**ASSIGNMENT-12**

**Project 1: USA Crime analysis**

**(US-CRIME-ANALYSIS)**

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

FBI code

dataset contains attributes related to crimes taking place in various areas like type of crime, FBI code related to that criminal case, arrest frequency, location of crime etc.

**crime1.pig**

Data = load '/home/acadgild/Downloads/Crimes\_-\_2001\_to\_present.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO\_MULTILINE', 'UNIX') 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, FBIcode:chararray, X:chararray, Y:chararray, year: int , updated\_on : chararray, lat: chararray, long: chararray, location:chararray);

GroupData = group Data by FBIcode;

Crimecount = foreach GroupData GENERATE group as FBIcode,COUNT(Data.FBIcode);

Store Crimecount into '/home/acadgild/session12/1' USING PigStorage(',');

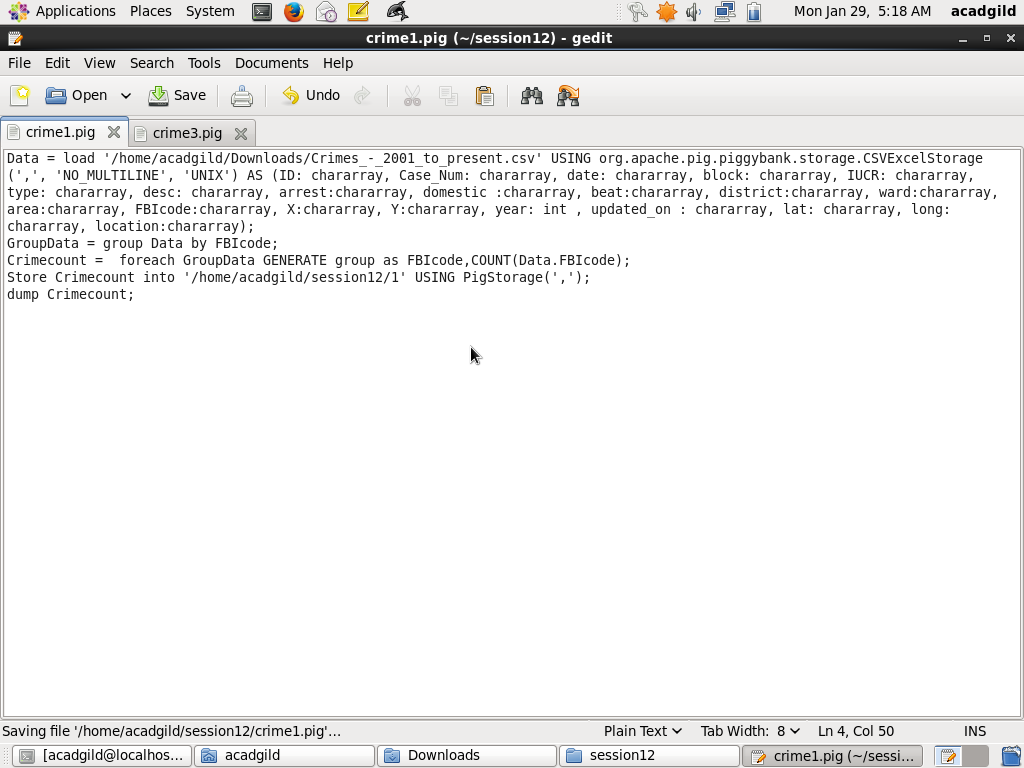
dump Crimecount;

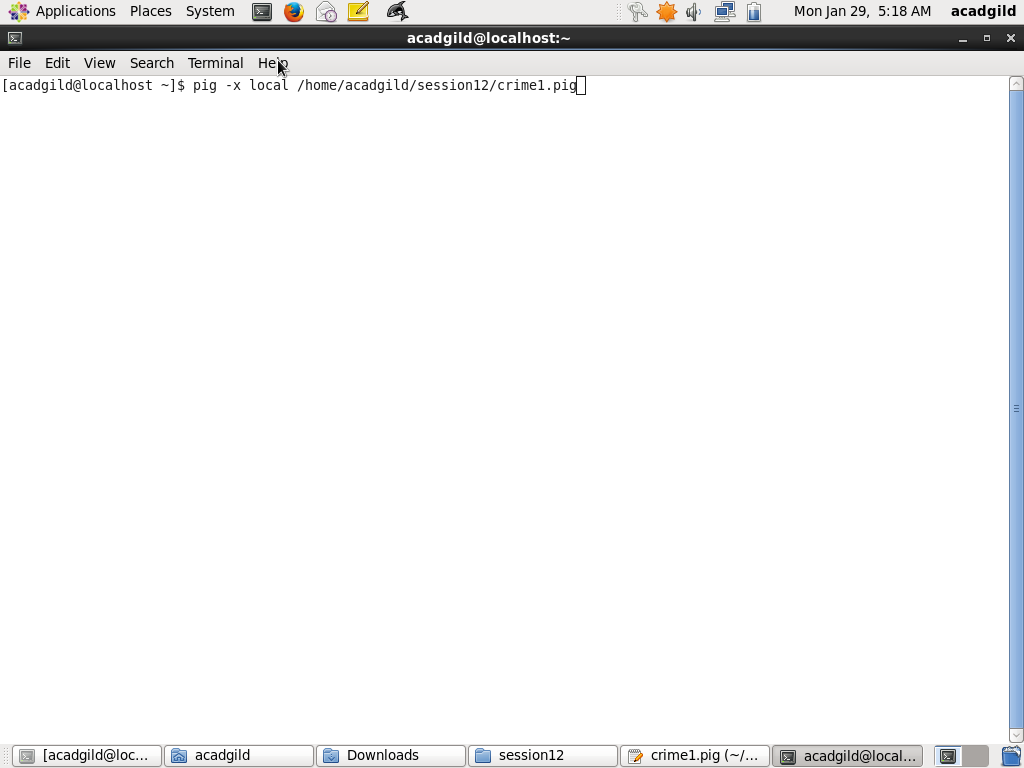
**Execution of code:**

Pig -x local /home/acadgild/session12/crime1.pig

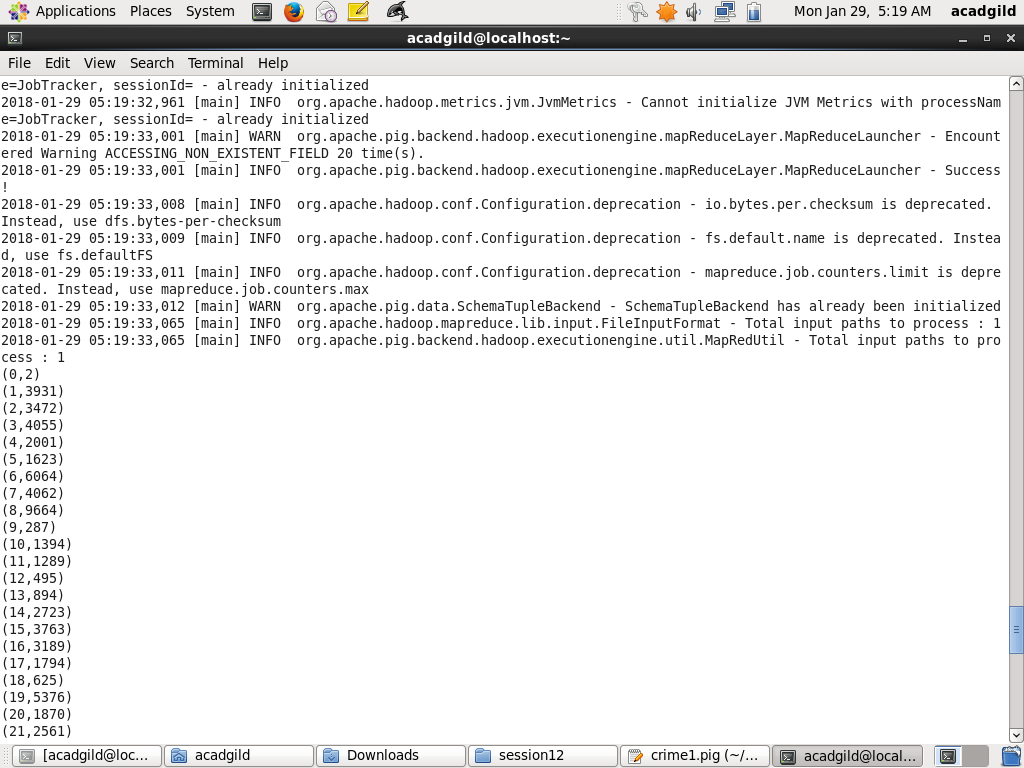
Loaded data into local path which are required to solve the problem.The input file provided is in csv format and is placed in /home/acadgild/Downloads.

The contents loaded from the csv file are grouped based on FBICode

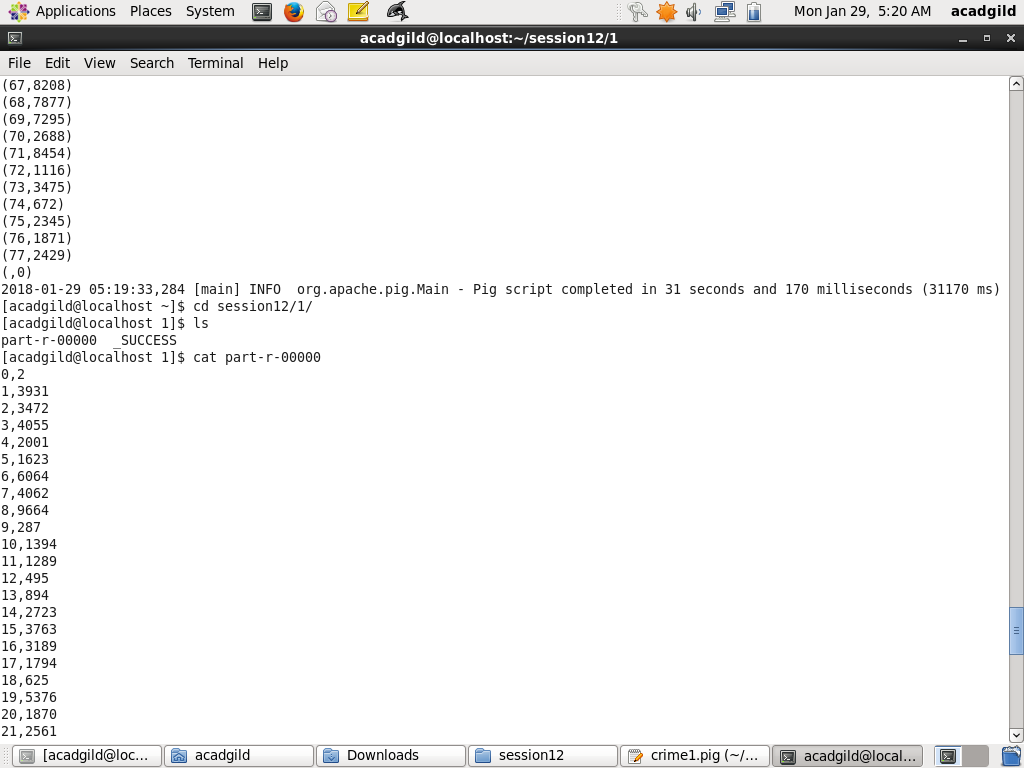




**Dump command dumps the final result after execution**



**Store command: Results are stored in the file it can be viewed using cat command**



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

**crime2.pig**

Data = load '/home/acadgild/Downloads/Crimes\_-\_2001\_to\_present.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO\_MULTILINE', 'UNIX') 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, FBIcode:chararray, X:chararray, Y:chararray, year: int , updated\_on : chararray, lat: chararray, long: chararray, location:chararray);

GroupData = group Data by FBIcode;

Crimecount = foreach GroupData GENERATE group as FBIcode,COUNT(Data.FBIcode);

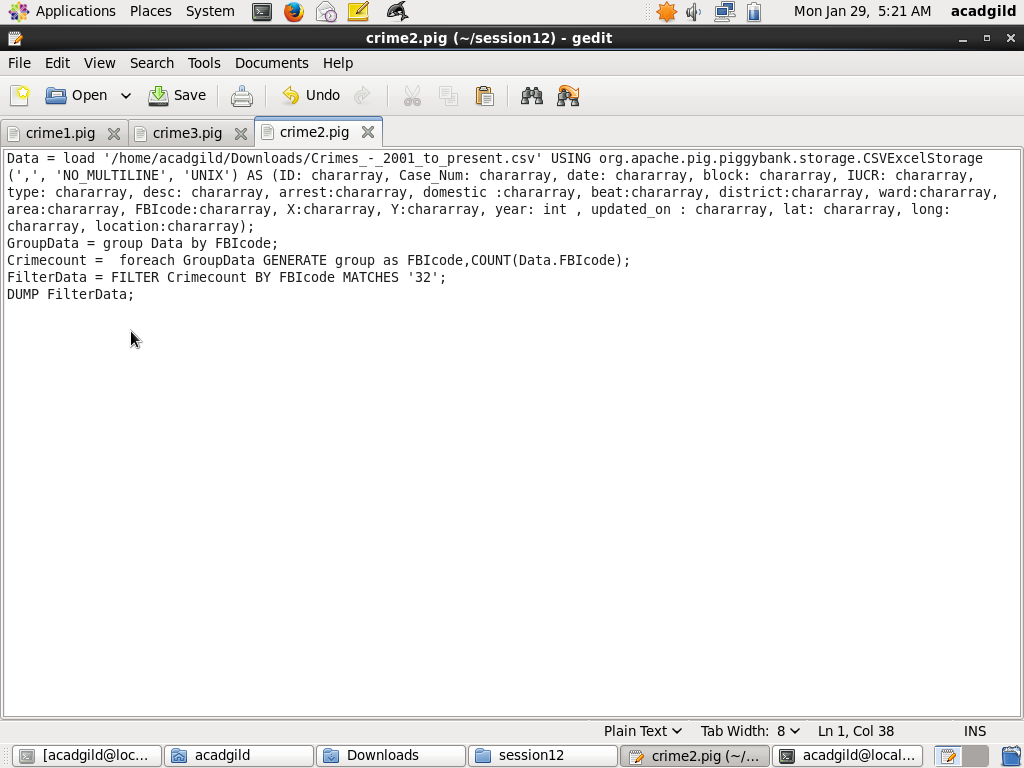
FilterData = FILTER Crimecount BY FBIcode MATCHES '32';

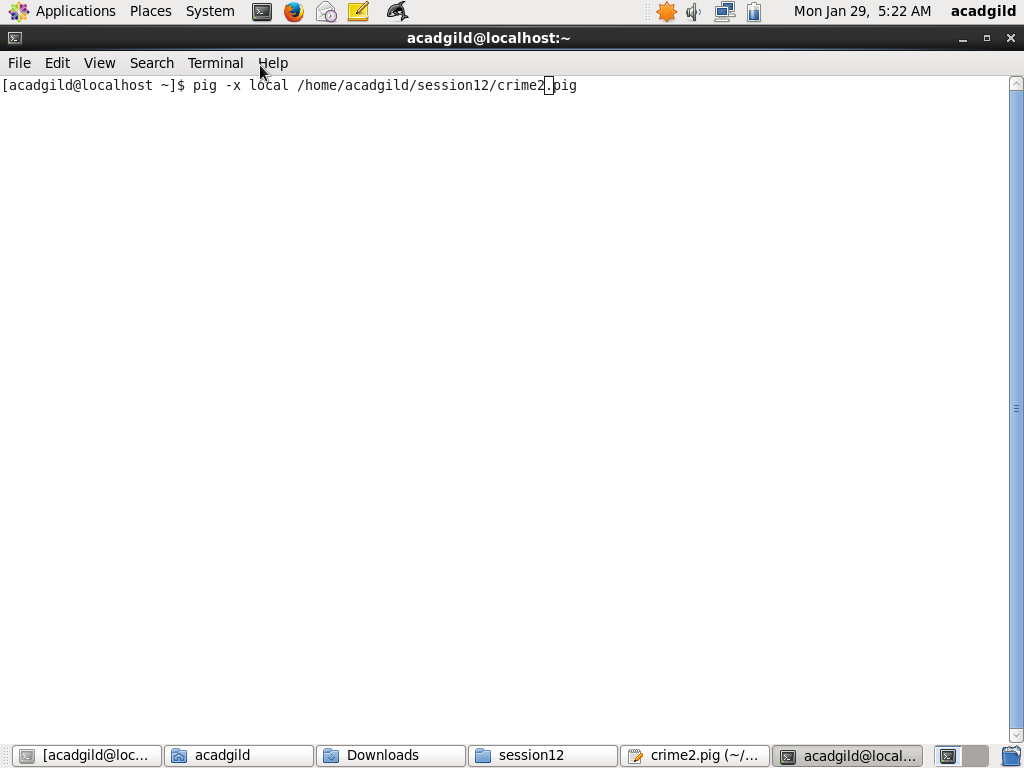
DUMP FilterData;

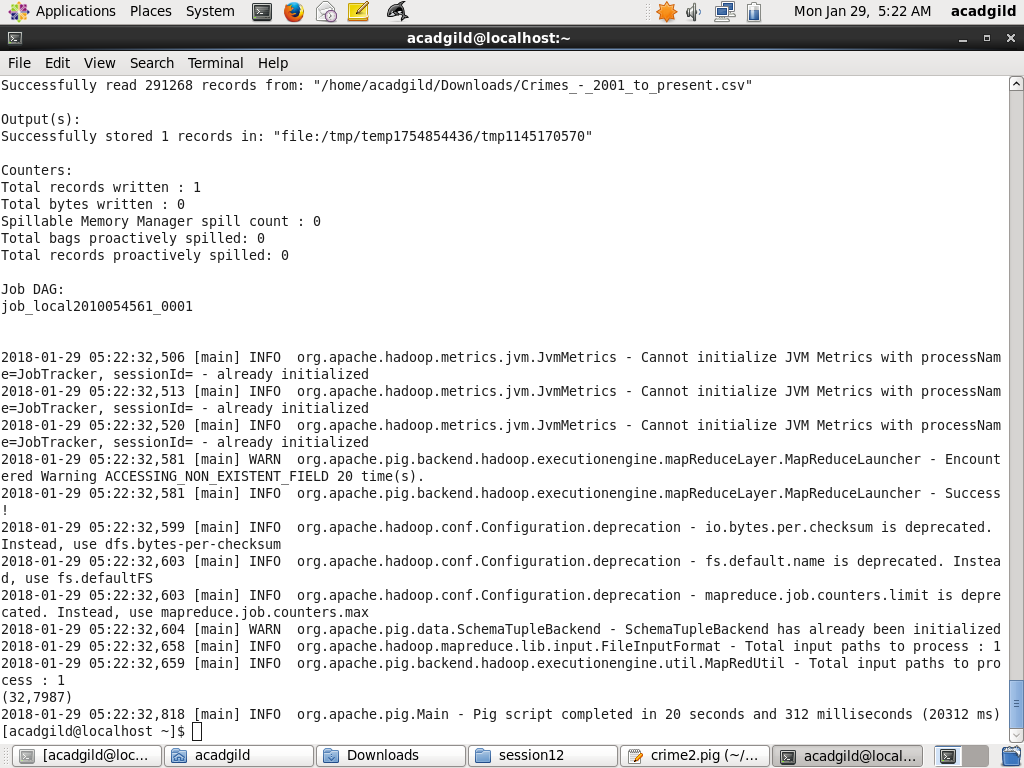
**Execution of Command:**

**Pig -x local /home/acadgild/session12/crime2.pig**

The loaded data is filtered by FBICode. The records with FBICode MATCHES 32 needs to be pulled out.







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

**Crime3.pig**

Data = load '/home/acadgild/Downloads/Crimes\_-\_2001\_to\_present.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO\_MULTILINE', 'UNIX') 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, FBIcode:chararray, X:chararray, Y:chararray, year: int , updated\_on : chararray, lat: chararray, long: chararray, location:chararray);

FilterData = FILTER Data BY type == 'THEFT';

GroupData = GROUP FilterData BY district;

crimecount = foreach GroupData GENERATE group as district,COUNT(FilterData.Case\_Num);

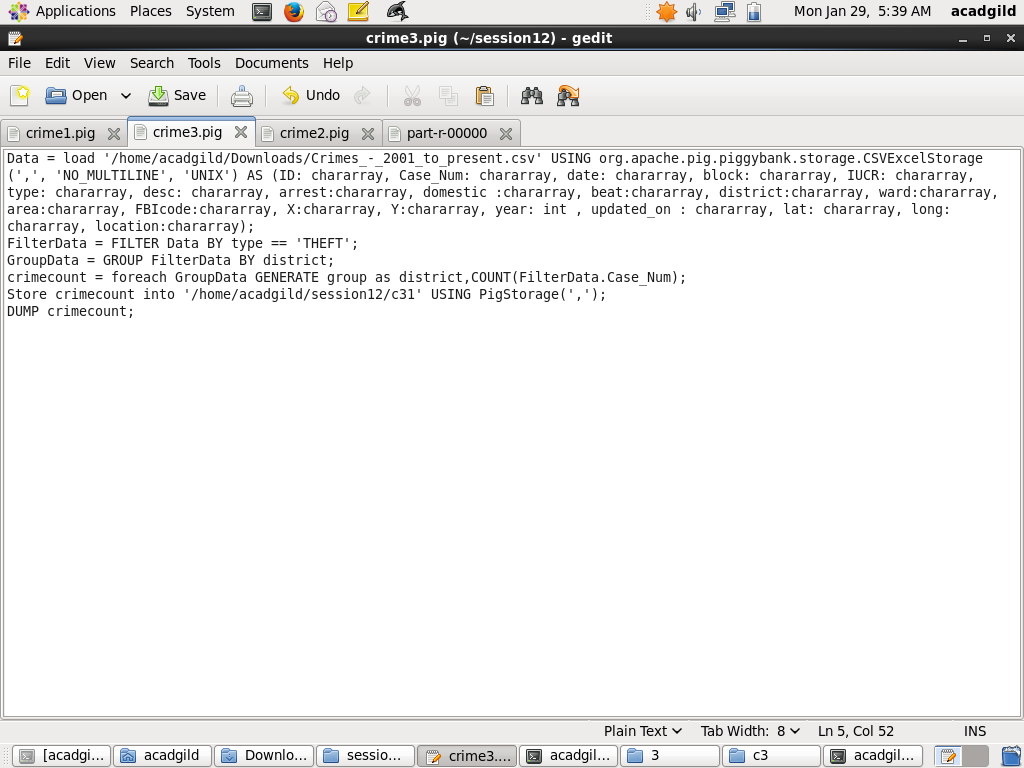
Store crimecount into '/home/acadgild/session12/c31' USING PigStorage(',');

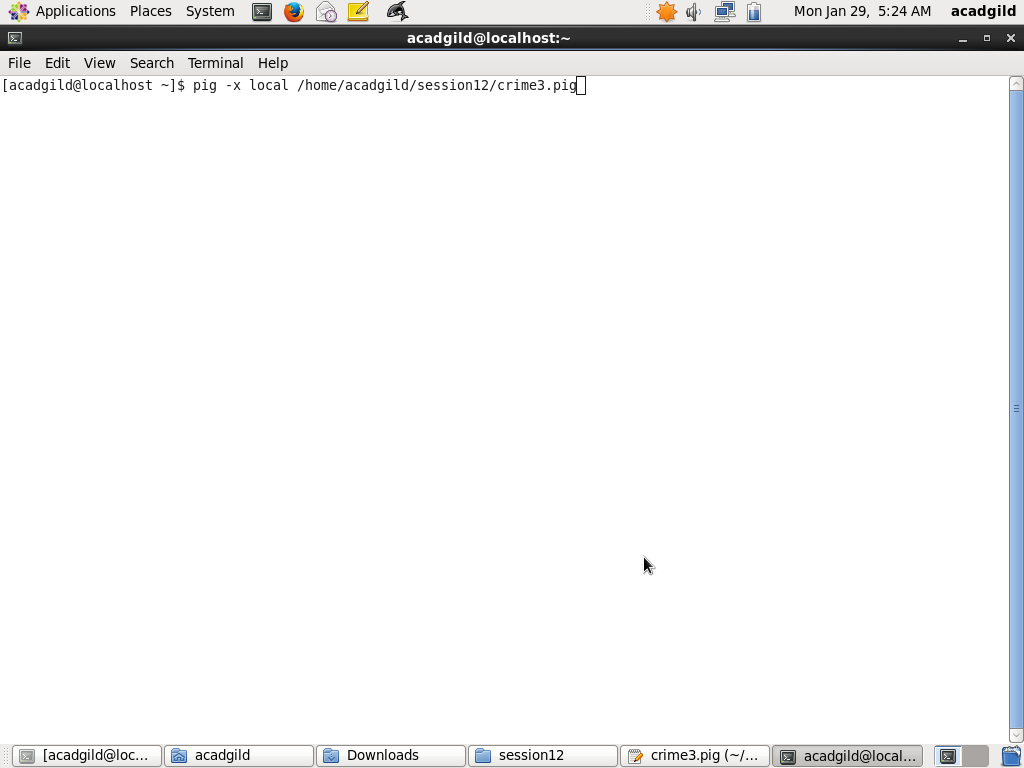
DUMP crimecount;

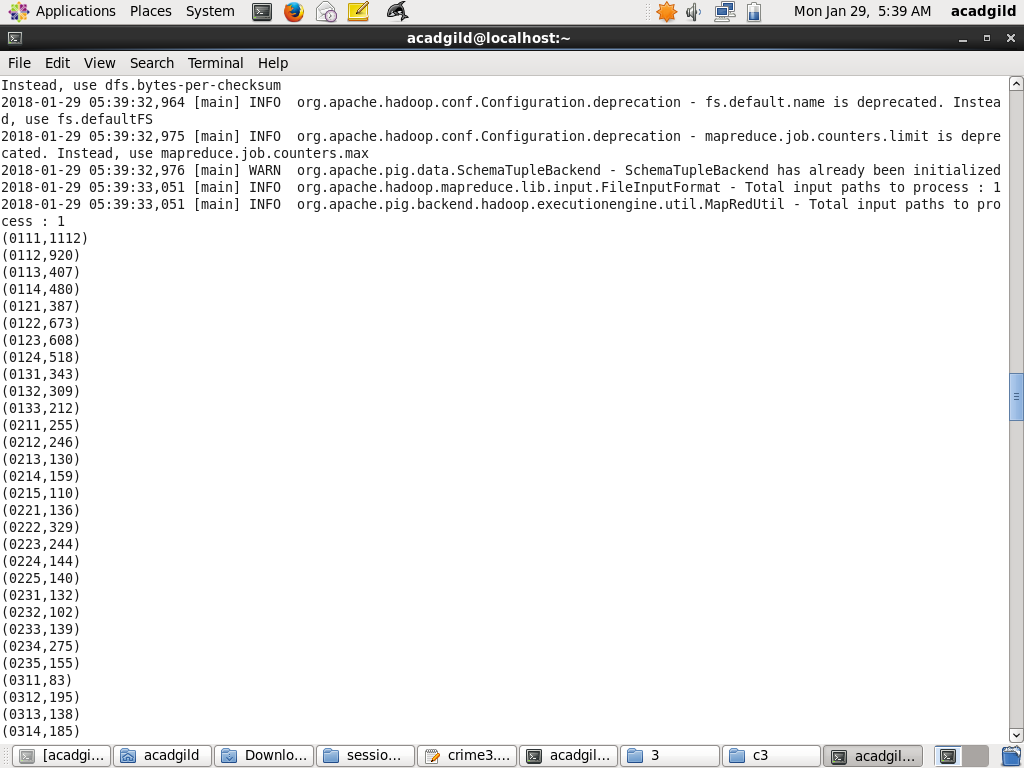
**Execution of Command:**

**Pig -x local /home/acadgild/session12/crime3.pig**

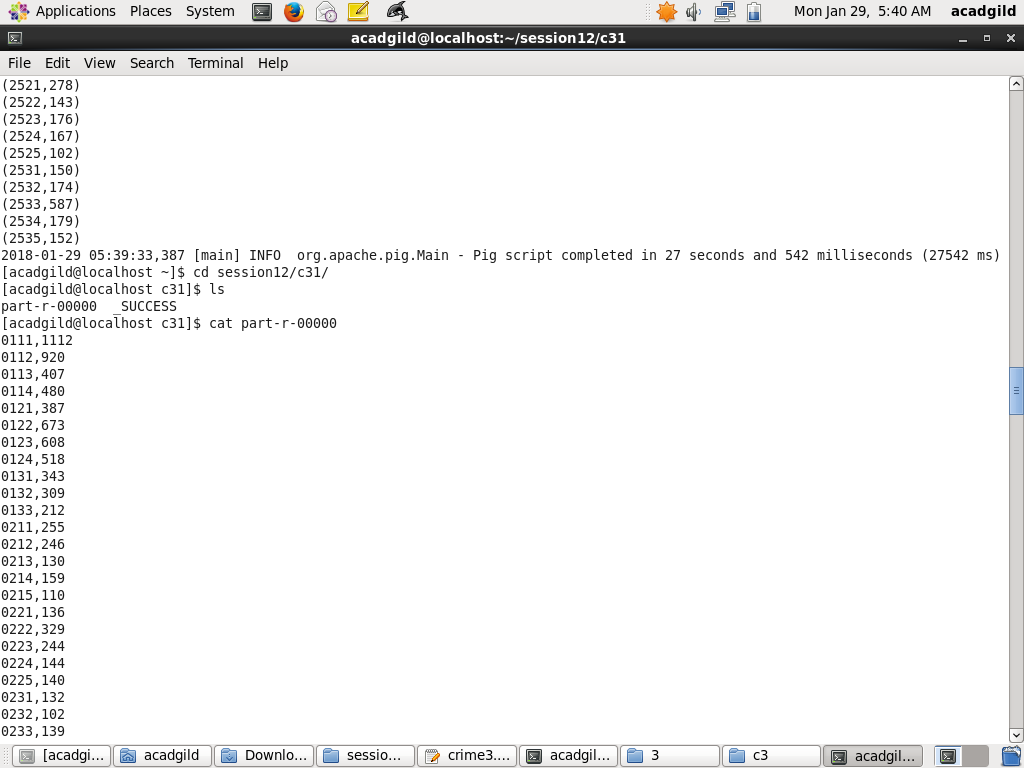
**Code Workflow:**





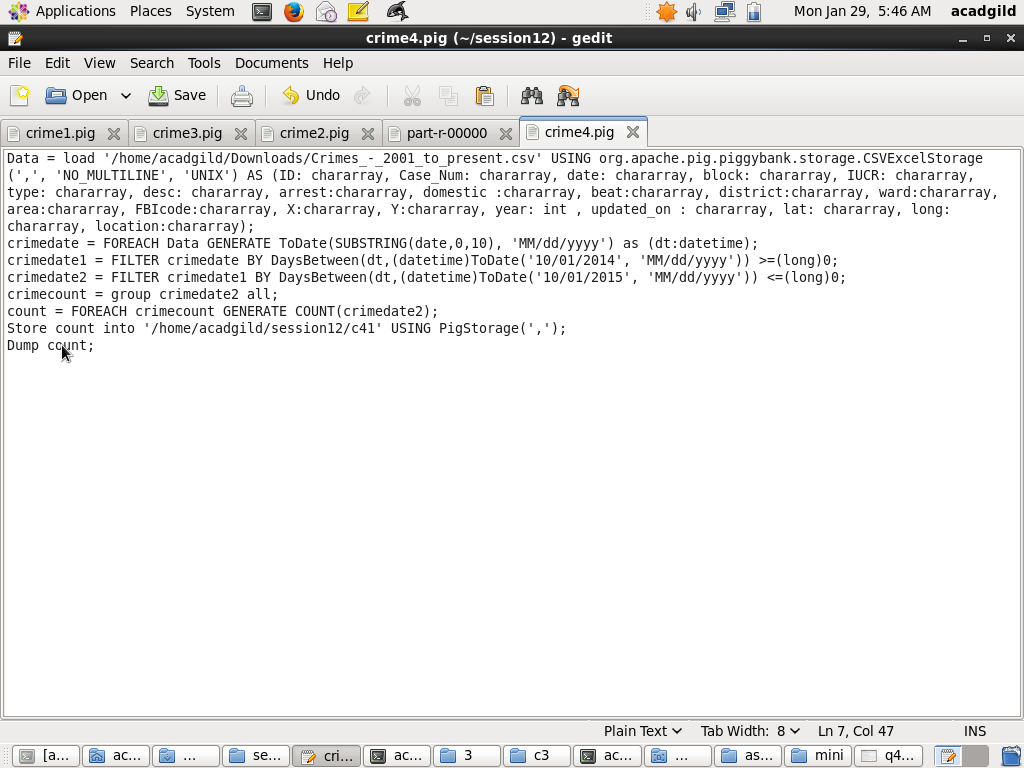


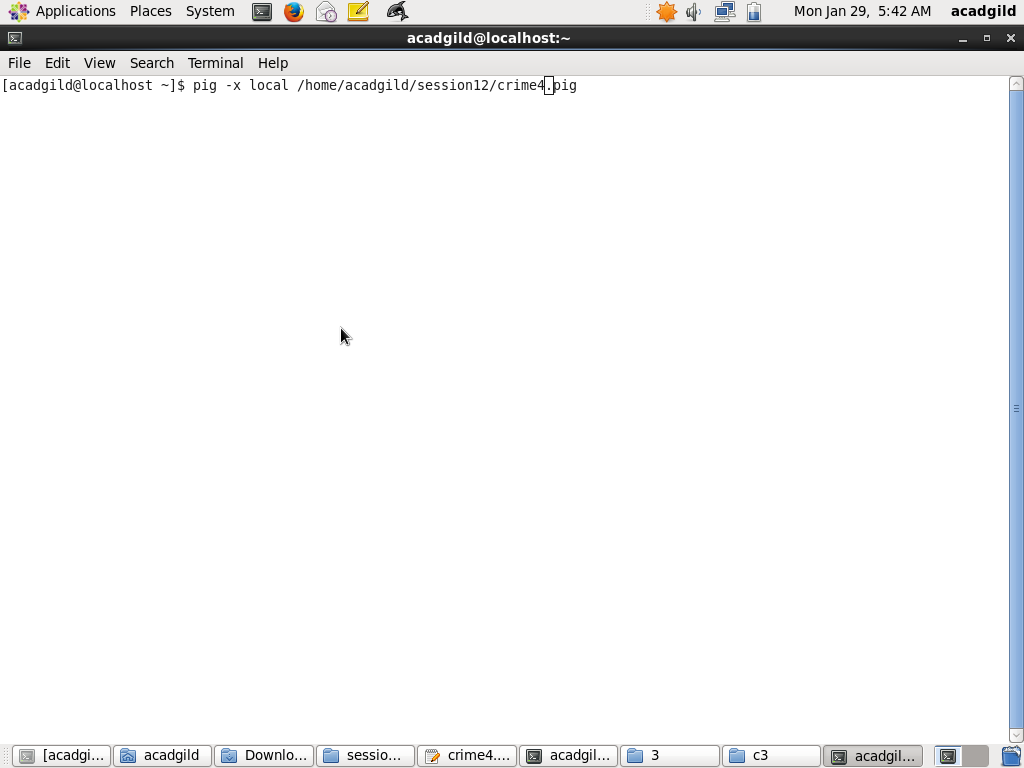
**Store command: Results are stored in the file it can be viewed using cat command**

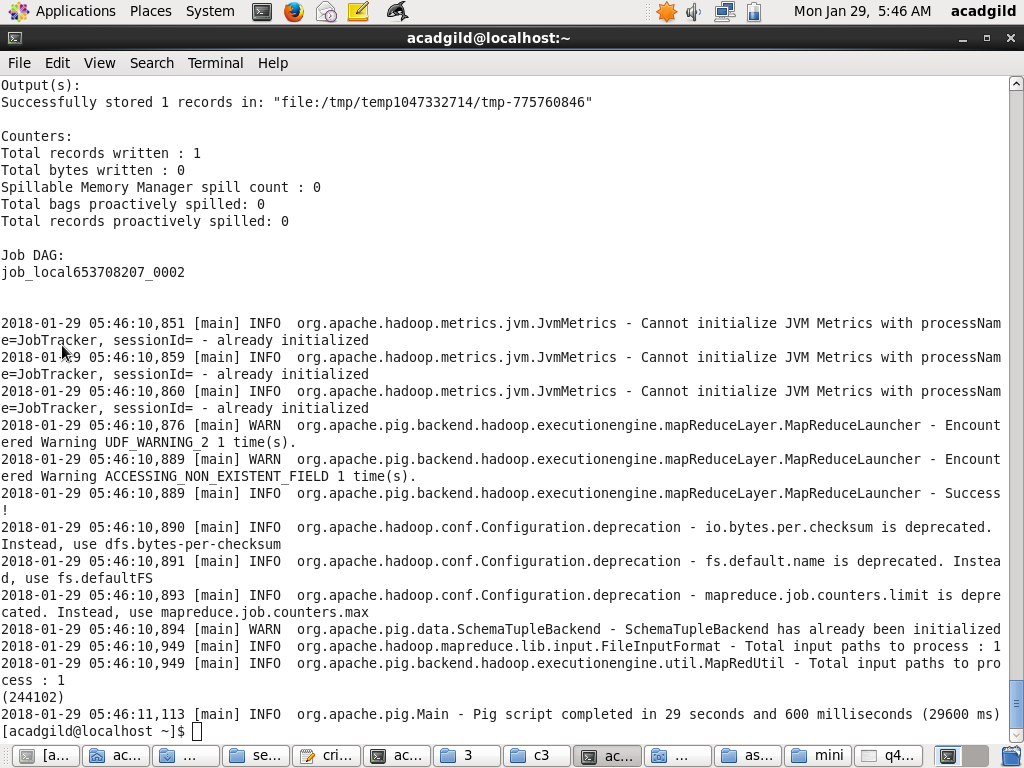


1. Write a MapReduce/Pig program to calculate the number of arrests done between October

2014 and October 2015.





 **Store command: Results are stored in the file it can be viewed using cat command**

