpresent.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',' , 'NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER') as (ID: chararray, Case\_Num: chararray, date: chararray, block: chararray, IUCR: chararray, type: chararray, desc: chararray, arrest:chararray, domestic :chararray, beat:chararray, district:chararray, ward:chararray, area:chararray, FBI\_Code:chararray, X:chararray, Y:chararray, year: int , updated\_on : chararray, lat: chararray, long: chararray, location:chararray);
* OCT = FOREACH A GENERATE ToDate(SUBSTRING(date,0,10), 'MM/dd/yyyy') as (dt:datetime);

grunt>FilterOCT14 = FILTER OCT BY DaysBetween(dt,(datetime)ToDate('10/01/2014', 'MM/dd/yyyy')) >=(long)0;

* FilterOCT15 = FILTER FilterOCT14 BY DaysBetween(dt,(datetime)ToDate('10/01/2015', 'MM/dd/yyyy')) <=(long)0;
* GroupArr = GROUP FilterOCT15 ALL;
* TOT = FOREACH GroupArr GENERATE COUNT(FilterOCT15) ;
* DUMP TOT;

Commnads:



Output:

