**USA CRIME ANALYSIS USE CASE 2**

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

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

code 32.

**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');**

**FBICODE\_CASE\_DATA = FOREACH US\_crime\_data GENERATE $14 as FBICODE,$1 as CASENUMBER;**

**FILTER\_FBICODE = FILTER FBICODE\_CASE\_DATA by FBICODE=='32';**

**GROUPBY\_FBICODE = group FILTER\_FBICODE by FBICODE;**

**Result = FOREACH GROUPBY\_FBICODE GENERATE group,COUNT(FILTER\_FBICODE.CASENUMBER);**

**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 (FBICODE, CASENUMBER) from relation US\_crime\_data and store it into another relation FBICODE\_CASE\_DATA.**

**Line 4: FILTER the data from relation FBICODE\_CASE\_DATA where FBICODE=='32' and store into another relation FILTER\_FBICODE.**

**Line 5: Now group the data from relation FILTER\_FBICODE on the basis of FBICODE and store into another relation GROUPBY\_FBICODE.**

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

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

