Project 1.1 USA Crime Analysis

Load data:

Copy data from Local to HDFS, and Load data to Pig

crime\_load= load 'hdfs:///miniproject/usacrimeanalysis/Crimes.csv' using PigStorage(',') as (

ID:int,

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,

XCoordinate:chararray,

YCoordinate:chararray,

Year:chararray,

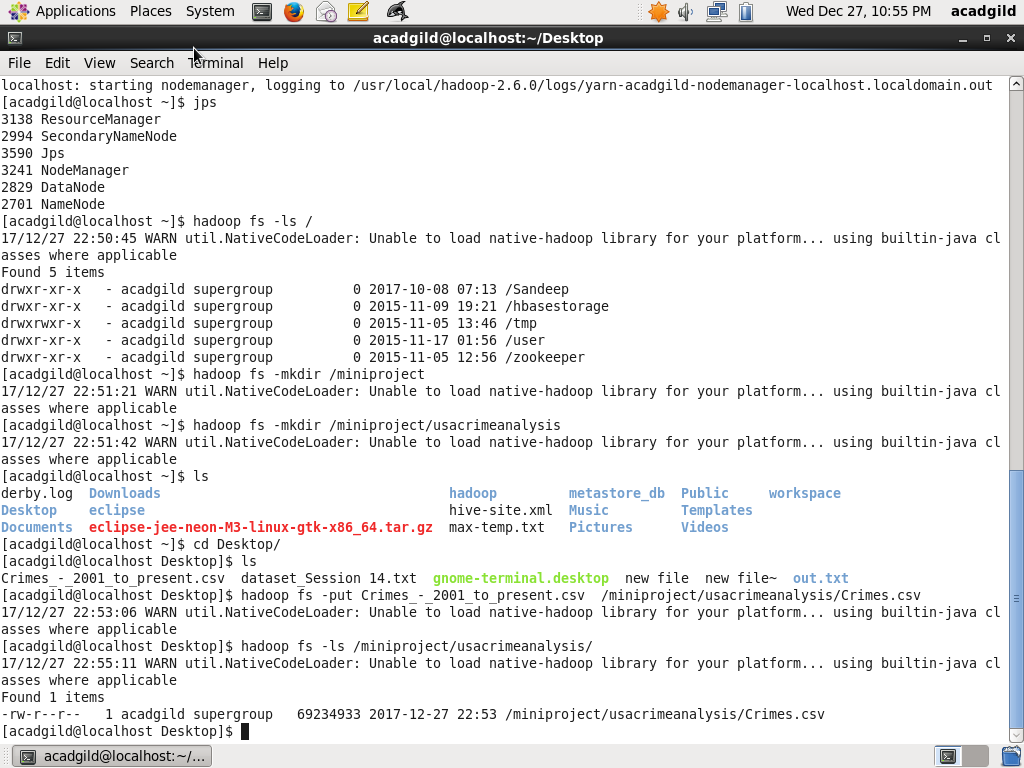
UpdatedOn:chararray,

Latitude:chararray,

Longitude:chararray,

Location:tuple(la:chararray,lo:chararray))

describe crime\_load;



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

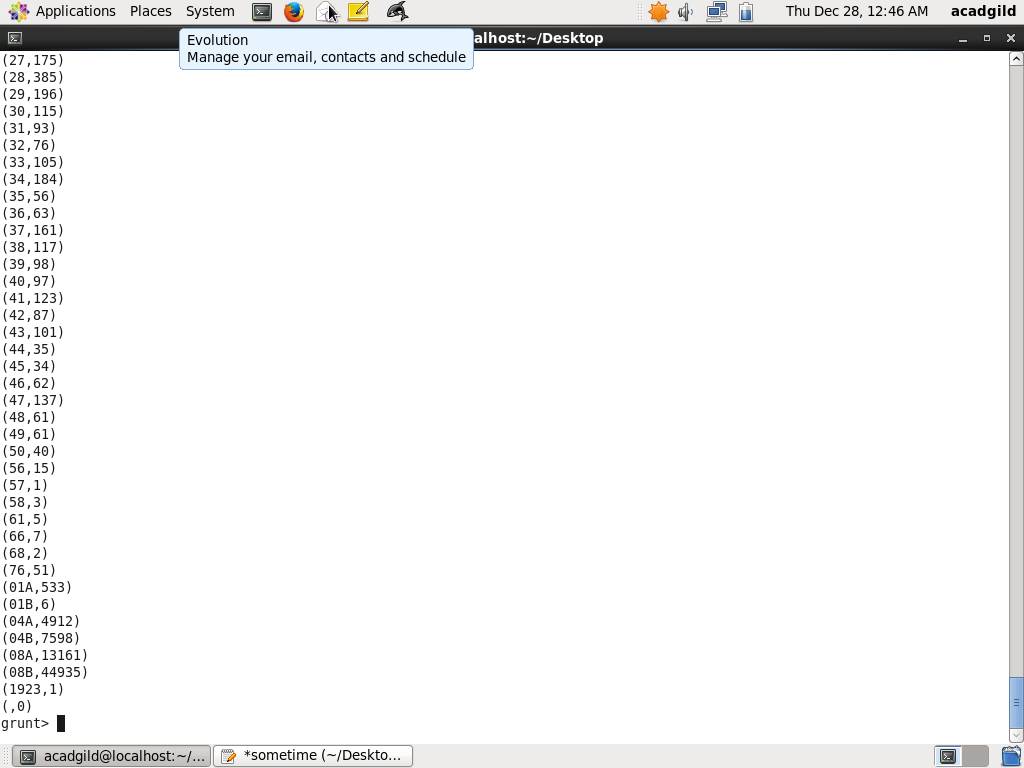
**Below script groups by FBI code and counts number of cases investigated under each FBI Code**

group\_fbicode= GROUP crime\_load BY FBICode;

cases= FOREACH group\_fbicode GENERATE group, COUNT(crime\_load.FBICode);

dump cases;

**Output:**



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

**code 32.**

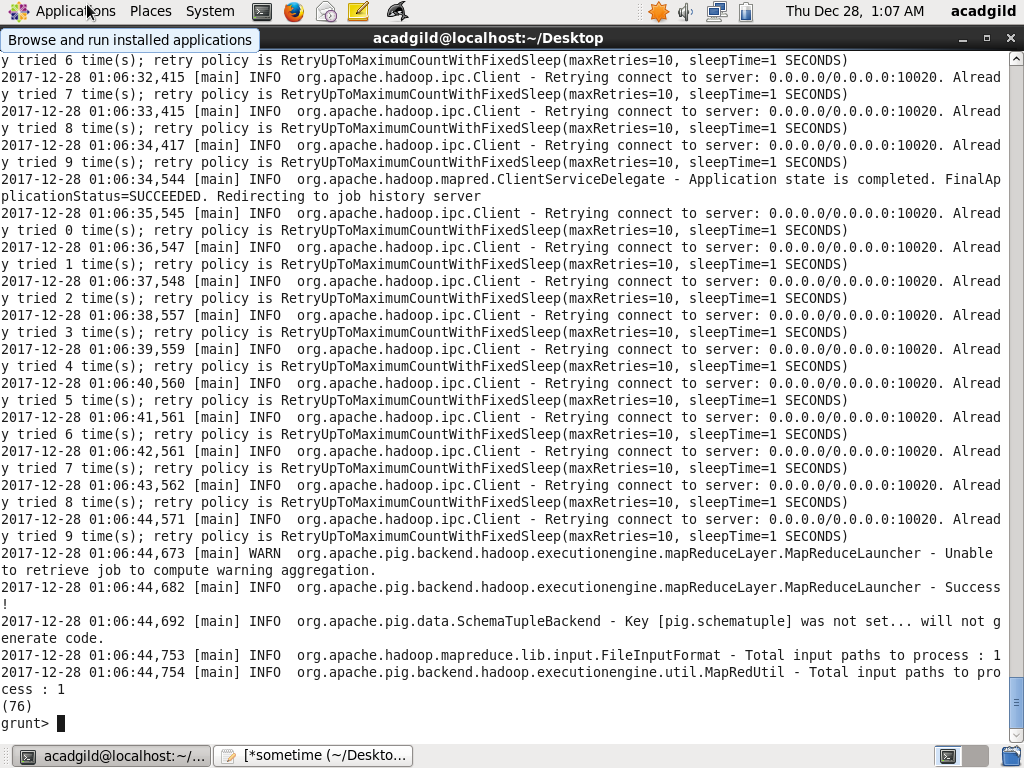
**Below script filters FBI code 32 and counts number of cases investigated**

filter\_32= FILTER crime\_load BY FBICode=='32';

group\_32= GROUP filter\_32 ALL;

cases\_32= FOREACH group\_32 GENERATE COUNT(filter\_32.FBICode);

**OUTPUT:**

****

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

**Below script filters crime by theft and arrest made, gets the number by district wise**

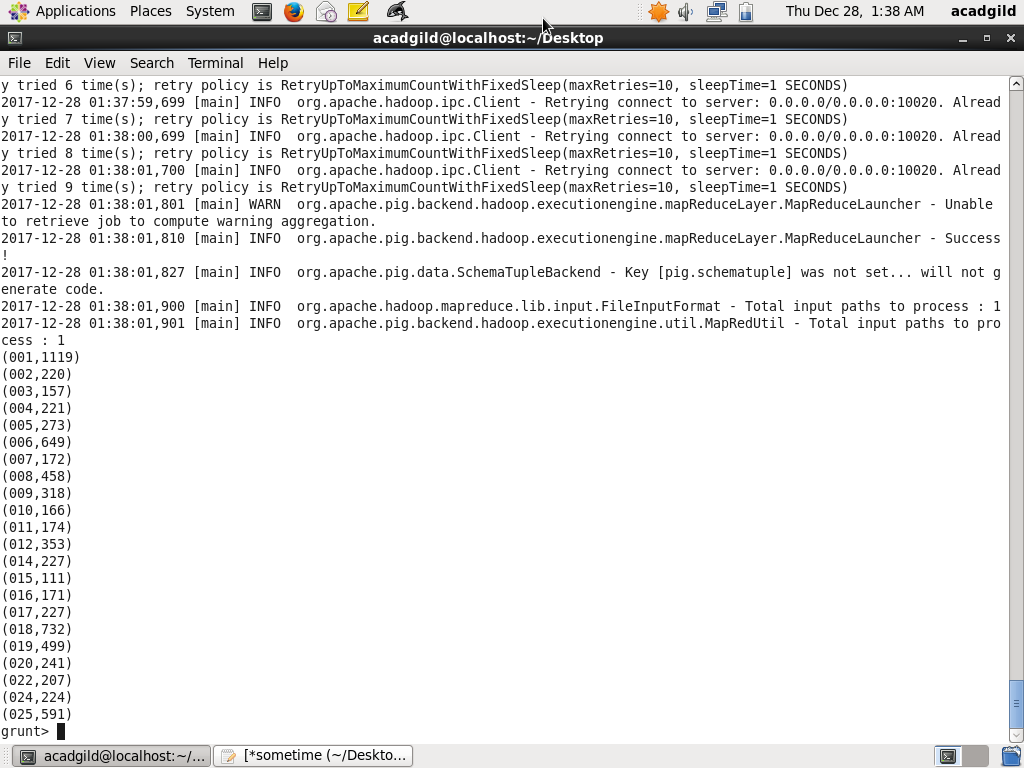
filter\_arrest\_by\_theft= FILTER crime\_load BY PrimaryType=='THEFT' AND Arrest=='true';

group\_arrest\_by\_theft = GROUP filter\_arrest\_by\_theft BY District;

cases\_arrest\_by\_theft = FOREACH group\_arrest\_by\_theft GENERATE group,COUNT(filter\_arrest\_by\_theft.District);

dump cases\_arrest\_by\_theft;

**OUTPUT:**

****

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

**Below script filters crime by date between October 2014 and 2015 and arrest made, gets the number**

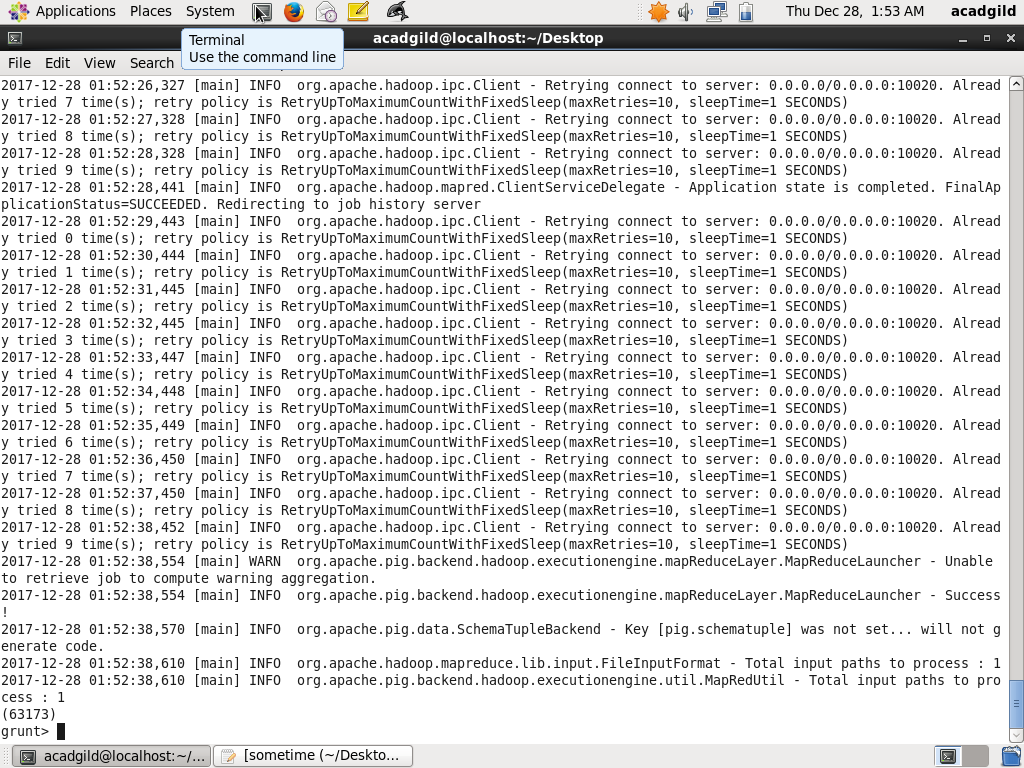
filter\_arrest\_by\_date = FILTER crime\_load BY ToDate(Date,'MM/dd/yyyy hh:mm:ss a') >= ToDate('10/01/2014 12:00:00 AM', 'MM/dd/yyyy hh:mm:ss a') AND ToDate(Date,'MM/dd/yyyy hh:mm:ss a') <= ToDate('10/31/2015 11:59:59 PM', 'MM/dd/yyyy hh:mm:ss a') AND Arrest=='true';

group\_arrest\_by\_date = GROUP filter\_arrest\_by\_date ALL;

count\_arrest\_by\_date = FOREACH group\_arrest\_by\_date GENERATE COUNT(filter\_arrest\_by\_date.Arrest)

dump count\_arrest\_by\_date;

**OUTPUT:**

****