**Project 1 - USA Crime Analysis**

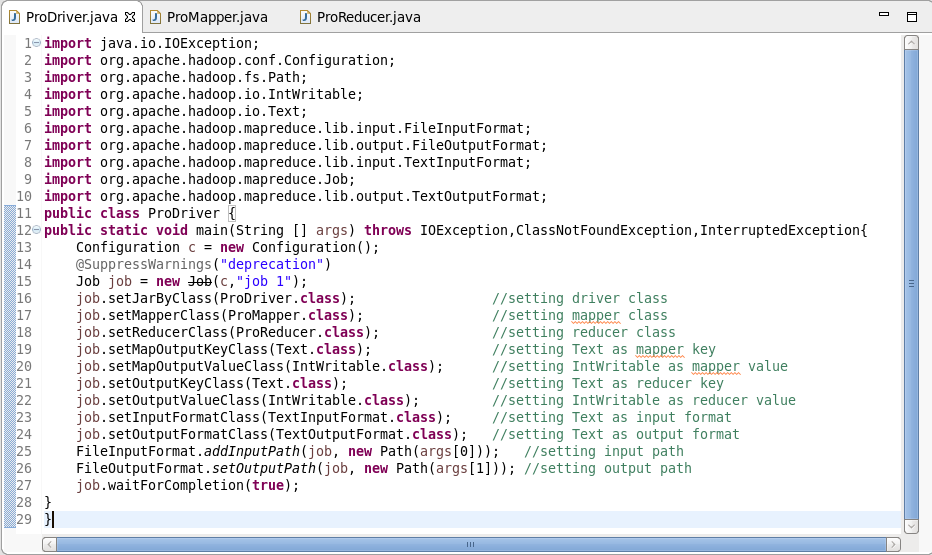
**Dataset Description:** ID, Case Number, Date, Block, IUCR, Primary Type, Description, Location Description, Arrest, Domestic, Beat, District, Ward, Community Area, FBICode, X Coordinate, Y Coordinate, Year, Updated On, Latitude, Longitude, Location.

**Part 1:**

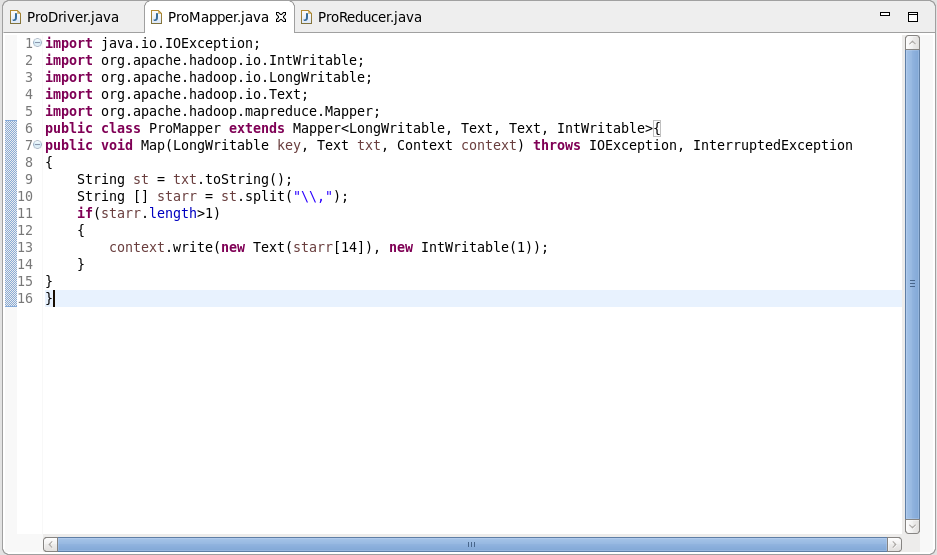
Write a MapReduce and pig program to calculate the number of cases investigated under each FBI code.

**MapReduce program:**

**Driver Class:**



**Mapper Class:**



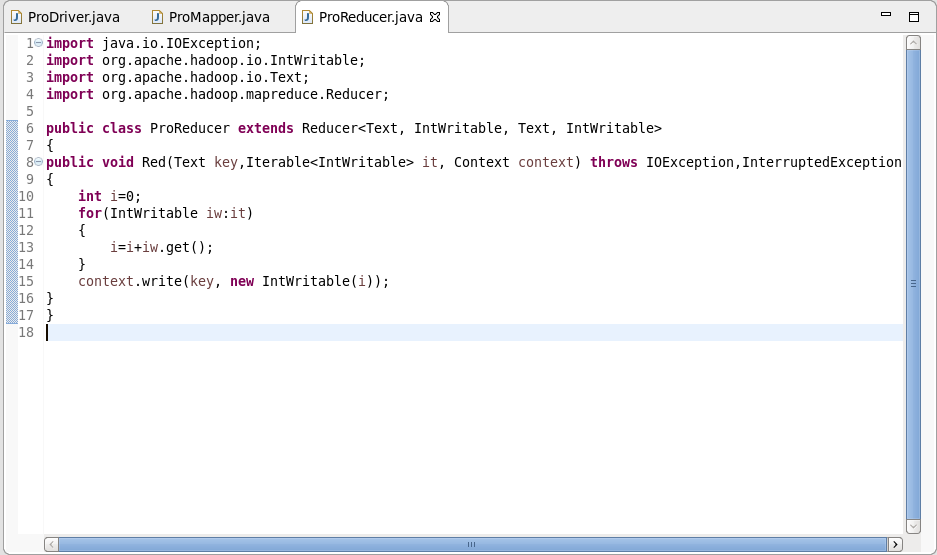
**Mapper Logic:**

LINE 7: Text offset from beginning is set as key and the entire dataset as value.

LINE 10: Separating the dataset with comma and storing the value in an array starr.

LINE 13: Setting FBI CODE as key and 1 as value in the context to be further processed by Reducer.

**Reducer Class:**



**Reducer Logic:**

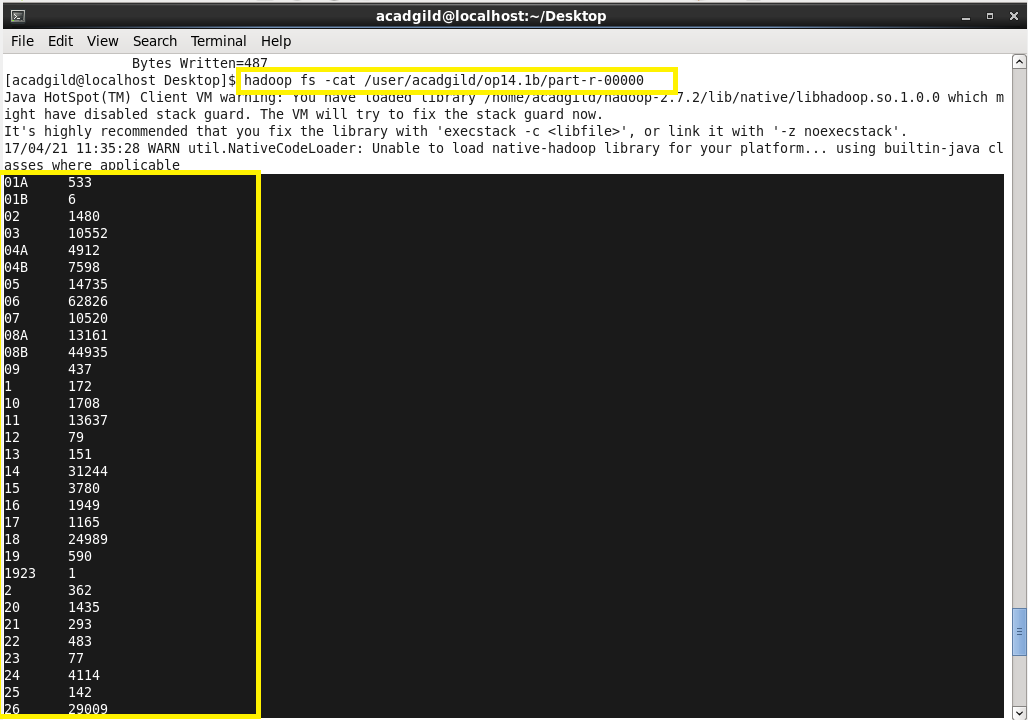
LINE 10: Initializing the integer i as 0.

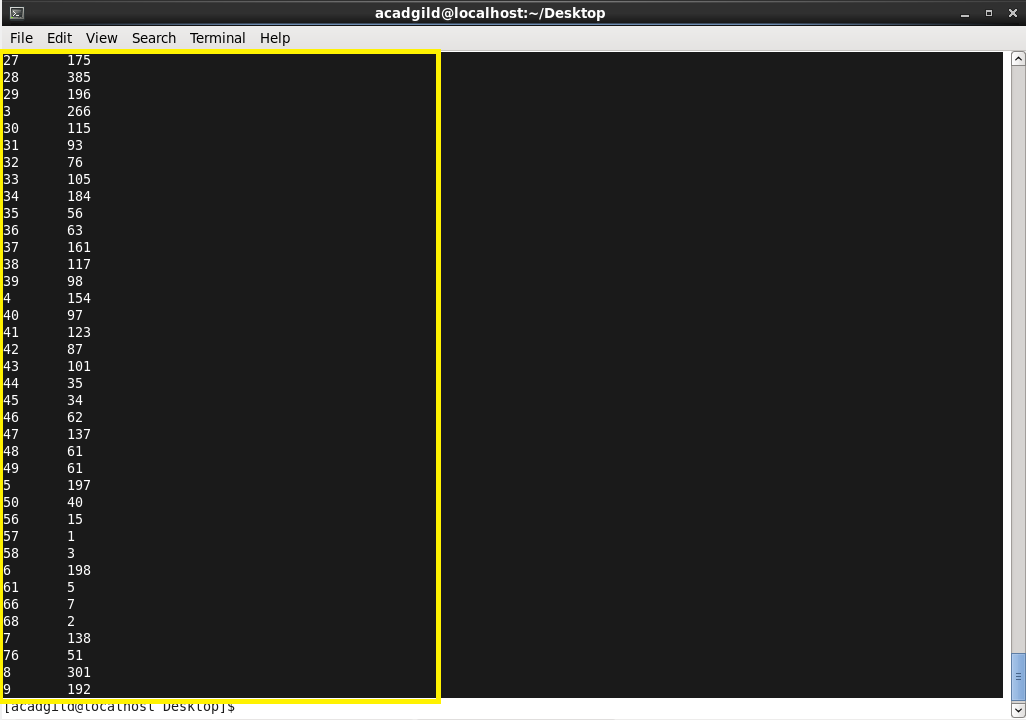
LINE 11: Running an infinite ‘for loop’.

LINE 13: i will add the iterative values.

LINE 15: Finally the same key is used as the key and the integer i is set as value.

**Outputs:**





**Pig Program:**

**Pig Command lines:**



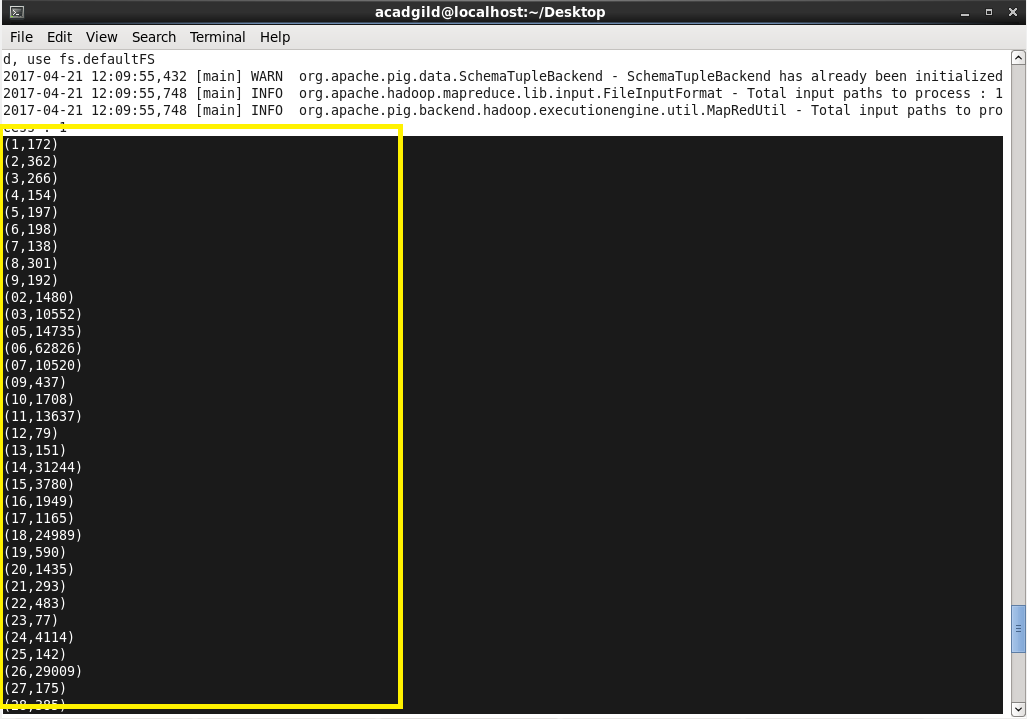
**Pig Command Logic:**

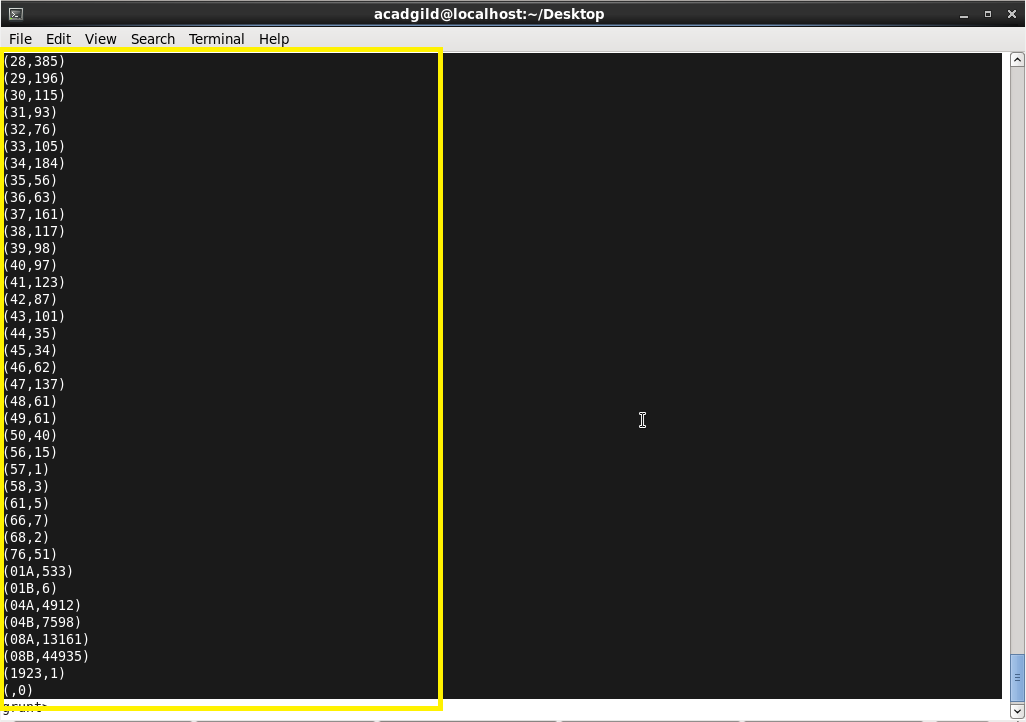
LINE 1: Loading the Dataset into the grunt shell using LOAD command.

LINE 2: Grouping the dataset by FBI code (column 15) using GROUP BY command.

LINE 3: Generating count of cases for each FBI CODE is done by using COUNT command.

**Outputs:**



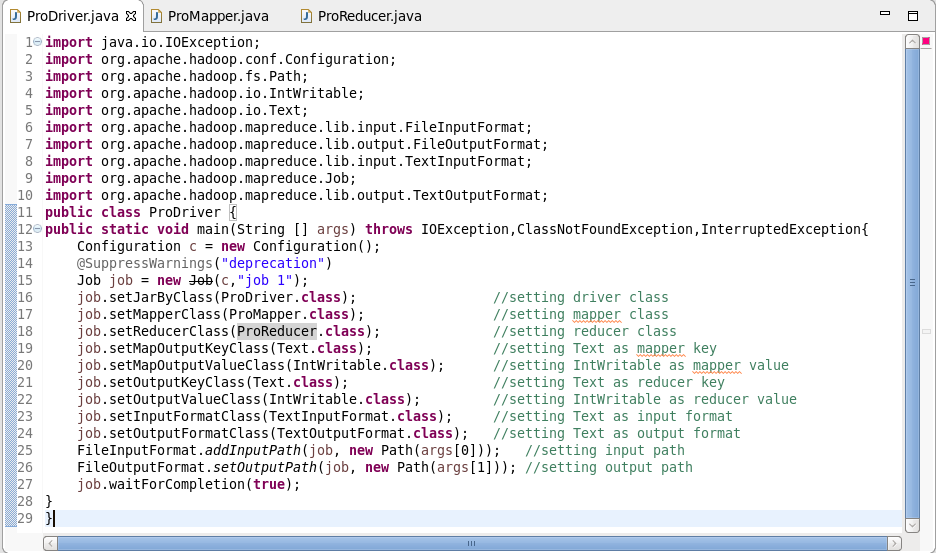


**Part 2:**

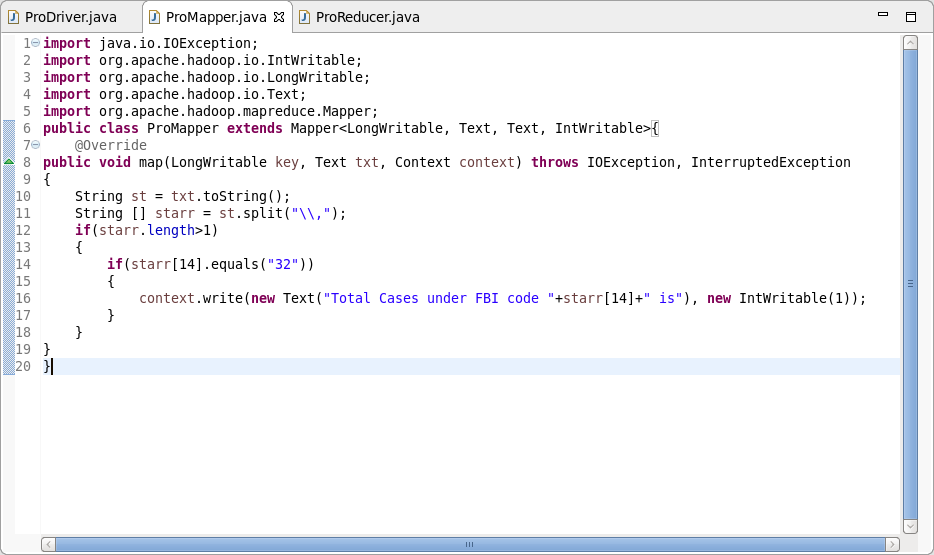
Write a MapReduce and pig program to calculate the number of cases investigated under FBI code 32.

**MapReduce Program:**

**Driver Class:**



**Mapper Class:**



**Mapper Logic:**

LINE 8: Text offset from beginning is set as key and the entire dataset is set as value.

LINE 11: Separating the dataset with comma and storing it in array starr.

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

LINE 18: Setting FBI CODE as key and 1 as value before sending it to Reducer.

**Reducer Class:**

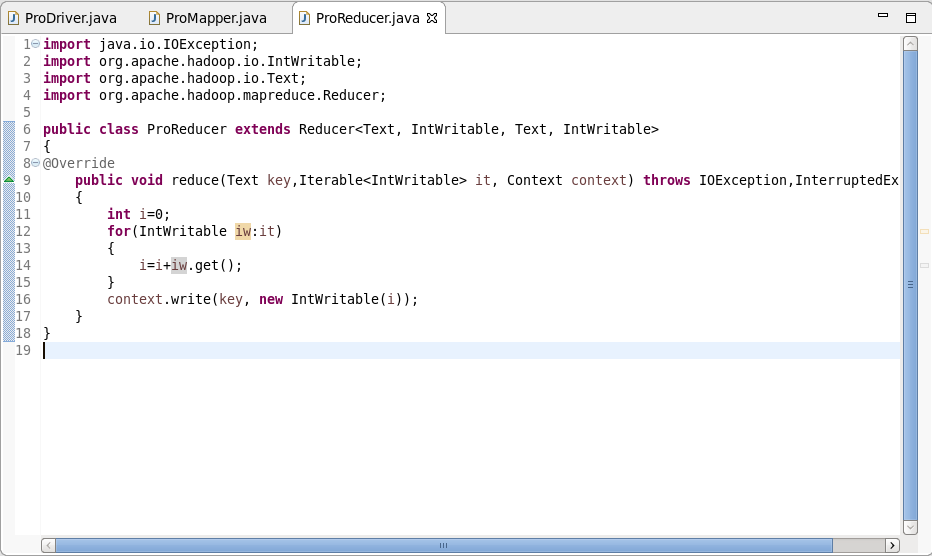
**Reducer Logic:**

LINE 11: Initializing the integer ‘i’ as 0.

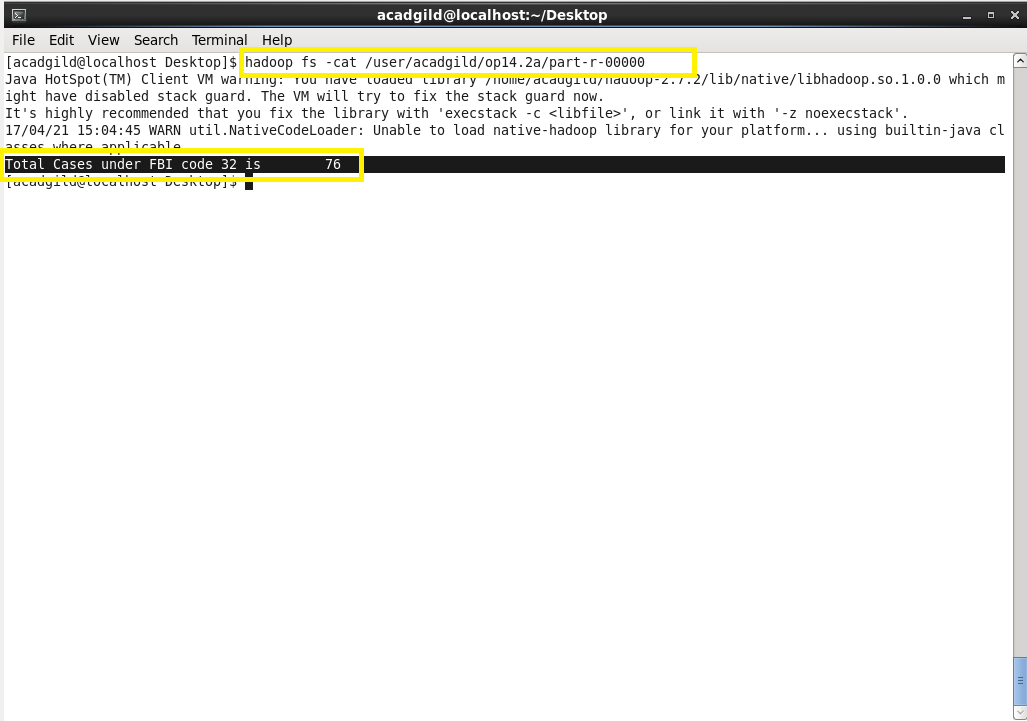
LINE 12: Running an infinite ‘for loop’

LINE 14: Integer ‘i’ will add the iterative values

LINE 16: Finally the same key is used as the key and the ‘i’ is set as value.



**Output:**



**Pig Program:**

**Pig Command lines:**



**Pig Command Logic:**

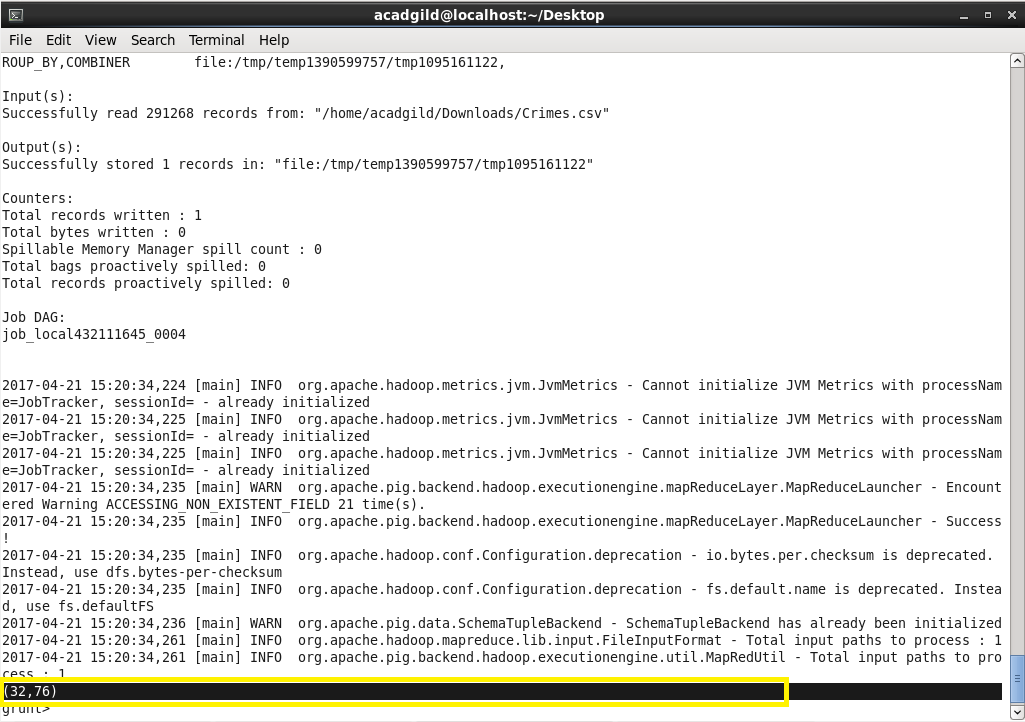
Line 1: Loading the dataset into grunt shell using LOAD command.

Line 2: Filtering the dataset by FBICODE=32 using FILTER command.

Line 3: Grouping the dataset by FBI Code (column 15) using GROUP BY command.

Line 4: Generating the count using COUNT command.

**Output:**

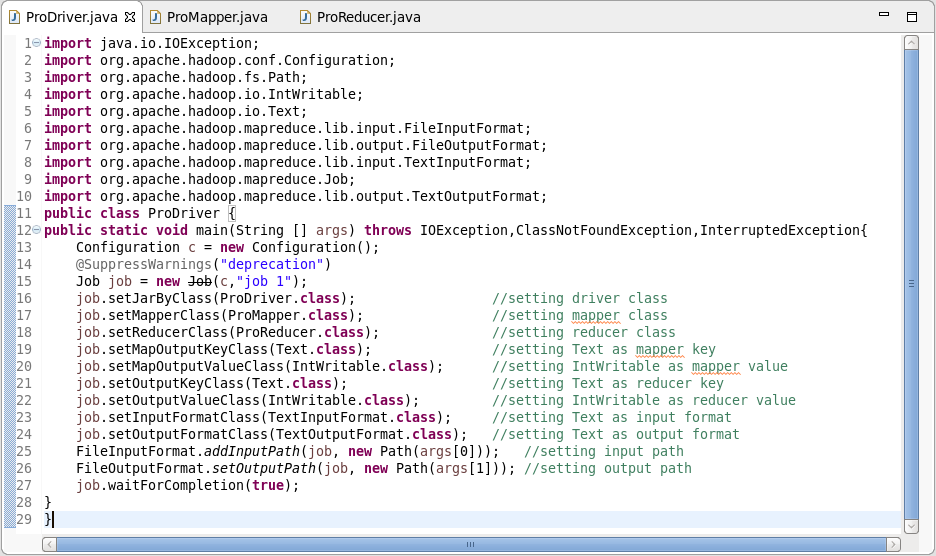


**Part 3:**

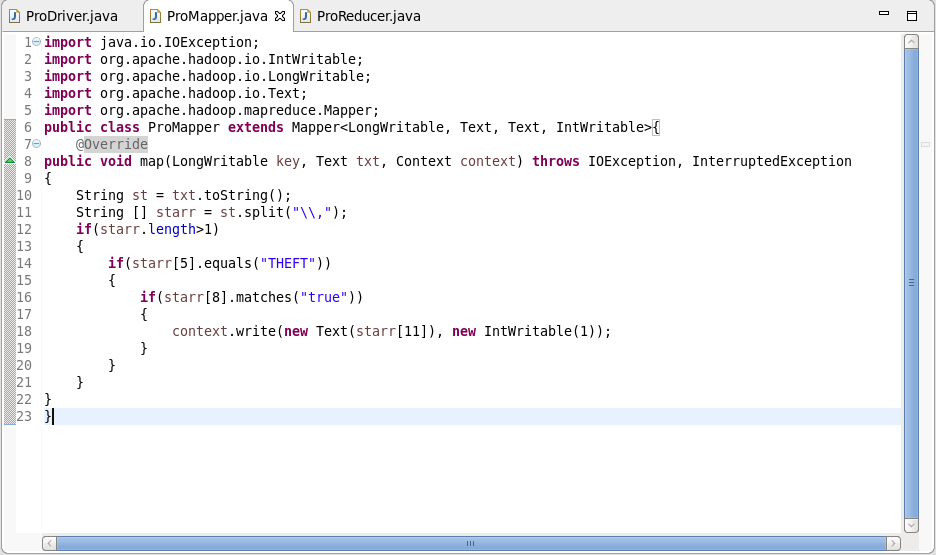
Write a MapReduce and pig program to calculate number of arrests in theft district wise.

**MapReduce Program:**

**Driver Class:**



**Mapper Class:**



**Mapper Logic:**

LINE 8: Text offset from beginning is set as key and the entire dataset is set as value.

LINE 10: Converting the text to string.

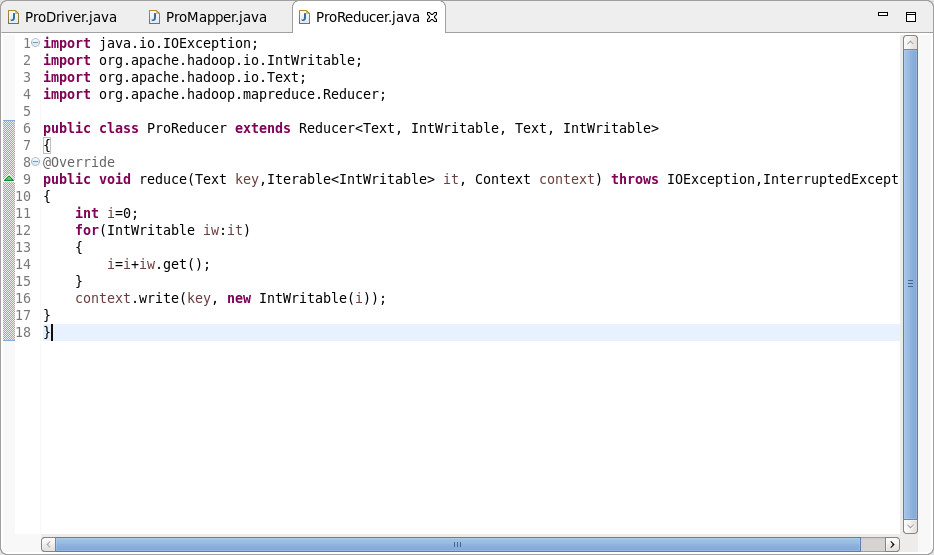
LINE 11: Splitting the string using comma and storing it in array starr.

LINE 14: Checking whether the primary type is ‘THEFT’.

LINE 16: Checking whether arrest is ‘true’.

LINE 18: Setting District as key and 1 as value.

**Reducer Class:**



**Reducer Logic:**

LINE 11: Initializing the integer ‘i’ as 0.

LINE 12: Running an infinite ‘for loop’.

LINE 14: The integer ‘i’ will add the iterative values.

LINE 16: Finally the same key is used as the key and the ‘i’ is set as value.

**Output:**



**Pig Program:**

**Pig Command lines:**



**Pig Command Logic:**

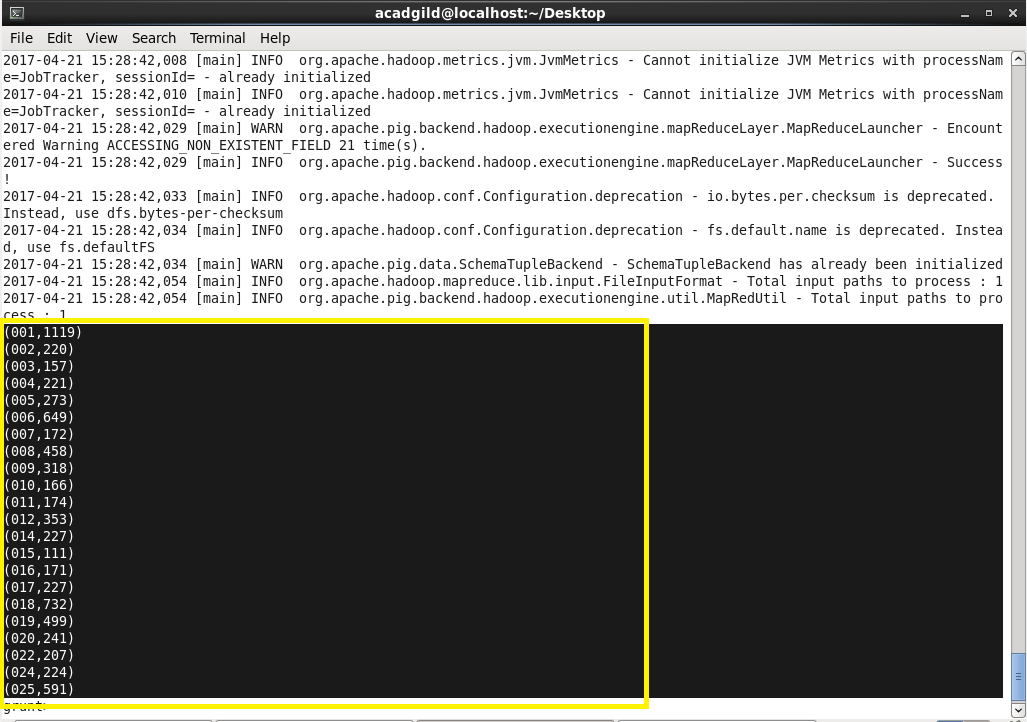
LINE 1: Loading the dataset into the Grunt shell using LOAD command.

LINE 2: Filtering the dataset by checking whether column 9 is ‘true’ and column 6 is ‘THEFT’.

LINE 3: Grouping the dataset by District ID using GROUP BY command.

LINE 4: Generating the count using COUNT command.

**Output:**

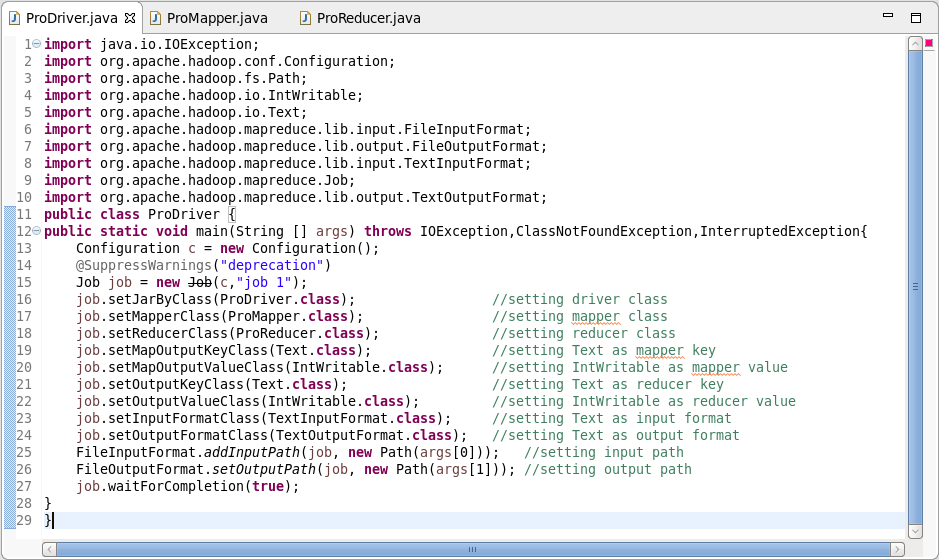


**Part 4:**

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

**MapReduce Program:**

**Driver Class:**



**Mapper Class:**

**Mapper Logic:**

LINE 13: converting the value to string and storing it in String st.

LINE 14: splitting the dataset in the String st with comma and storing it in an array starr.

LINE 16: Using Simple Date Format the date format is specified.

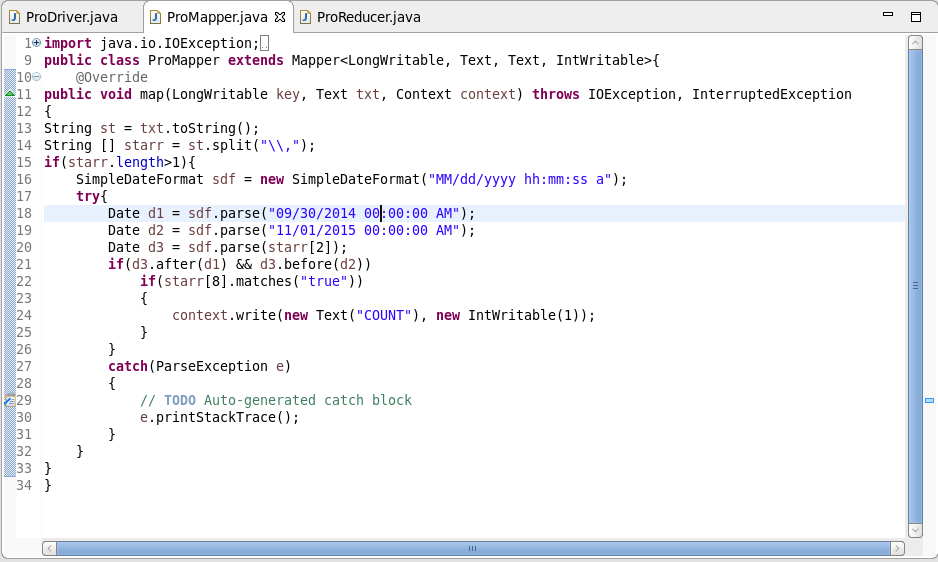
LINE 17: try..catch block is used for any exception in date declaration.

LINE 18: AND 19: Using parse we are declaring the string to date d1 and d2.

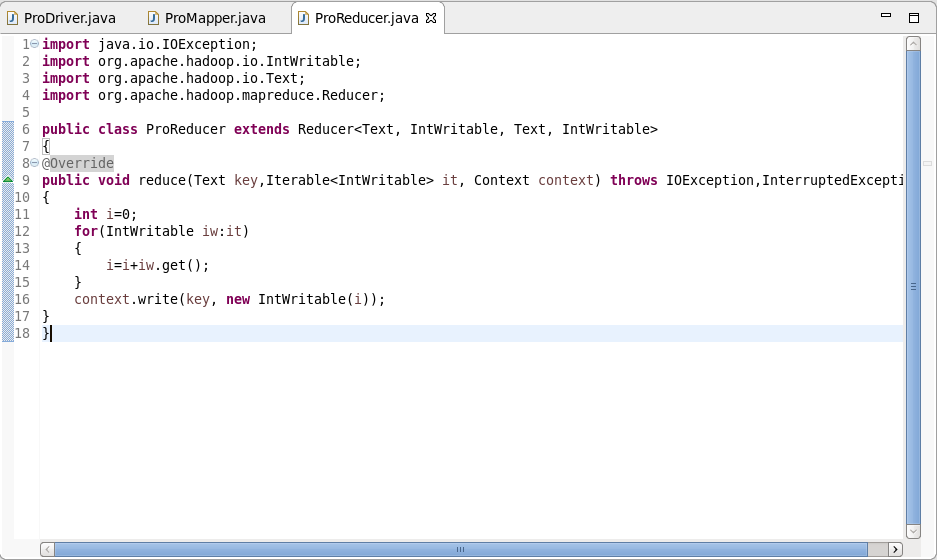
LINE 20: Then string starr[2] is stored into date d3 by parsing.

LINE 21: Checking the condition (from 1st October 2014 to 31st October 2015).

LINE 24: Checking if arrest is ‘true’ (starr[8] is ‘true’).



**Reducer Class:**



**Reducer Logic:**

LINE 11: Initializing the integer ‘i’ as 0.

LINE 12: Running an infinite ‘for loop’.

LINE 14: The integer ‘i’ will add the iterative values.

LINE 16: Finally the same key is used as the key and the ‘i’ is set as value.

**Output:**



**Pig Program:**

**Pig Command lines:**



**Pig Command Logic:**

Line 1: Loading the dataset into the Grunt shell using LOAD command.

Line 2: Converting the date column in the dataset to date format using ToDate command.

Line 3: Filtering the dataset by specifying the period (from 1st of October 2014 to 31st of October 2015) and checking whether arrest is ‘true’.

Line 4: Grouping the resultant dataset by arrest column.

Line 5: Counting the grouped dataset by COUNT command.

**Output:**

