USA Crime Analysis

Step1:

Load into PIG schema with alias crime\_details with fields ID, CaseNumber, Date, Block, IUCR, PrimaryType, Description, LocationDescription, Arrest, Domestic, Beat, District, Ward, CommunityArea, FBICode, extra, XCordinate, YCoordinate, Year, UpdatedOn, Latitude, Longitude, Location)

The PIG script is as below:

crime\_details = LOAD '/home/acadgild/project1/Crimes\_-\_2001\_to\_present.csv'

USING PigStorage(',') AS (ID:long, CaseNumber:chararray, Date:chararray,

Block:chararray, IUCR:chararray, PrimaryType:chararray, Description:chararray,

LocationDescription:chararray,Arrest:chararray, Domestic:chararray,

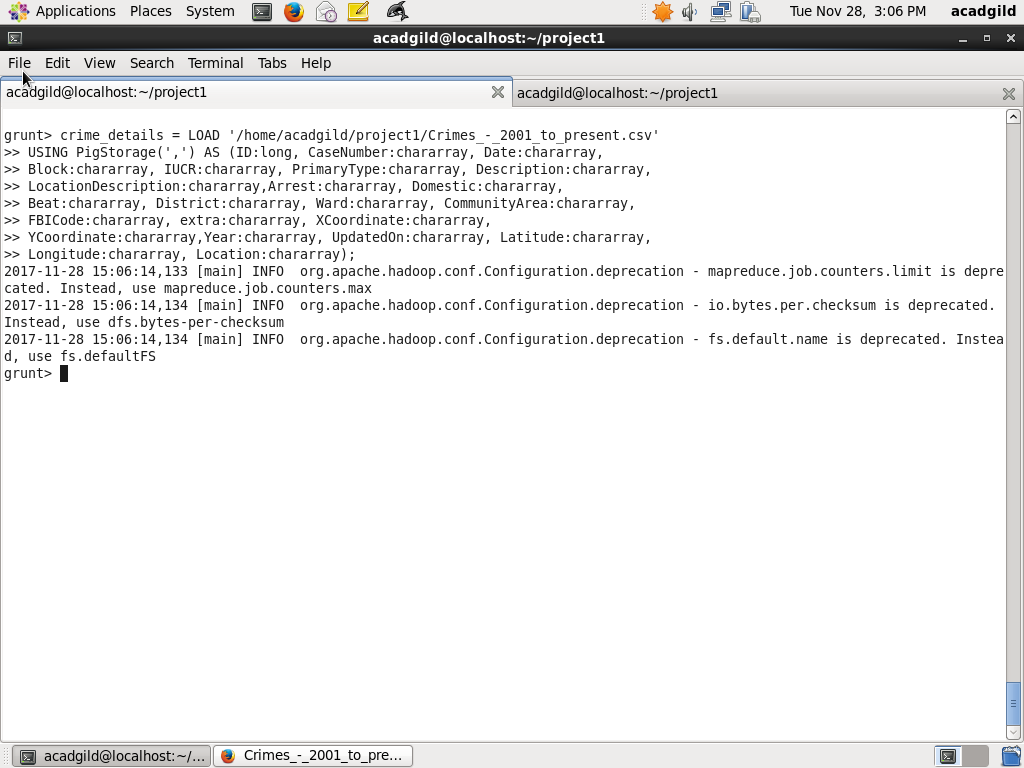
Beat:chararray, District:chararray, Ward:chararray, CommunityArea:chararray,

FBICode:chararray, extra:chararray, XCoordinate:chararray,

YCoordinate:chararray,Year:chararray, UpdatedOn:chararray, Latitude:chararray,

Longitude:chararray, Location:chararray);

Screenshot is as below:



Task1: Write a MapReduce/Pig program to calculate the number of cases investigated under each

FBI code

1. Group crime\_details by FBICode

crime\_details\_group\_by\_fbi\_code = GROUP crime\_details BY FBICode;

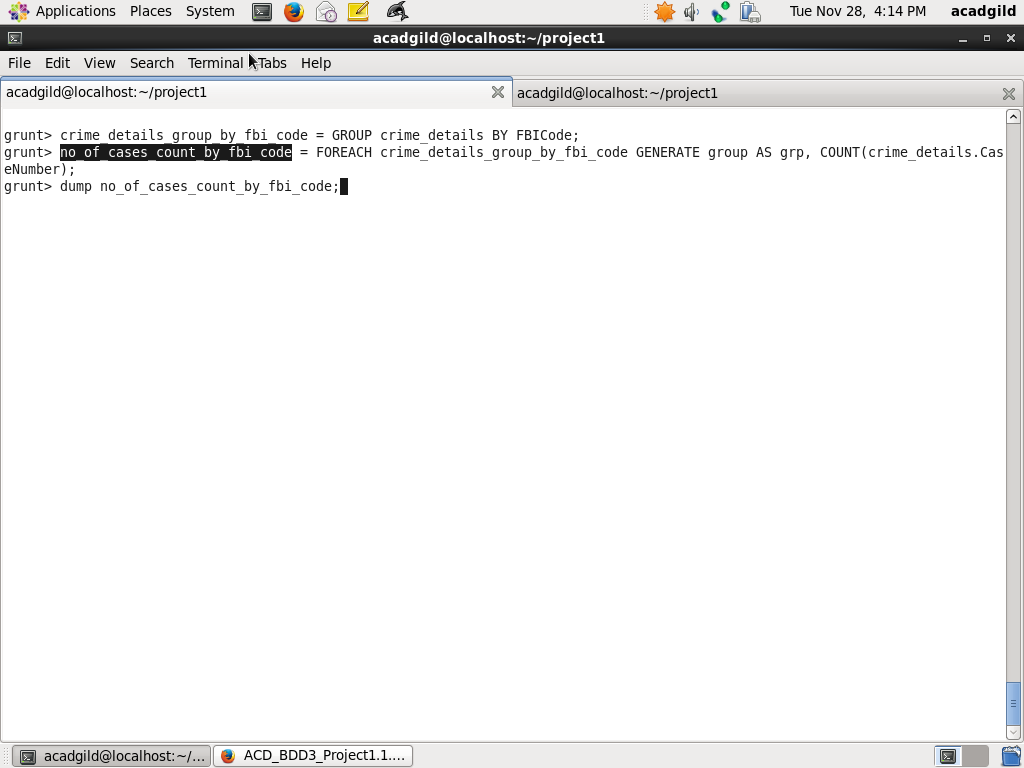
1. For each group count number of cases registered using CaseNumber and put to alias no\_of\_cases\_count\_by\_fbi\_code

no\_of\_cases\_count\_by\_fbi\_code = FOREACH crime\_details\_group\_by\_fbi\_code GENERATE group AS grp, COUNT(crime\_details.CaseNumber);

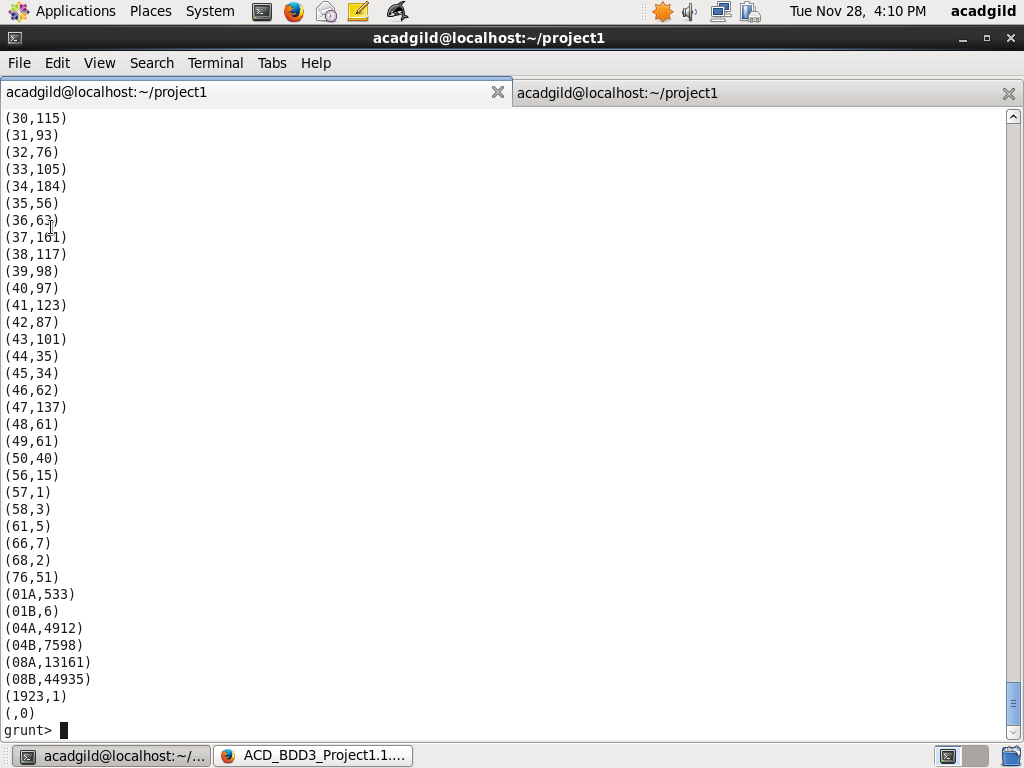
1. Dump the alias no\_of\_cases\_count\_by\_fbi\_code

dump no\_of\_cases\_count\_by\_fbi\_code;

Screenshot is as below:



Output is given in screenshot below:



Task2. Write a MapReduce/Pig program to calculate the number of cases investigated under FBI

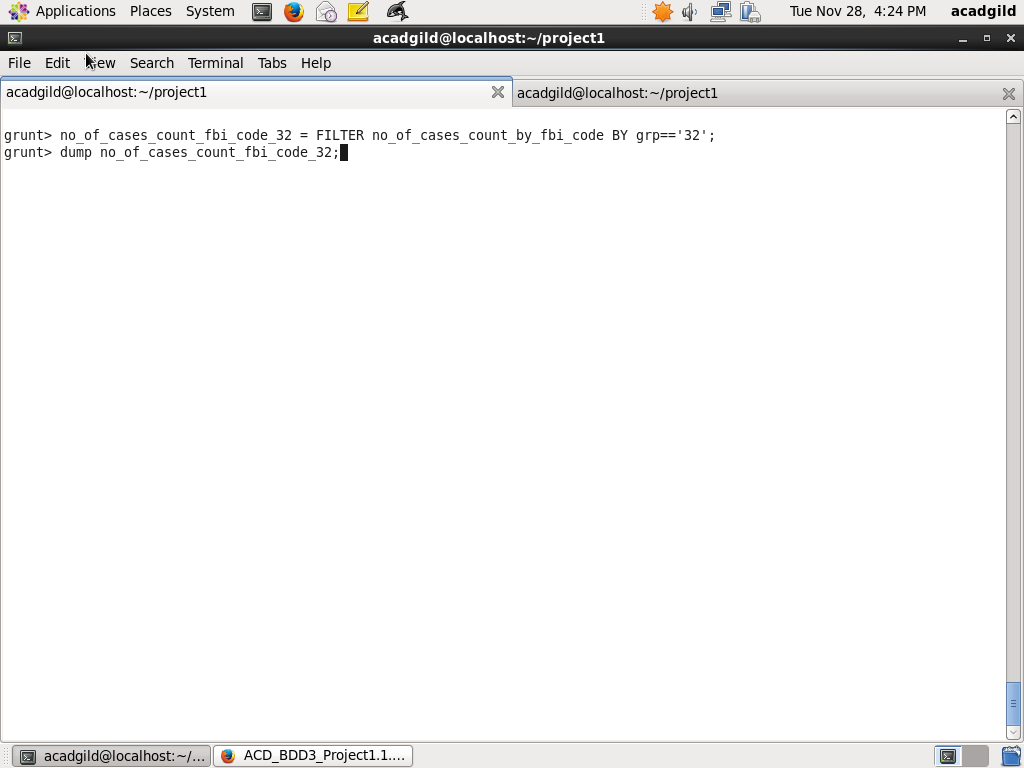
code 32.

Filter the cases count by FBI Code 32 and put to alias no\_of\_cases\_count\_fbi\_code\_32

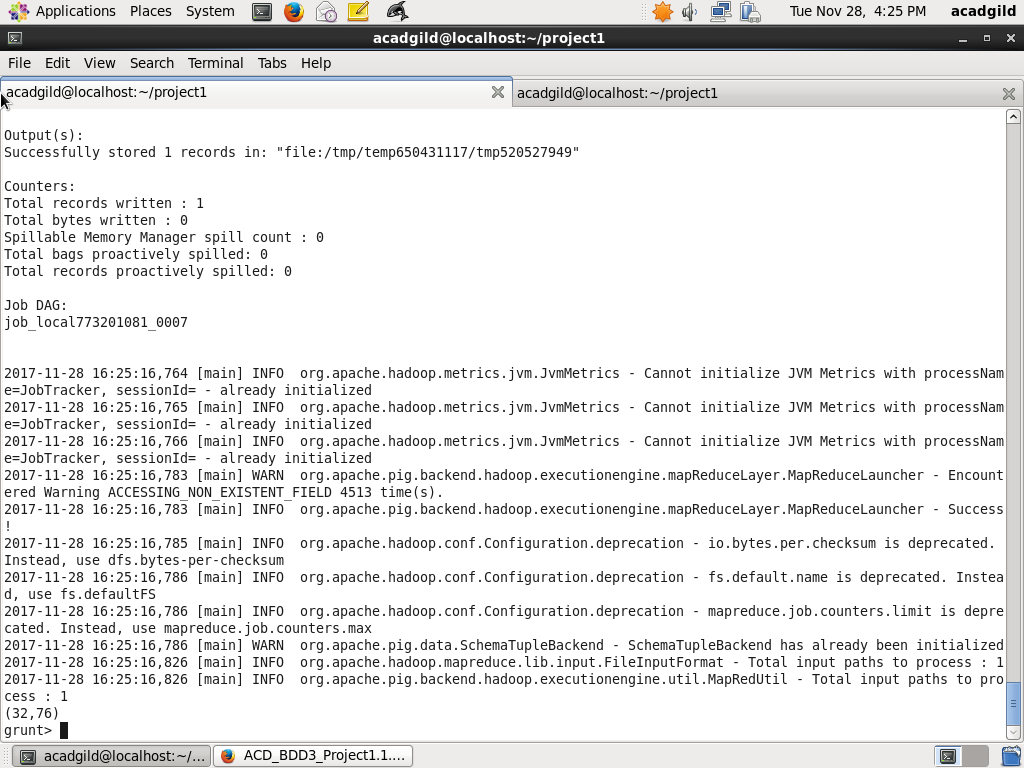
no\_of\_cases\_count\_fbi\_code\_32 = FILTER no\_of\_cases\_count\_by\_fbi\_code BY grp=='32';

dump no\_of\_cases\_count\_fbi\_code\_32;

Screenshot is as below:



Output is as below:



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

i. Filter crime\_details for which PrimaryType is THEFT and ARREST is true and put to alias crime\_details\_theft\_arrested. PIG script is as below:

crime\_details\_theft\_arrested = FILTER crime\_details BY PrimaryType == 'THEFT' AND Arrest == 'true';

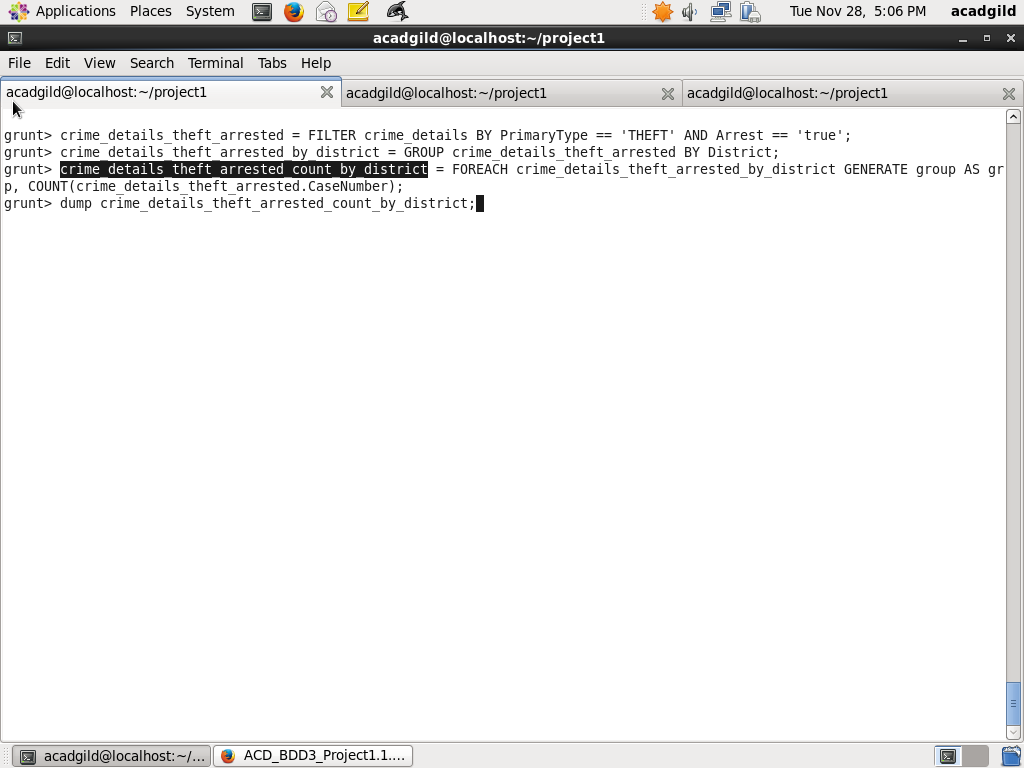
ii. Group crime\_details\_theft\_arrested by District and put to alias crime\_details\_theft\_arrested\_by\_district

crime\_details\_theft\_arrested\_by\_district = GROUP crime\_details\_theft\_arrested BY District;

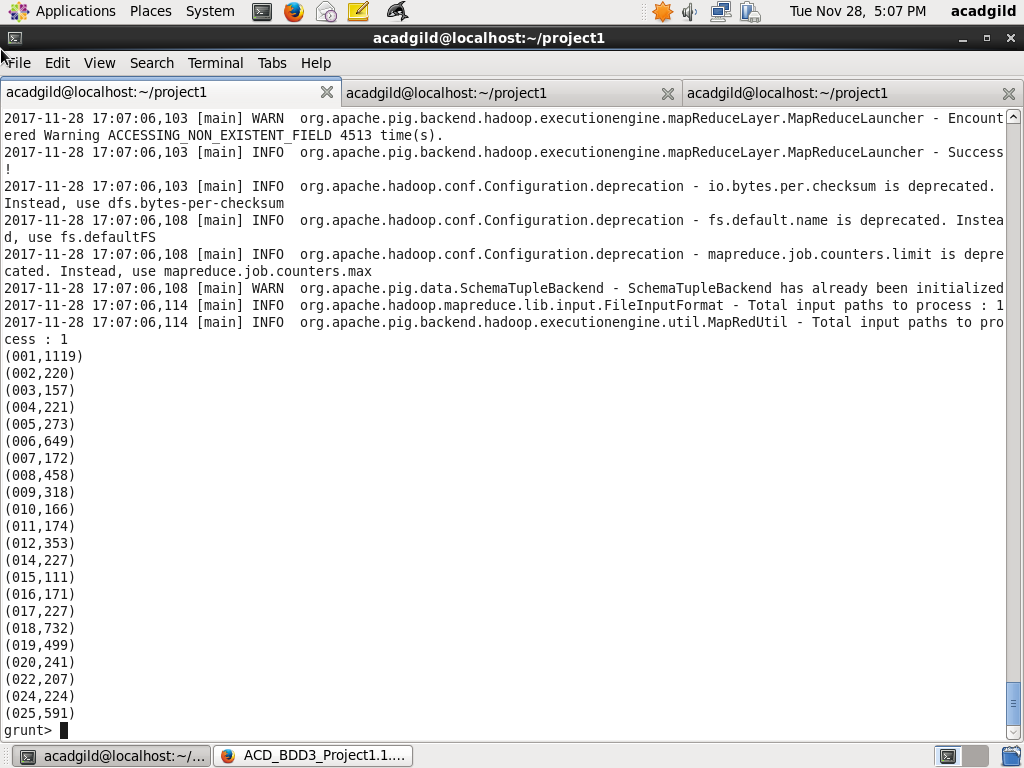
iii. For each group in crime\_details\_theft\_arrested\_count\_by\_district, determine number of cases

crime\_details\_theft\_arrested\_count\_by\_district = FOREACH crime\_details\_theft\_arrested\_by\_district GENERATE group AS grp, COUNT(crime\_details\_theft\_arrested.CaseNumber);

Screenshot is as below:



Output is as below:



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

2014 and October 2015.

1. Find list of arrested by filtering Arrest for true and put to alias crime\_details\_arrested

crime\_details\_arrested = FILTER crime\_details BY Arrest == 'true';

1. Find list of arrests between October 1, 2014, 12:00:00 AM and October 31, 2015, 11:59:59PM and put to alias crime\_details\_arrested\_between\_dates

crime\_details\_arrested\_between\_dates = FILTER crime\_details\_arrested BY ToDate(Date, 'MM/dd/yyyy hh:mm:ss a') >= ToDate('10/01/2014 12:00:00 AM', 'MM/dd/yyyy hh:mm:ss a') AND ToDate(Date, 'MM/dd/yyyy hh:mm:ss a') <= ToDate('10/31/2015 11:59:59 PM', 'MM/dd/yyyy hh:mm:ss a');

1. Group crime\_details\_arrested\_between\_dates by ALL

crime\_details\_arrested\_between\_dates\_all = GROUP crime\_details\_arrested\_between\_dates ALL;

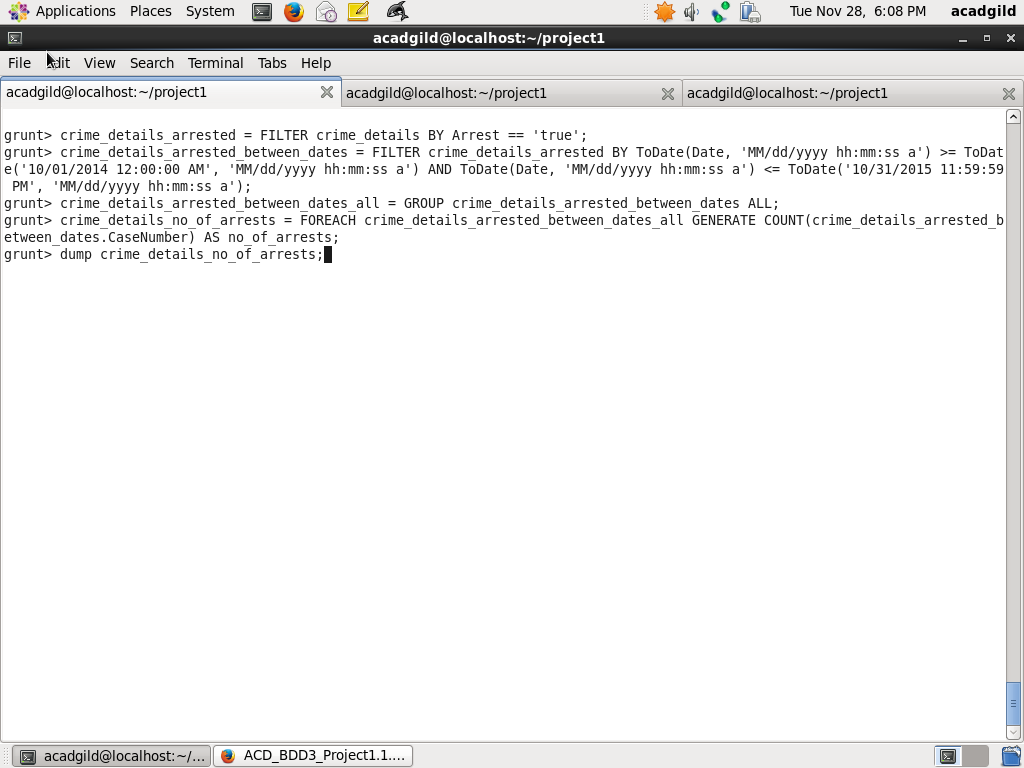
1. Find number of arrests using COUNT function on CaseNumber field and put to alias crime\_details\_no\_of\_arrests

crime\_details\_no\_of\_arrests = FOREACH crime\_details\_arrested\_between\_dates\_all GENERATE COUNT(crime\_details\_arrested\_between\_dates.CaseNumber) AS no\_of\_arrests;

1. Dump crime\_details\_no\_of\_arrests

dump crime\_details\_no\_of\_arrests;

Screenshot is as below:



Output is as below:

