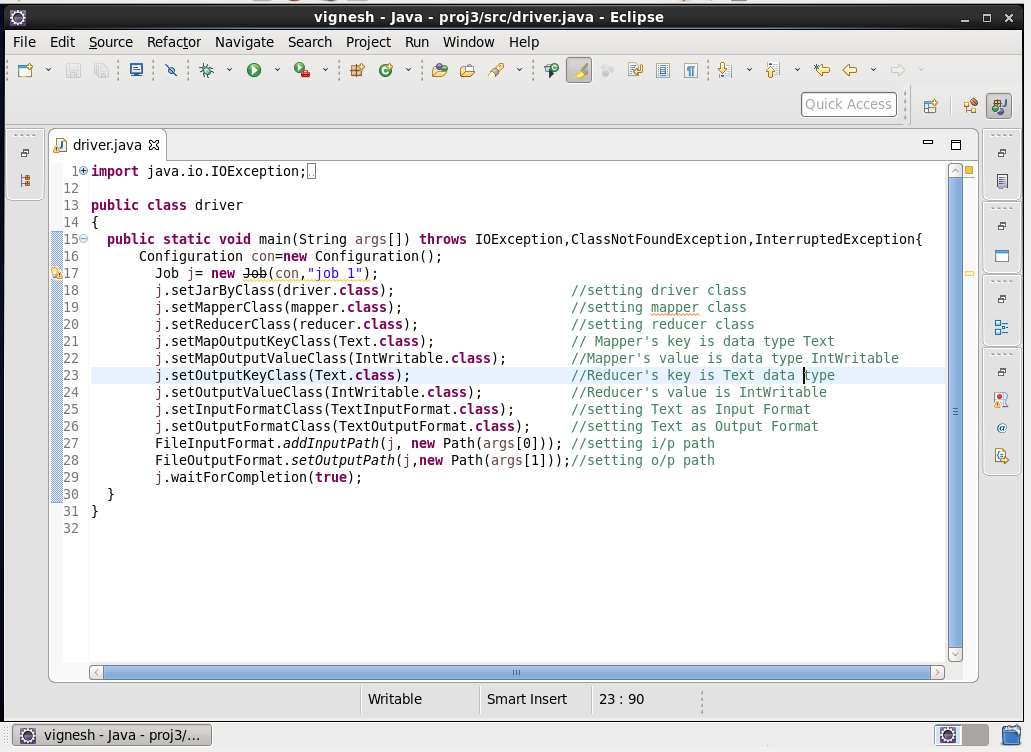
**ASSIGNMENT 13.1 : PROJECT 1**

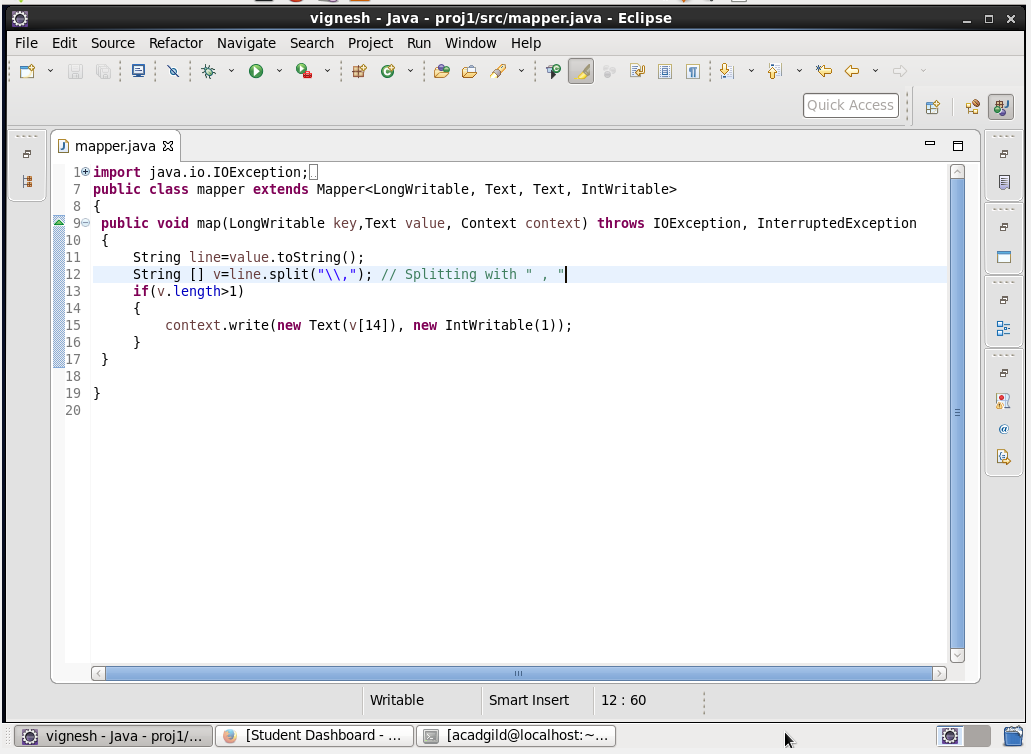
**USA CRIME ANALYSIS**

**1. WRITE A MAPREDUCE AND PIG PROGRAM TO CALCULATE THE NUMBER OF CASES INVESTIGATED UNDER EACH FBI CODE.**

**DRIVER CLASS:**

****

**MAPPER CLASS:**

****

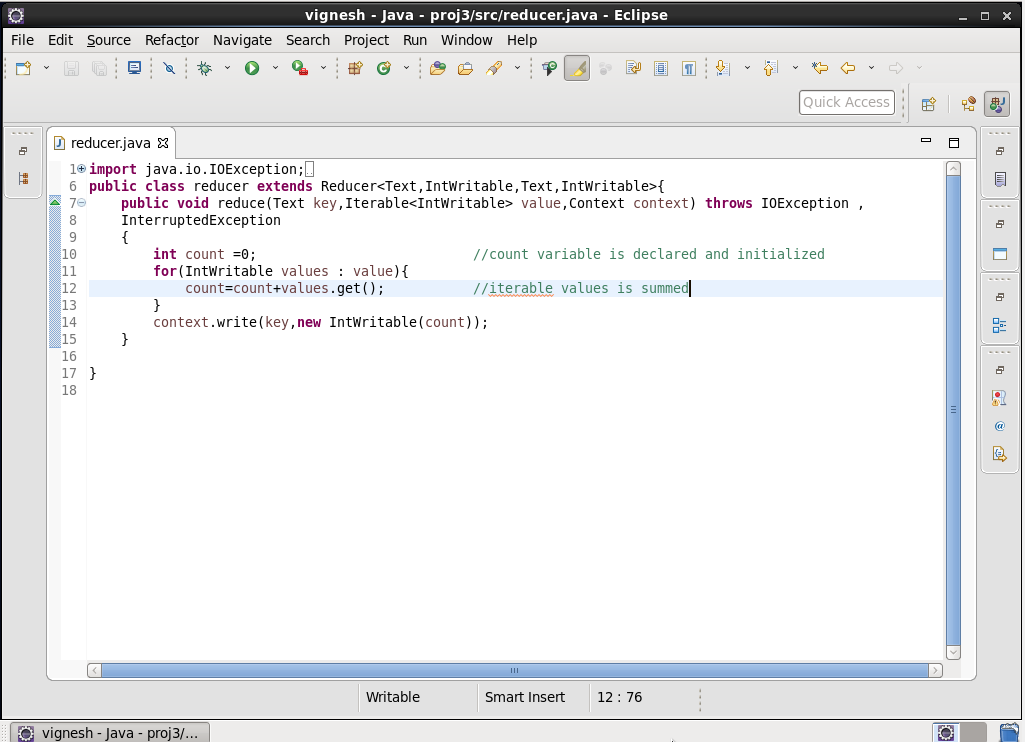
**MAPPER LOGIC:**

LINE 9 : Text offset from beginning is set as key and the entire dataset is set as value.

LINE 12 : Separating the dataset with comma and storing it in array v.

LINE 15 : Setting FBI CODE as key and 1 as value to be further processed by reducer.

**REDUCER CLASS:**

****

**REDUCER LOGIC:**

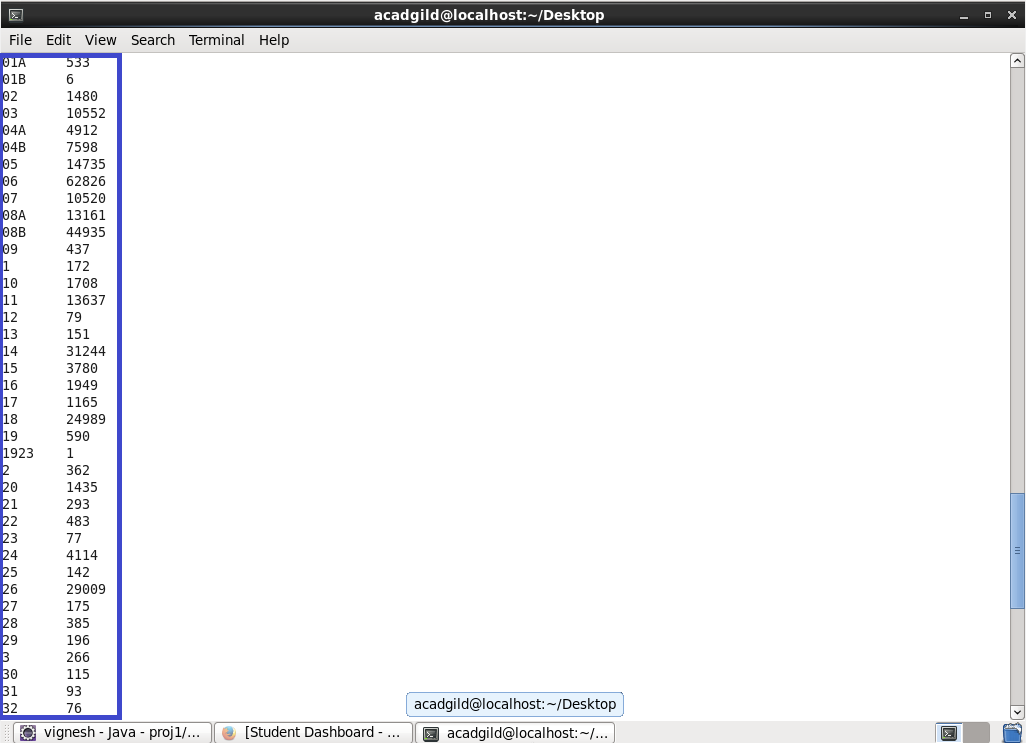
*LINE 10: Initializing the count as 0.*

*LINE 11 : Running an infinite for loop.*

*LINE 12: count will add the iterative values.*

*LINE 14: Finally the same key is used as the key and the count is set as value.*

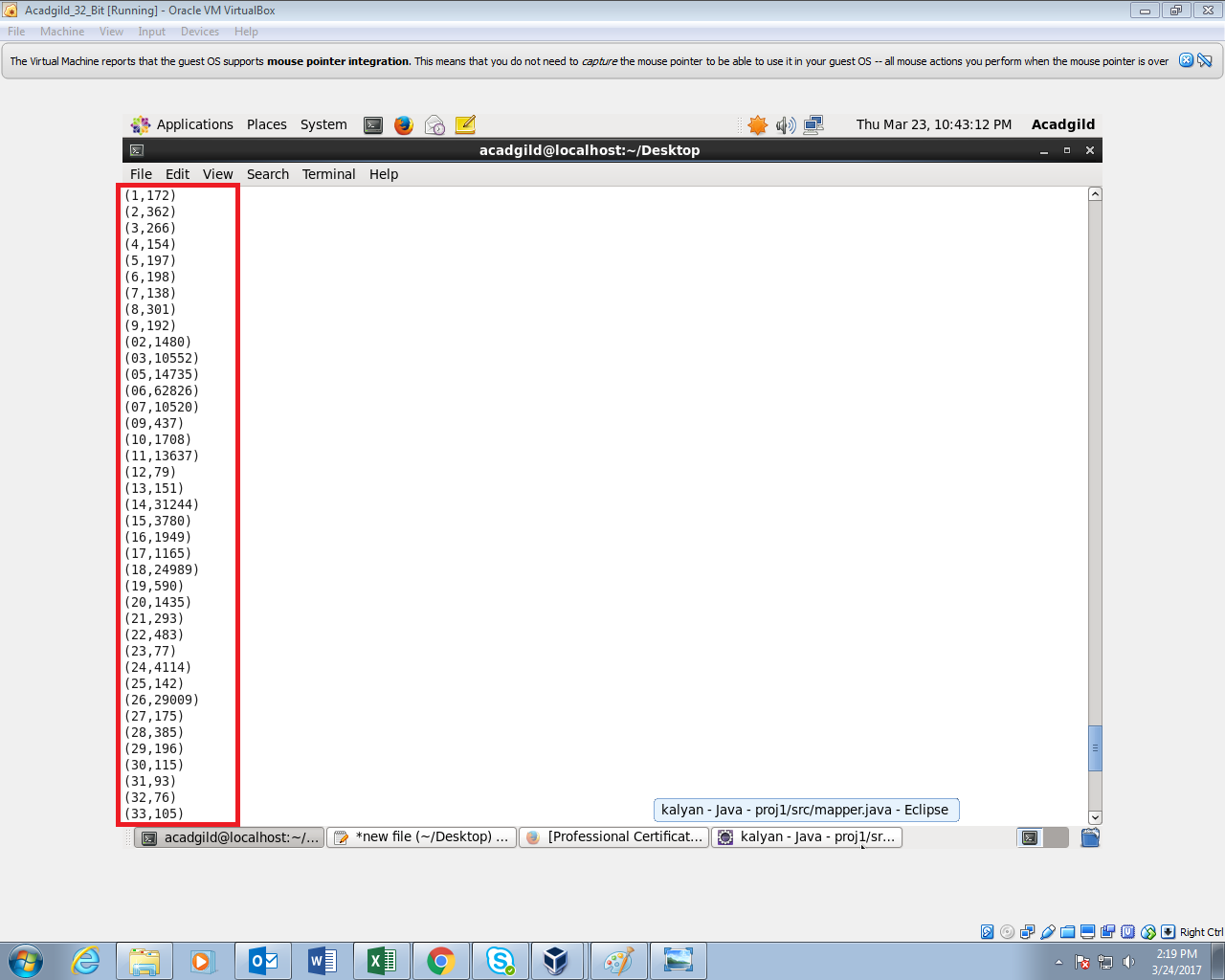
**OUTPUT:**

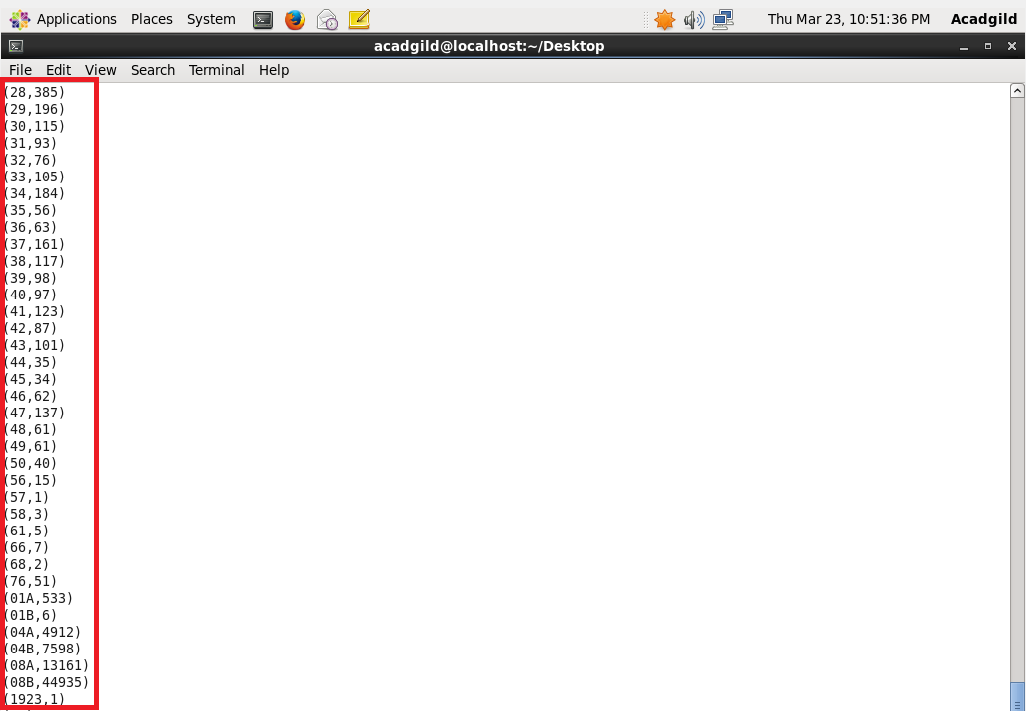
****

**PIG PROGRAM :**

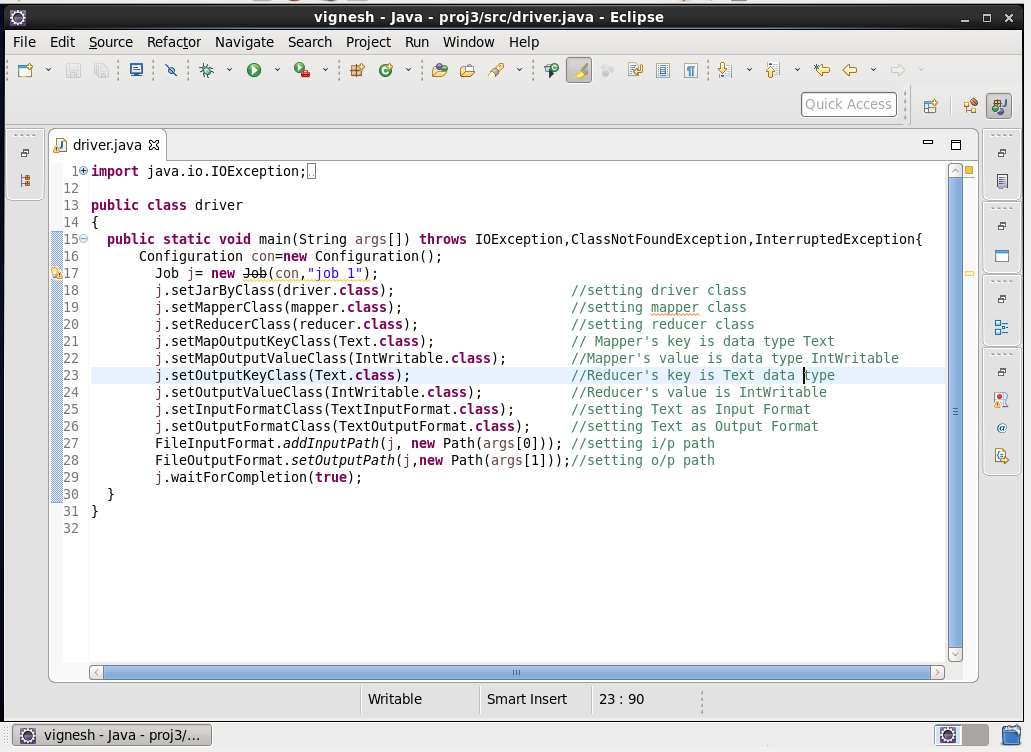
**LOGIC:**

* *Loading the Dataset into the grunt shell using LOAD command.*
* *Grouping by FBI code (column 15) using GROUP BY command.*
* *Generating count of cases for each FBI CODE is done by COUNT command.*
* *Dumping the output to get the count.*

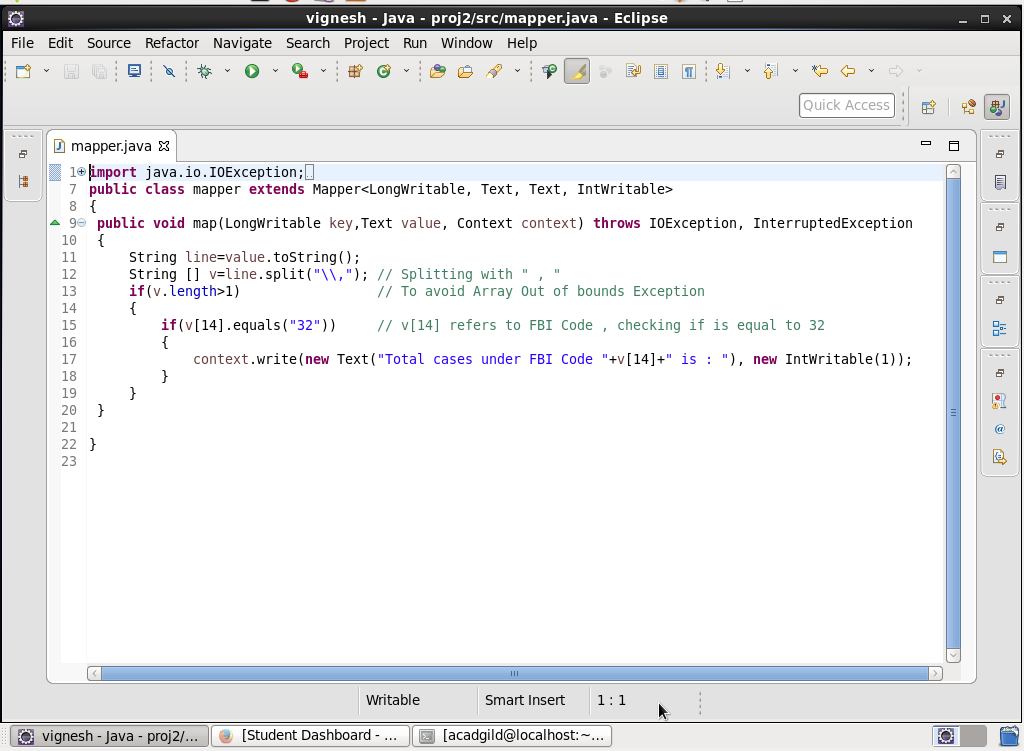
**PIG CODE:OUTPUT:**

****

**2) WRITE A MAPREDUCE AND PIG PROGRAM TO CALCULATE THE NUMBER OF CASES INVESTIGATED UNDER FBI CODE 32.**

**DRIVER CLASS:**

**MAPPER CLASS:**

****

**MAPPER LOGIC:**

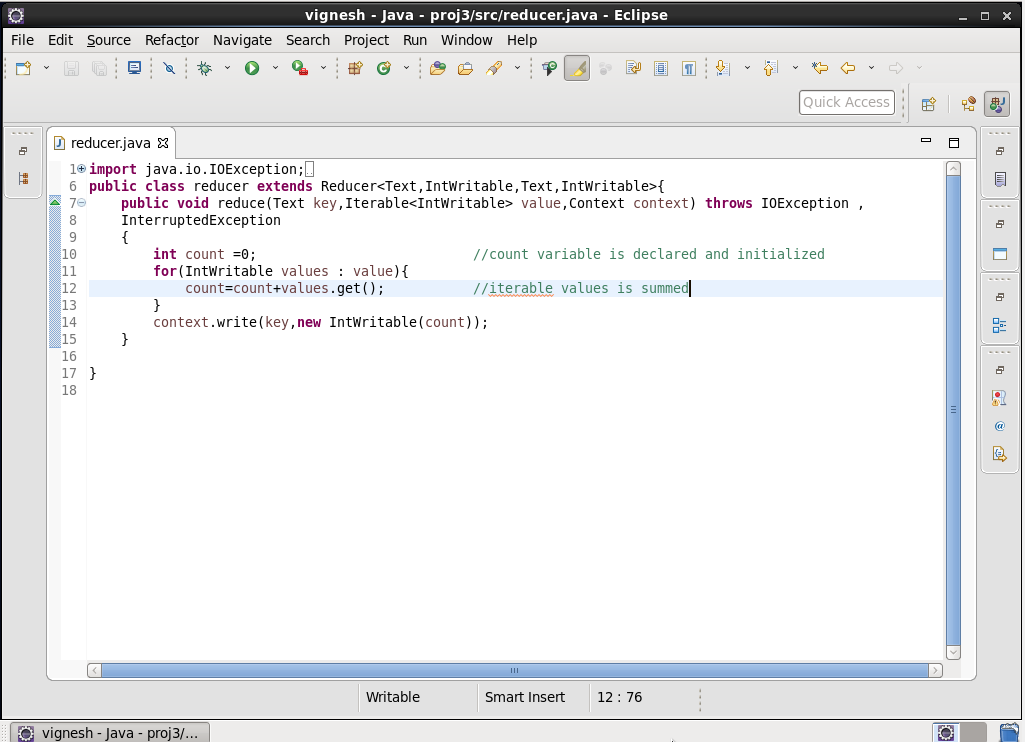
*LINE 9 : Text offset from beginning is set as key and the entire dataset is set as value.*

*LINE 12 : Separating the dataset with comma and storing it in array v.*

*LINE 15: Checking if FBI CODE v[14] is 32(Filtering the FBI CODE=32).*

*LINE 17: Setting FBI CODE as key and 1 as value.*

**REDUCER CLASS:**

****

**REDUCER LOGIC:**

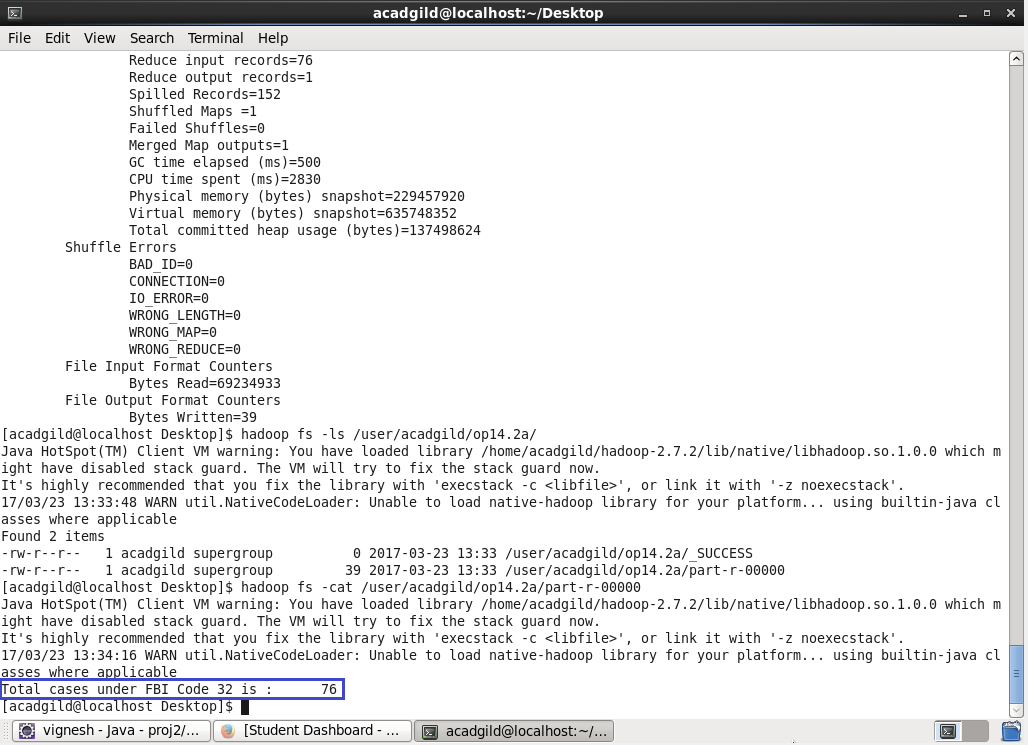
LINE 10: Initializing the count as 0.

LINE 11: Running an infinite for loop.

LINE 12: count will add the iterative values.

LINE 14: Finally the same key is used as the key and the count is set as value.

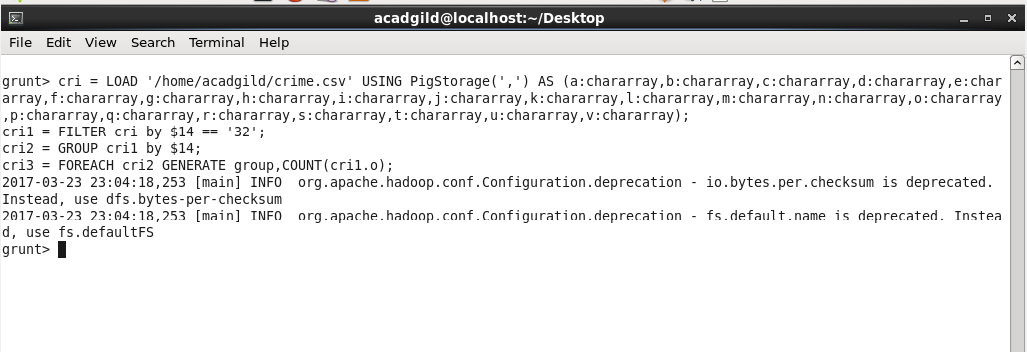
**OUTPUT:**

****

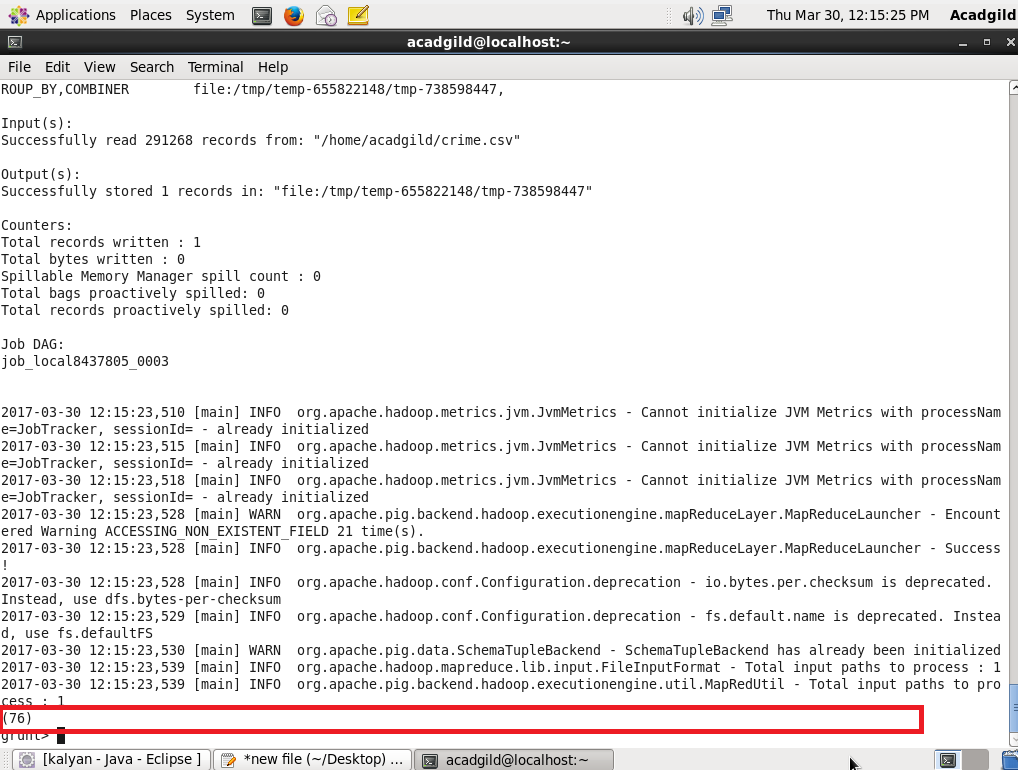
**PIG PROGRAM :**

**LOGIC:**

* *Loading the Dataset into grunt shell using LOAD command.*
* *Filtering the Dataset by FBICODE=32 using FILTER command.*
* *Grouping by FBI code (column 15) using GROUP BY command.*
* *Generating the count using COUNT command.*

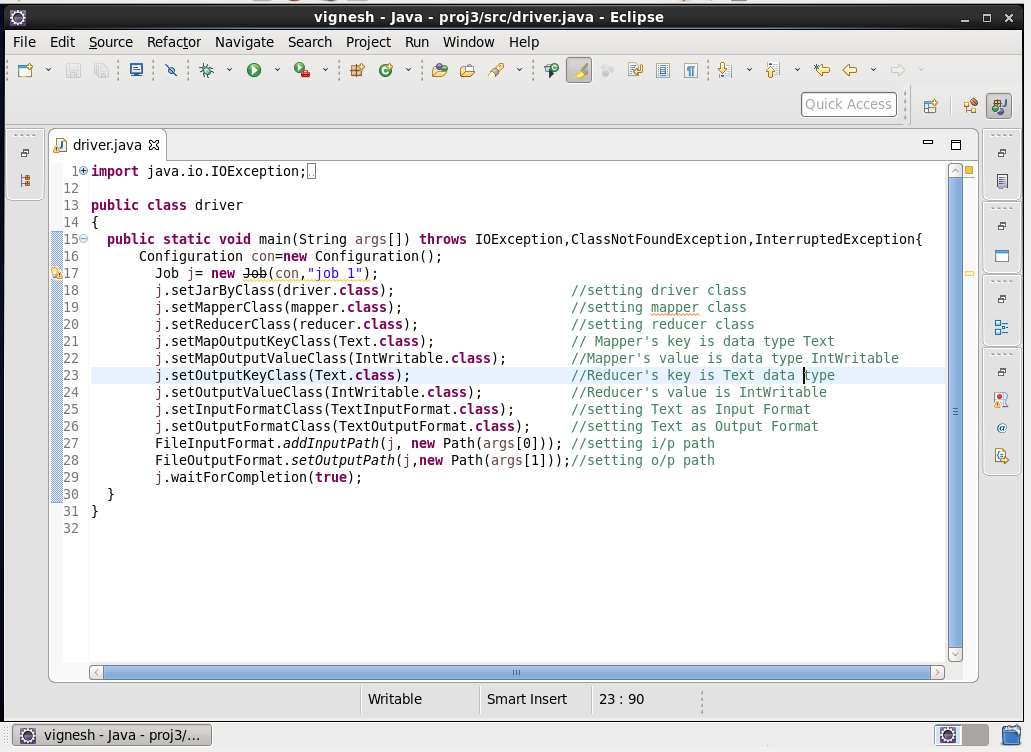
**PIG COMMANDS:**

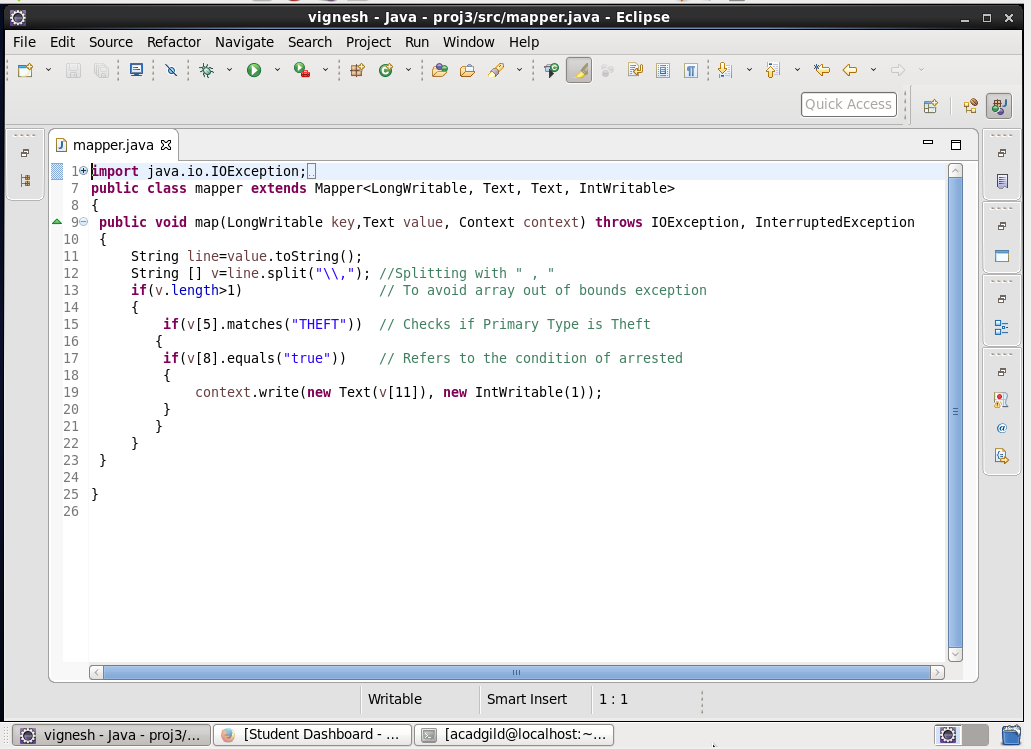
**PIG OUTPUT:**

****

**3) WRITE A MAPREDUCE AND PIG PROGRAM TO CALCULATE THE NUMBER OF ARRESTS IN THEFT DISTRICT WISE.**

**DRIVER CLASS:**

****

**MAPPER CLASS:**

**MAPPER LOGIC:**

*LINE 11: Converting the value to string.*

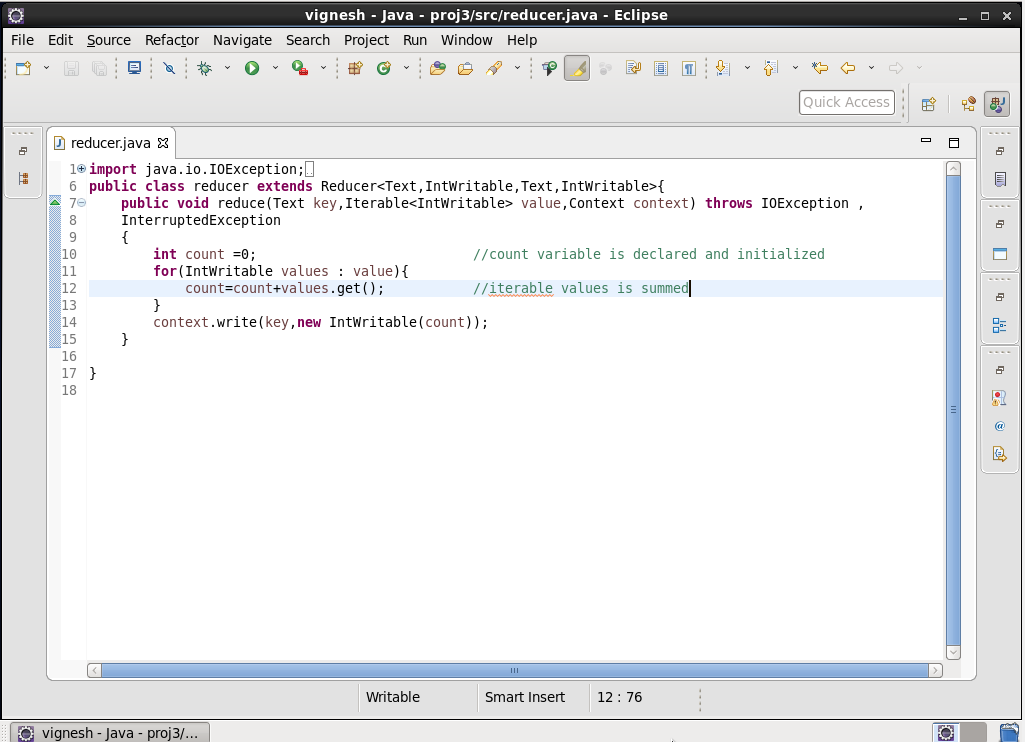
*LINE 12: Splitting the string using comma and storing it in array v.*

*LINE 15 : First Checking whether the primary type is THEFT.*

*LINE 17: Then checking whether arrest is done.*

*LINE 19 : Setting District as key and 1 as value.*

**REDUCER CLASS:**

****

**REDUCER LOGIC:**

*LINE 10: Initializing the count as 0.*

*LINE 11: Running an infinite for loop.*

*LINE 12: Count will add the iterative values.*

*LINE 14: Finally the same key is used as the key and the count is set as value.*

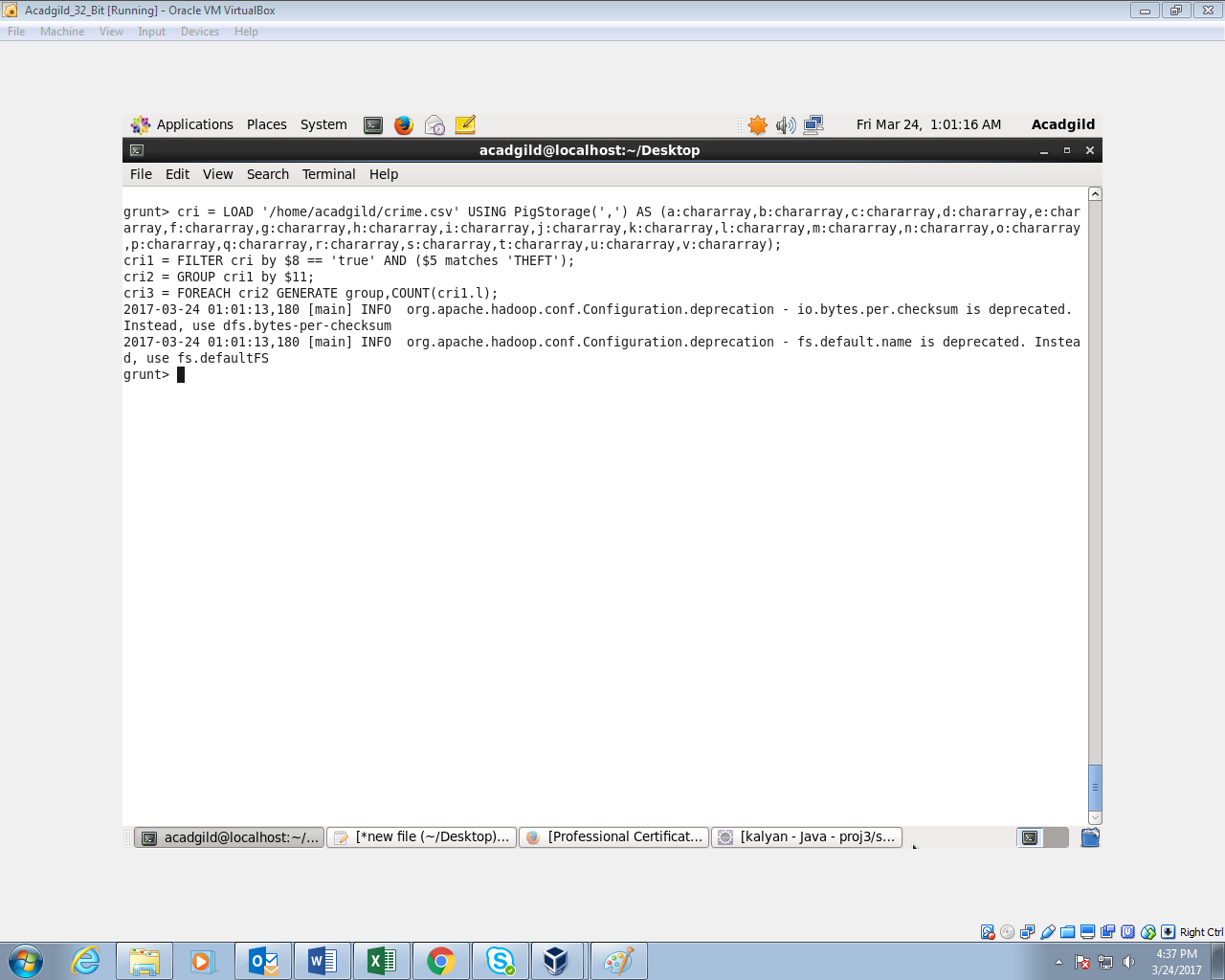
**OUTPUT:**

****

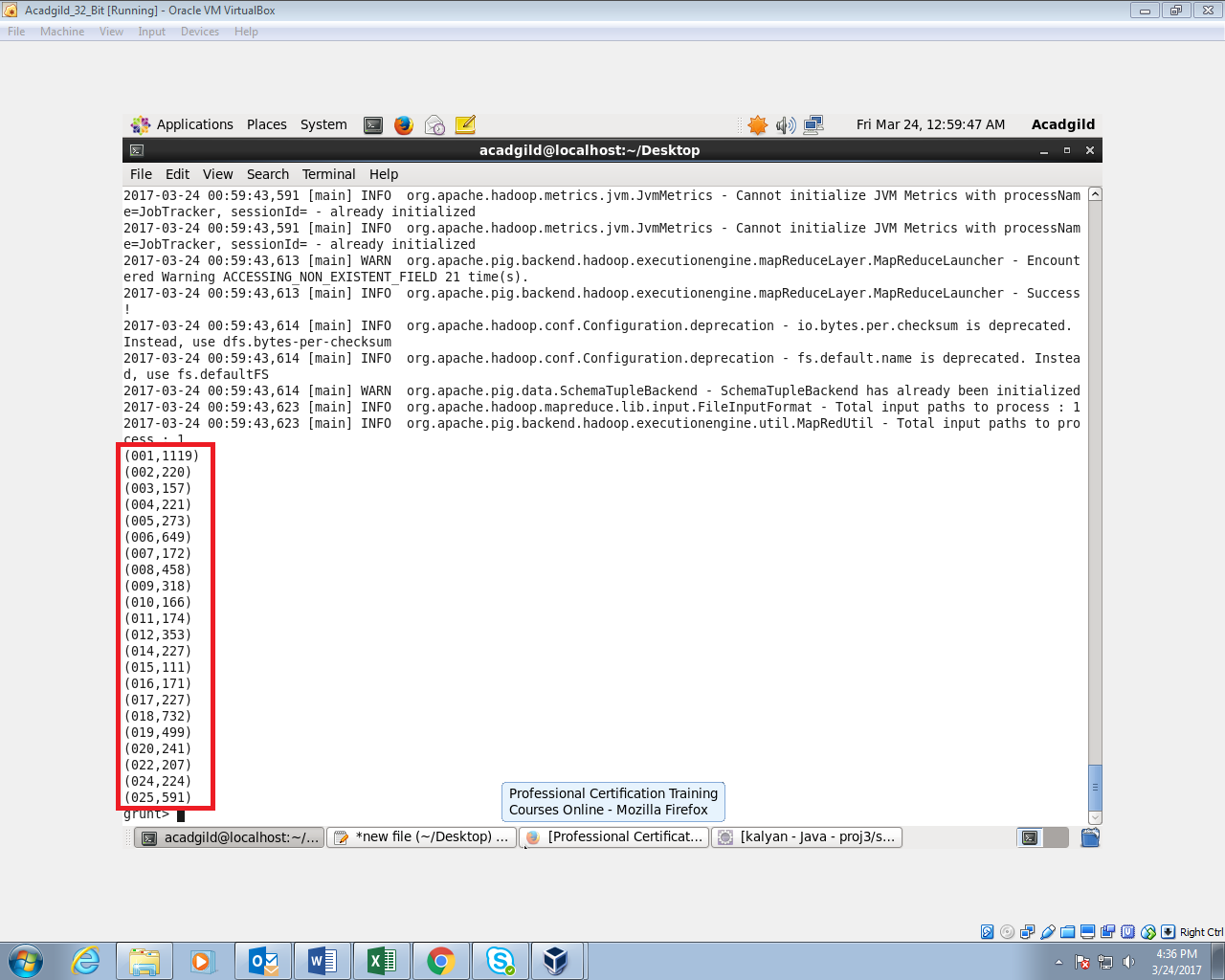
**PIG PROGRAM :**

**LOGIC :**

* *Loading the Dataset into the Grunt shell using LOAD command.*
* *Filter the Dataset by checking whether column 9 is true and column 6 is theft.*
* *Grouping by District ID using GROUP BY command.*
* *Generating the count using COUNT command.*

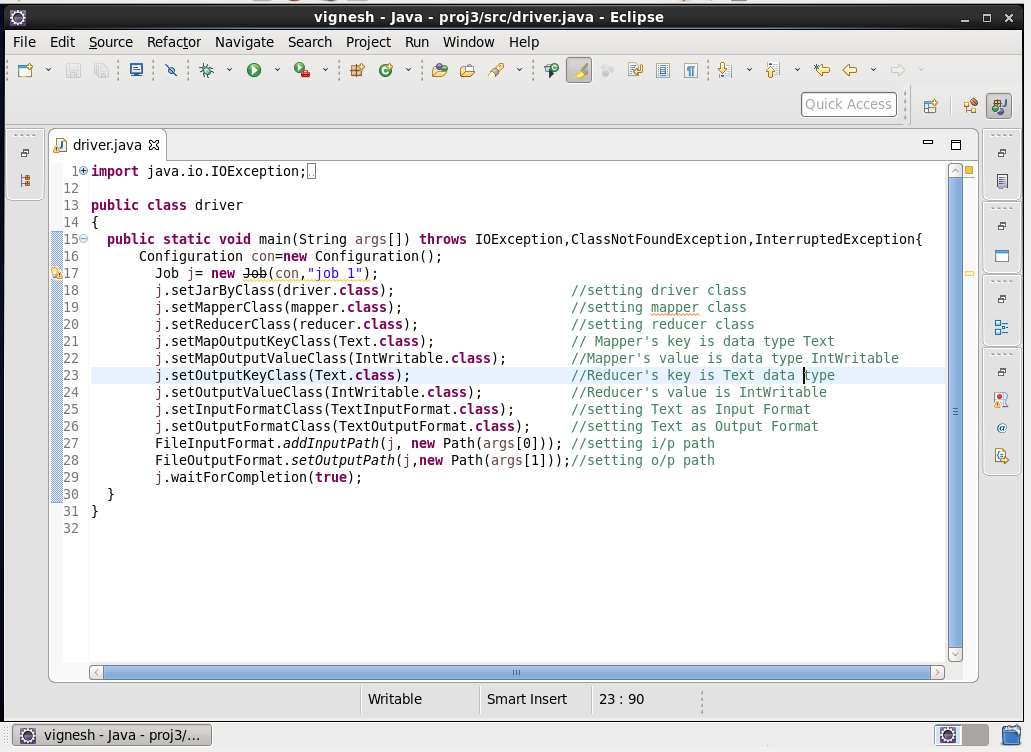
**PIG COMMAND:**

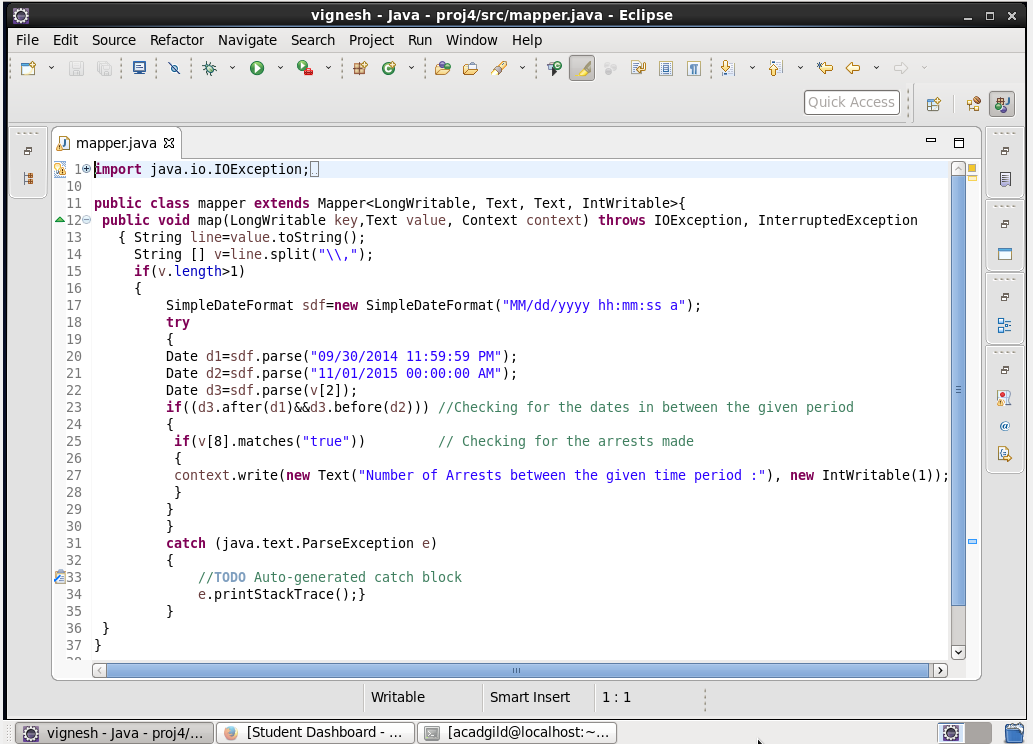
**PIG OUTPUT:**

****

**4) WRITE A MAPREDUCE AND PIG PROGRAM TO CALCULATE THE NUMBER OF ARRESTS DONE BETWEEN OCTOBER 2014 AND OCTOBER 2015.**

**DRIVER CLASS:**

****

**MAPPER CLASS:**

**MAPPER LOGIC:**

*LINE 13: Converting the value to string and storing it in string line.*

*LINE 14: Splitting the dataset with comma and storing it in array v.*

*LINE 17 : Using simple date format we are specifying the date format.*

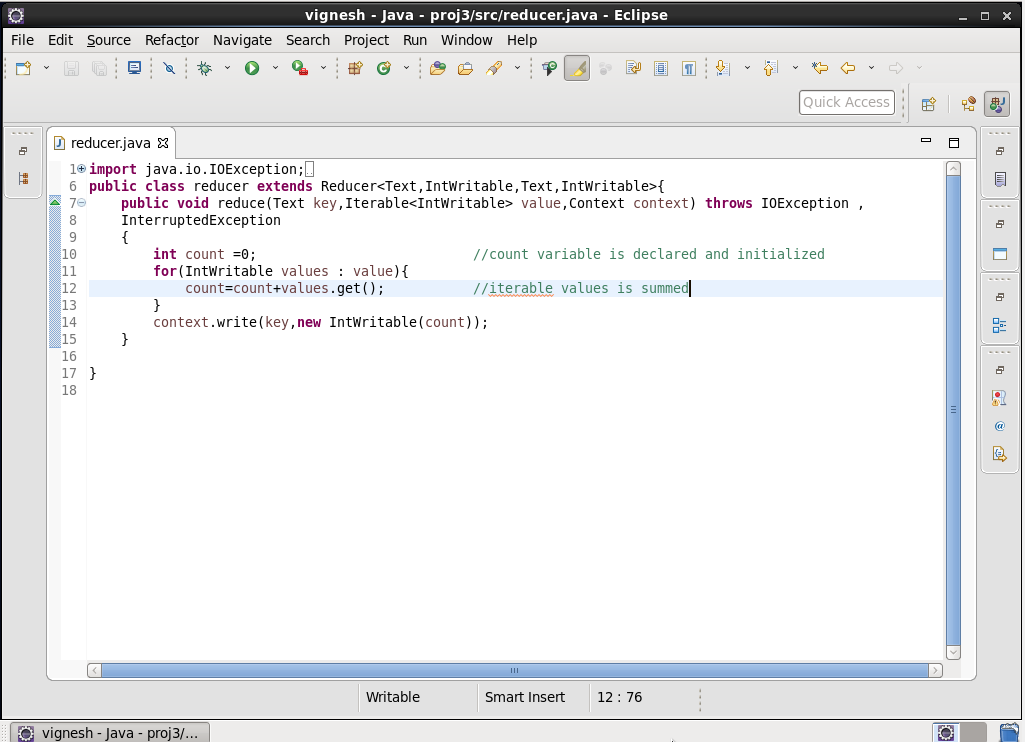
*LINE 20 AND 21: Using parse we are converting the string to date d1 and d2.*

*LINE 22 :**Parsing v[2] by converting it to date d3.*

*LINE 23 : Checking the condition (from 1st October 2014 to 31st October 2015 ).*

*LINE 25: Checking whether arrest is TRUE (v[8] is true).*

**REDUCER CLASS:**

****

**REDUCER LOGIC:**

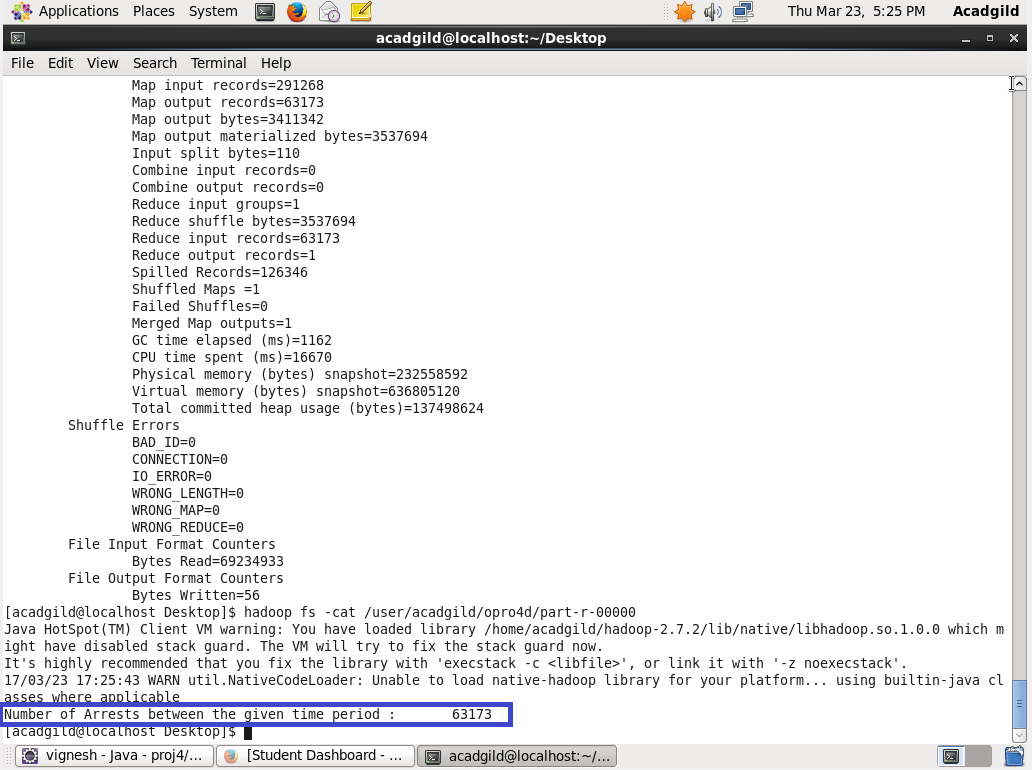
*LINE 10: Initializing the count as 0.*

*LINE 11: Running an infinite for loop.*

*LINE 12: count will add the iterative values.*

*LINE 14: Finally the same key is used as the key and the count is set as value.*

**OUTPUT:**

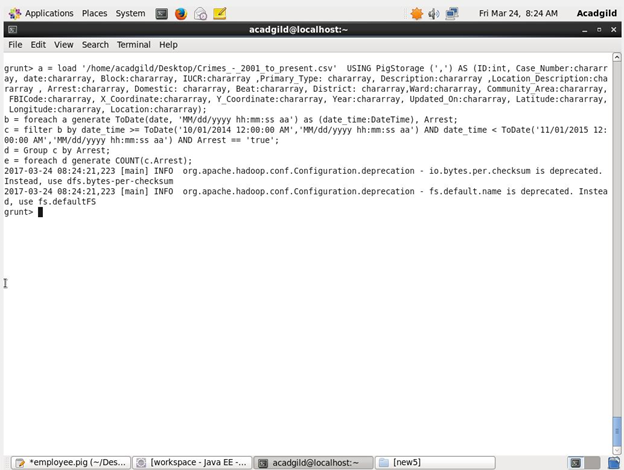
****

**PIG PROGRAM :**

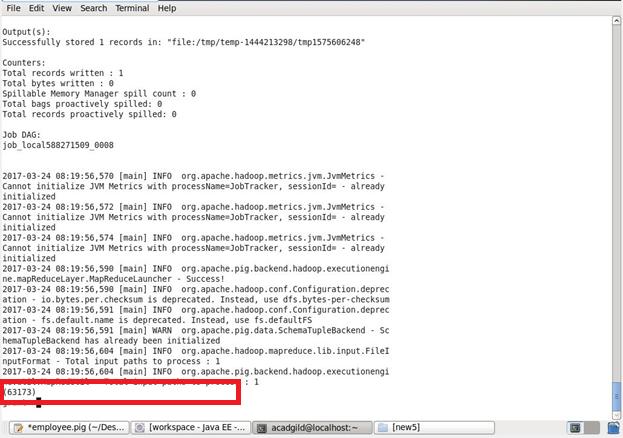
**LOGIC:**

1. *Loading the Dataset into the Grunt shell using LOAD command.*
2. *Converting the date column to date format using ToDate command.*
3. *Filtering the dataset by specifying the period (from 1st of October 2014 to 31st of October 2015) and checking whether arrest is done.*
4. *Counting the filtered dataset by count command.*

**PIG COMMAND:**

****

**PIG OUTPUT:**

****