**Project 1.1**

**Project 1.1 - USA Crime Analysis**

**1. 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.

**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

**Problem Statement**

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

FBI code

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

code 32.

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

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

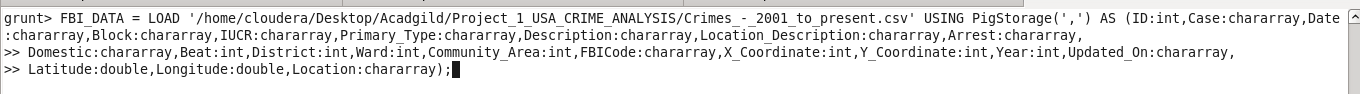
2014 and October 2015

**Problem Statement**

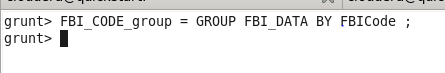
1. Write a MapReduce/Pig program to calculate the number of cases investigated under each

FBI code

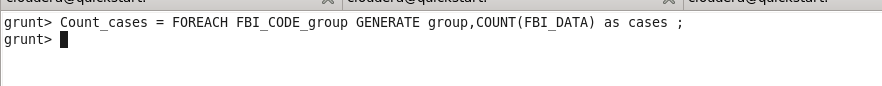
**LOADING THE DATA**



**GROUPING THE DATA ON BASIS OF FBI CODE**

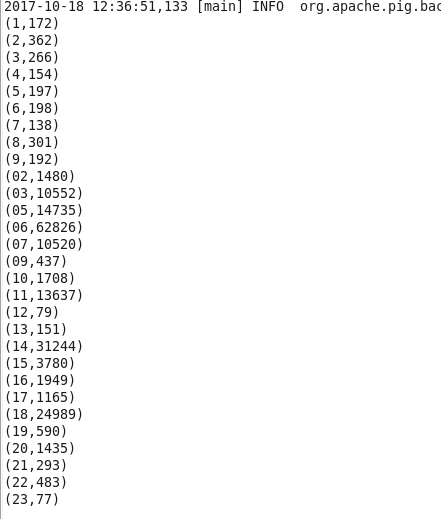


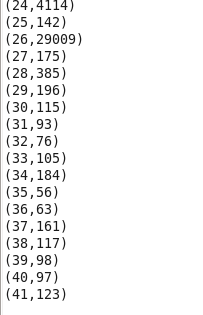
**COUNTING THE CASES ON THE BASIS OF FBI CODE**

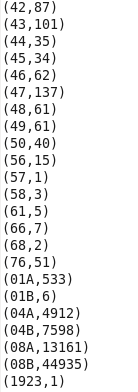


**FINAL RESULT**

**DUMP Count\_cases;**





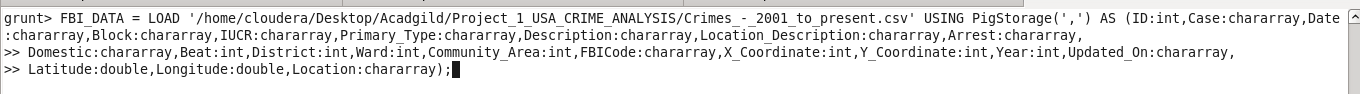


**Problem**

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

**code 32**.

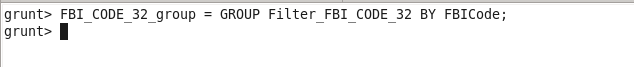
**LOADING THE DATA**



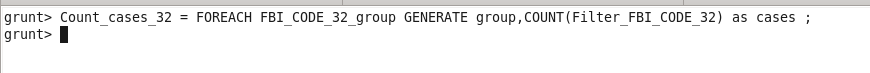
**FILTERING THE DATA ON BASIS OF FBI CODE EQUALS TO 32**



**GROUPING THE DATA ON BASIS OF FBI CODE**

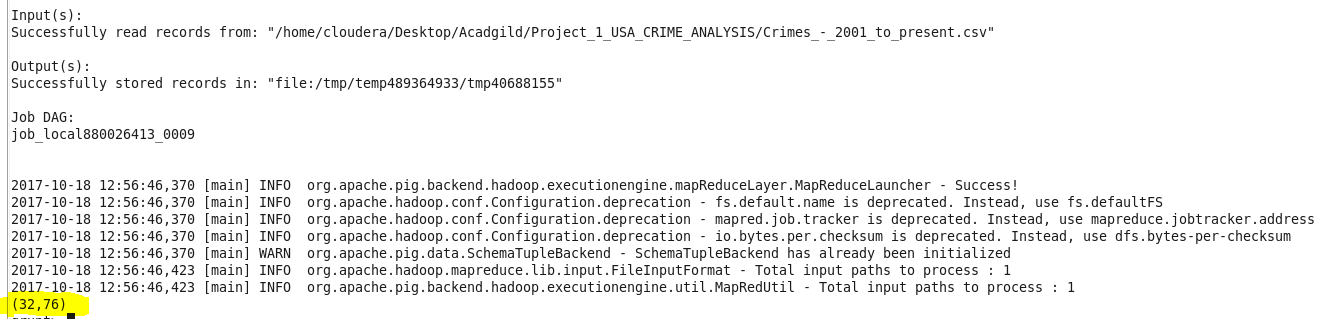


**COUNTING THE CASES ON THE BASIS OF FBI CODE equals to 32**



**FINAL OUTPUT**

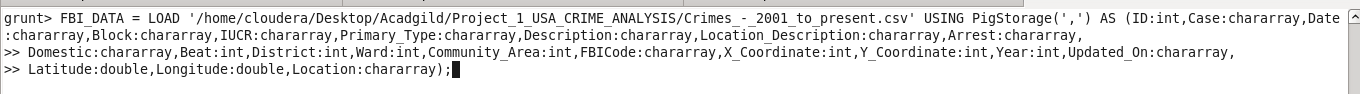




**PROBLEM**

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

**LOADING THE DATA**



**FILTERING THE DATA WITH ARRESTS EQUAL TO TRUE AND TYPE OF CRIME AS THEFT**



**GROUPING THE DATA ON BASIS OF DISTRICT**

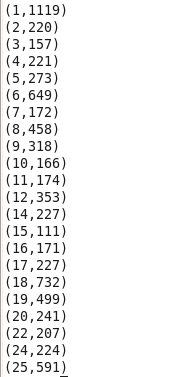


**GENERATING THE NUMBER OF ARRESTS**



**FINAL OUTPUT**





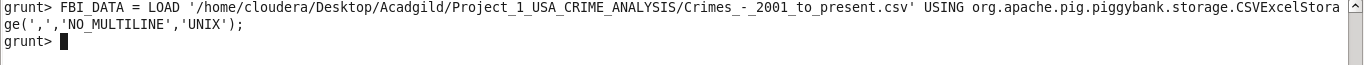
**4. Write a MapReduce/Pig program to calculate the number of arrests done between October**

**2014 and October 2015.**

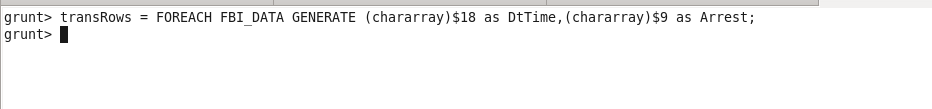
**Registering the Piggbank.jar**



**LOADING THE DATA**



**SELECITNG ROWS ON WHICH WE HAVE TO WORK**



**Filtering the rows on basis of arrest and time**



**CHANGING DATE INTO ONE FORMAT**



**GETTING DATEWISE ARREST**



**CALCULATING TOTAL ARREST BASIS ON OUR CONDITION**



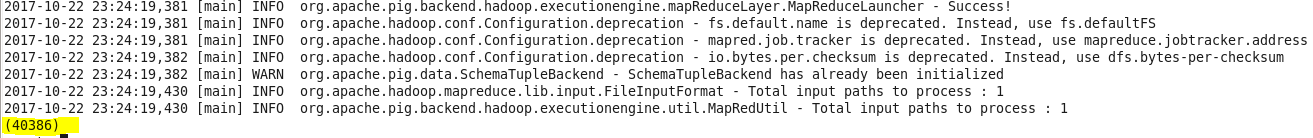
**GROUPING OUR DATA**



**CALCULATING FINAL OUTPUT**



**FINAL OUTPUT**



**NOTES: GITHUB REPOSITORY CONTAINS**

**1)DOCUMENT**

**2)CODE**

**3)Jar file**