*Big Data and Hadoop Development*

**A C A D G I L D**

**Project 1.1**

**Page 1**

*Big Data and Hadoop Development*

**A C A D G I L D**

*Project 1.1 - USA Crime Analysis*

**Table of Contents**

1. Introduction .......................................................................................................................................... 3

2. Objective ............................................................................................................................................... 3

3. Prerequisites ......................................................................................................................................... 3

4. Associated Data Files ............................................................................................................................ 3

5. Problem Statement ............................................................................................................................... 3

6. Approximate Time to Complete Task ................................................................................................... 3

**Page 2**

*Big Data and Hadoop Development*

**A C A D G I L D**

**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.

**2. Objective**

**3. Prerequisites**

You should have Hadoop cluster installed in your system.

**4. 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

**5. 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

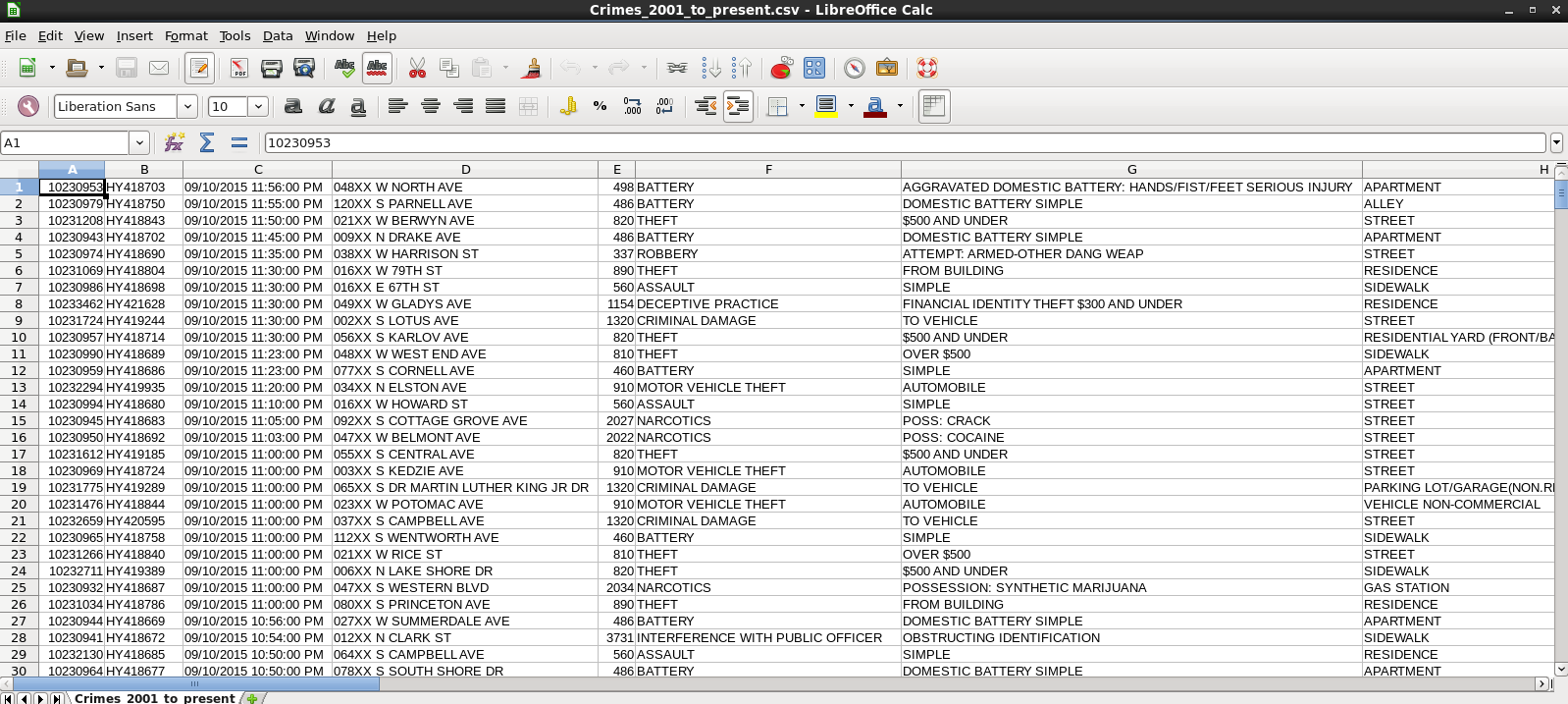
code 32.

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.

**Solution :**

dataset:



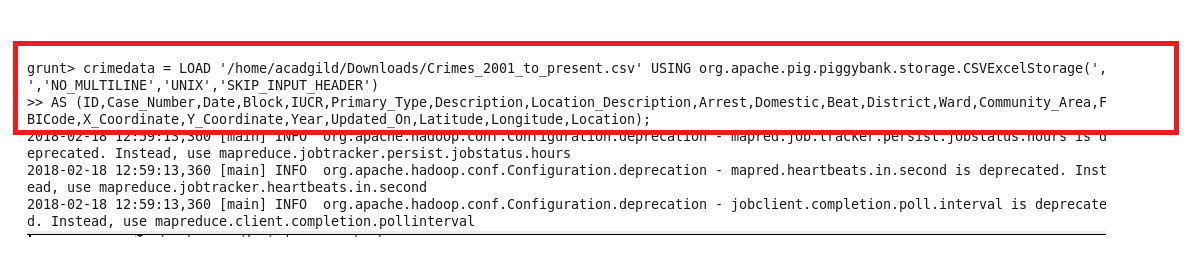
**Dataset Description:**

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

1. Write a Pig program to calculate the number of cases investigated under each FBI code



* Group all the data in the data-loaded relation crimeDataset by **FBICode** in the relation **groupFBICodes**. This will result in groups of record for each FBICode

crimedata\_x = FOREACH crimedata GENERATE $0 AS ID:int,

$1 AS Case\_Number:int,

$2 AS Date:chararray,

$3 AS Block:chararray,

$4 AS IUCR:chararray,

$5 AS Primary\_Type:chararray,

$6 AS Description:chararray,

$7 AS Location\_Description:chararray,

$8 AS Arrest:boolean,

$9 AS Domestic:boolean,

$10 AS Beat:int,

$11 AS District:int,

$12 AS Ward:int,

$13 AS Community\_Area:int,

$14 AS FBICode:chararray,

$15 AS X\_Coordinate:int,

$16 AS Y\_Coordinate:int,

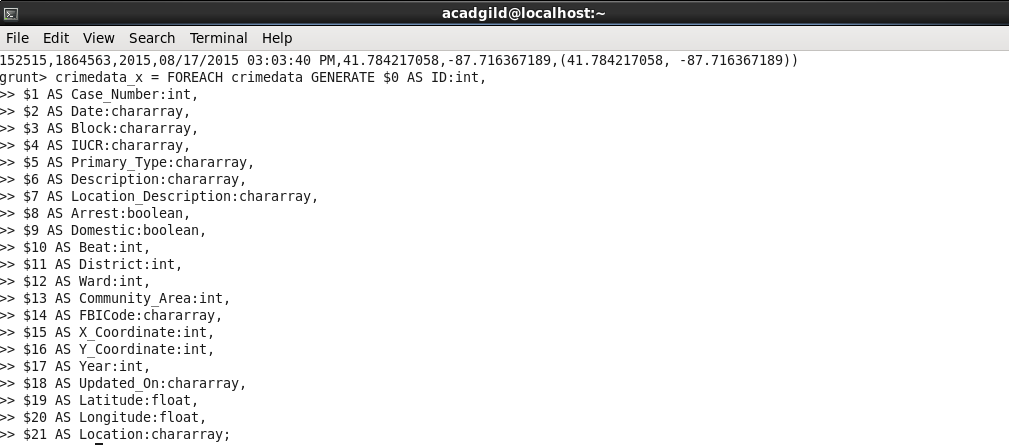
$17 AS Year:int,

$18 AS Updated\_On:chararray,

$19 AS Latitude:float,

$20 AS Longitude:float,

$21 AS Location:chararray;





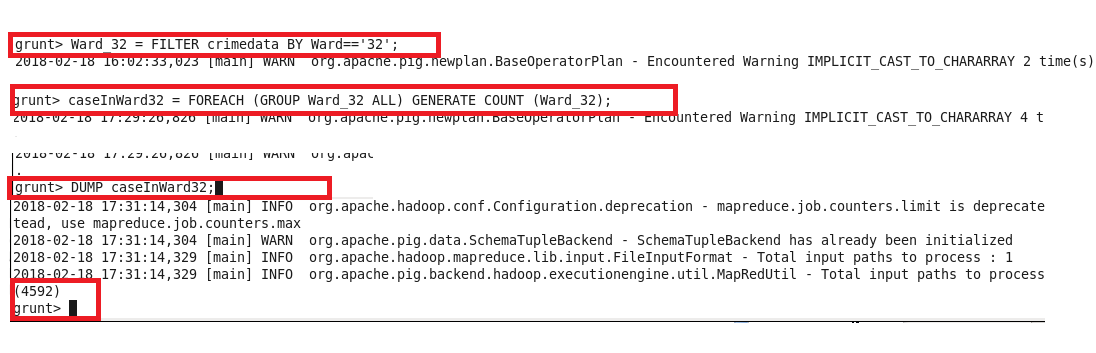


Output:

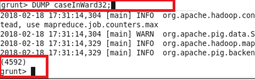


2. Write a Pig program to calculate the number of cases investigated under FBI

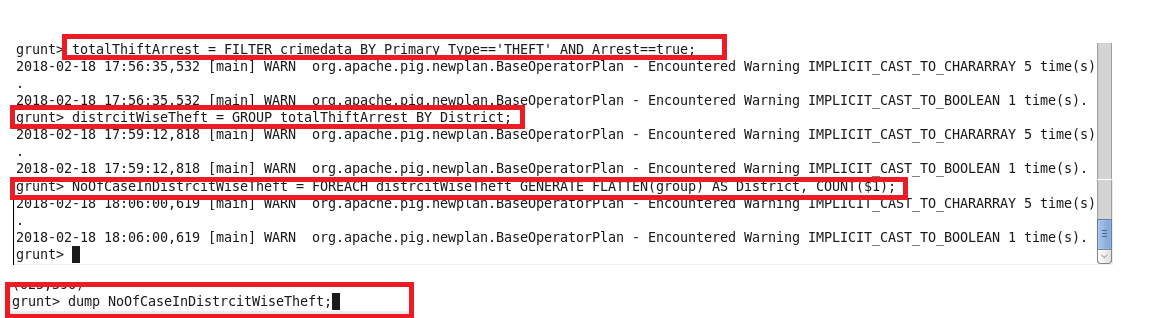
code 32



output:



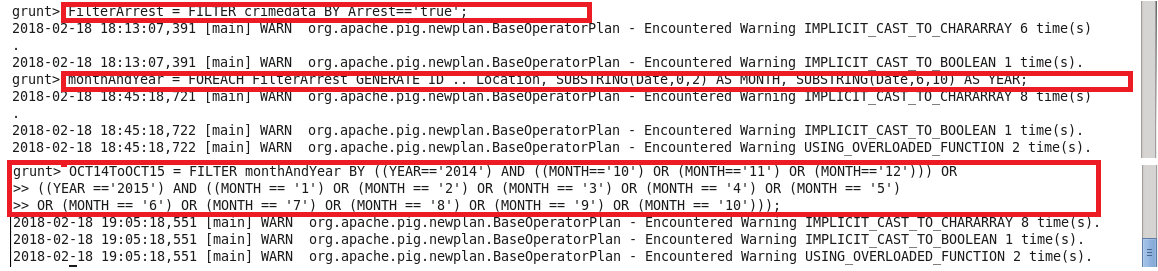
3. Pig program to calculate the number of arrests in theft district wise.

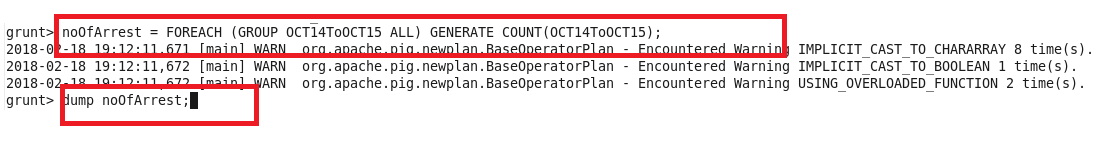


Output:



4.Pig program to calculate the number of arrests done between October 2014 and October 2015.





**OutPut:**

