Project1:

Introduction

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

5. Problem Statement

**crime\_table1 = LOAD '/home/acadgild/azhar/ProjectUSACRIME/Crimes\_-\_2001\_to\_present.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER') AS(id:int,casenumber:chararray,Date:chararray,block:chararray,IUCR:chararray,primarytype:chararray,description:chararray,locationdesc:chararray,arrest:chararray,domestic:chararray,beat:chararray,district:chararray,ward:chararray,community:chararray,fbicode:chararray,xcoord:chararray,ycoord:chararray,year:chararray,updatedon:chararray,latitude:chararray,lagitude:chararray,location:chararray);**

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

FBI code

**Code:**

**grp\_crime = group crime\_table1 by fbicode;**

**total\_case\_fbi = foreach grp\_crime generate group,COUNT(crime\_table1.casenumber);**

**store total\_case\_fbi into '/home/acadgild/azhar/ProjectUSACRIME/TOTALCASEFBICODE';**

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

code 32.

**Code:**

**filter26= filter crime\_table1 by fbicode == '26';**

**grp\_filter26 = group filter26 by fbicode;**

**total\_case\_fbi26 = foreach grp\_filter26 generate group,COUNT(filter26.casenumber);**

**store total\_case\_fbi26 into '/home/acadgild/azhar/ProjectUSACRIME/TOTALCASEFBICODE26'**

**Note: FBICODE 36 is not available. That is why I used code 26.**

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

**Code:**

**theft\_case = filter crime\_table1 by primarytype == 'THEFT';**

**grp\_district = group theft\_case by district;**

**total\_theft\_district = foreach grp\_district generate group, COUNT(theft\_case.primarytype);**

**store total\_theft\_district into '/home/acadgild/azhar/ProjectUSACRIME/totaltheftdistrict';**

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

2014 and October 2015.

**Code:**

**type\_cast = foreach crime\_table1 generate ToDate(SUBSTRING(Date,0,10),'MM/dd/yyyy') as (dt:datetime)**

**arrest\_done1 = filter type\_cast by DaysBetween(dt,ToDate('10/1/2014','MM/dd/yyyy')) >= (long)0;**

**arrest\_done2 = filter arrest\_done1 by DaysBetween(dt,ToDate('10/1/2015','MM/dd/yyyy')) <= (long)0;**

**grp\_arrest\_done = group arrest\_done2 all;**

**count\_all = foreach grp\_arrest\_done generate COUNT(arrest\_done2);**

**store count\_all into '/home/acadgild/azhar/ProjectUSACRIME/arrestedcase';**