This 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.

2. Objective

3. Prerequisites

You should have Hadoop cluster installed in your system.

4. Associated Data Files

https://drive.google.com/file/d/0B1QaXx7tpw3SaUJHOHBZclBXWG8/view?usp=sharing

Dataset Description:

ID,Case Number,Date,Block,IUCR,Primary Type,Description,Location

Description,Arrest,Domestic,Beat,District,Ward,Community Area,FBICode,X Coordinate,Y

Coordinate,Year,Updated On,Latitude,Longitude,Location

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

***REGISTER '/home/acadgild/project1/piggybank.jar';***

***-- register the piggybank jar file to import the csv files***

***A = load '/home/acadgild/project1/Crimes.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');***

***-- load csv file.***

***B = foreach A generate (long)$0 as id,(chararray)$14 as fbicode;***

***-- pickup the necessary coloumns with the type casting. $0 is id and $14 is fbi code***

***D = group B by fbicode;***

***-- grouping the releation D based on fbicode***

***E = foreach D generate group, COUNT(B.id);***

***-- finding the count for cases for each fbicode***

***F= order E by $1 DESC;***

***-- sorting the records based on cases solved by fbicode***

***store F into '/home/acadgild/project1/output1/';***

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

***REGISTER '/home/acadgild/project1/piggybank.jar';***

***-- register the piggybank jar file to import the csv files***

***A = load '/home/acadgild/project1/Crimes.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');***

***-- load csv file.***

***B = foreach A generate (long)$0 as id,(chararray)$14 as fbicode;***

***-- pickup the necessary coloumns with the type casting. $0 is id and $14 is fbi code***

***D = FILTER B BY fbicode == '32';***

***store D into '/home/acadgild/project1/output2/';***

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

***REGISTER '/home/acadgild/project1/piggybank.jar';***

***-- register the piggybank jar file to import the csv files***

***A = load '/home/acadgild/project1/Crimes.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');***

***-- load csv file.***

***B = foreach A generate (chararray)$5 as primarytype,(boolean)$8 as arrest, (int)$11 as dist;***

***-- pickup the necessary coloumns with the type casting. $5 primary type. type of crime. $8 arrested or not. $11 district***

***-- dump B;***

***C = group B by dist;***

***describe C;***

***D = FOREACH C {***

***S = FILTER B BY primarytype == 'THEFT' AND arrest == true;***

***GENERATE COUNT (S.$2), $0;***

***}***

***store D into '/home/acadgild/project1/output3/';***

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

2014 and October 2015.

***REGISTER '/home/acadgild/project1/piggybank.jar';***

***loadData = LOAD '/home/acadgild/project1/Crimes.csv' USING***

***org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX');***

***selectCols = FOREACH loadData GENERATE (chararray)$2 as DtTime,(chararray)$8 as Arrest;***

***filterSelectCols = FILTER selectCols BY (DtTime is not null) AND (Arrest == 'true');***

***dateSubstring = FOREACH filterSelectCols GENERATE***

***ToDate(SUBSTRING(DtTime,0,19),'MM/dd/yyyy hh:mm:ss') as Dt,Arrest;***

***dateWiseArrest = FOREACH dateSubstring GENERATE GetMonth(Dt) as Month,GetYear(Dt) as Year,Arrest;***

***totalArrest = FILTER dateWiseArrest BY (Month>9 AND Year == 2014) OR (Month<11 and Year == 2015);***

***groupTotalArrest = GROUP totalArrest ALL;***

***countTotalArrest = FOREACH groupTotalArrest GENERATE COUNT(totalArrest.Arrest) as TotalArrests;***

***store countTotalArrest into '/home/acadgild/project1/output4/';***