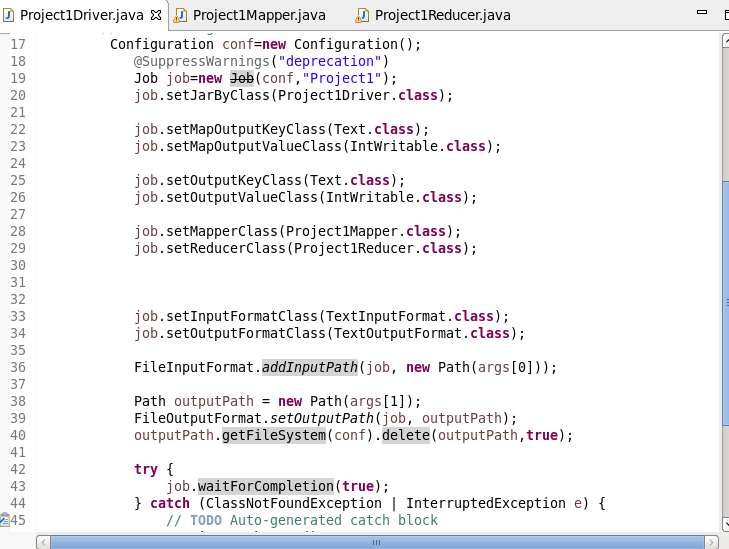
**PROJECT 1 TASK-1**

**Sample dataset:**

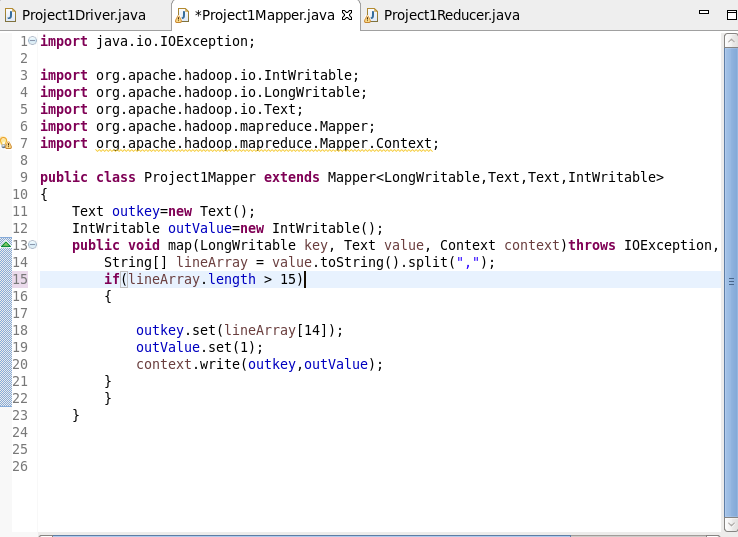


* Write a mapreduce program to calculate the number of cases investigated under each FBI code
* **Driver class:**

****

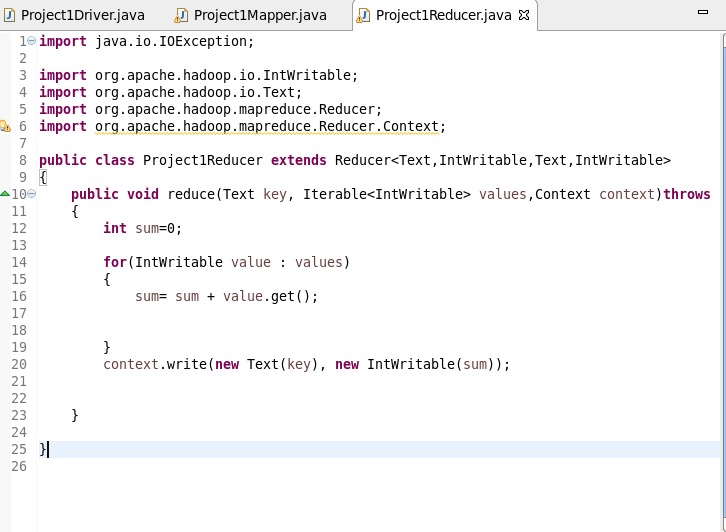
* **Mapper class:**

Sending fbi code as key and 1 as output

****

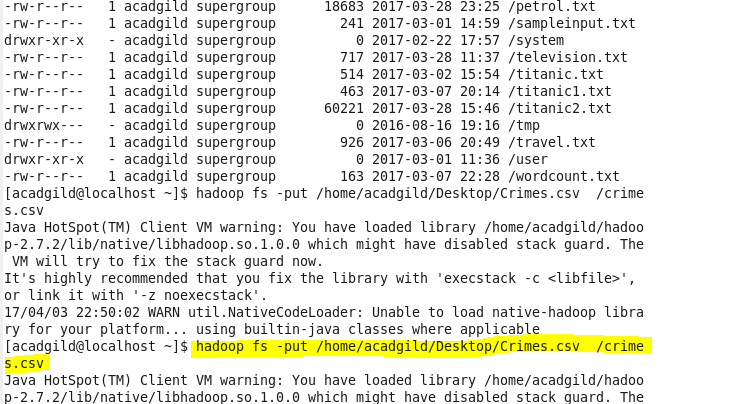
* **Reducer class:**

Counting no of cases for each fbi code

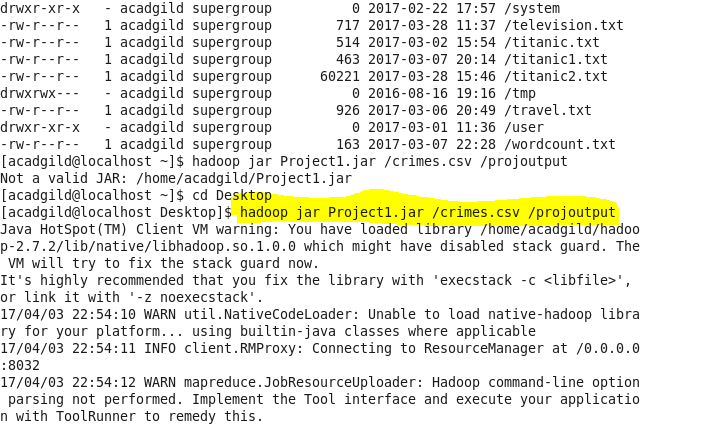


**Output:**

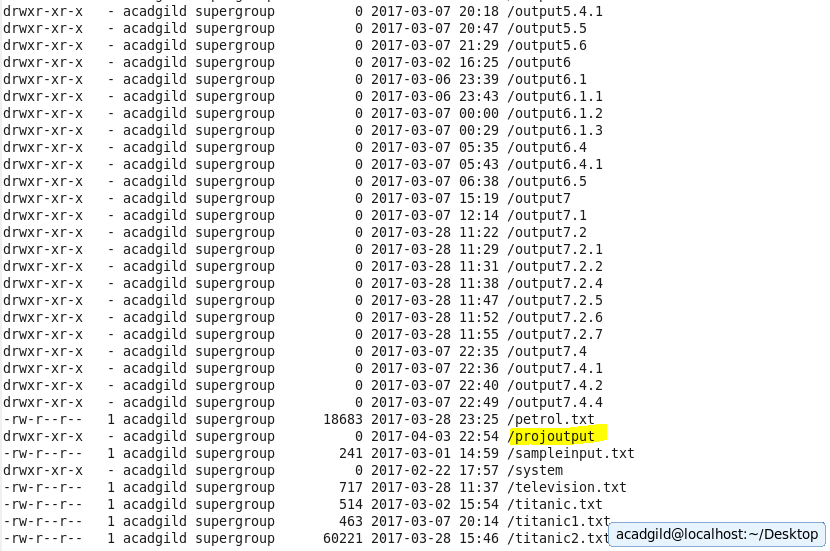
Copying file to hdfs

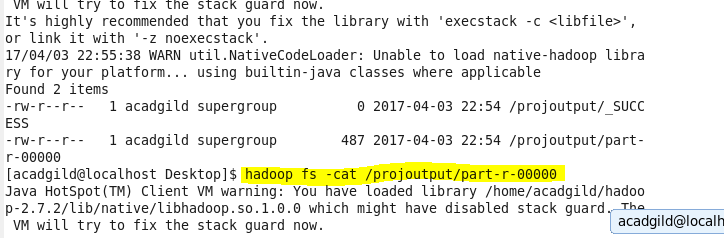


Running jar file:

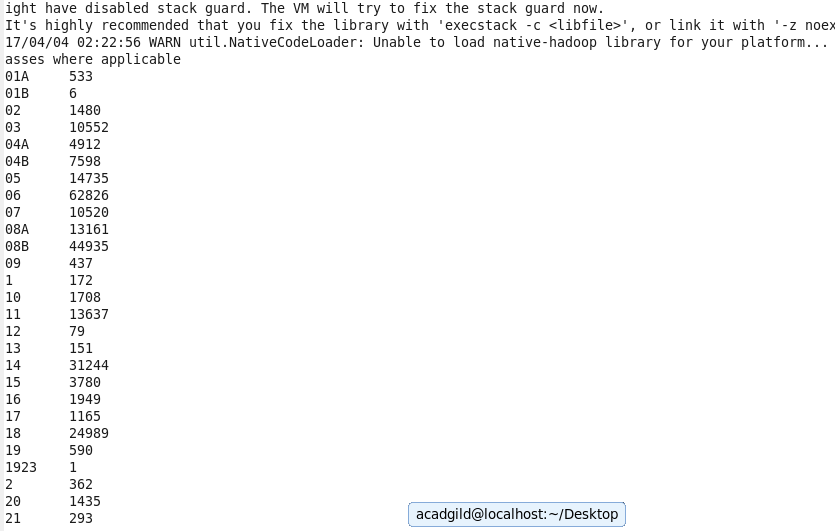


Output file:



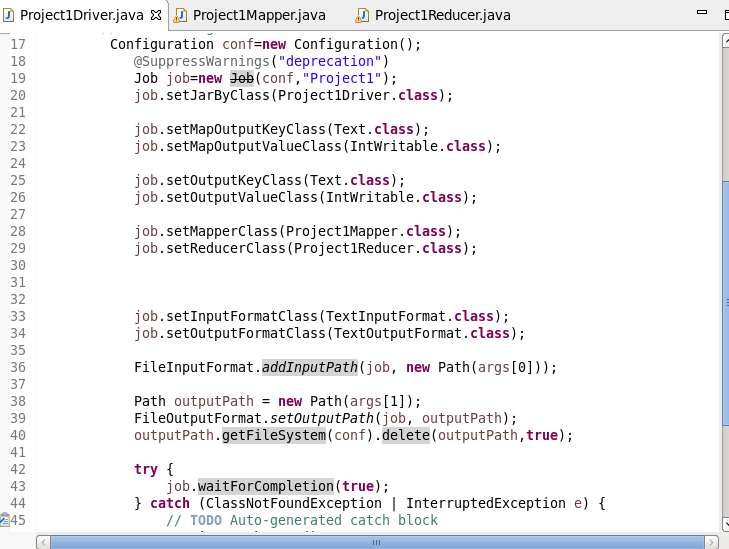


Final result

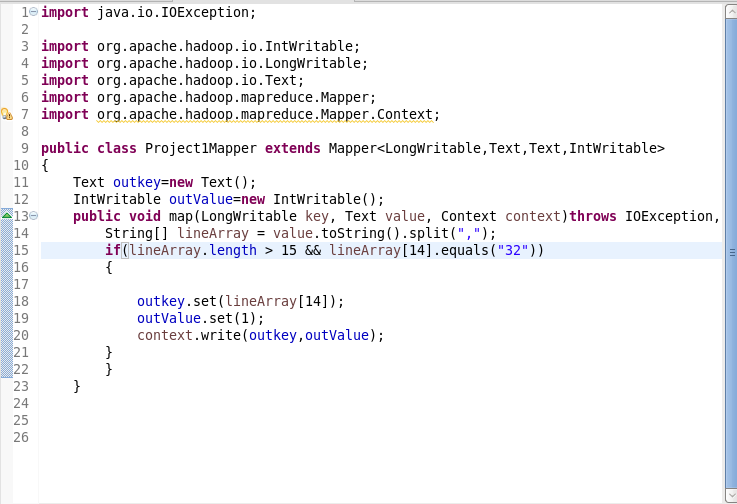


* **Write a mapreduce program to calculate the number of cases investigated under FBI code 32**
* **Driver class**

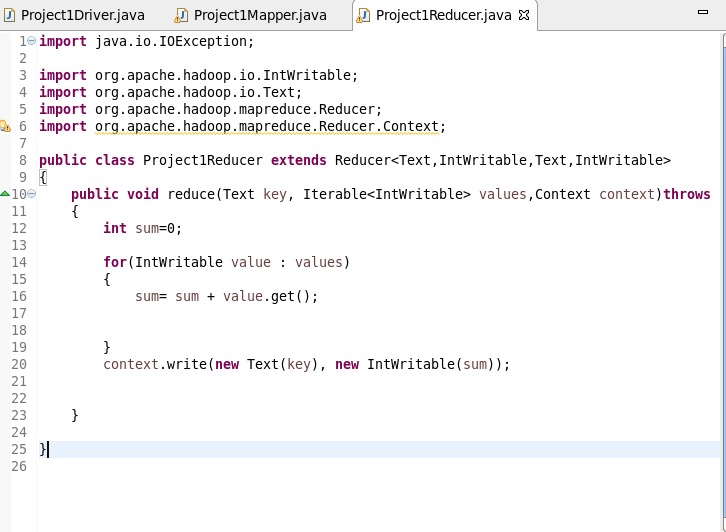
Set the configuration

****

* **Mapper class:**

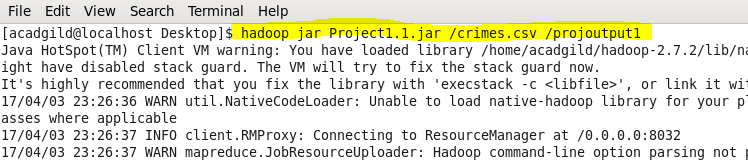
****

* **Reducer class:**

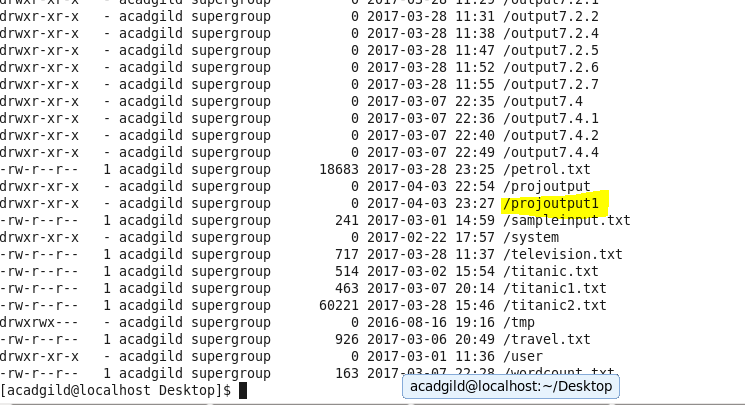
****

**Output:**

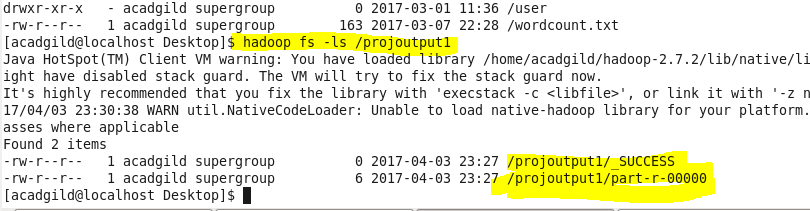
**Running jar file**

****

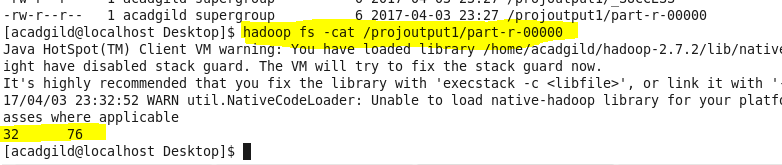
**Output directory created:**

****

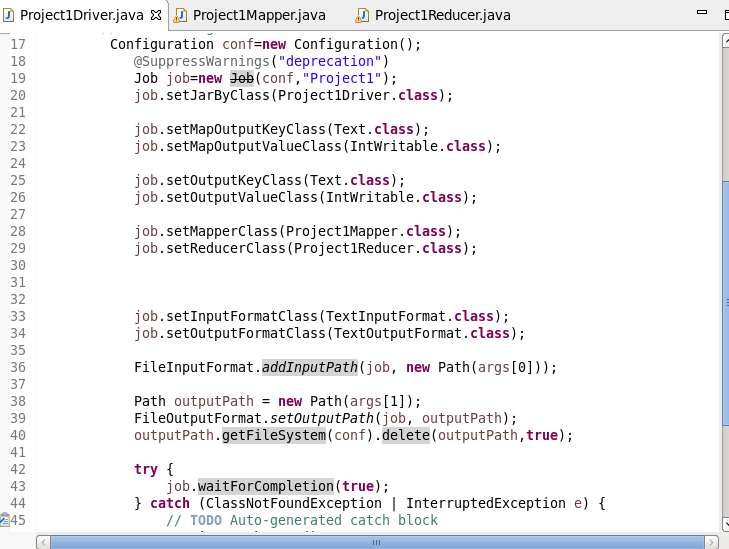
**Listing files in output directory**

****

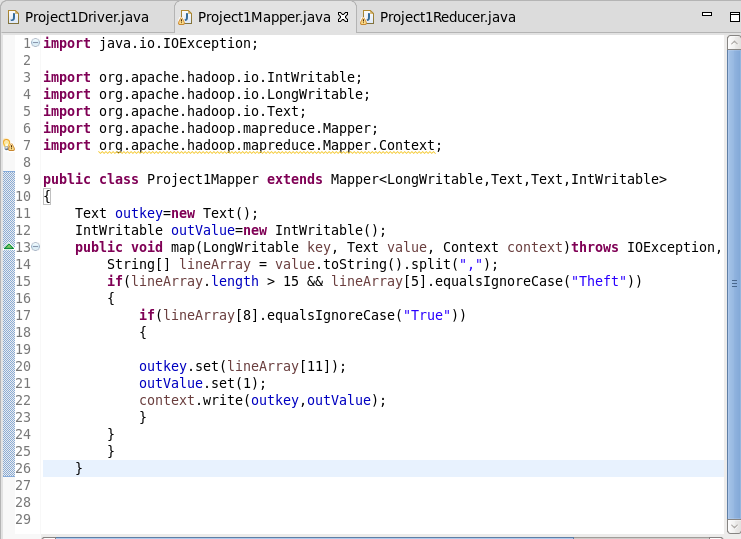
**Final output**

****

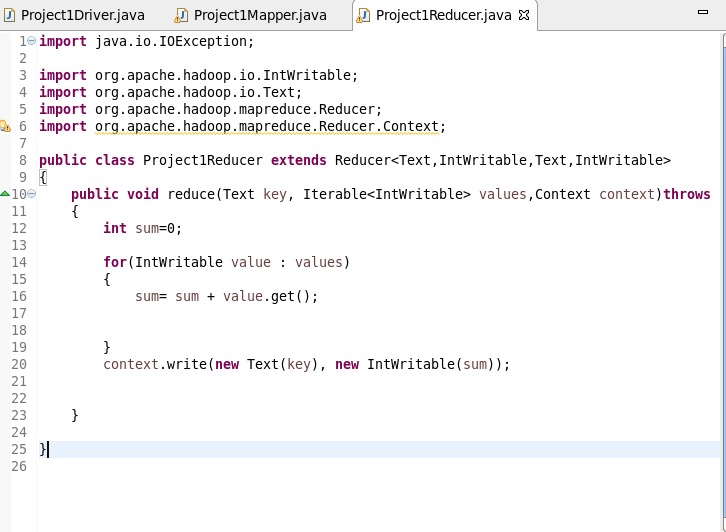
* **Write a mapreduce program to calculate the number of arrests in theft district wise**.
* **Driver class:**

****

* **Mapper class:**

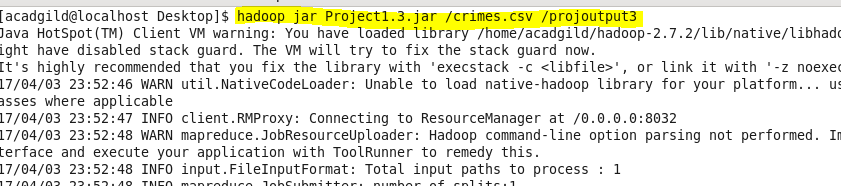
****

* **Reducer class:**

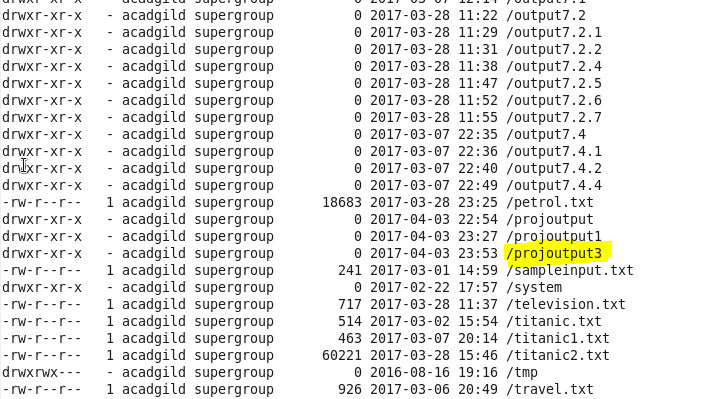
****

**Output:**

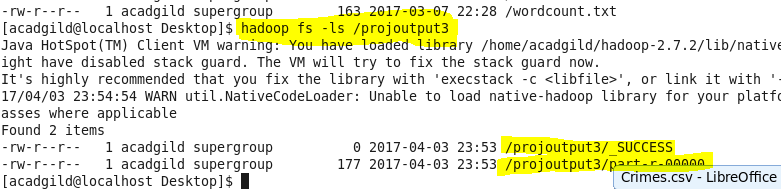
**Running jar file**

****

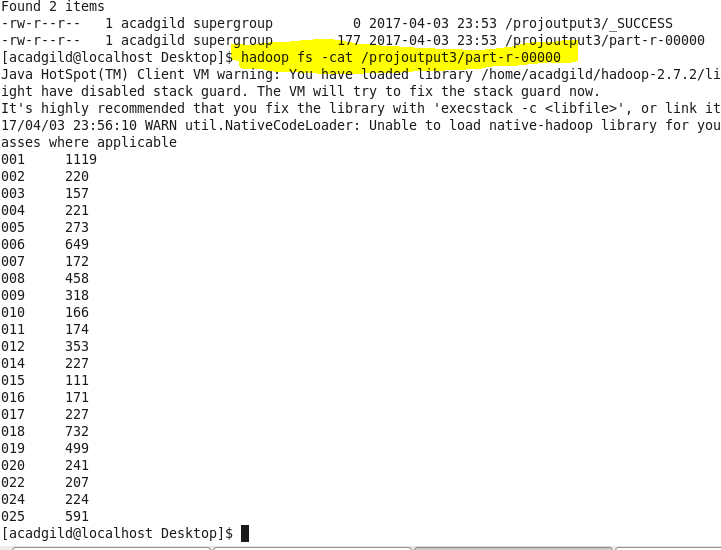
**Output file created**

****

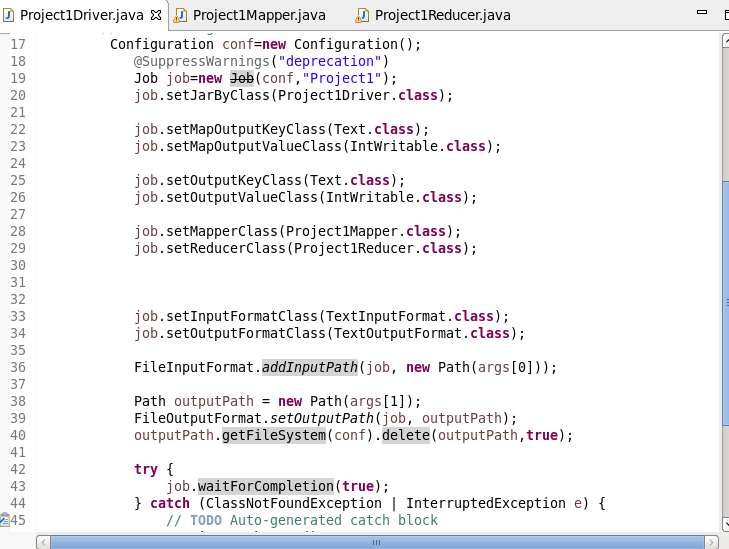
**Listing files in output directory**

****

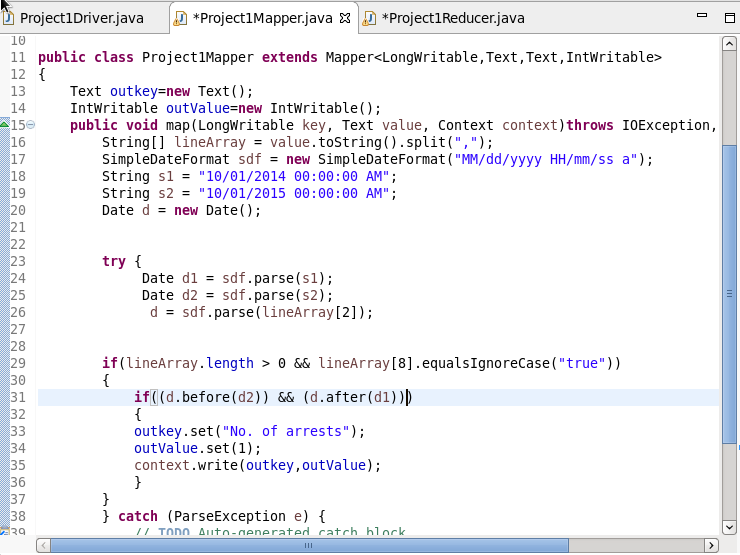
**Final output:**

****

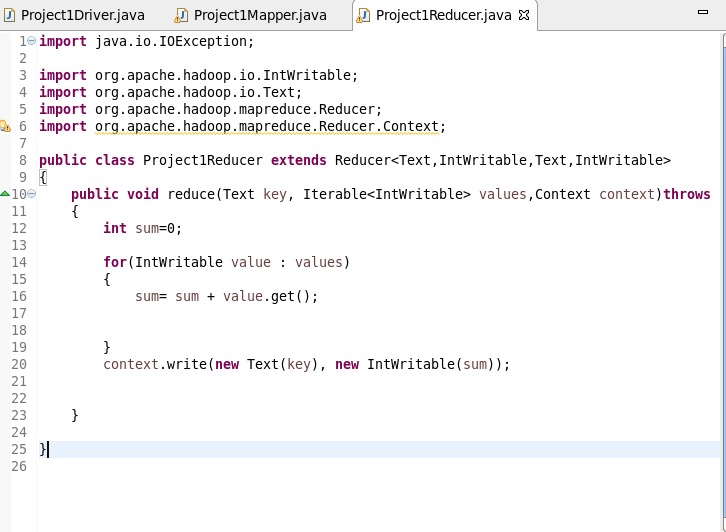
* **Write a mapreduce program to calculate the number of arrests done between October 2014 and October 2015.**
* **Driver class:**

****

* **Mapper class:**

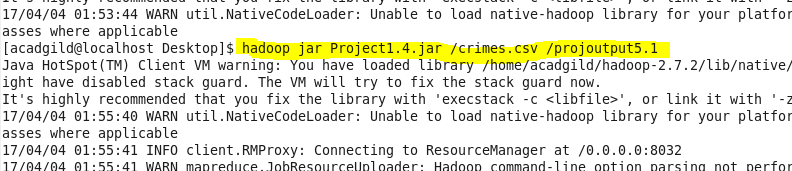
****

* **Reducer class:**

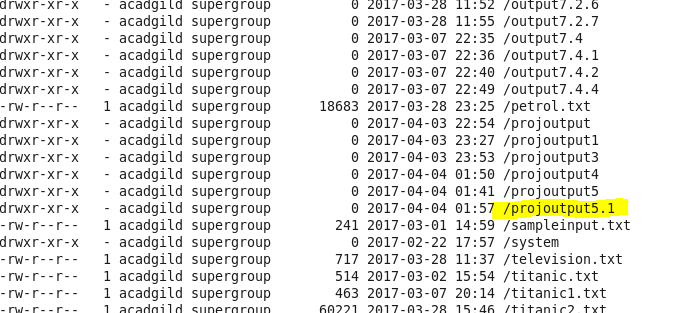
****

**Output:**

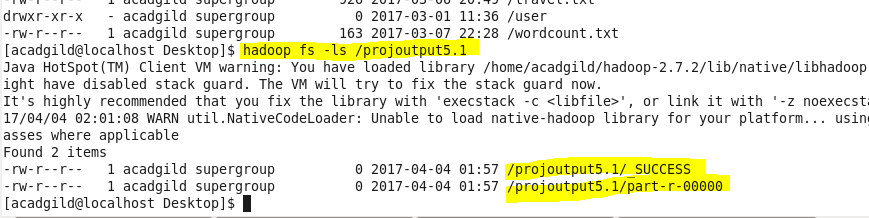
**Running jar file:**

****

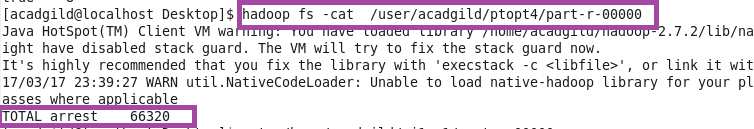
**Output file created :**

****

**Listing files of output directory**

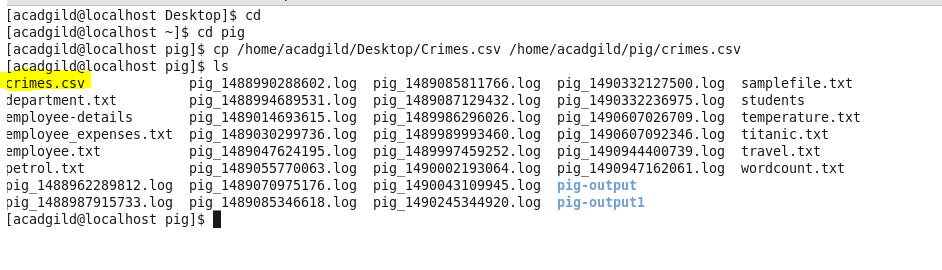
****

**Final Output:**

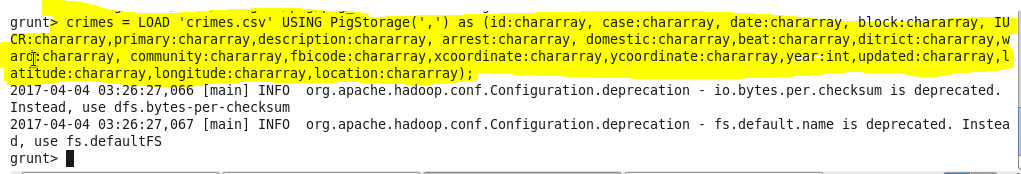
****

* **Write a pig program to calculate the number of cases investigated under each FBI code**

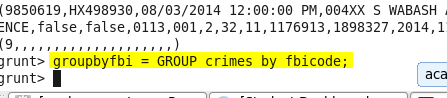
**Sample file(crimes.csv)**



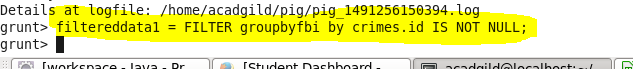
**Loading file:**



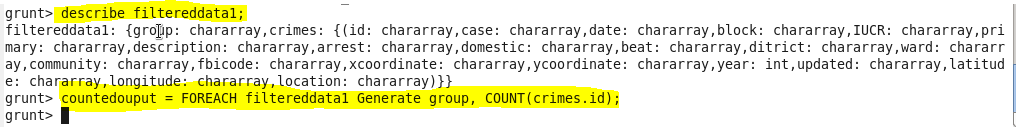
**Grouping by fbi code**



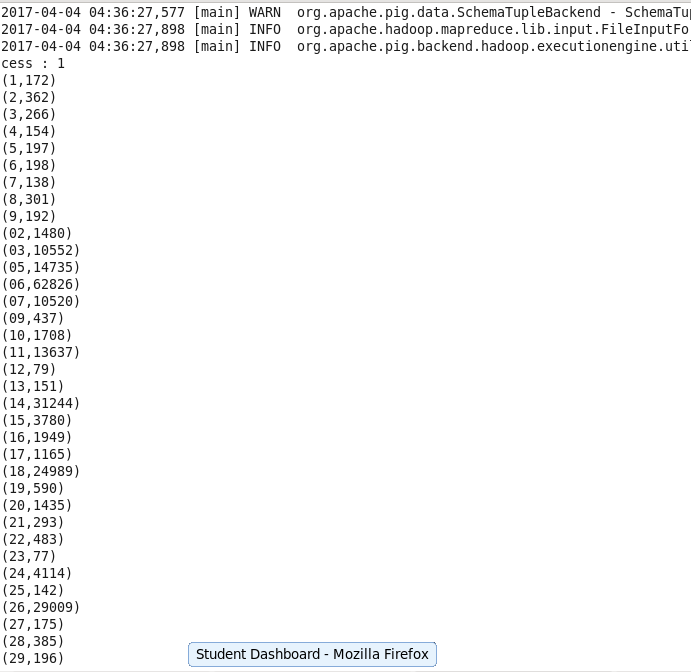
**Filtering out null values**



**Counting cases for each fbi code:**

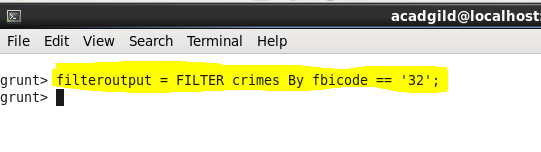


**Output:**

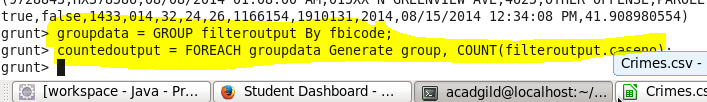


* **Write a pig program to calculate the number of cases investigated under FBI code 32**

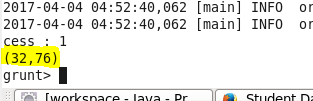
**Filter data by fbicode = 32**



**Group data by fbicode and counting dataset**



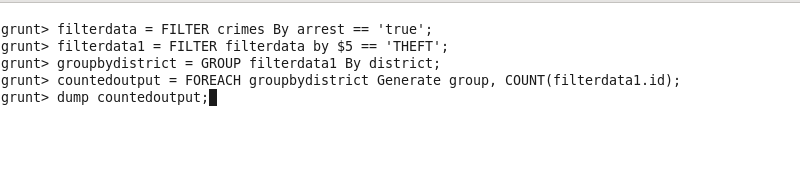
**Output:**



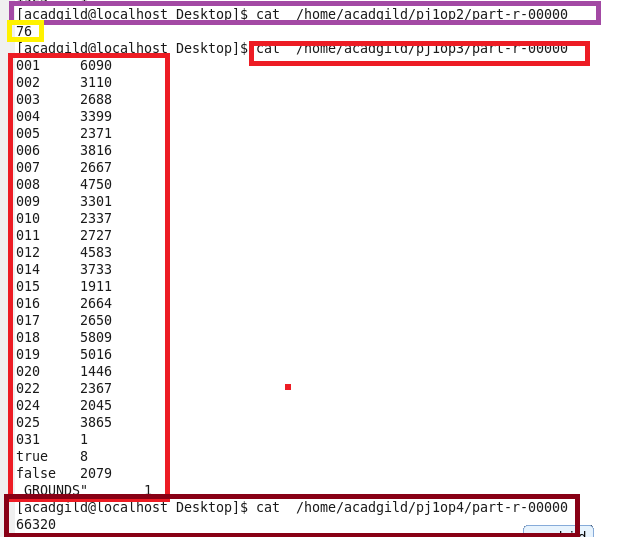
* **Write a pig program to calculate the number of arrests in theft district wise.**

**Filter data by arrest=true and crime primarytype = theft.**

**Group the result by district and generate count for each district;**

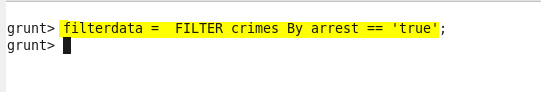


**Output:**

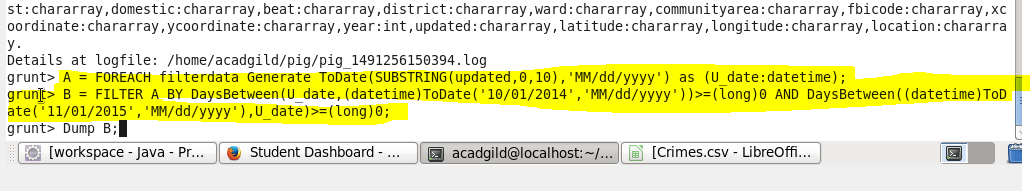
****

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

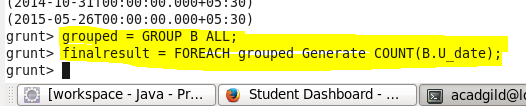
**Filter dataset where arrest has been done**



**Filtering data on basis of date of crime**



**Counting the cases between October 2014 and October 2015**



**Final output**

