**Project 1**

**USA Crime Analysis**

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.

Dataset Description:

Columns:

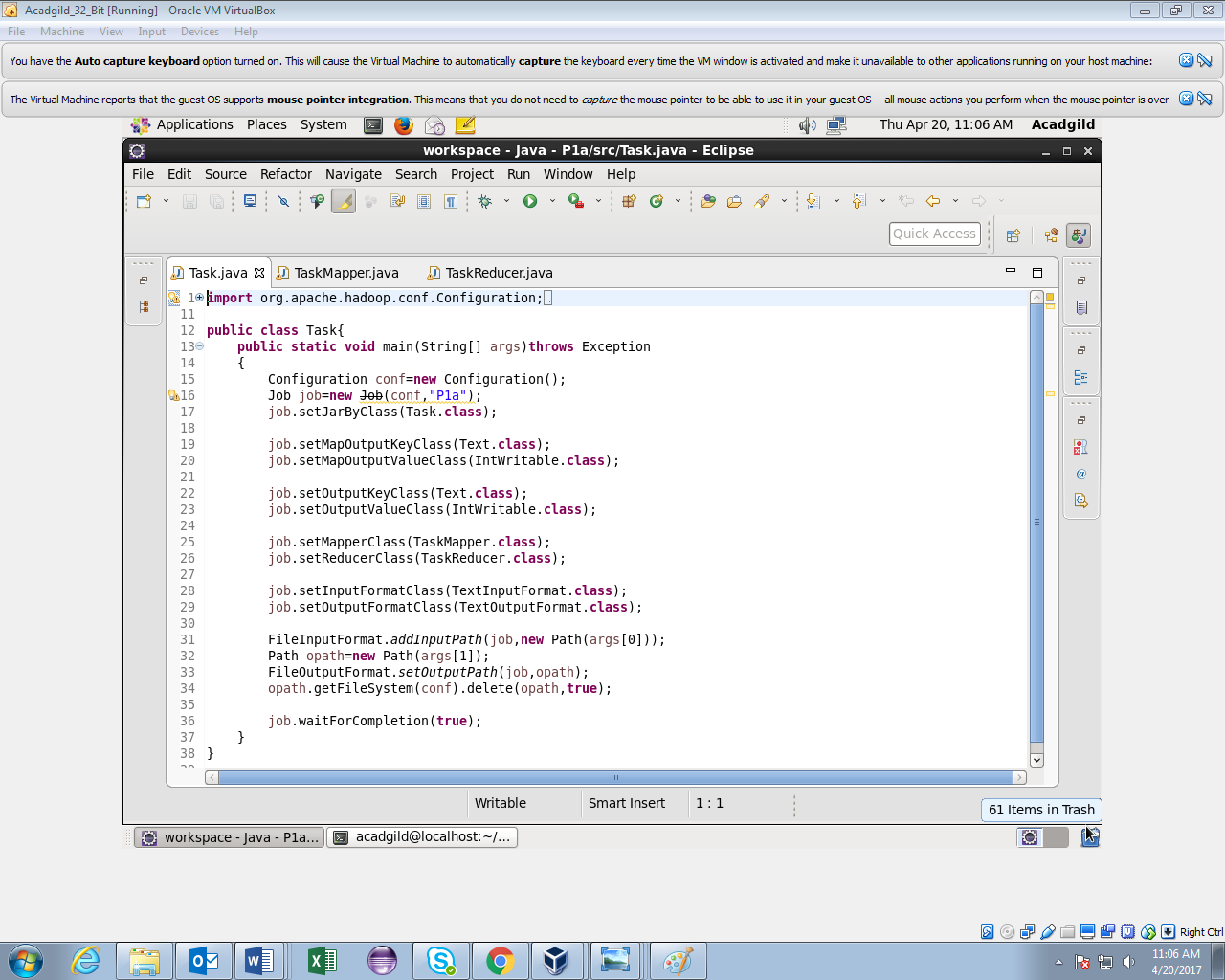
ID,Case Number,Date,Block,IUCR,Primary Type,Description,Location Description,Arrest,Domestic,Beat,District,Ward,Community Area,FBICode,X Coordinate,Y Coordinate,Year,Updated On,Latitude,Longitude,Location

Problem Statement

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

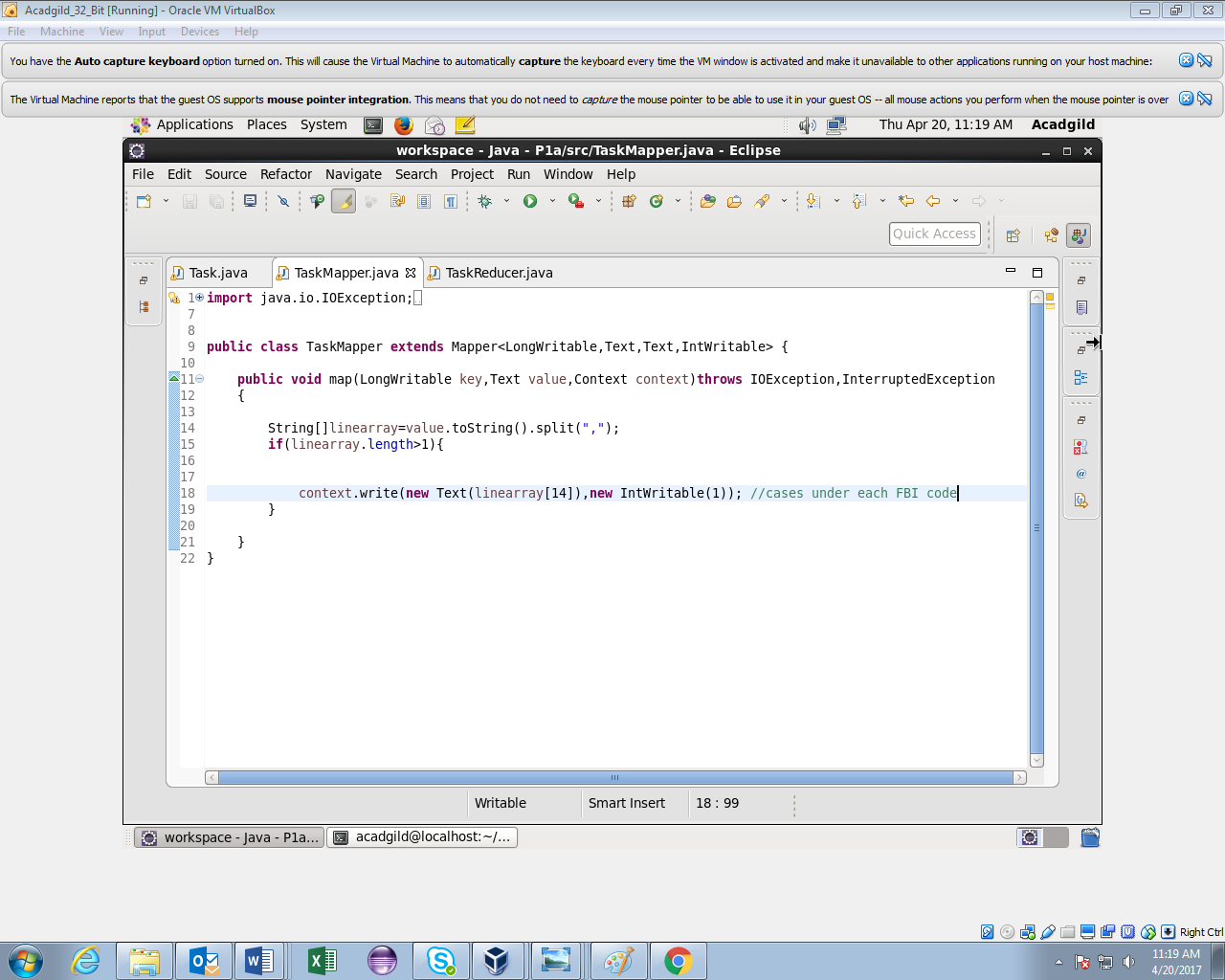
**Mapreduce**

**Driver Class**



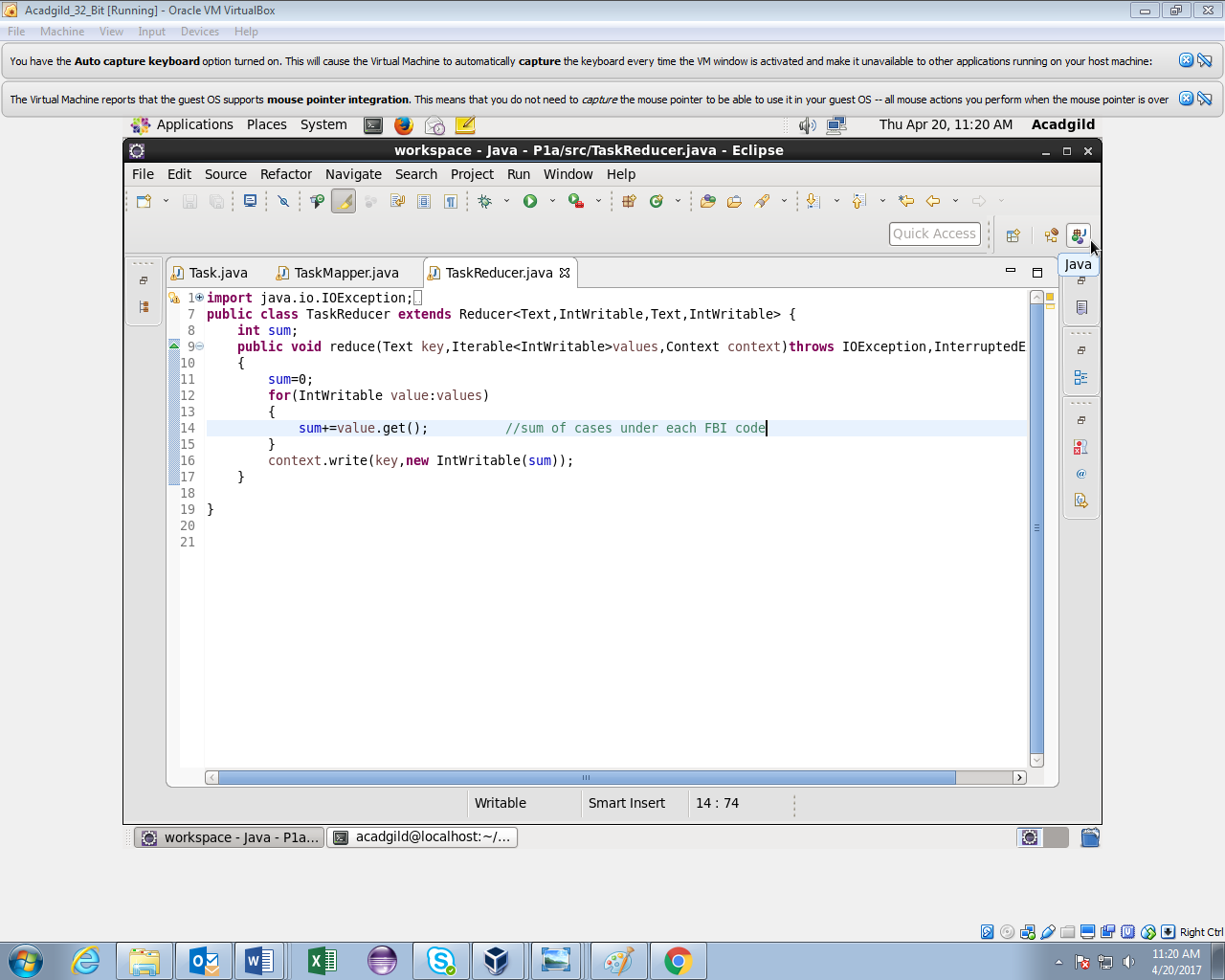
**Mapper class**

FBI code is given as key and count of 1 as value for each case.

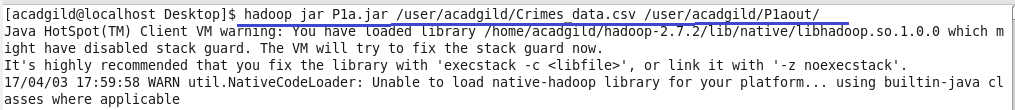


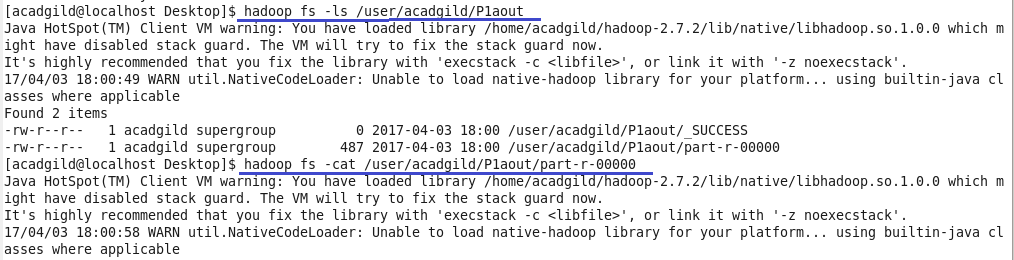
**Reducer class**

Counted the number of cases for each FBI code.

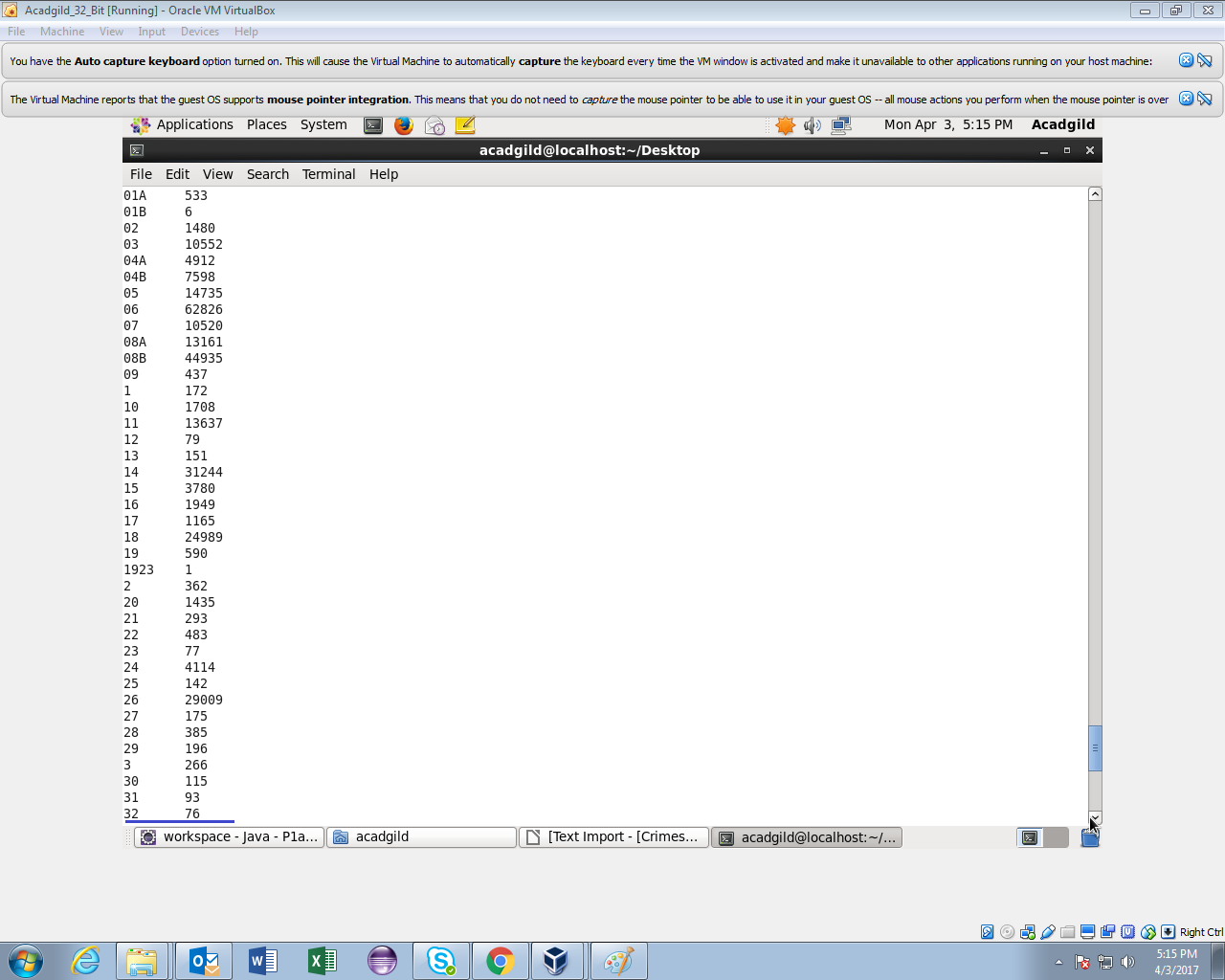


**Running jar**





**Output**



**Output(contd.)**

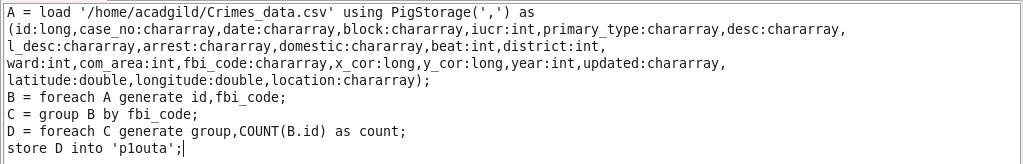


**Pig Script**

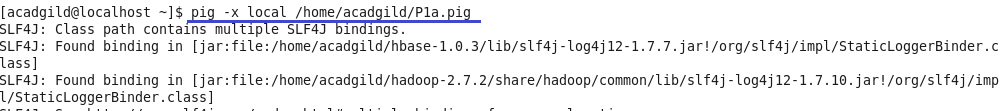
Logic:

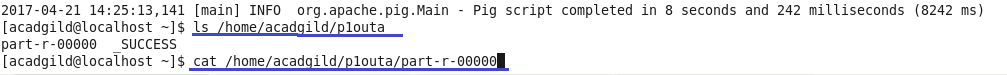
1.Loaded the dataset

2.grouped by fbi code and made a count of case id.

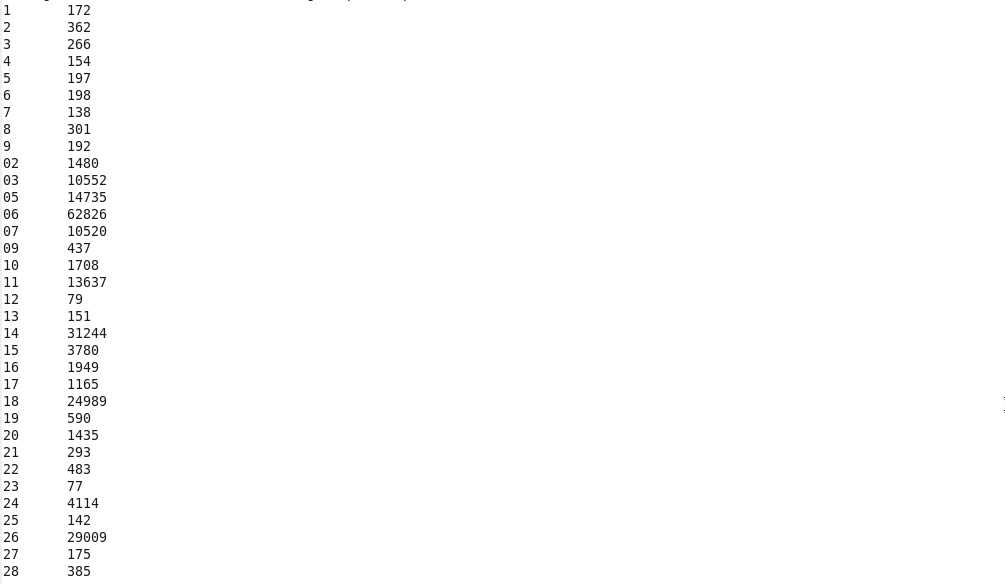


**Running Pig Script**





**Output**

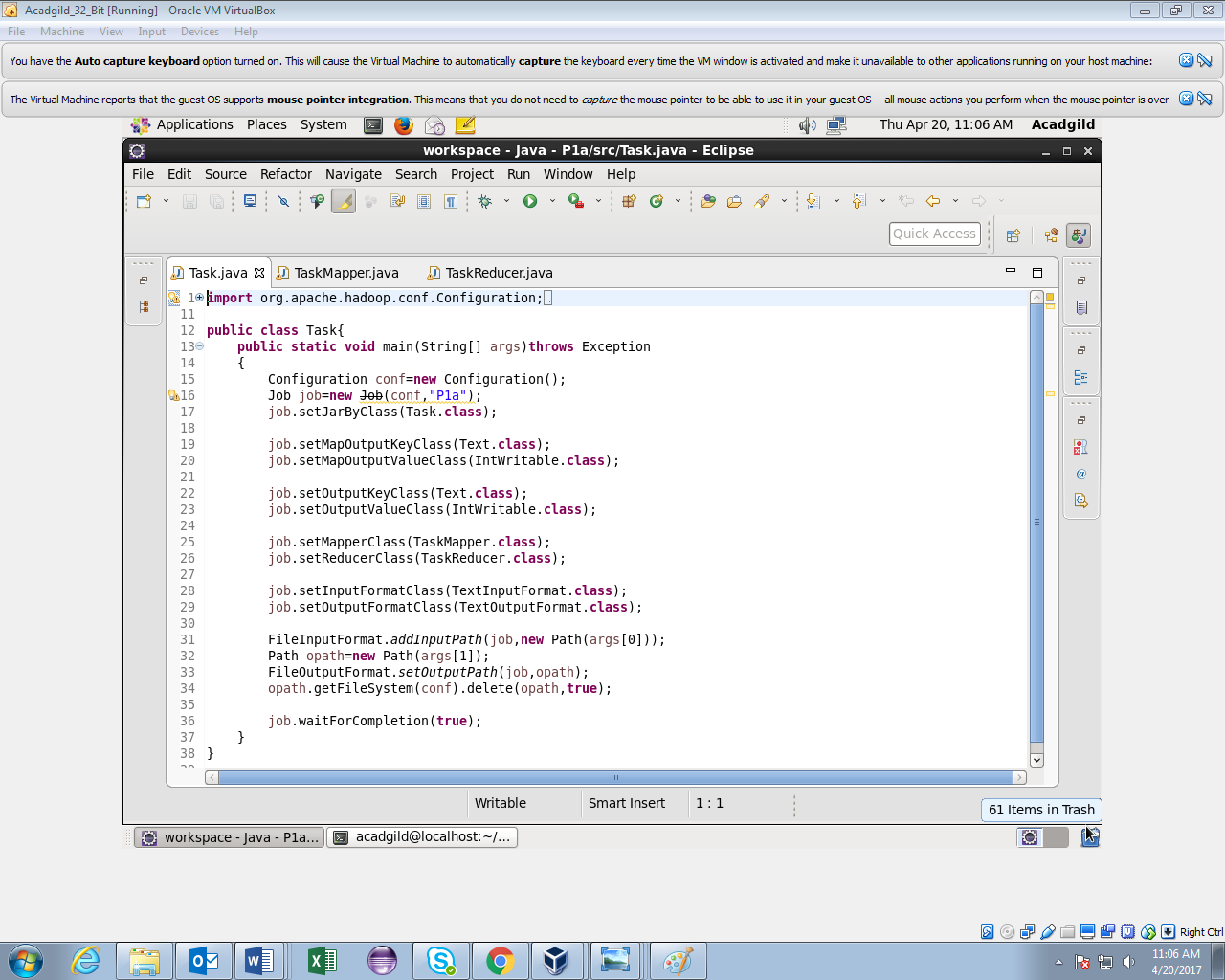




1. Write a mapreduce and pig program to calculate the number of cases investigated under FBI code 32.

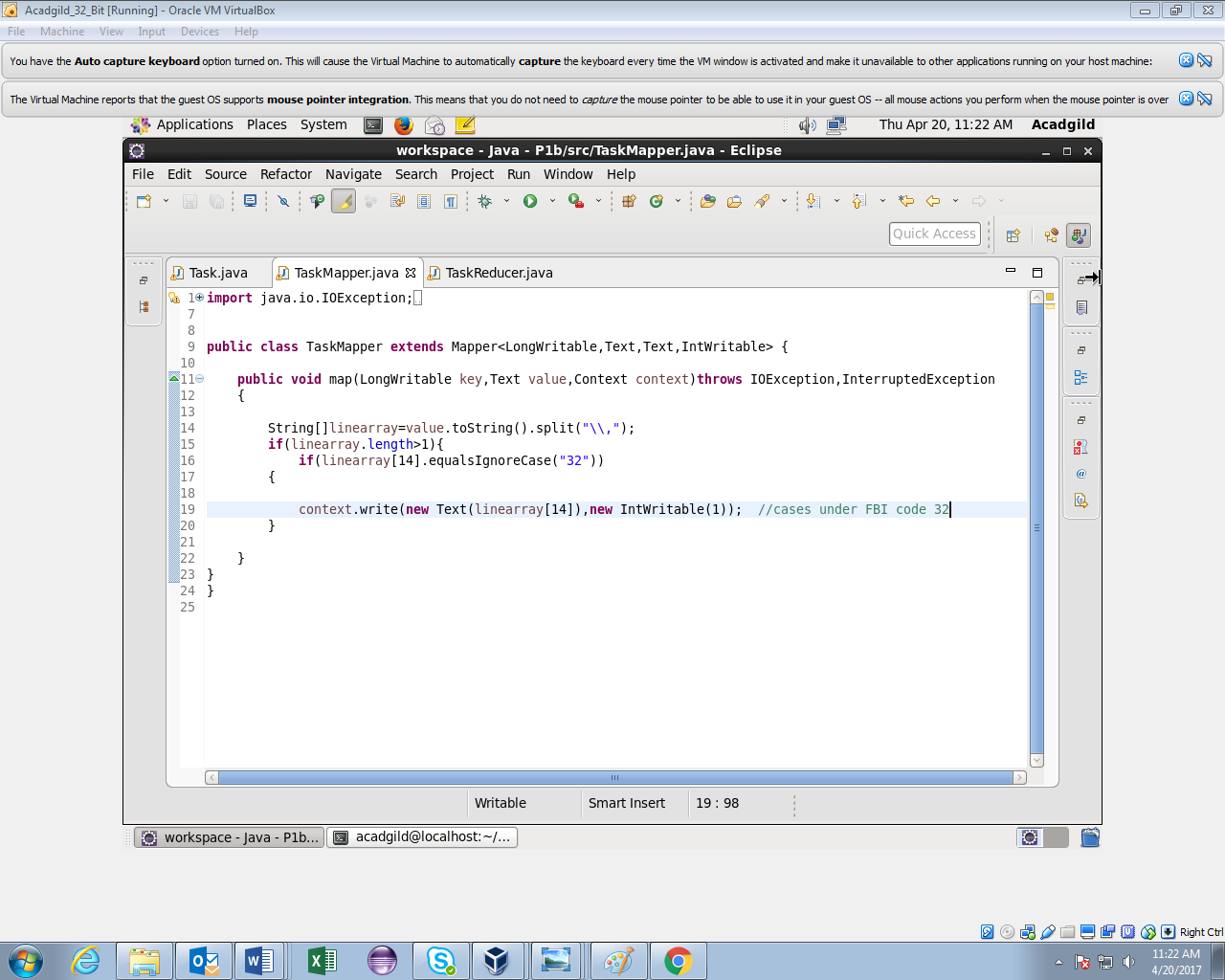
**Mapreduce**

**Driver class**



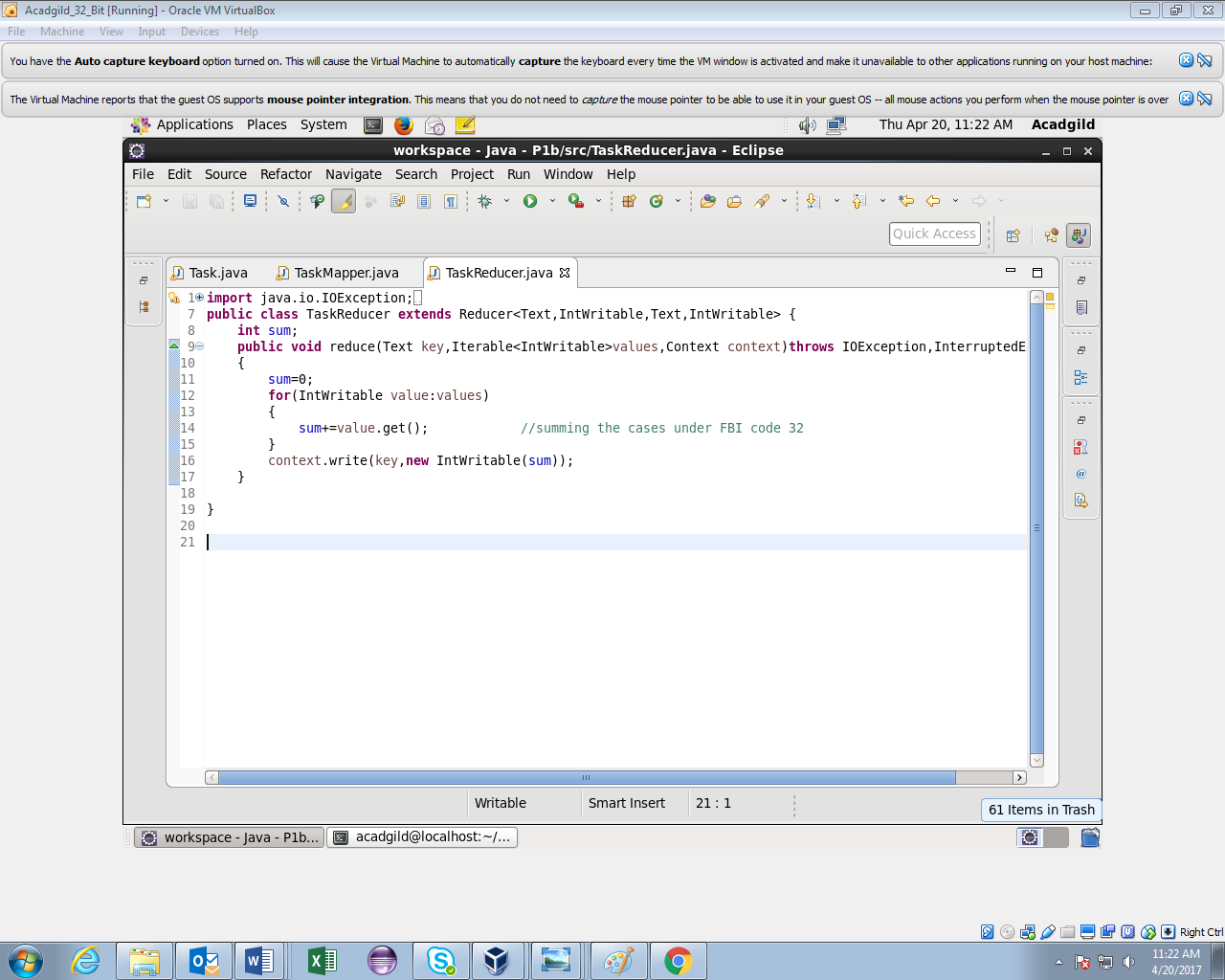
**Mapper class**

Filtered by FBI code=32 and given it as key and a count of 1 as value for each case.

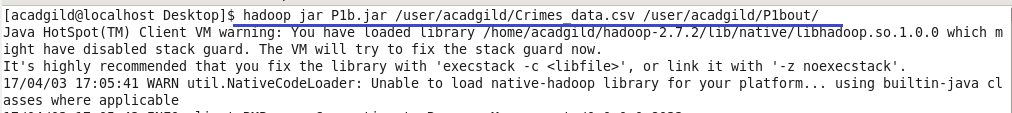


**Reducer class**

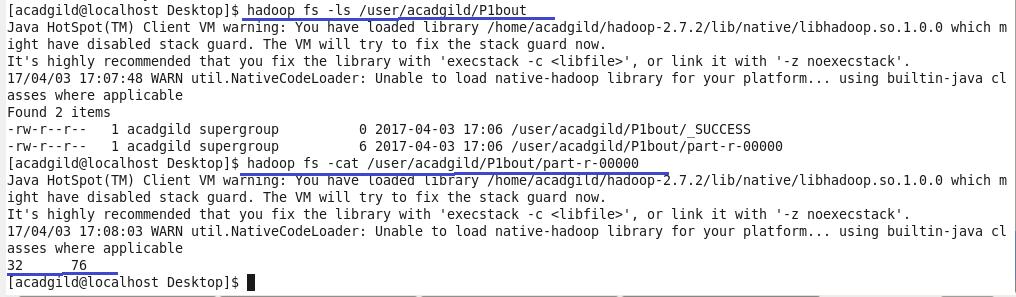
Counted the number of cases for fbi code 32.



**Running jar**



**Output**



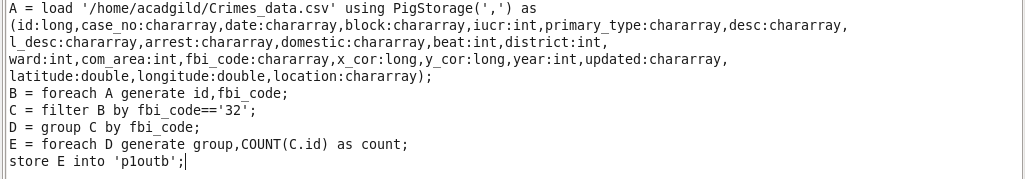
**Pig Script**

Logic

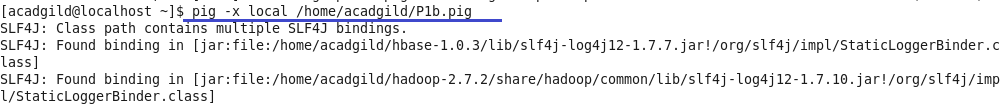
1.Loaded the dataset.

2.filtered by fbi code =32

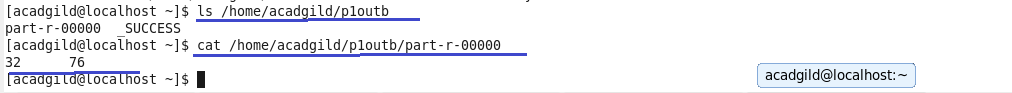
3.grouped by fbi code and made a count of case id.



**Running Pig Script**



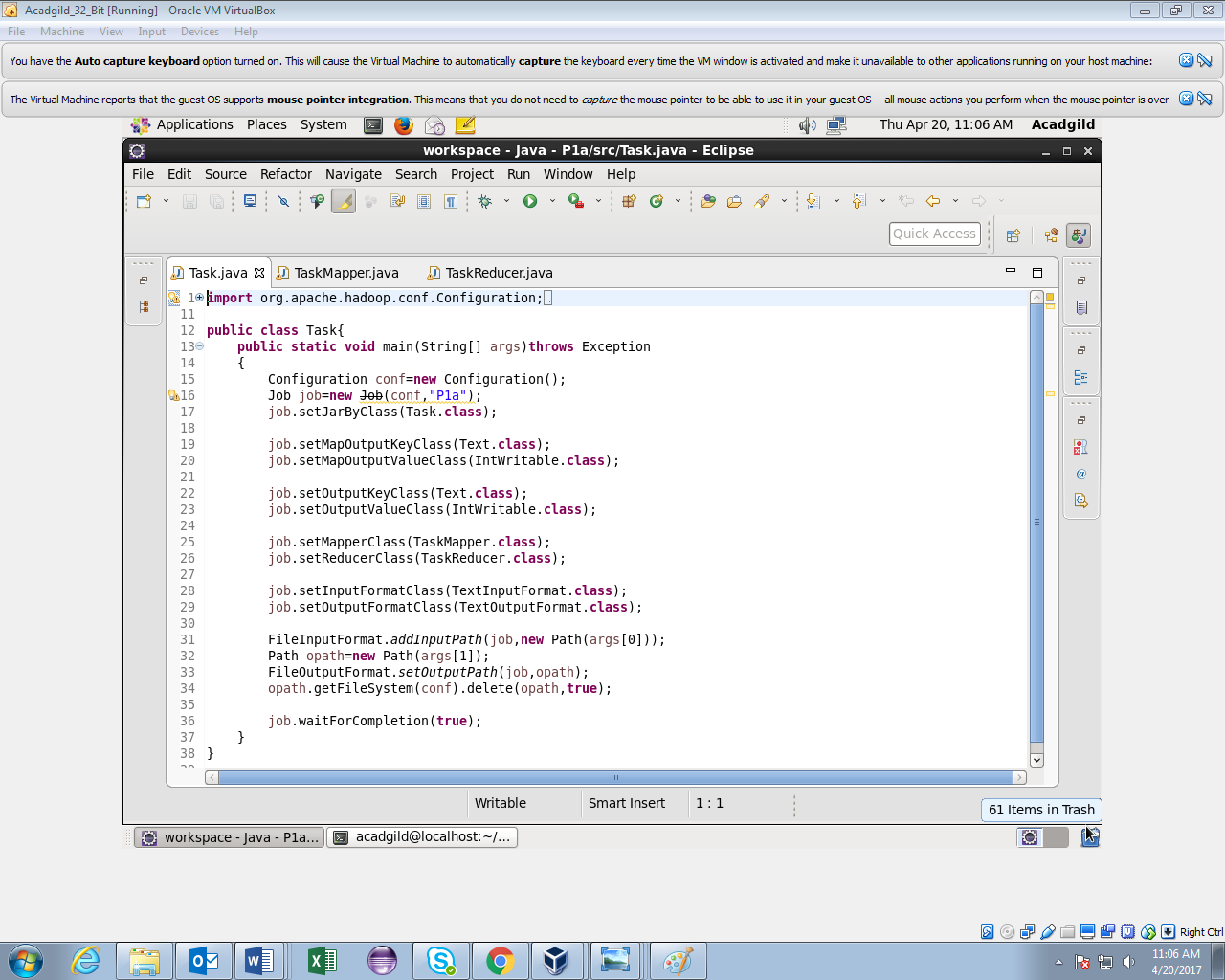
**Output**



1. Write a mapreduce and pig program to calculate the number of arrests in theft district wise.

**Mapreduce**

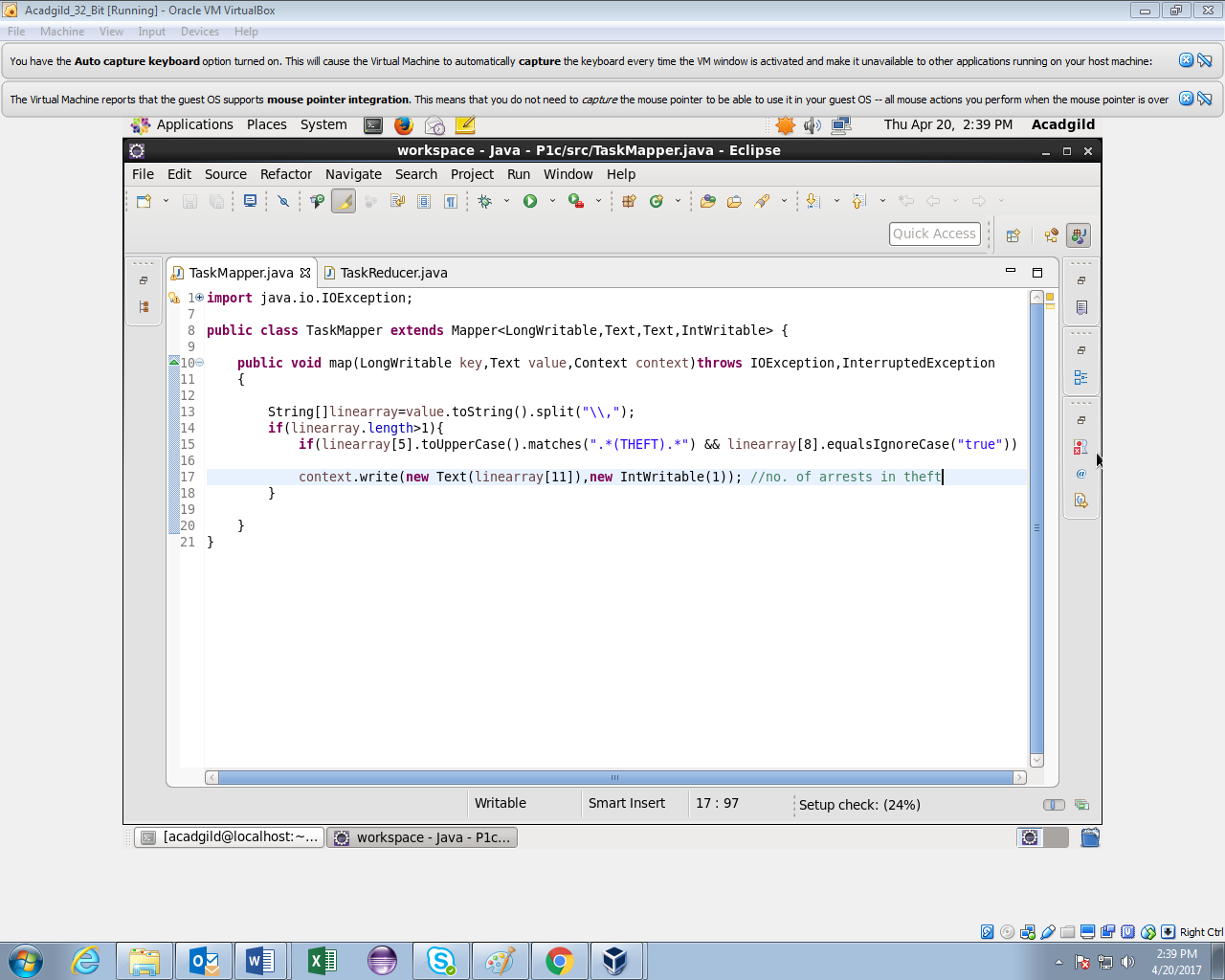
**Driver class**



**Mapper class**

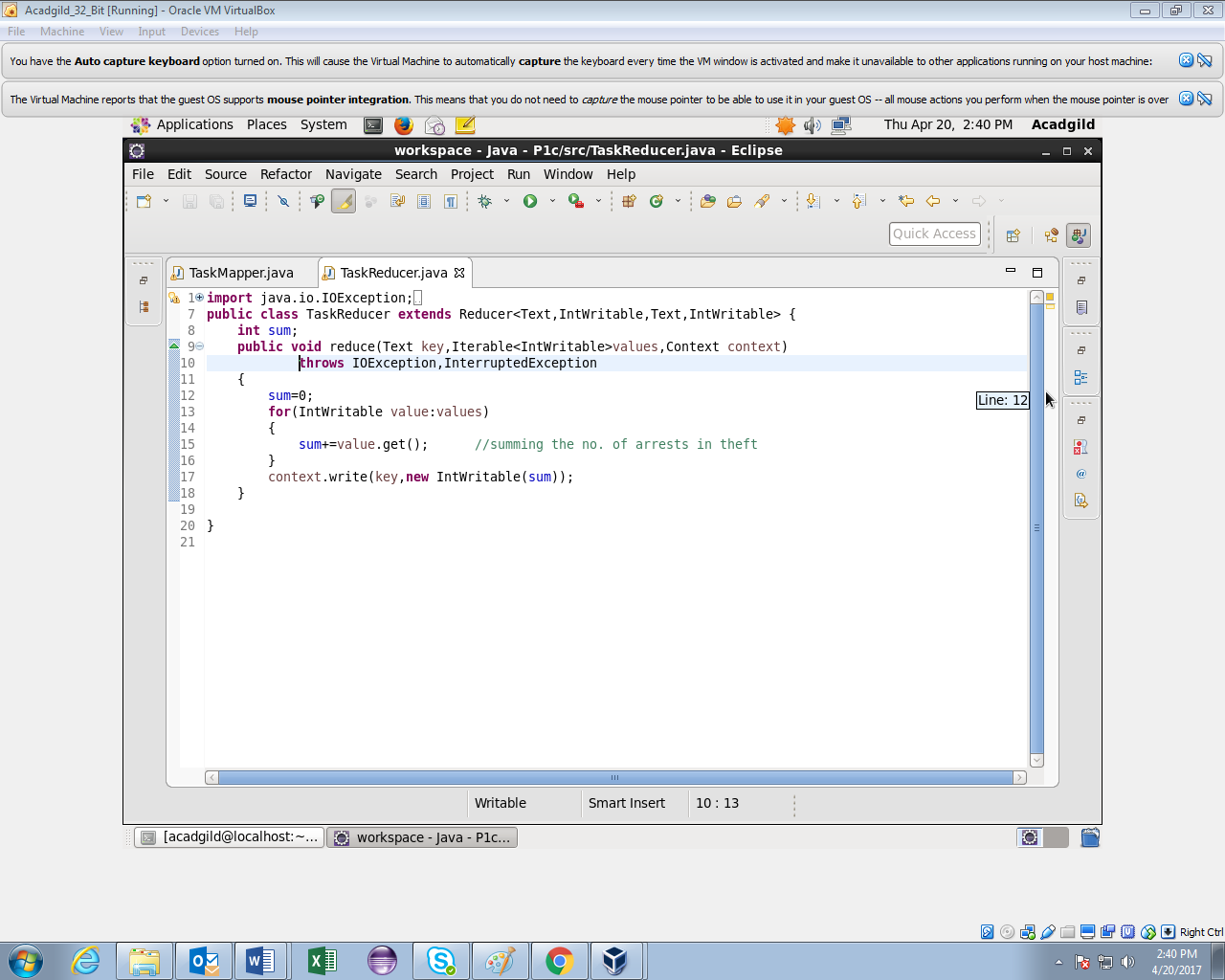
Logic:

Filtered by primary type theft and arrest = true.District is given as key and count of 1 as value for each arrest.

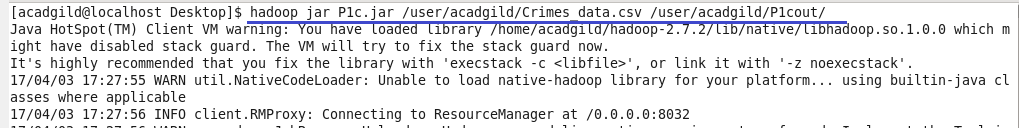


**Reducer class**

Counted the no. of arrests in theft for each district.



**Running jar**



**Output**



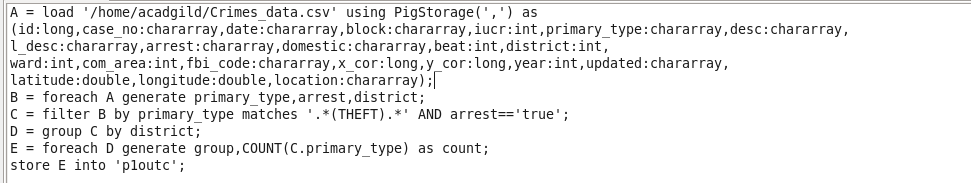
**Pig Script**

Logic:

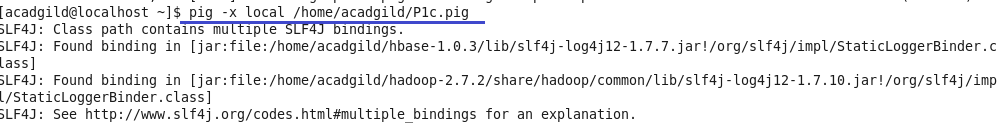
1.loaded the dataset.

2.filtered by primary type theft using matches function and arrest=true

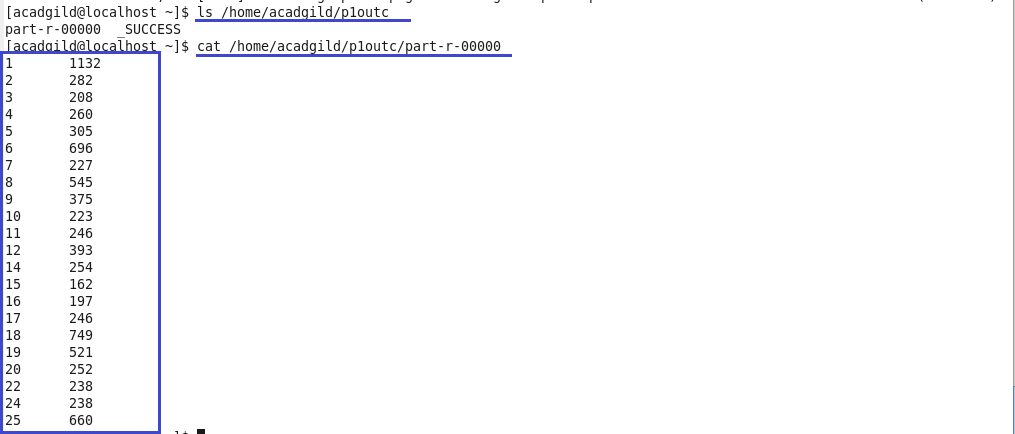
3.grouped by district and made count of primary type.



**Running Pig Script**

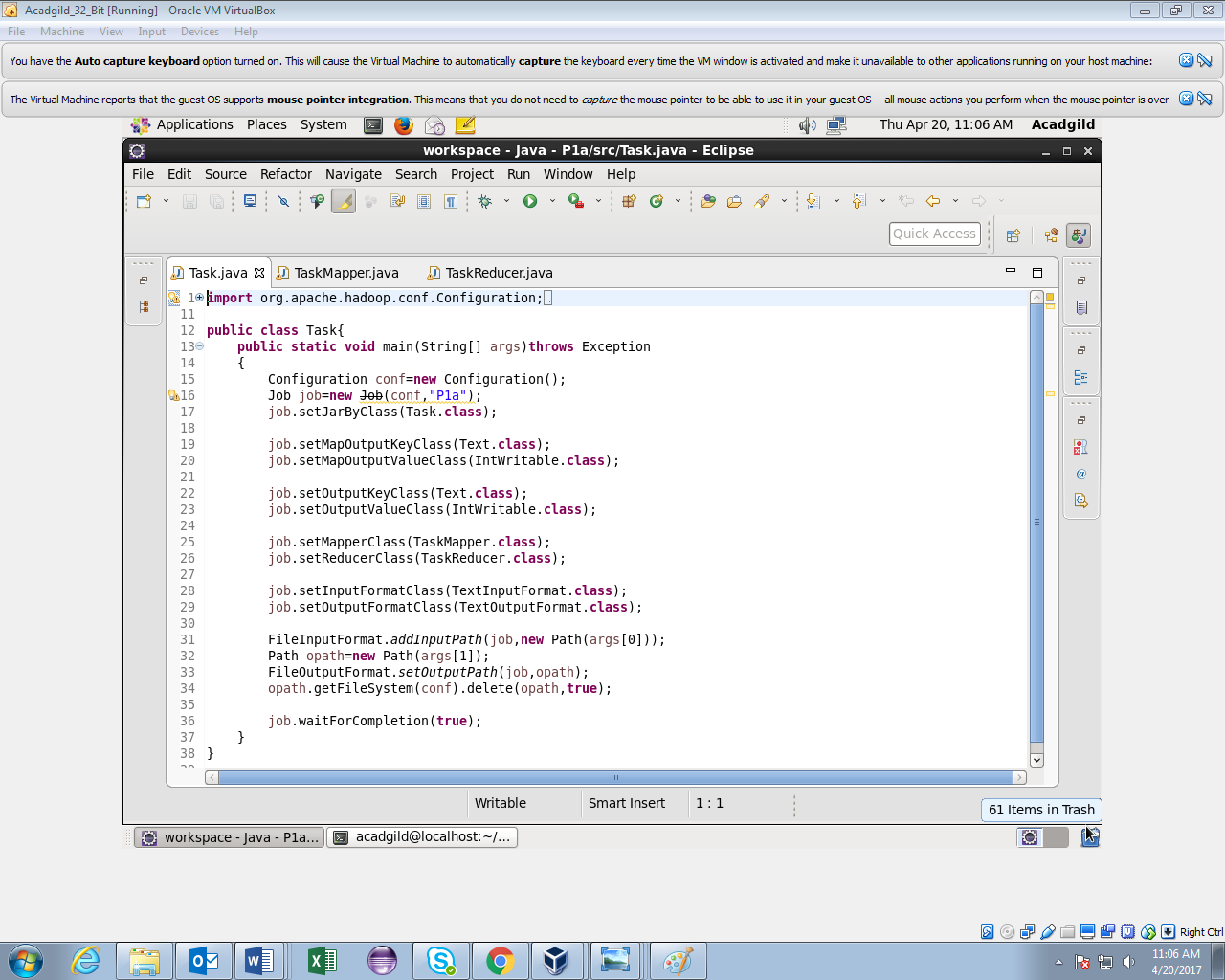


**Output**



1. Write a mapreduce and pig program to calculate the number of arrests done between October 2014 and October 2015

**Driver class**

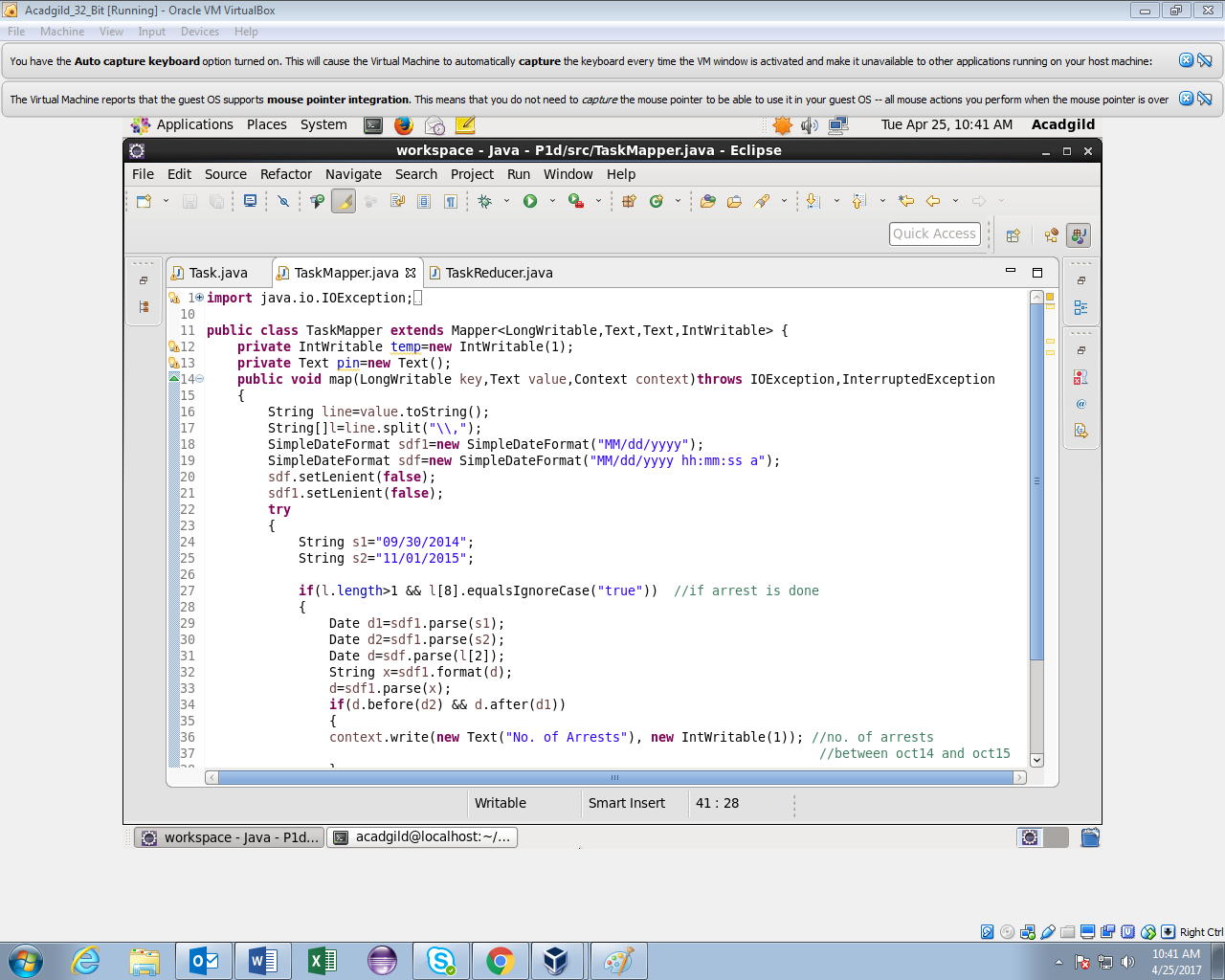


**Mapper class**

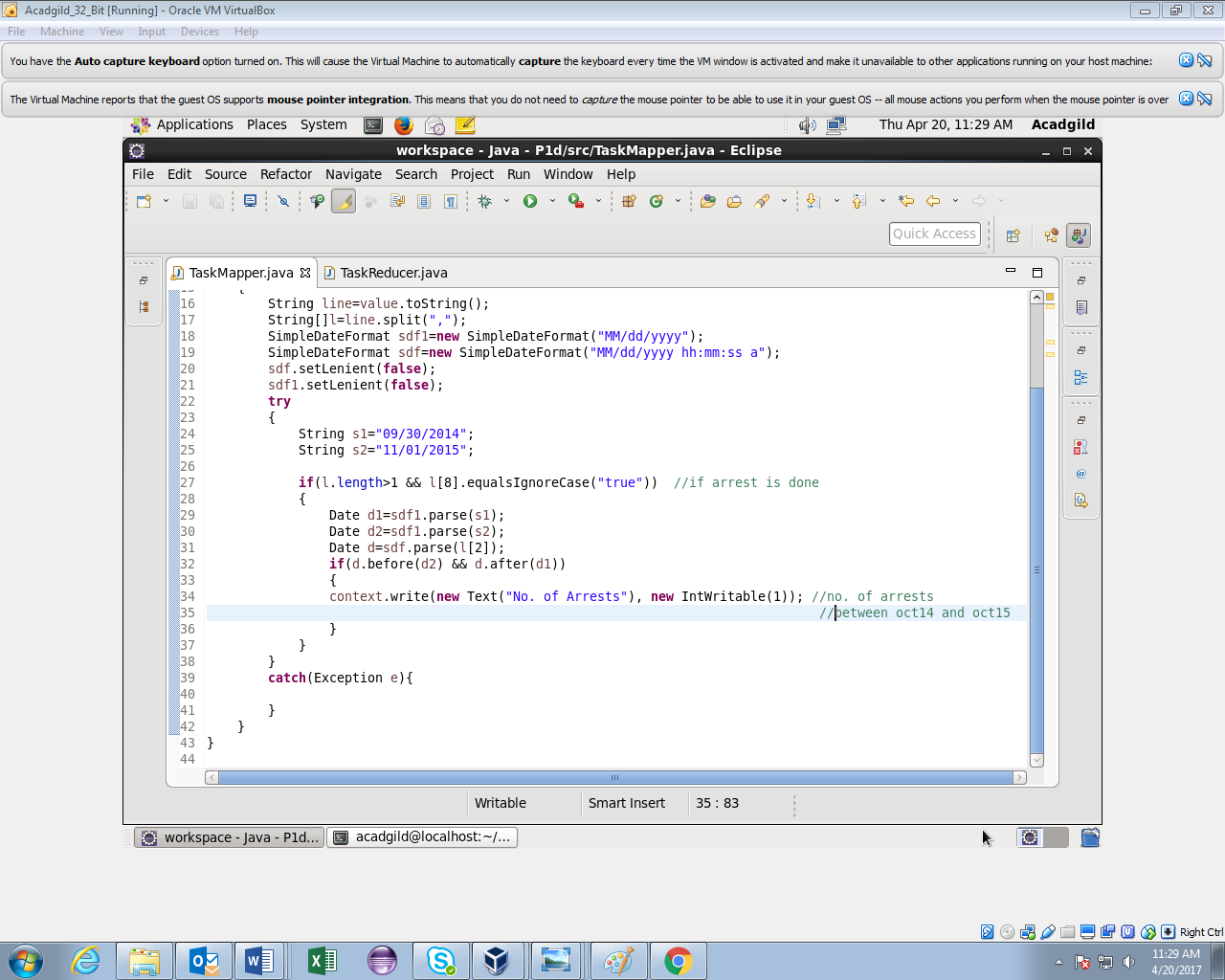
Logic

Date function is used to convert string into dates and filtered for arrest=true.

No. of arrests given as key and count of 1 as value for each arrest between the dates 30 sept 2014 and 1 nov 2015 for the period given oct 2014 to oct 2015.

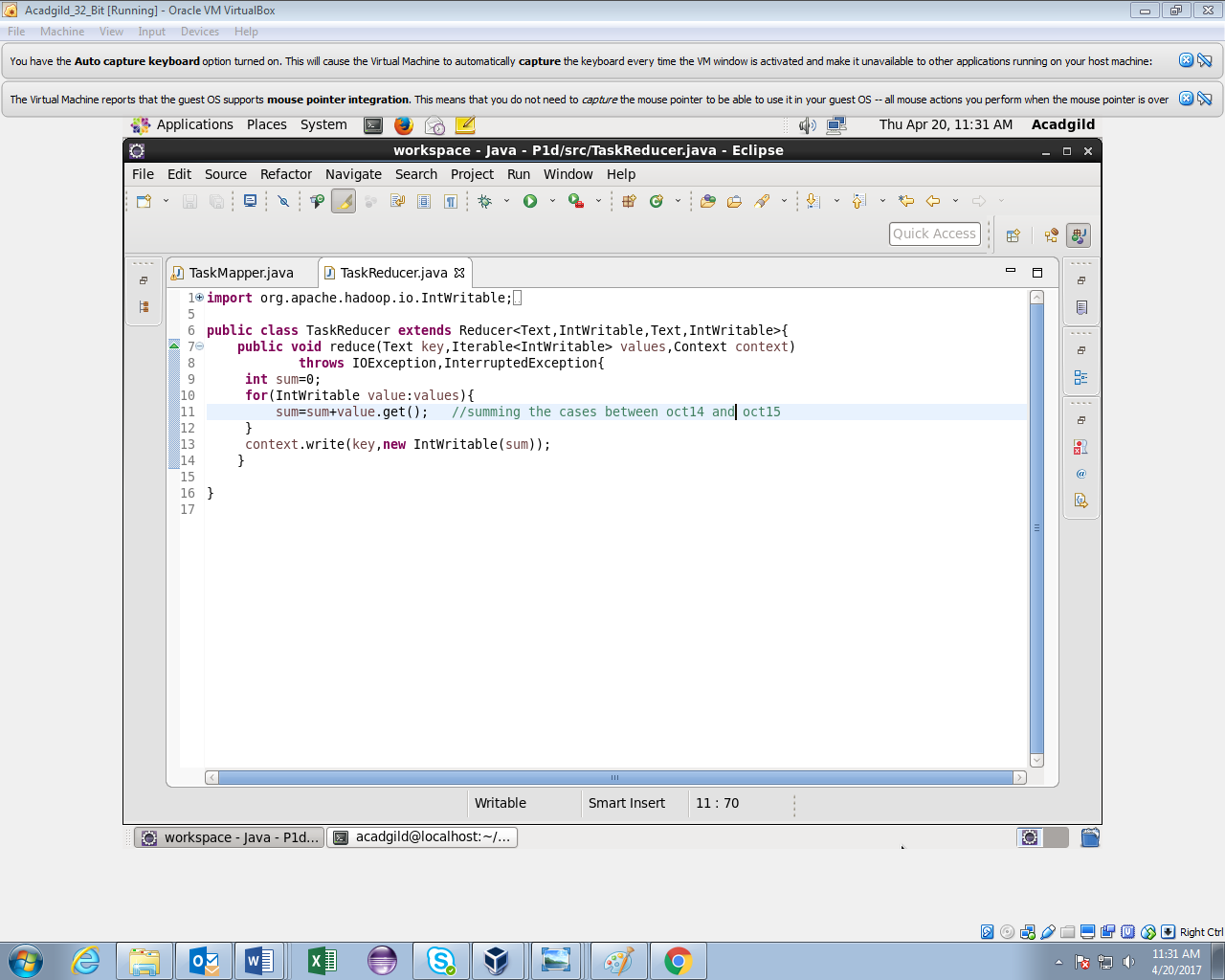


**Mapper class(contd..)**

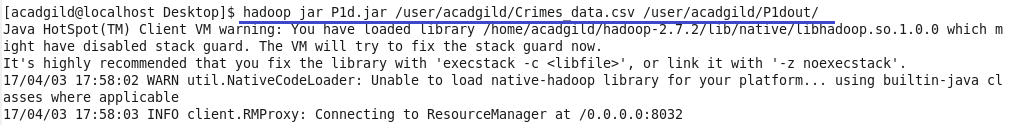


**Reducer class**

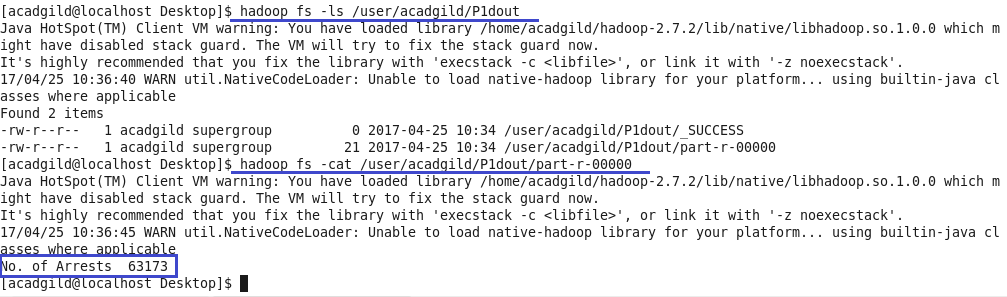
Counted the no of. Arrests for the given period.



**Running jar**



**Output**

****

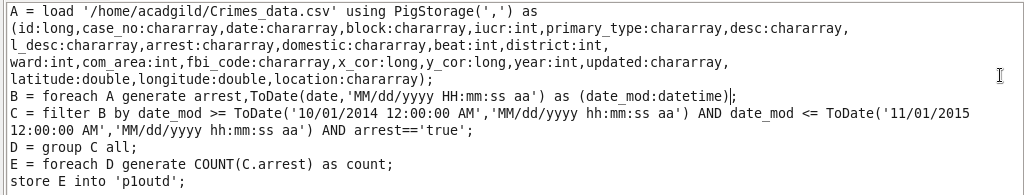
**Pig Script**

Logic

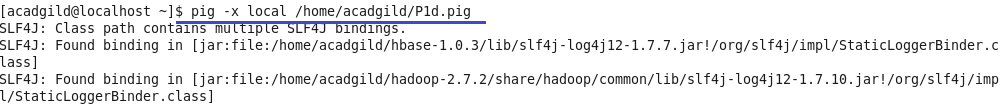
1.loaded the dataset

2.ToDate function is used to convert into dates and filtered for the given dates and arrest=true.

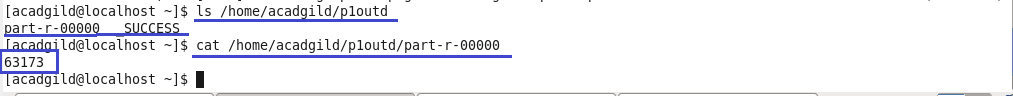
3.grouped by all and made count of arrests.



**Running Pig Script**



**Output**

****