**USA CRIME ANALYSIS USE CASE 4**

Below is the Dataset URL:

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

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.

**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

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

2014 and October 2015.

**REGISTER '/home/acadgild/pig/ProjectFiles/piggybank.jar';**

**US\_crime\_data = load '/home/acadgild/pig/ProjectFiles/Crimes\_-\_2001\_to\_present.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');**

**ARREST\_DATA = FOREACH US\_crime\_data GENERATE ToDate(SUBSTRING($2,0,10),'MM/dd/yyyy') as ArrestDate,$8 as Arrest;**

**FILTER\_ARREST\_DATA = FILTER ARREST\_DATA by (Arrest=='true') AND (ArrestDate>=ToDate('2014-10-01','yyyy-MM-dd') AND ArrestDate<=ToDate('2015-10-31','yyyy-MM-dd'));**

**GROUPBY\_ARREST\_DATA = group FILTER\_ARREST\_DATA by Arrest;**

**Result = FOREACH GROUPBY\_ARREST\_DATA GENERATE COUNT(FILTER\_ARREST\_DATA.Arrest);**

**Dump Result;**

**Description of code:**

**Line 1: Register piggbank.jar file**

**Line 2: Create relation US\_crime\_data using CSVExcelStorage from input file.**

**Line 3: Extract mandatory columns (ArrestDate, Arrest) from relation US\_crime\_data and store it into another relation ARREST\_DATA.**

**Line 4: FILTER data from relation ARREST\_DATA using the conditions Arrest=='true', date between October 2014 to October 2015 and store it into another relation FILTER\_ARREST\_DATA.**

**Line 5: Group the data of relation FILTER\_ARREST\_DATA on the basis of Arrest field and store it into relation GROUPBY\_ARREST\_DATA.**

**Line 6: Now for each GROUPBY\_ARREST\_DATA generate group and count the Arrest cases.**

**Line 7: Dump the result or store the result and check the output.**



