**USA CRIME ANALYSIS USE CASE 3**

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

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

**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');**

**DISTRICT\_DATA = FOREACH US\_crime\_data GENERATE $5 as Description,$12 as DISTRICT;**

**FILTER\_THEFT\_DATA = FILTER DISTRICT\_DATA by Description=='THEFT';**

**GROUPBY\_DISTRICT\_THEFT\_DATA = group FILTER\_THEFT\_DATA by DISTRICT;**

**Result = FOREACH GROUPBY\_DISTRICT\_THEFT\_DATA GENERATE group,COUNT(FILTER\_THEFT\_DATA.Description);**

**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 (Description, DISTRICT) from relation US\_crime\_data and store it into another relation DISTRICT\_DATA.**

**Line 4: FILTER the data from relation DISTRICT\_DATA on the basis of condition Description=='THEFT' and store it into another relation FILTER\_THEFT\_DATA.**

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

**Line 6: Now for each GROUPBY\_DISTRICT\_THEFT\_DATA generate group and count the number of arrests for theft.**

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



