**USA Crime Analysis**

Loading data:

REGISTER /usr/local/pig/lib/piggybank.jar

**crimesData** = LOAD ‘crimes.csv’ USING org.apache.pig.piggybank.storage.CSVExcelStorage(‘,’,’NO\_MULTILINE’,’UNIX’,’SKIP\_INPUT\_HEADER’);

**1)**

grouped= GROUP crimesData by fbicode;

result= FOREACH grouped GENERATE group,COUNT(crimesData);

DUMP result;

**Output:**

(1,172)

(2,1842)

(3,10818)

(4,154)

(5,14932)

(6,63024)

(7,10658)

(8,301)

(9,629)

(10,1708)

(11,13637)

(12,79)

(13,151)

(14,31244)

(15,3780)

(16,1949)

(17,1165)

(18,24989)

(19,590)

(20,1435)

(21,293)

(22,483)

(23,77)

(24,4114)

(25,142)

(26,29009)

(27,175)

(28,385)

(29,196)

(30,115)

(31,93)

(32,76)

(33,105)

(34,184)

(35,56)

(36,63)

(37,161)

(38,117)

(39,98)

(40,97)

(41,123)

(42,87)

(43,101)

(44,35)

(45,34)

(46,62)

(47,137)

(48,61)

(49,61)

(50,40)

(56,15)

(57,1)

(58,3)

(61,5)

(66,7)

(68,2)

(76,51)

(01A,533)

(01B,6)

(04A,4912)

(04B,7598)

(08A,13161)

(08B,44935)

(1923,1)

(,1)

**2)**

filtered= FILTER crimesData BY fbicode=='32';

grouped= GROUP filtered all;

counted= FOREACH grouped GENERATE COUNT(filtered);

DUMP counted;

**Output:**

(76)

**3)**

filtered= FILTER crimesData BY (primary\_type=='THEFT' AND arrest=='true')

grouped= GROUP filtered BY district;

result= FOREACH grouped GENERATE group,COUNT(filtered);

DUMP result;

**Output:**

(1,1119)

(2,220)

(3,157)

(4,221)

(5,273)

(6,649)

(7,172)

(8,458)

(9,318)

(10,166)

(11,174)

(12,353)

(14,227)

(15,111)

(16,171)

(17,227)

(18,732)

(19,499)

(20,241)

(22,207)

(24,224)

(25,591)

**4)**

dateFilter= FILTER crimesData BY (date1>'10/01/2014' AND date1<'10/01/2015') AND arrest=='true';

DUMP dateFilter;

grouped= GROUP dateFilter all;

count\_res= FOREACH grouped GENERATE COUNT(dateFilter);

DUMP count\_res;

**Output:**

(243)