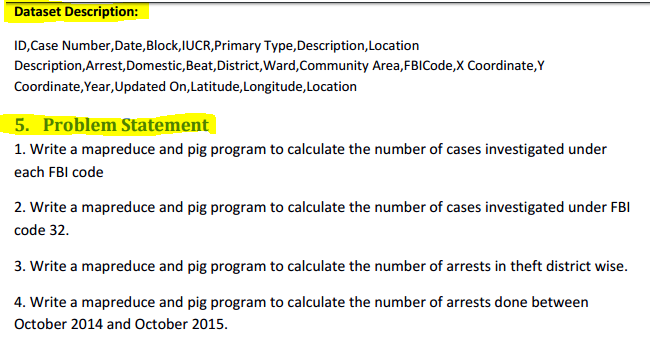
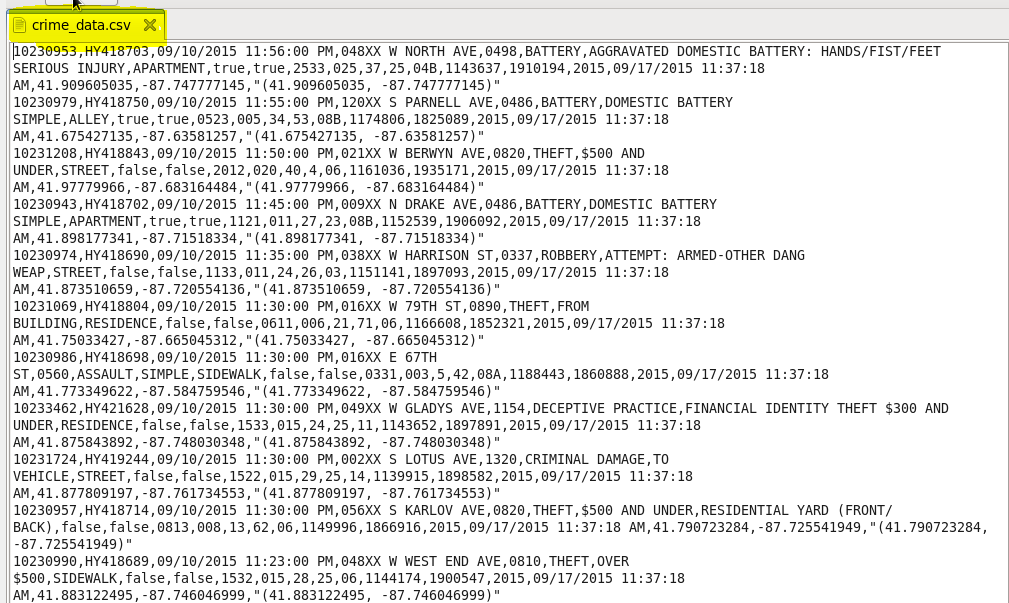
**Assignment\_13.1**

**Project – USA Crime Analysis**

****

**Data Set**

****

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

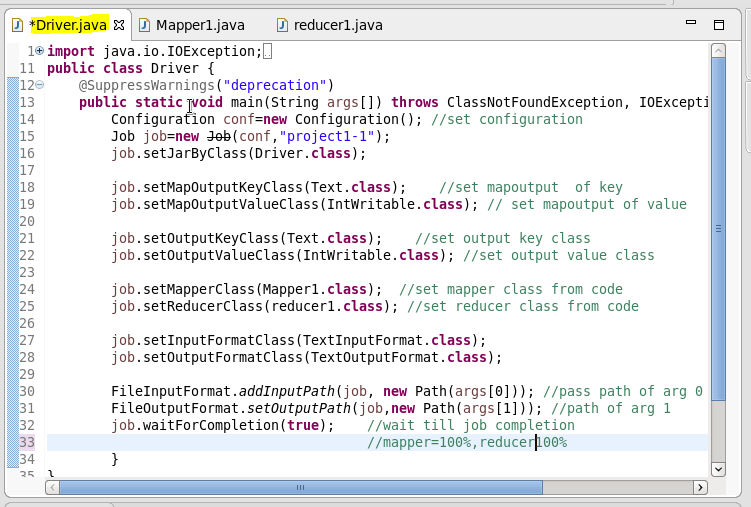
**It is expected to have the same results from both the operations map Reduce or pig script on data set.**

**MapReduce:**

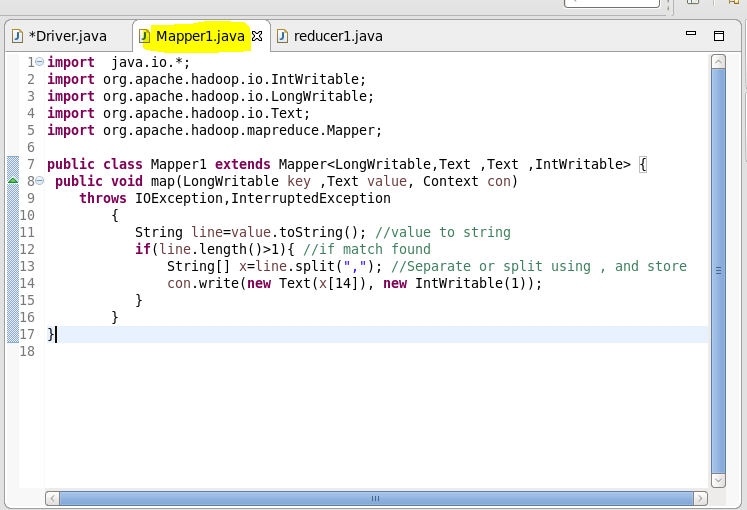
In this method we will map the pair in the form of key and value. The driver class, is used to set configuration, Mapper and reducer class and data format of key value. In Mapper class for a particular key, the value. The value of value will be updated for each iteration and repetition of the same FBIcode. The count will get updated and will be summed together in reducer class.

1. **Program-**

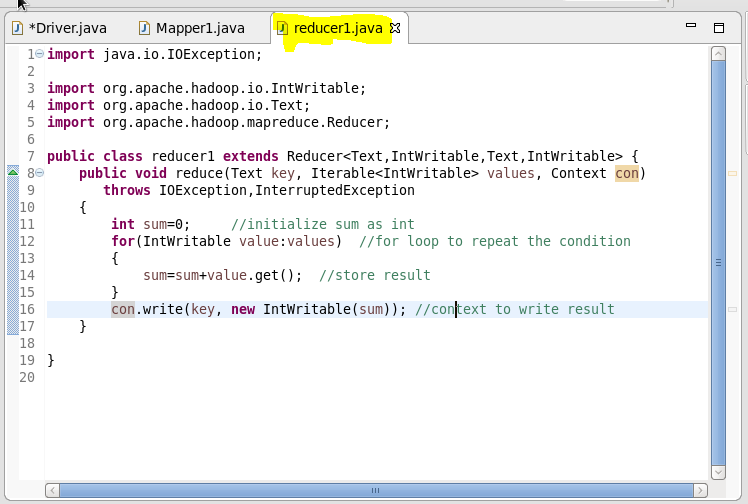
**Driver:-**

****

**Mapper:-**

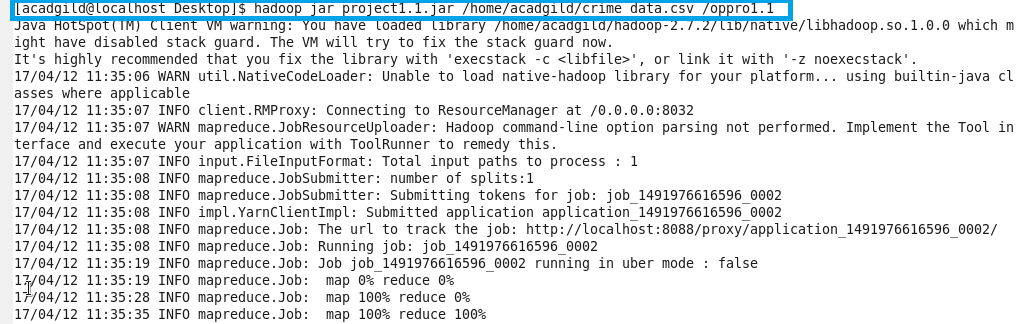
****

**Reducer:-**

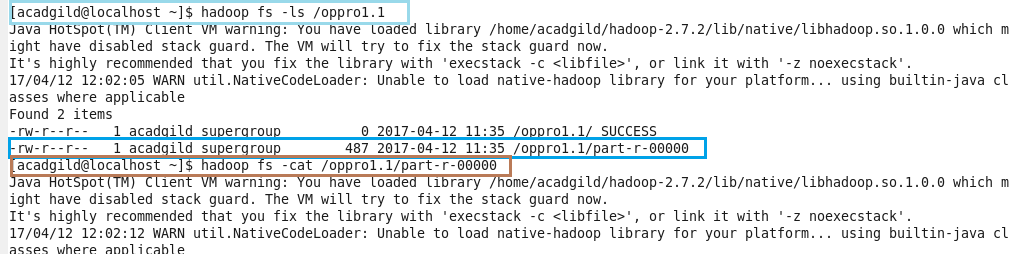
****

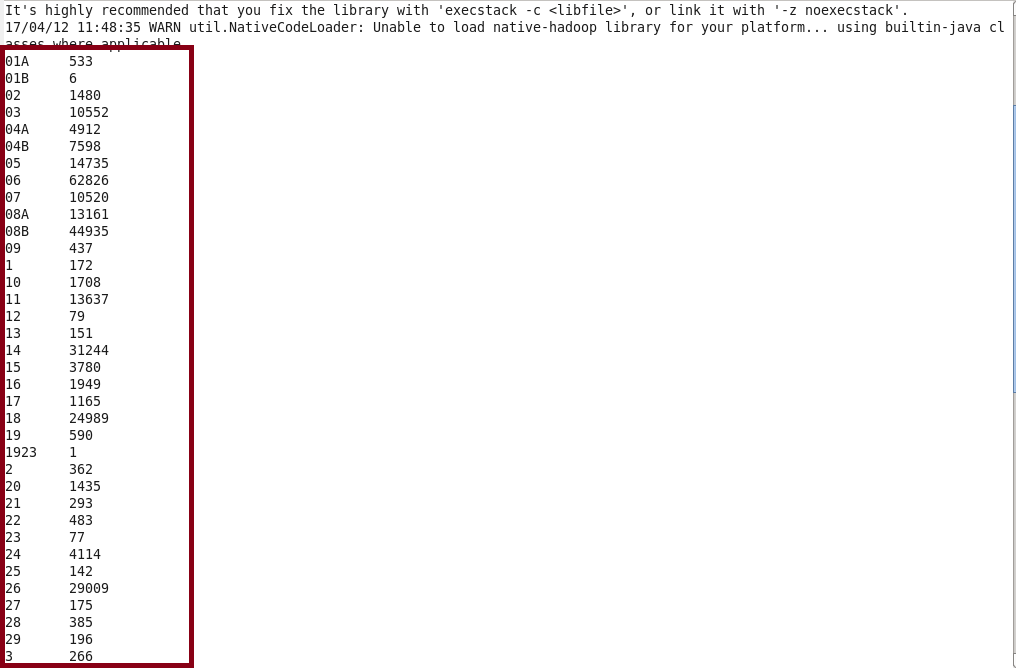
**Result:**

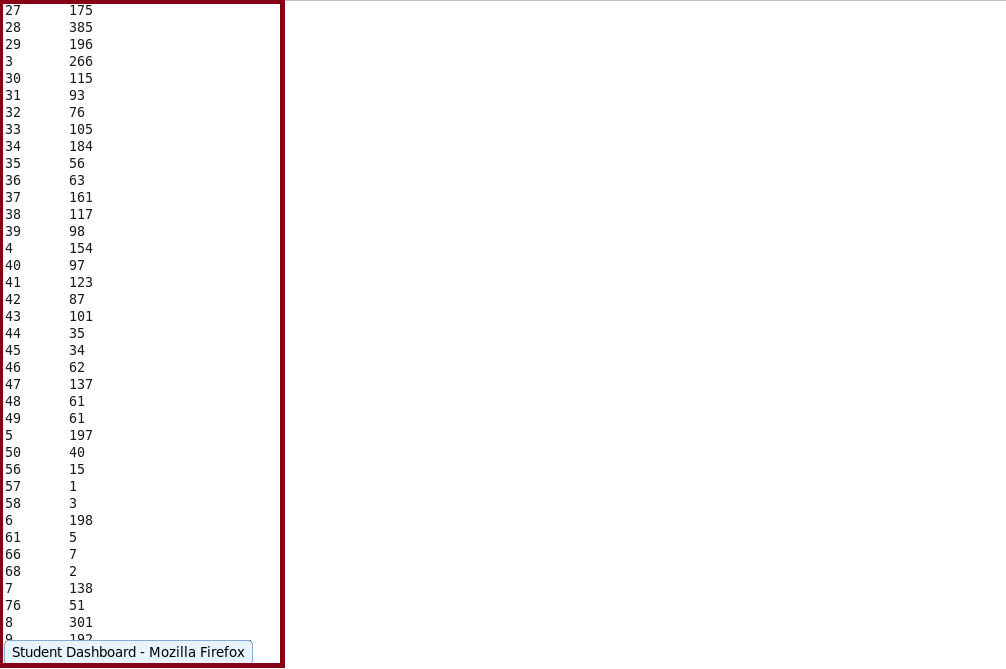
**Execution of jar File:-**

****

**Display Result:-**

****

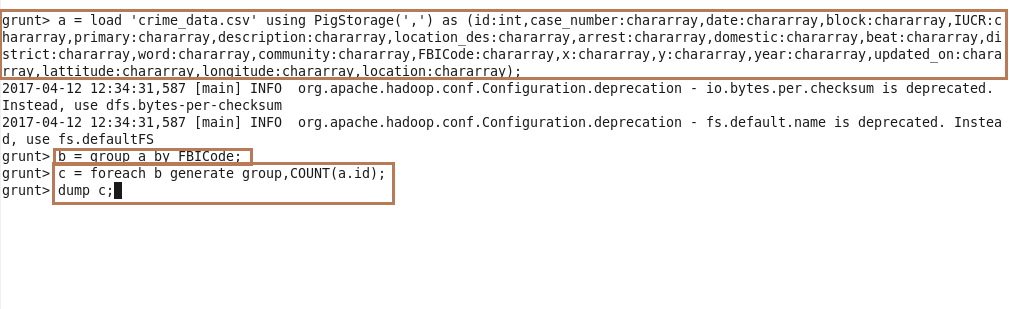
****

****

1. **Pig Script:-**

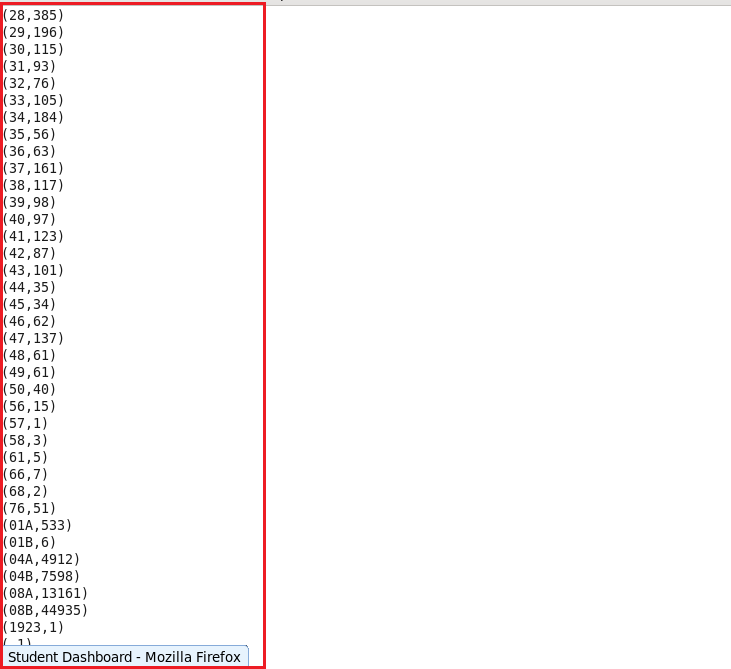
The same problem can be solved using Pig Script commands. First we will load that file into grunt shell and will group by FBICode, then later will count the count of case numbers and will display both.

**Commands:-**

****

**Result:-**

****

****

**If we Compare both the results by map reduce and pig they are exactly same.**

**Que.2) Write a mapreduce and pig program to calculate the number of cases investigated under FBI code 32.**

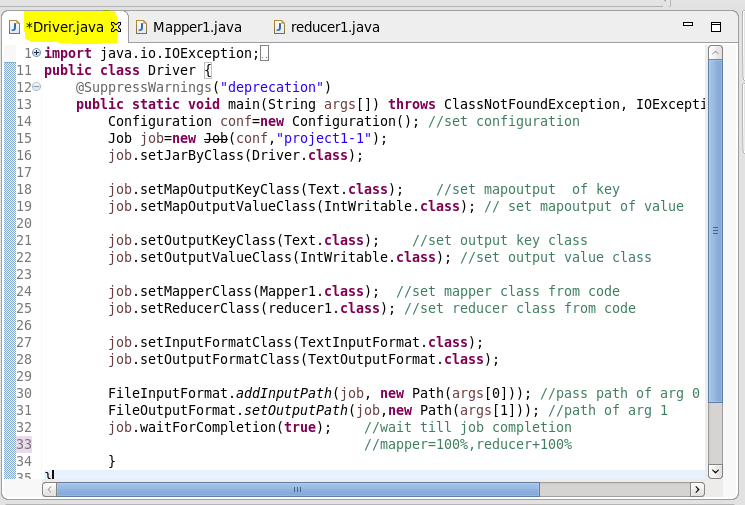
**It is expected to have the same results from both the operations map Reduce or pig script on data set.**

**1. MapReduce:**

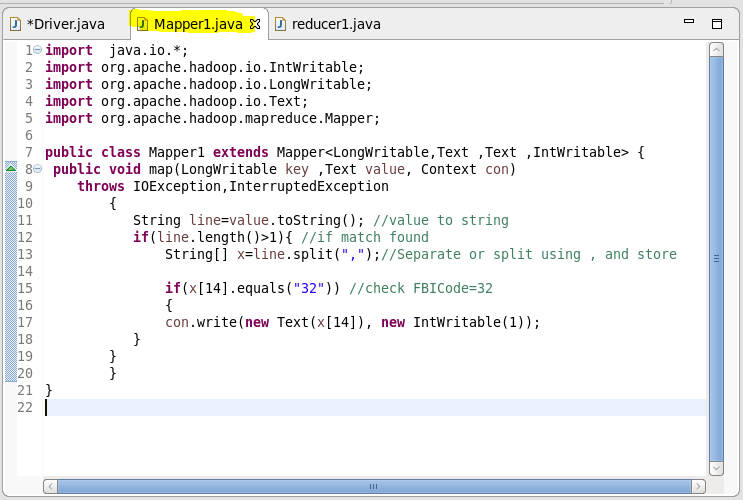
In this method we will map the pair in the form of key and value. The driver class, is used to set configuration, Mapper and reducer class and data format of key value. In Mapper class for a particular key, the value. The value of value (FBICode) will be checked for each iteration and repetition of the FBIcode having value as **‘32’**. The count will get updated and will be summed together in reducer class.

**Program-**

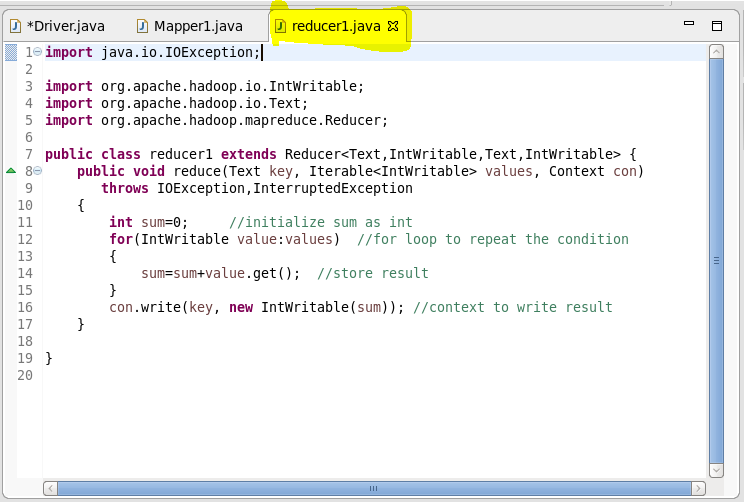
**Driver:-**

****

**Mapper:-**

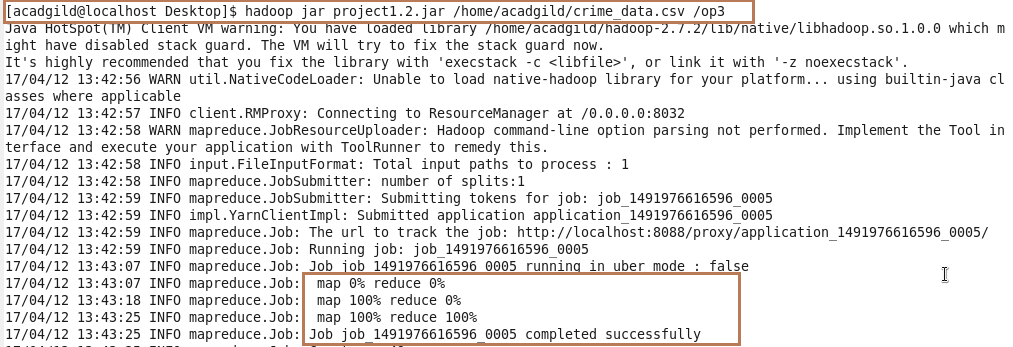
****

**Reducer:-**

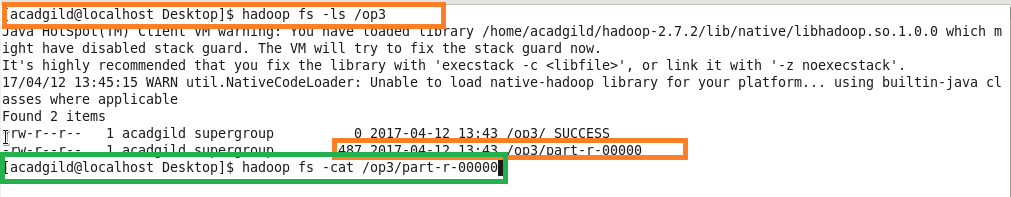
****

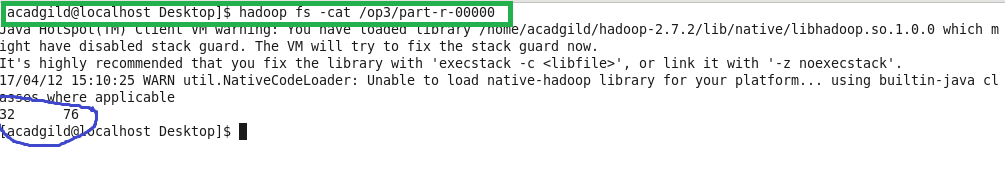
**Result:-**

**Execution of jar file:**

****

**Result:**

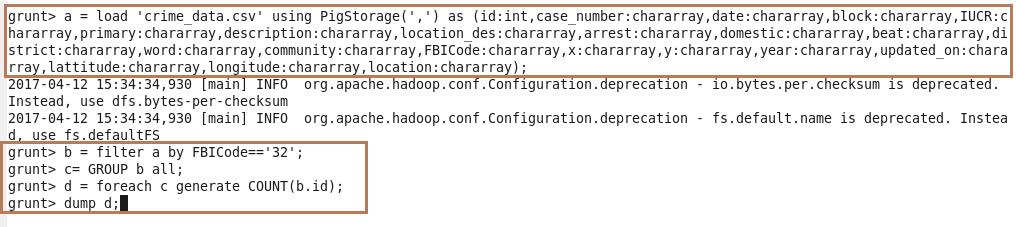
****

****

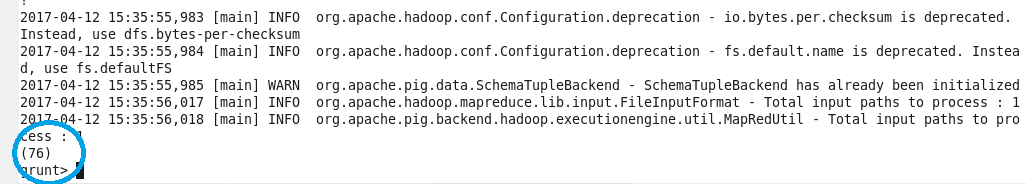
**2. Pig Script:**

The same problem can be solved using Pig Script commands. First we will load that file into grunt shell and will group by FBICode having code as **32**, then later will count the count of case numbers satisfying the condition and will display both code and its count.

**Commands:**

****

**Result**:



On comparing both the results we see, both the results are same. For FBICode=32 , the count is 76 .

**Que.3) Write a mapreduce and pig program to calculate the number of arrests in theft district wise**

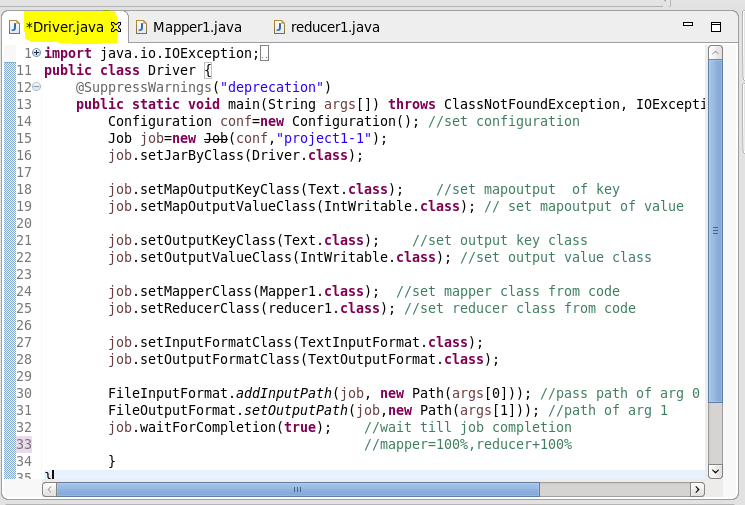
**It is expected to have the same results from both the operations map Reduce or pig script on data set.**

**1. MapReduce:**

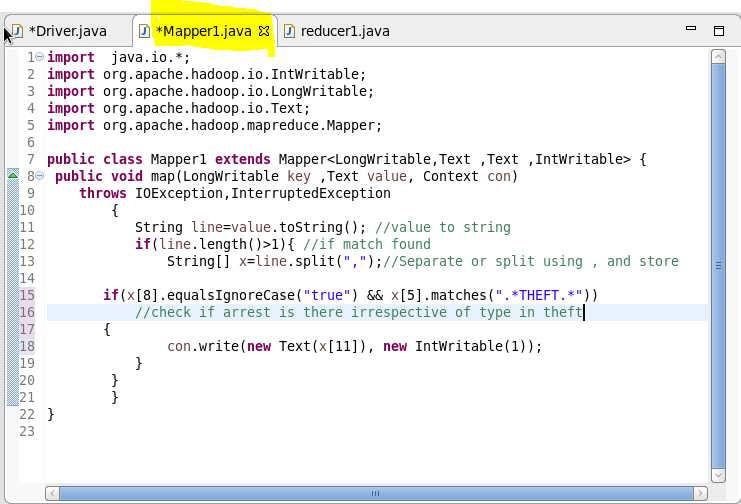
In this method we will map the pair in the form of key and value. The driver class, is used to set configuration, Mapper and reducer class and data format of key value. In Mapper class for a particular key (district), the value (arrests). The value of value will be checked for each iteration and repetition of the arrests having value. The count will get updated and will be summed together in reducer class.

**Program-**

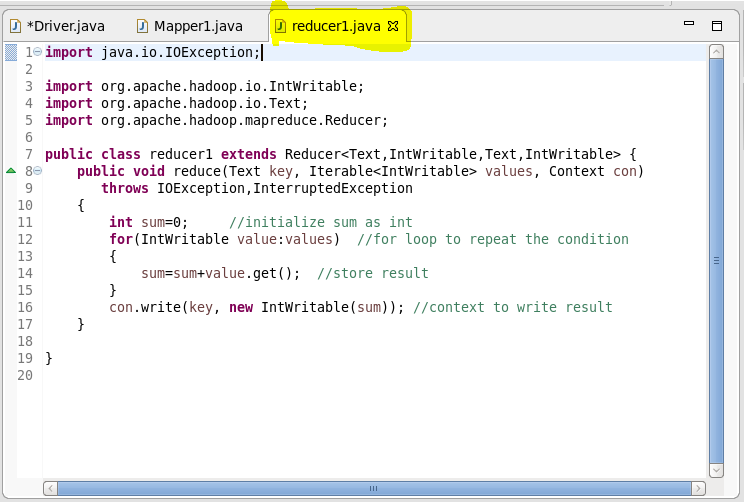
**Driver:-**

****

**Mapper:-**

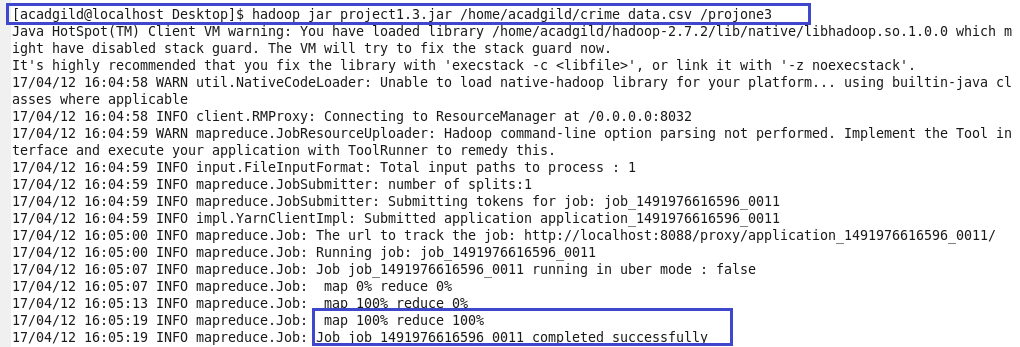
****

**Reducer:-**

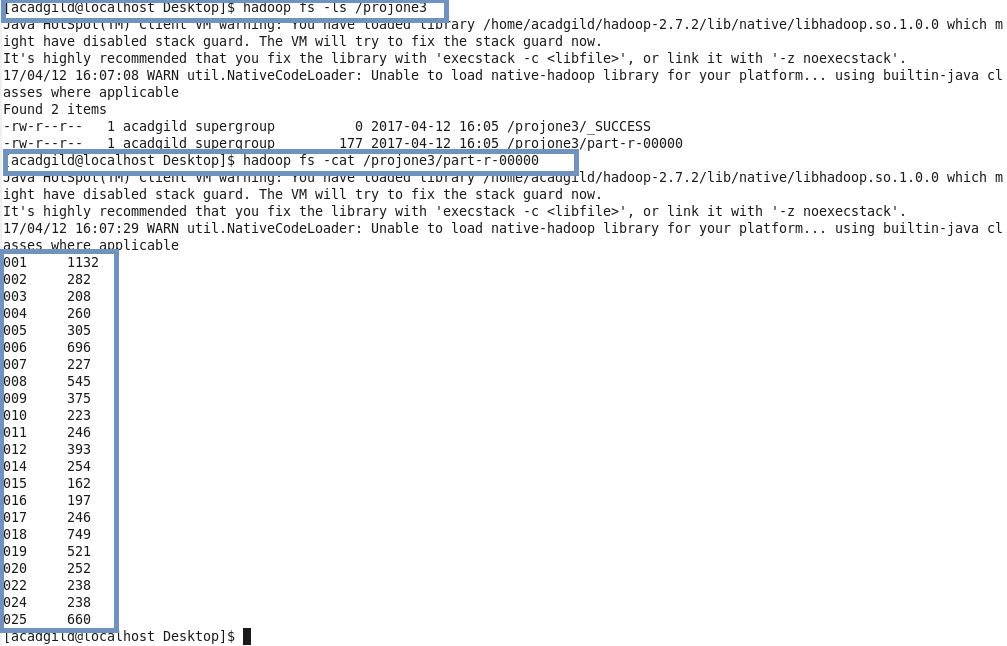
****

**Result:-**

**Execution of jar file:**

****

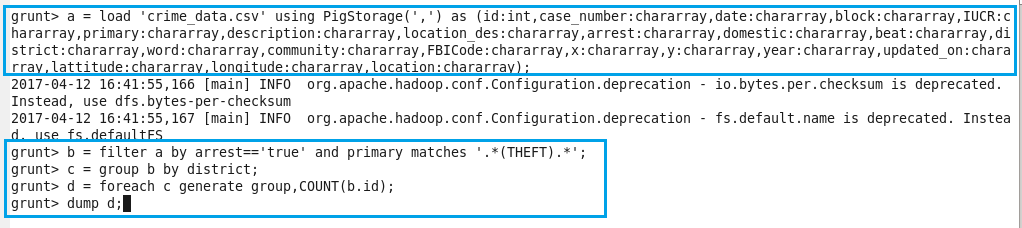
**Result:**

****

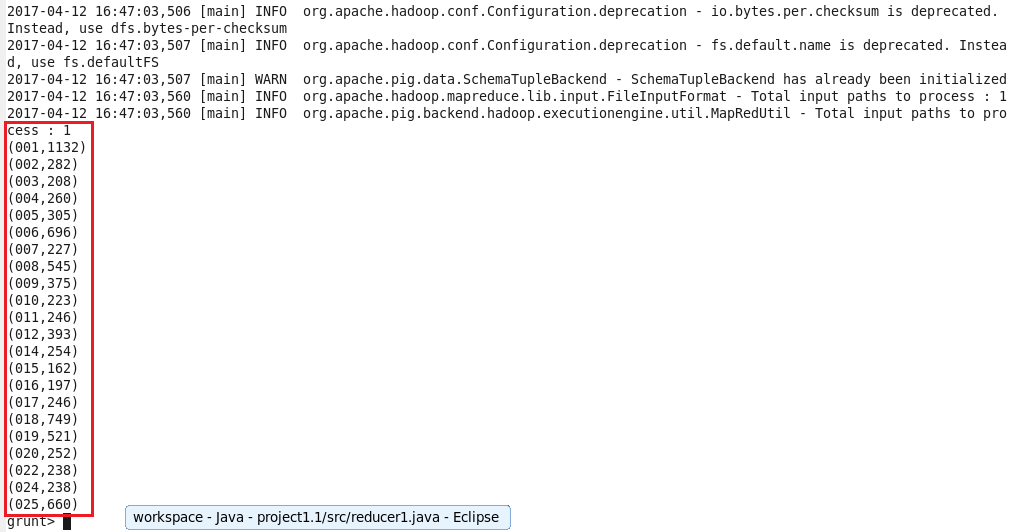
**2. Pig Script:**

The same problem can be solved using Pig Script commands. First we will load that file into grunt shell and will filter data by conditions satisfying arrest=true and primary column matches any theft, then later will count the count of case numbers satisfying the condition and will display both code and its count.

**Commands:**

****

**Result**:



On comparing both the results we see, both the results are same.

**Que.4)** **Write a mapreduce and pig program to calculate the number of arrests done between October 2014 and October 2015.**

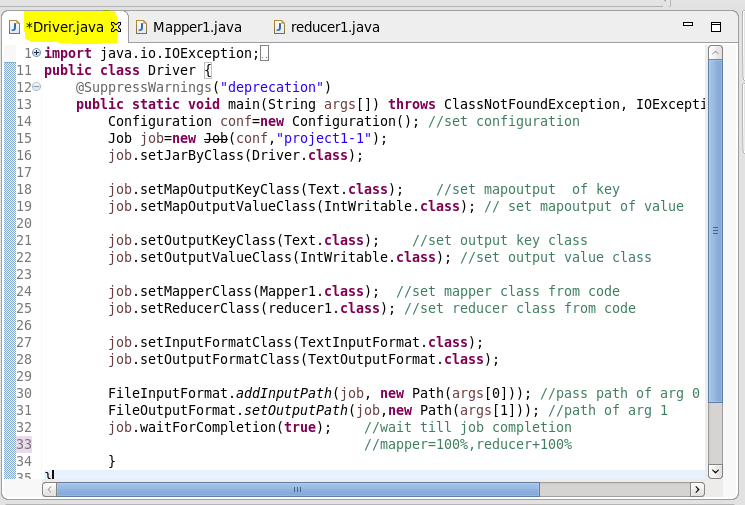
**It is expected to have the same results from both the operations map Reduce or pig script on data set.**

**1. MapReduce:**

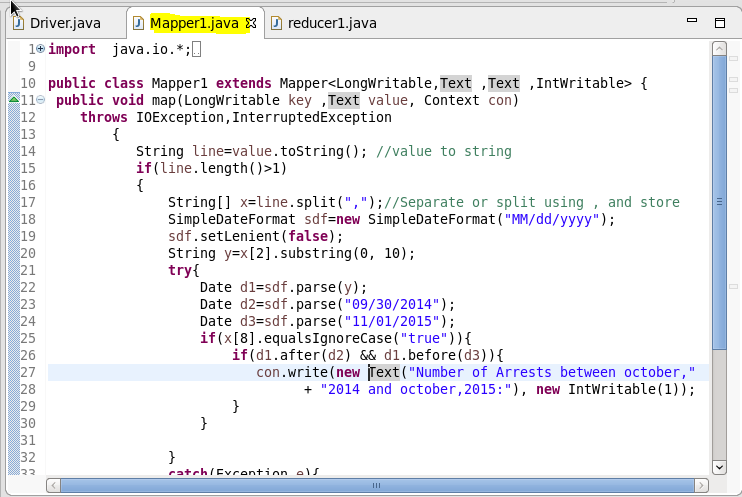
In this method we will map the pair in the form of key and value. The driver class, is used to set configuration, Mapper and reducer class and data format of key value. In Mapper class for a particular key (date), the value (arrests). The value of value will be checked for each iteration and repetition of the arrests having value. The condition is that date has to fall between months of **October 2014 to October 2015**. The count will get updated and will be summed together in reducer class.

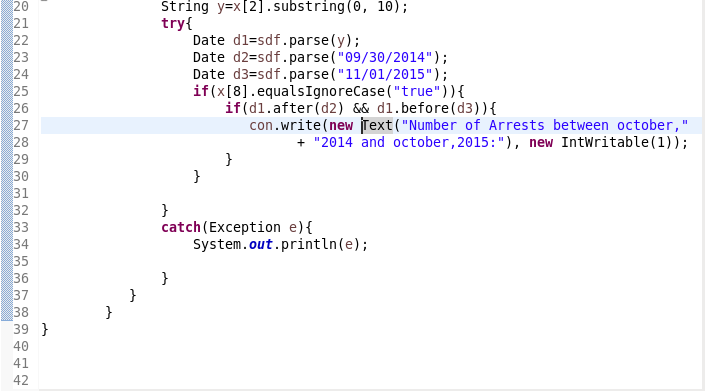
**Program-**

**Driver:-**

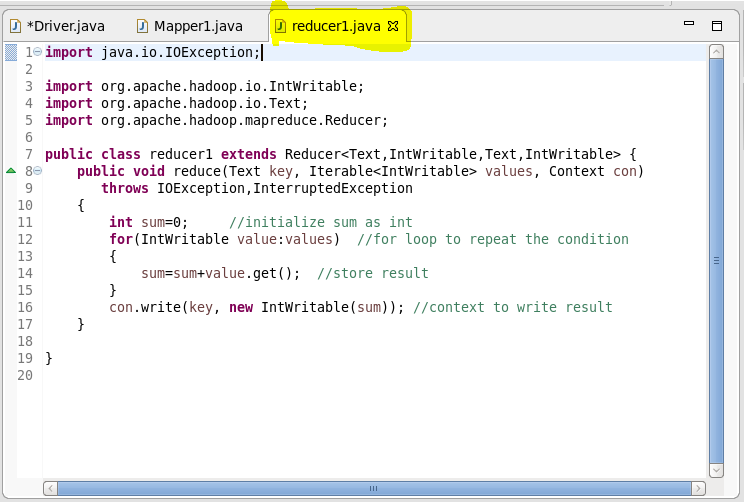
****

**Mapper:-**

****

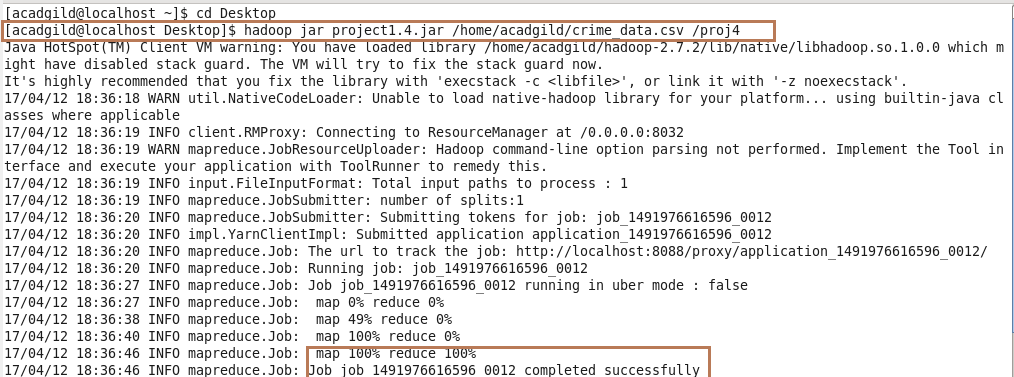
****

**Reducer:-**

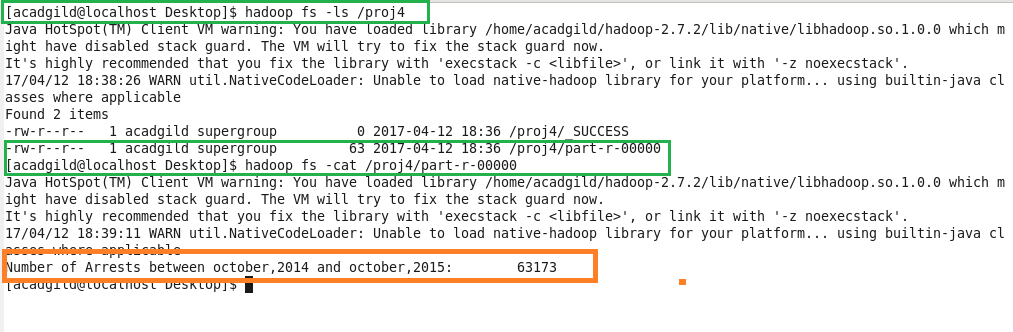
****

**Result:-**

**Execution of jar file:**

****

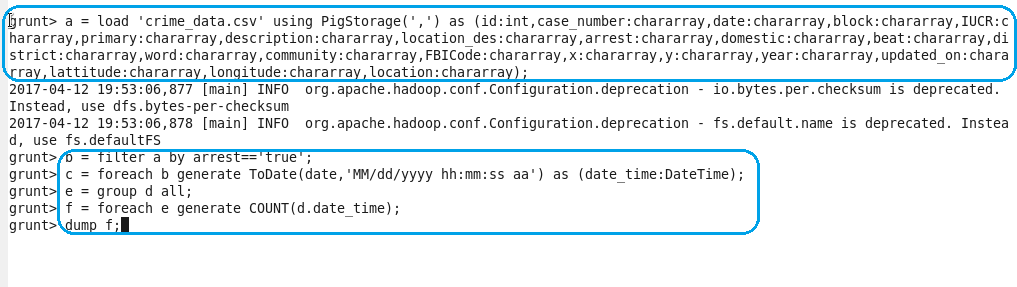
**Result:**

****

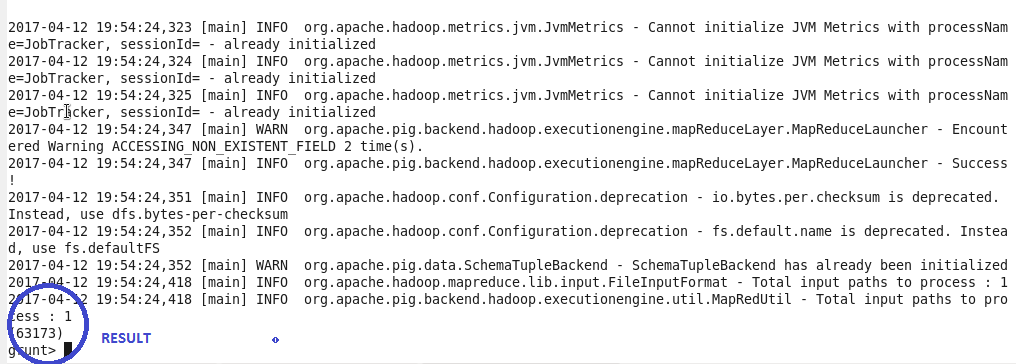
**2. Pig Script:**

The same problem can be solved using Pig Script commands. First we will load that file into grunt shell and will filter data by conditions satisfying arrest=true and date of arrest between oct 2014 to oct 2015, then later will count the count of case numbers satisfying the condition and will display both code and its count.

**Commands:**

****

**Result**:



On comparing both the results we see, both the results are same.