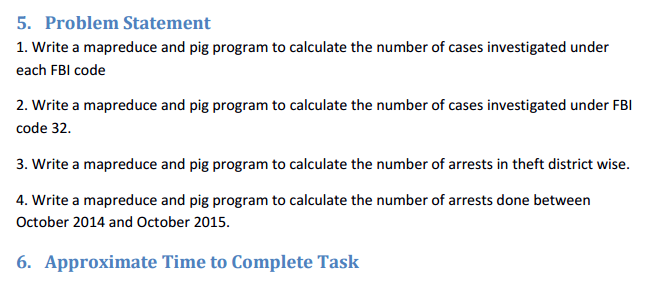
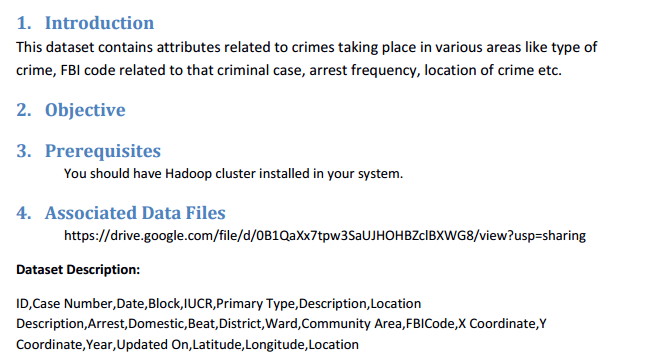
**Project 1**

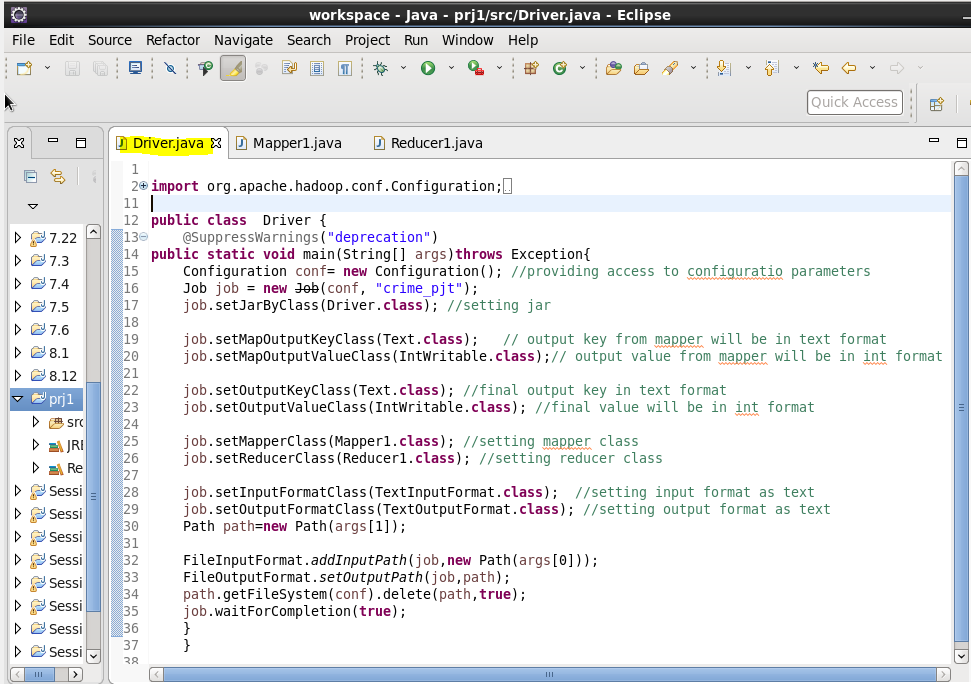
**QUESTION:-**

****

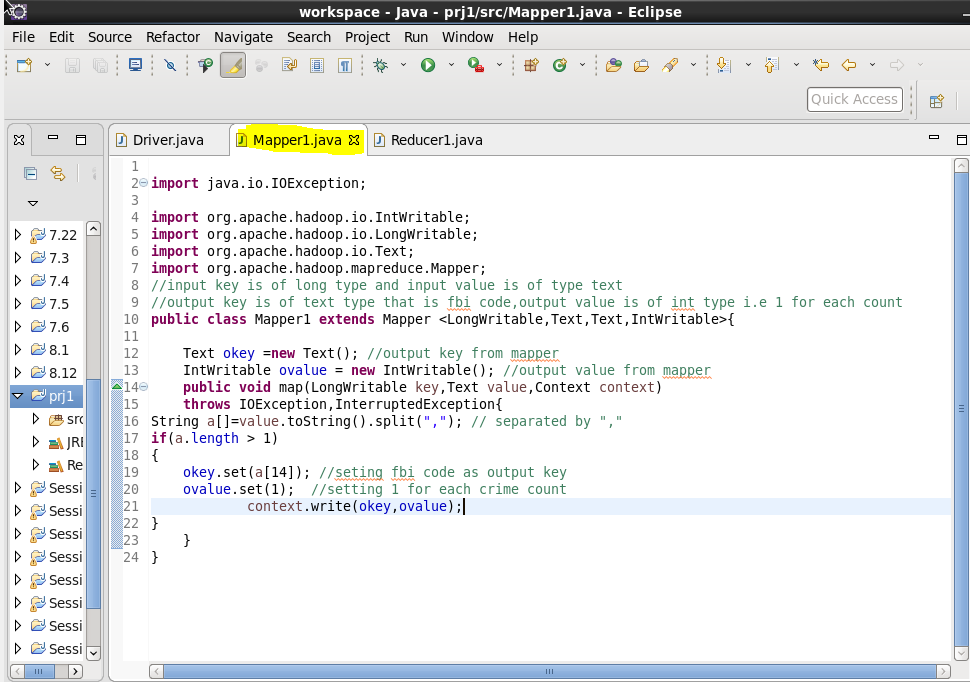
1. **Number of cases investigated under each FBI code :**

**USING MAPREDUCE :**

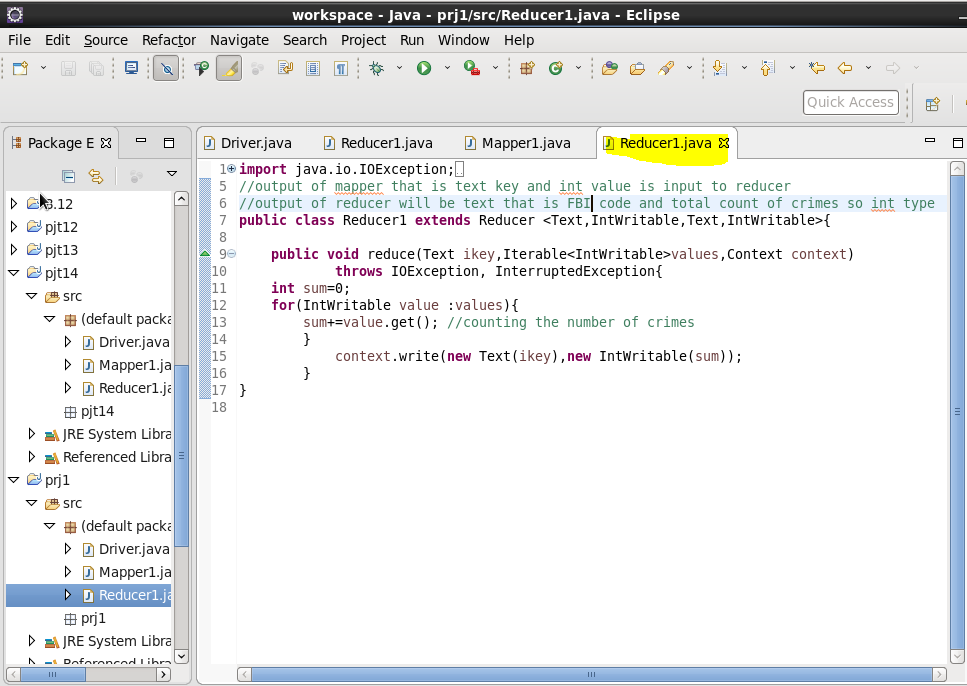
1. **Driver class :**

****

1. **Mapper class :**

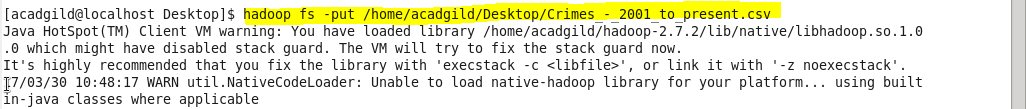
****

1. **Reducer Class:**

****

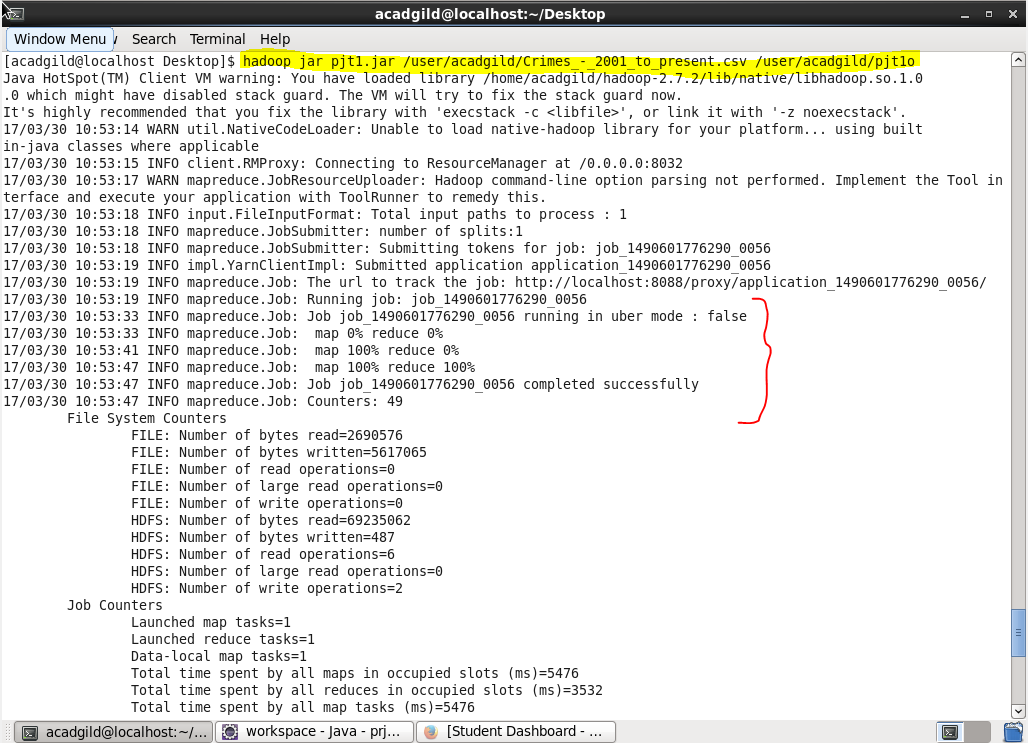
**First we will put our database file to HDFS:**

We used put command ,so as to copy our file to HDFS :

****

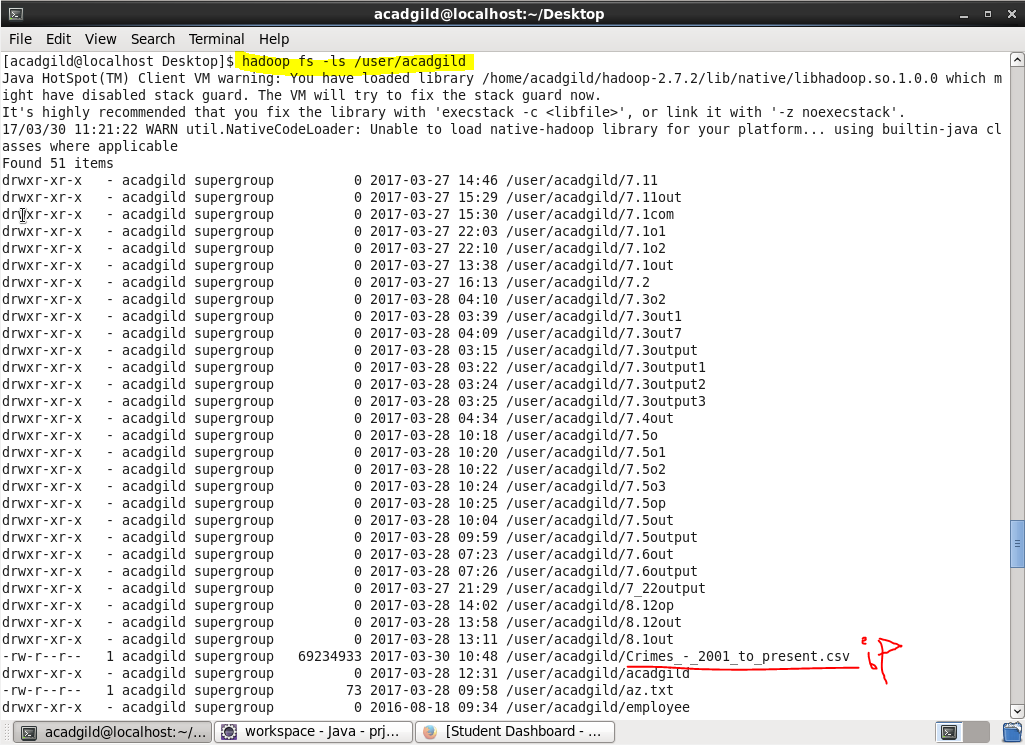
**Running Jar file :**

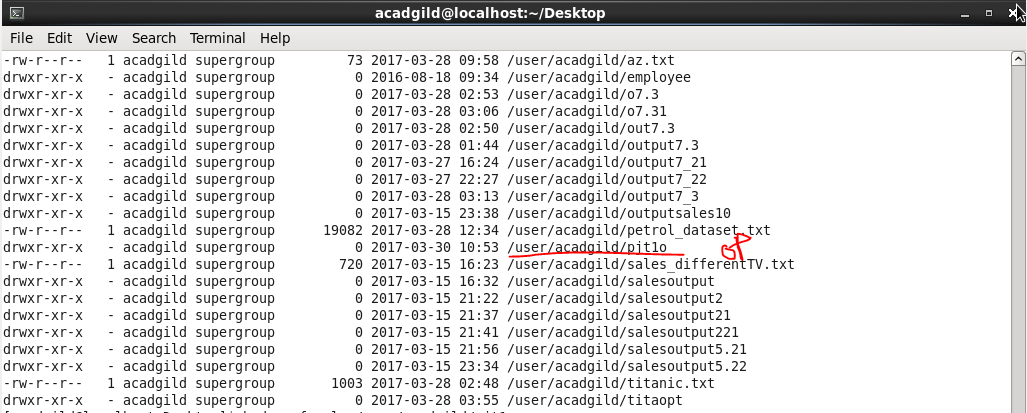
After Creating jar file of ourjava code we will run the jar file as following :

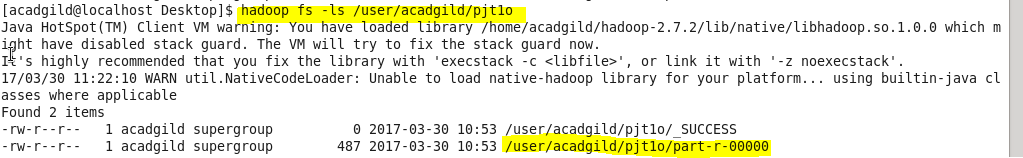
****

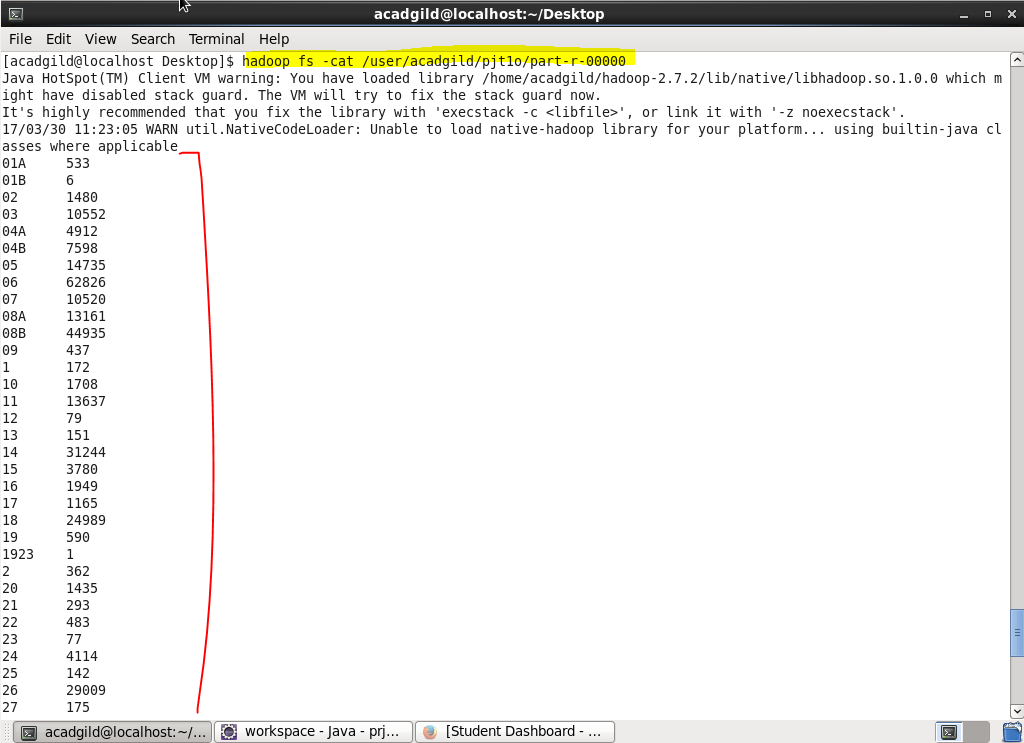
On successfully running jar file we will check weather out output is there in given location or not

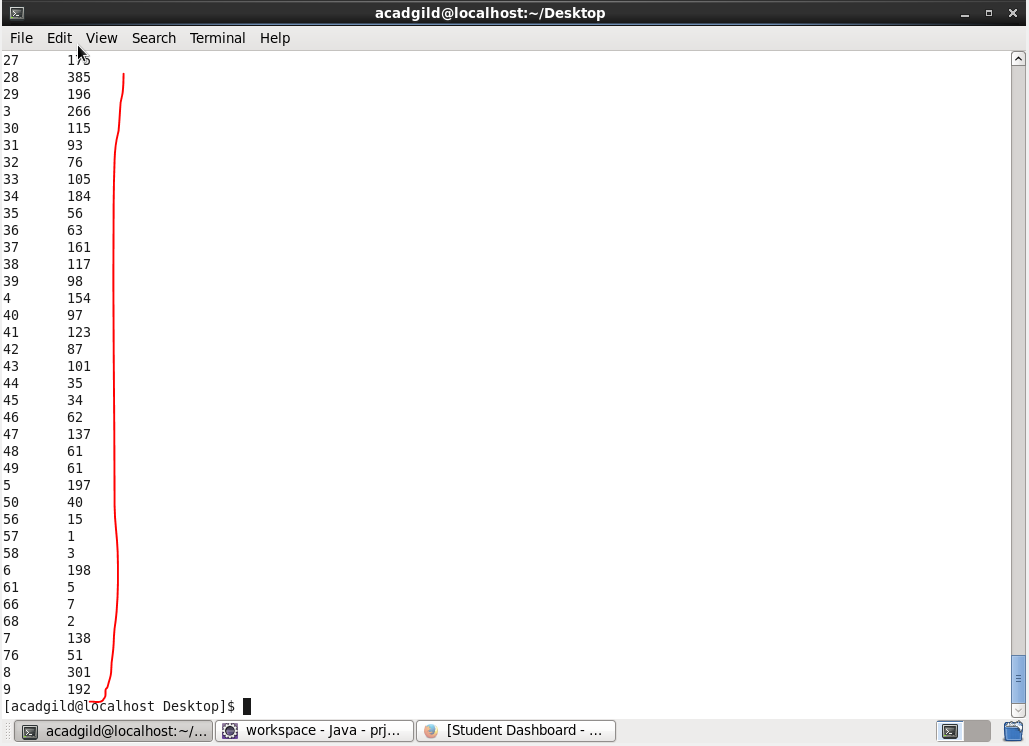
So we will list the files in HDFS as:

****

****

****

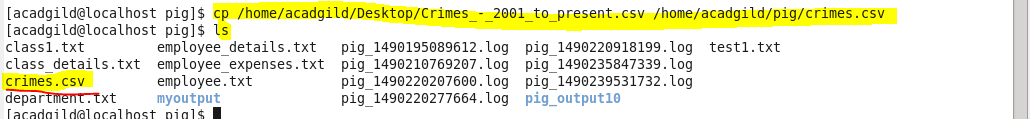
Now we got our file in particular location,so now we will display the contents of out output using cat command as :



Above screenshots shows the output of our mapreduce program that shows the FBI Code in first column and corresponding number of crimes in second column.

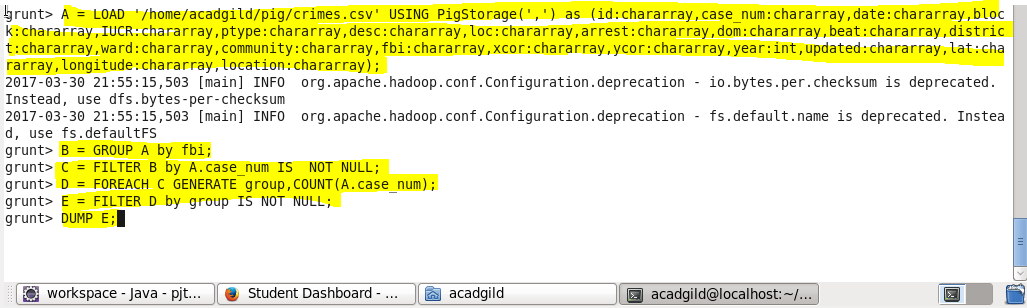
1. **USING PIG PROGRAM :**

**First we will copy the file from old location to pig folder where we are storing our files-**

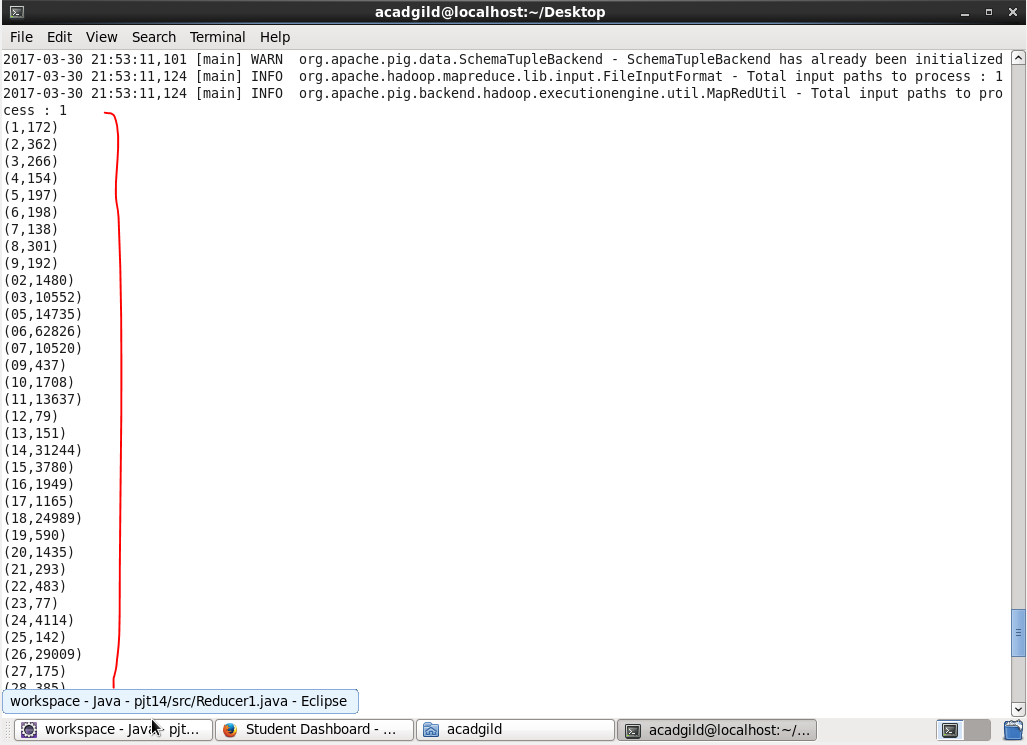
****

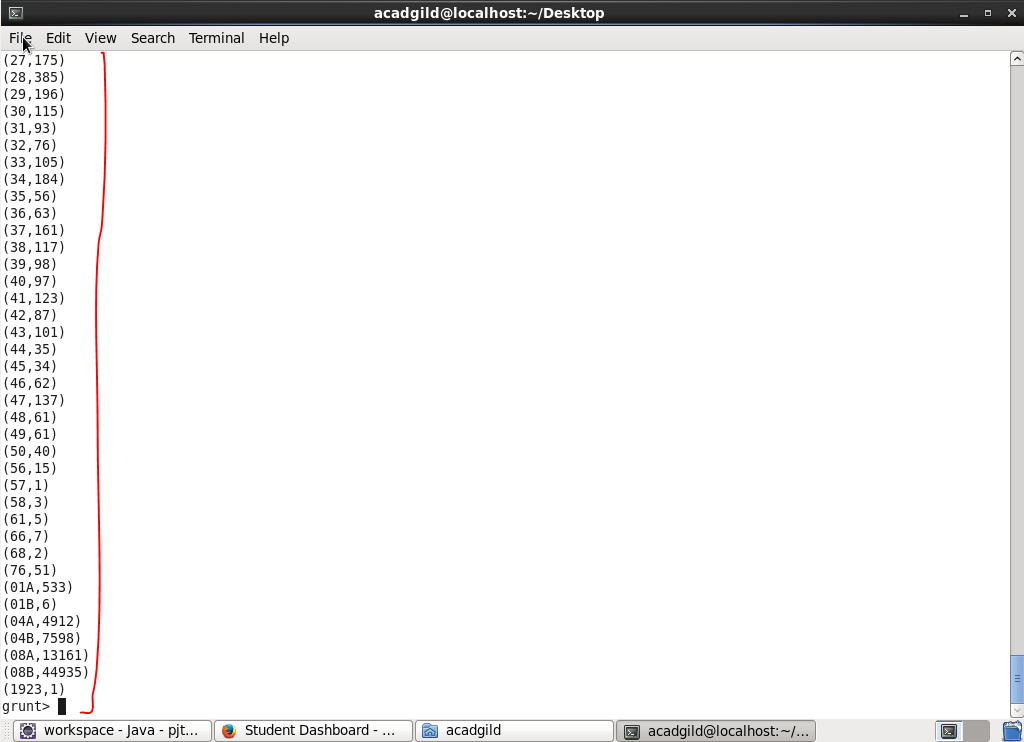
**Pig script :**

* First we will load our file using LOAD command then we will group our data by fbi codes as we want the count of crimes in each FBI code . Then we will filter our output by the case number if it is there or not checking that if not null then that’s a case. Then we will generate the count using COUNT function then again we will filter if group is not null .

****

**Output:**

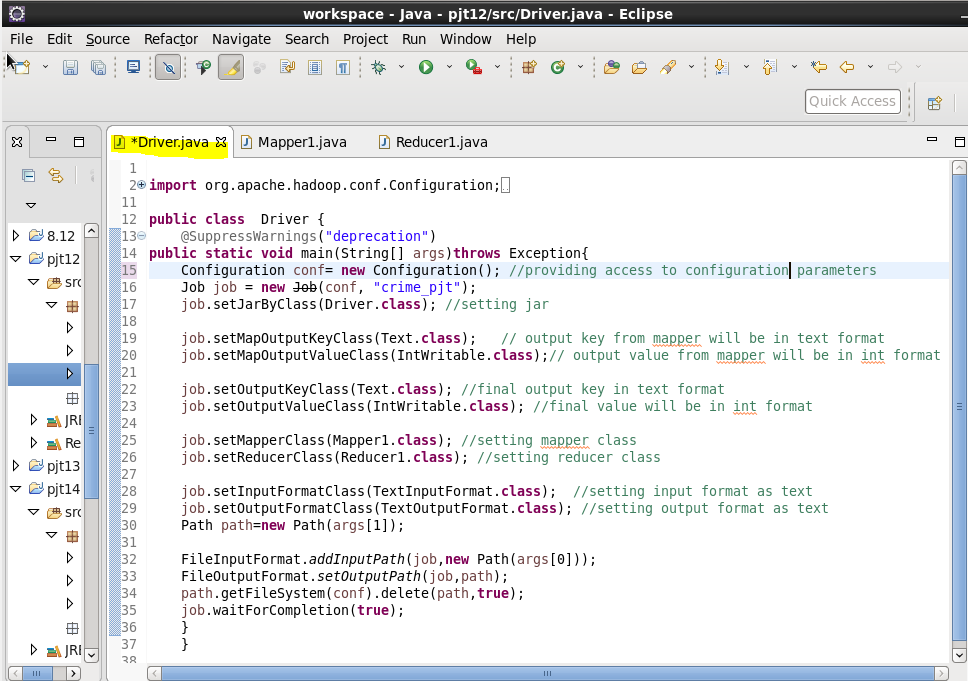
****

****

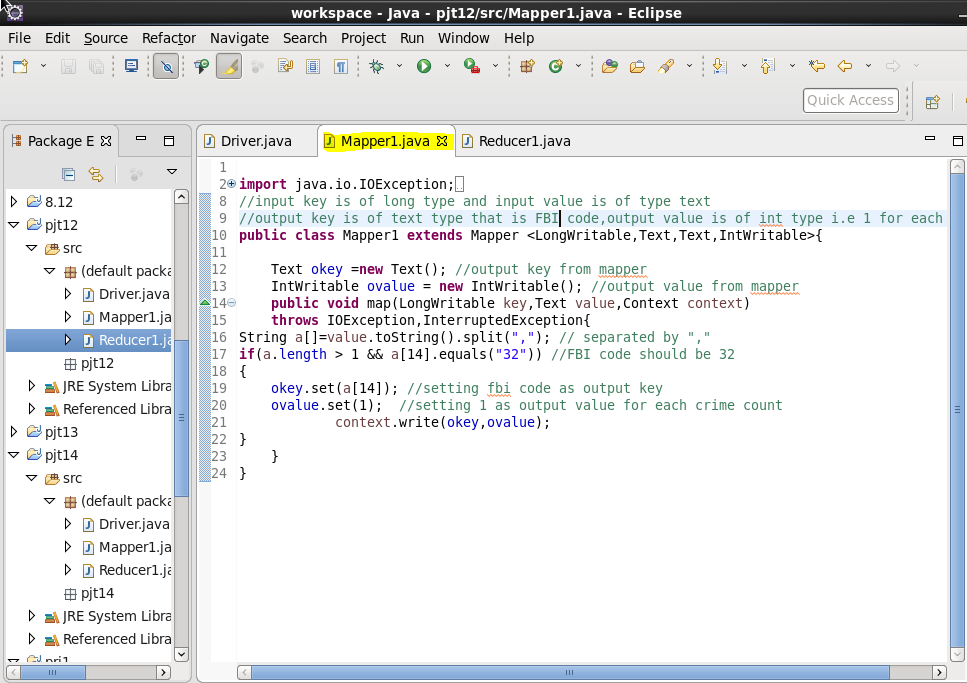
**So,we got the same result from mapreduce and pig hence verified.**

**2 - Number of cases investigated under FBI code 32 :**

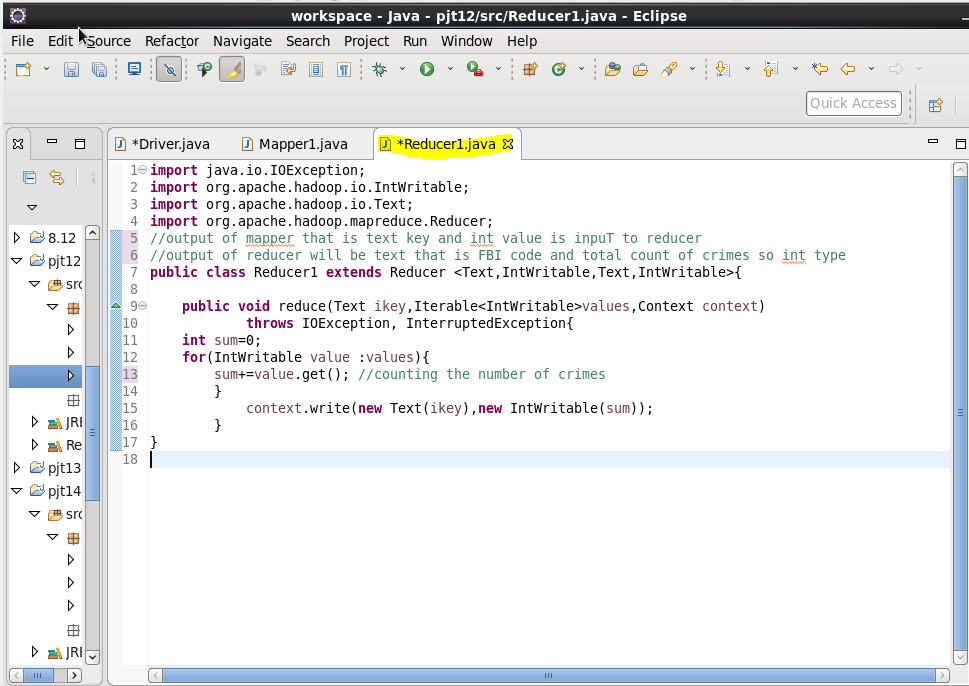
1. **Driver class :**

****

1. **Mapper class :**

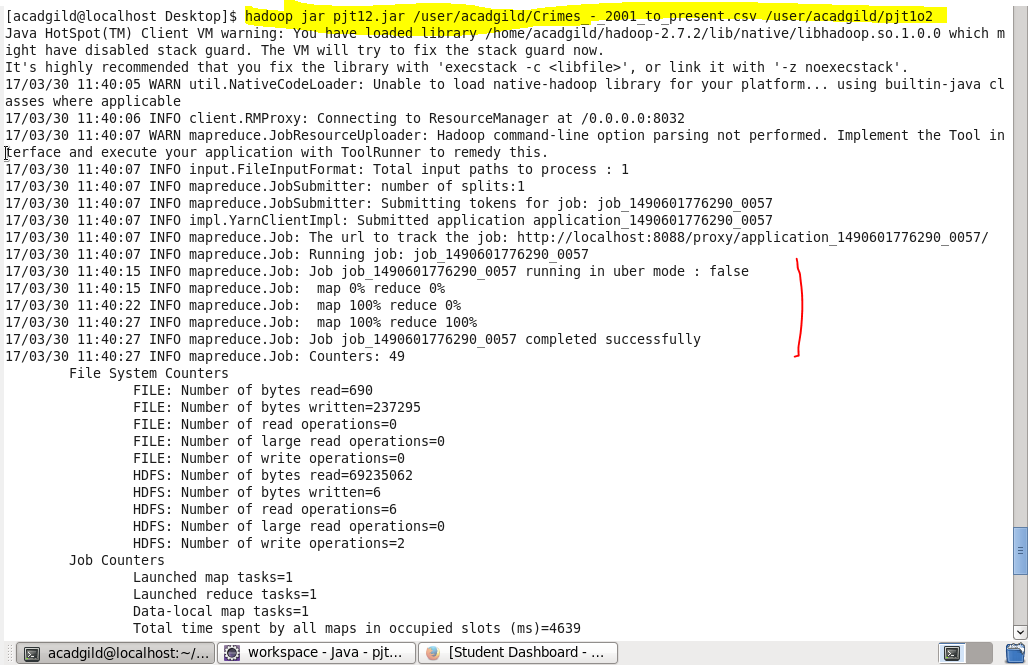
****

1. **Reducer Class:**

****

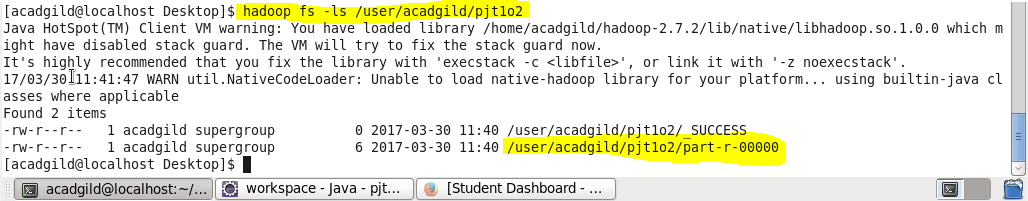
**Running Jar file :**

After Creating jar file of our java code we will run the jar file as following :

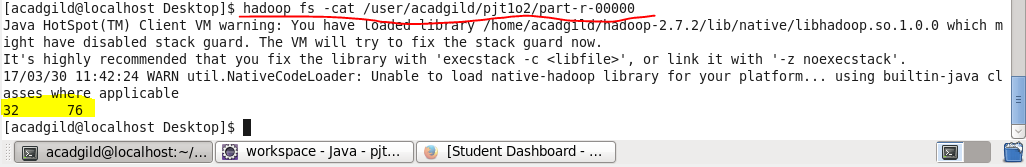
****

After successfully running jar file we will check weather out output is there in given location or not

So we will list the files in HDFS as:



Now we got our file in particular location,so now we will display the contents of out output using cat command as :



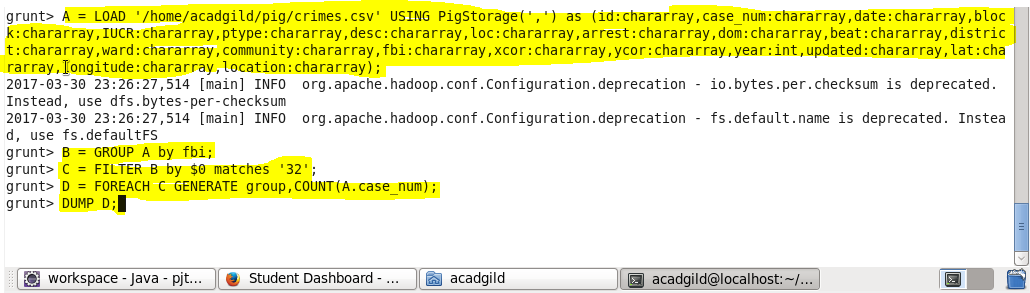
Above screenshots shows the output of our mapreduce program that shows the FBI Code is 32 in first column and corresponding number of cases investigated in second column.

**USING PIG PROGRAM :**

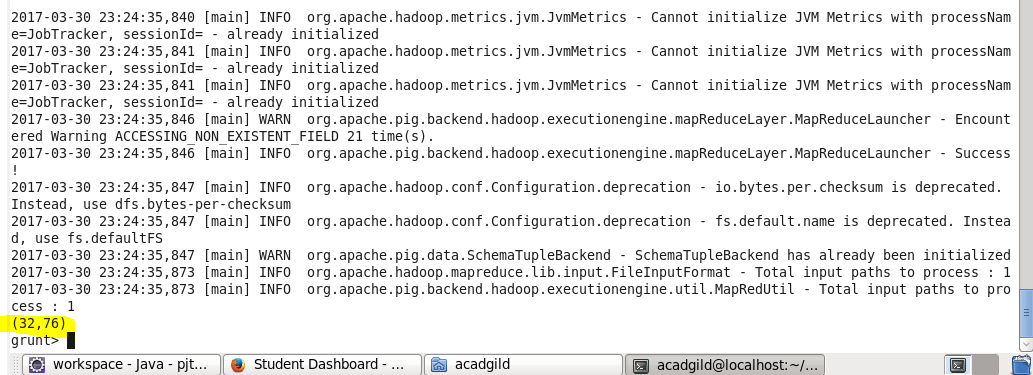
**Pig script :**

* First we will load our file using LOAD command then we will group our data by fbi codes as we want the count of cases investigated for FBI code equal to 32 . Then we will filter our output by the fbi code that is it should be 32.

Then we will generate the count using COUNT function .Then using DUMP command we will show the result

****

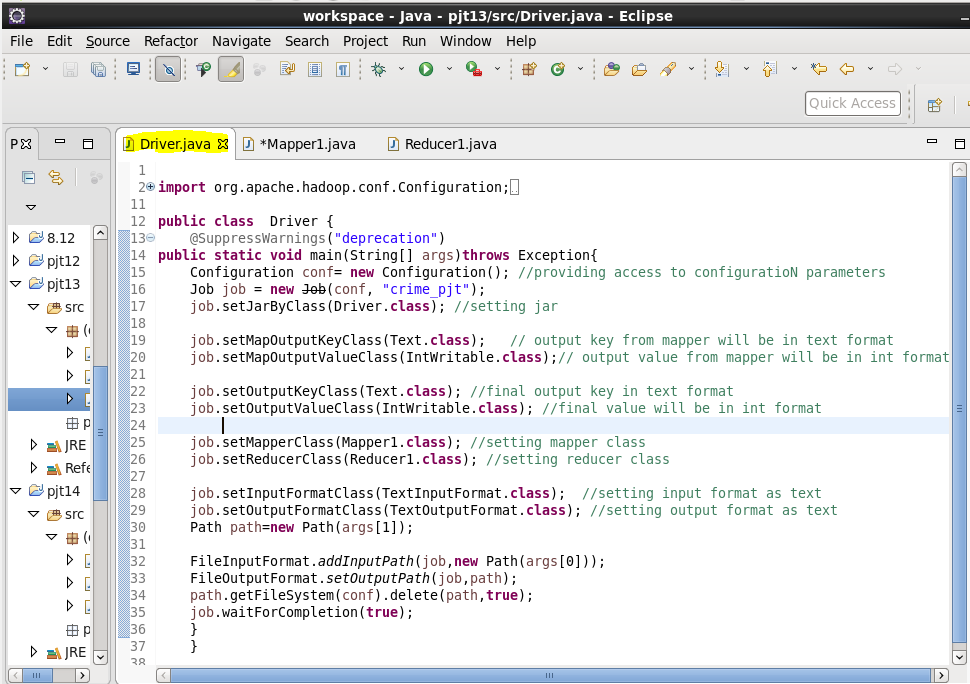
**Output :**

****

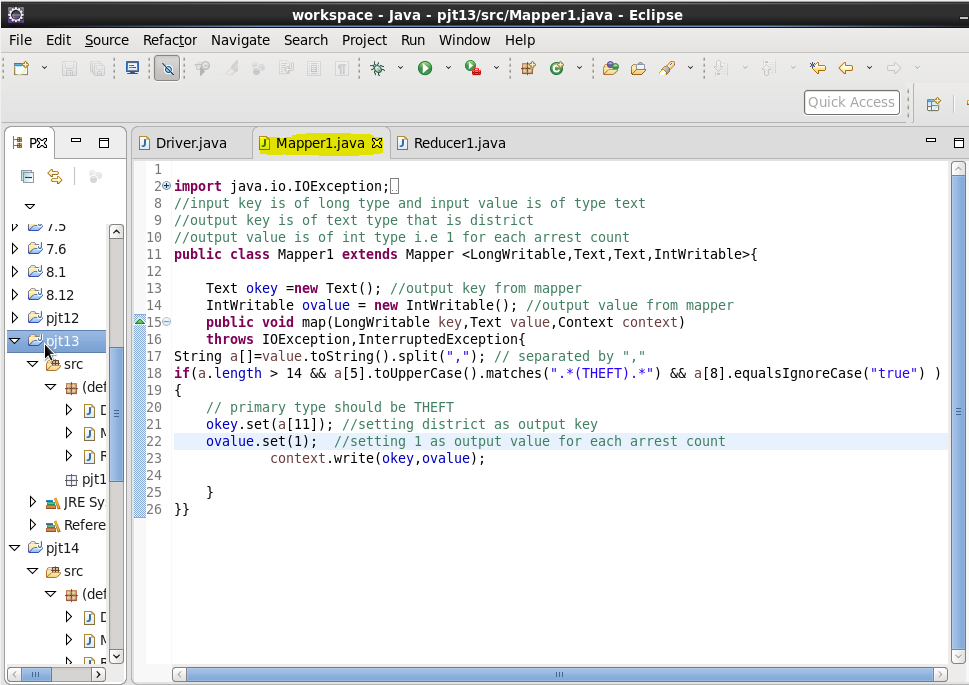
**So,we got the same result from mapreduce and pig hence verified.**

**3- Number of arrests in theft district wise :**

1. **Driver class :**

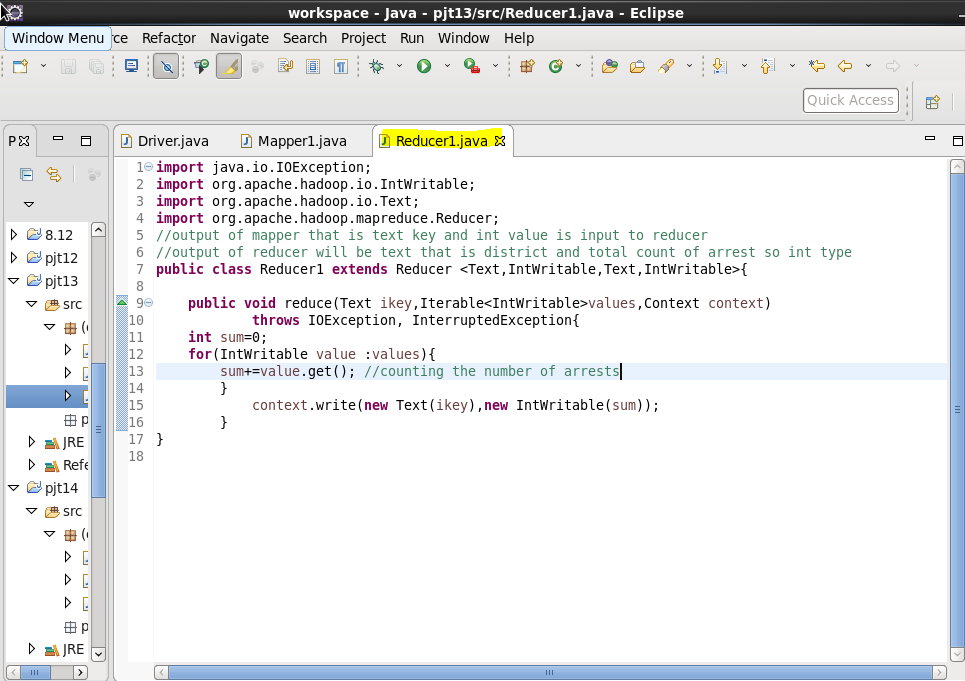
****

1. **Mapper class :**

****

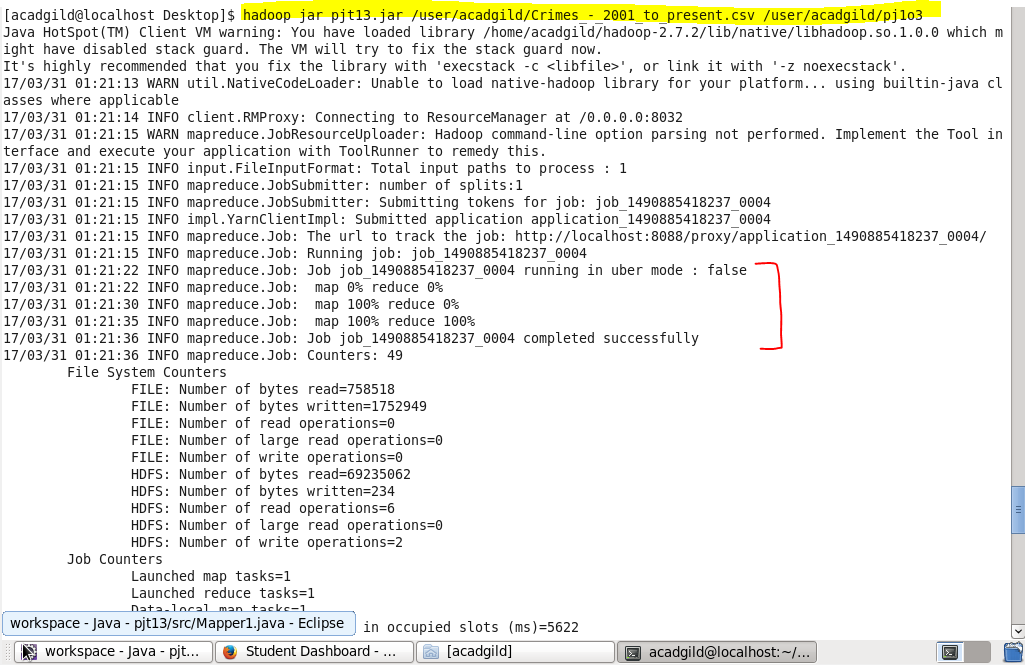
* we have given a[8] equals to true that is we are selecting those data where arrest is done.

1. **Reducer Class:**

****

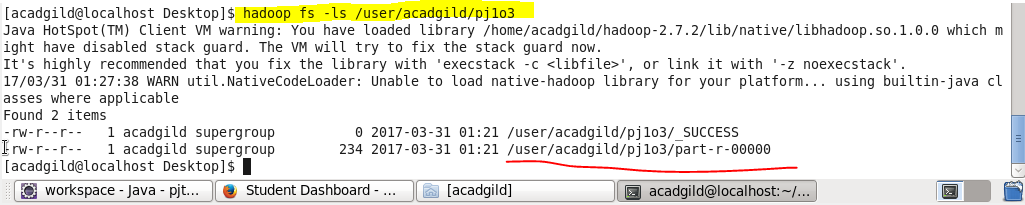
**Running Jar file :**

After Creating jar file of our ja va code we will run the jar file as following :

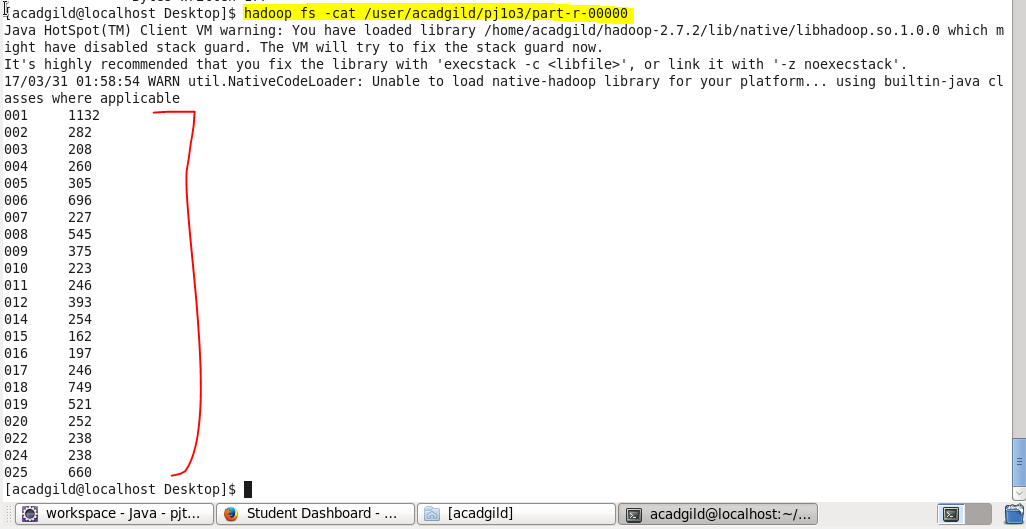


After successfully running jar file we will check weather out output is there in given location or not

So we will list the files in HDFS as:



Now we got our file in particular location,so now we will display the contents of out output using cat command as :

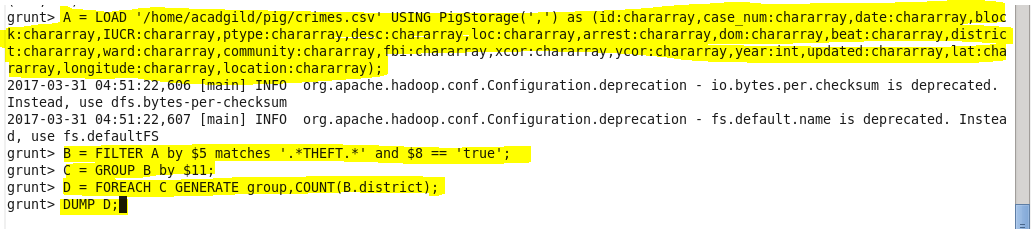


Above screenshots shows the output of our mapreduce program that shows the district in first column and corresponding number of arrests in second column.

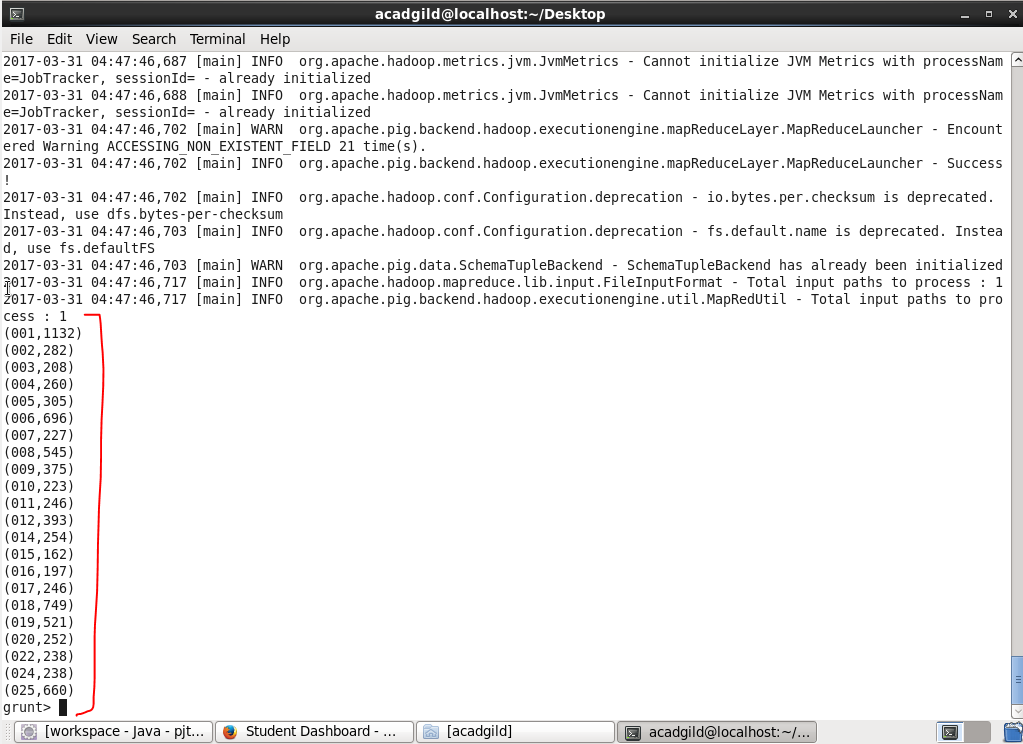
**USING PIG PROGRAM :**

**Pig script :**

* First we will load our file using LOAD command then we will filter our data by primary type matches ‘THEFT’ as we want the count of arrests so we will filter our data with arrest equals ‘true’. Then we will group our data by district column and then we will generate the count using COUNT function.Then using DUMP command we will show the result



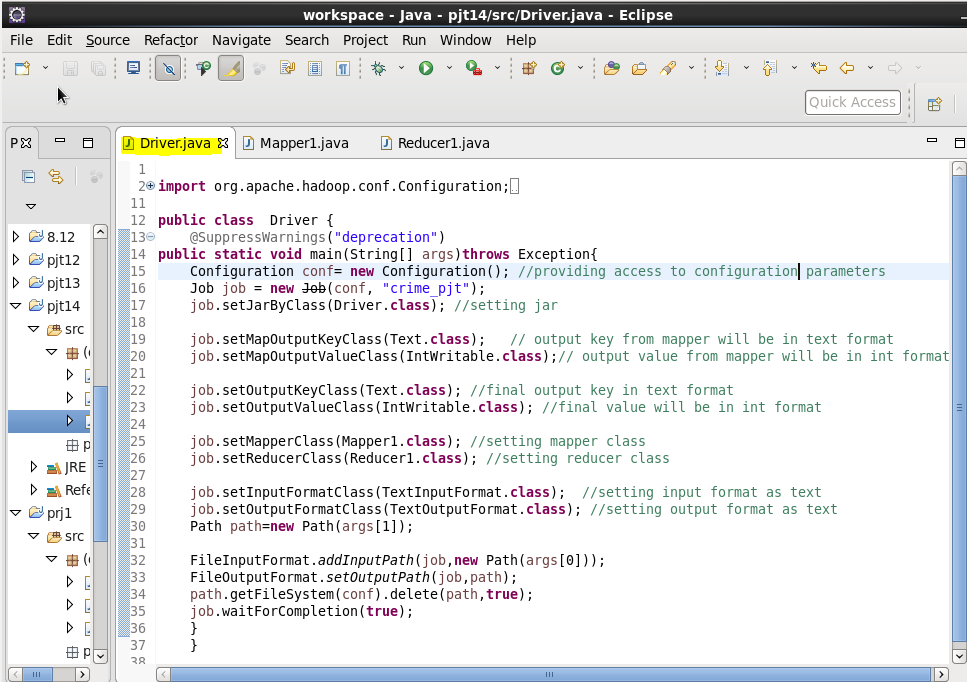
**Output :**

****

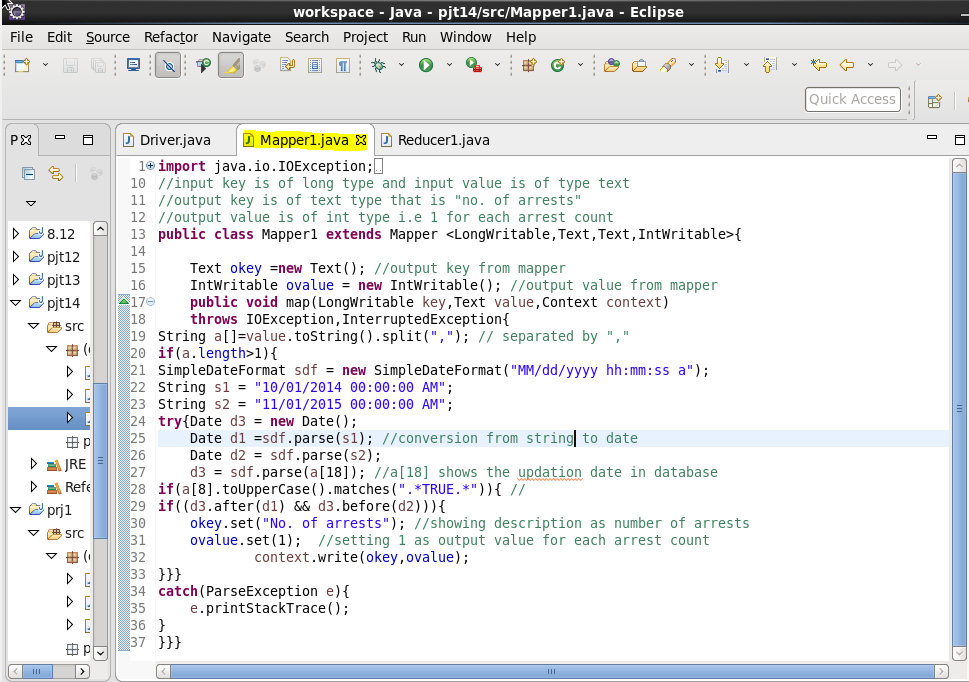
**So,we got the same result from mapreduce and pig hence verified.**

**4- Number of arrests done between October 2014 and October 2015 :**

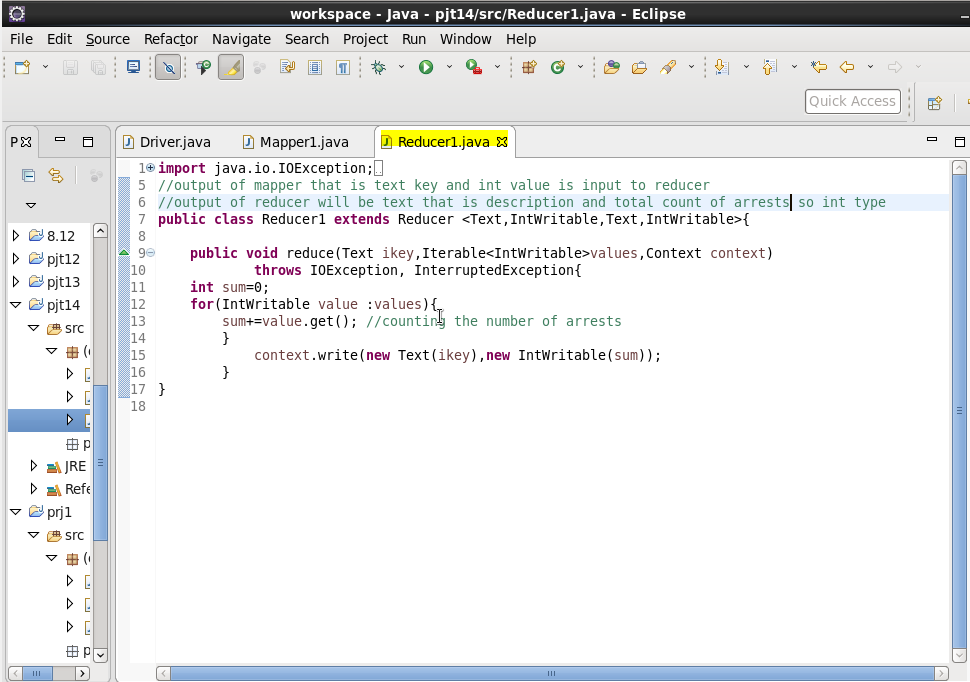
1. **Driver class :**

****

1. **Mapper class :**

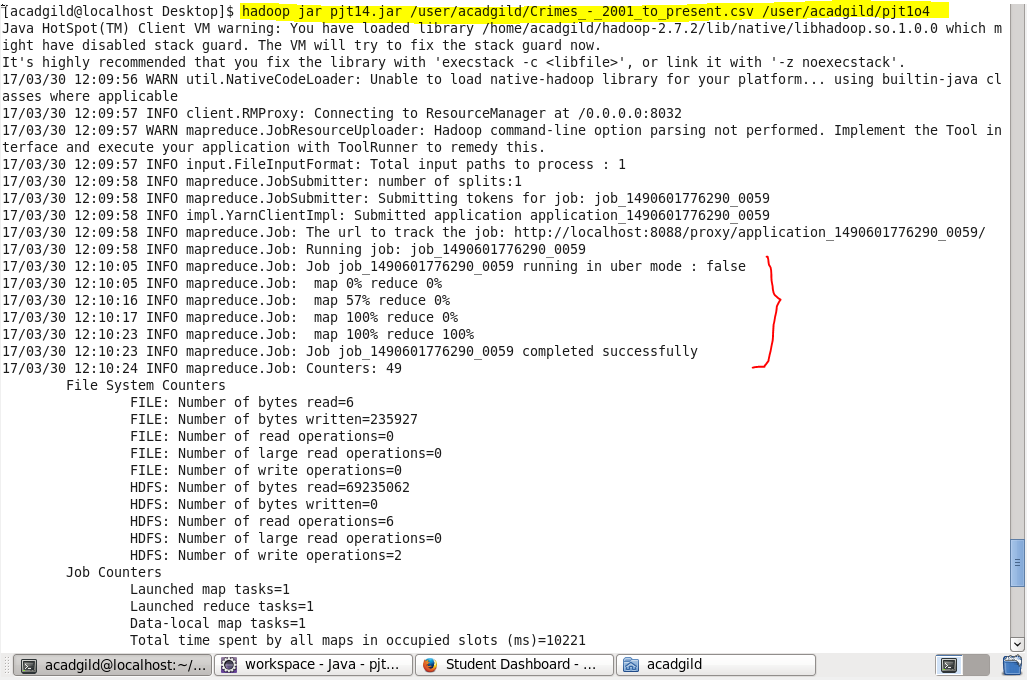
****

1. **Reducer Class:**

****

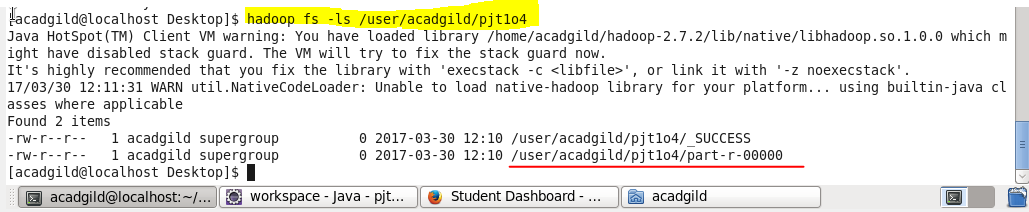
**Running Jar file :**

After Creating jar file of our java code we will run the jar file as following :

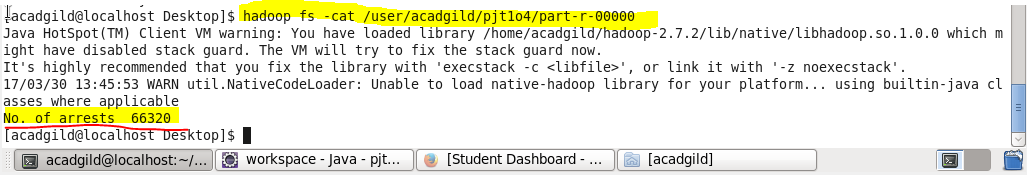


After successfully running jar file we will check weather out output is there in given location or not

So we will list the files in HDFS as:



Now we got our file in particular location,so now we will display the contents of out output using cat command as :

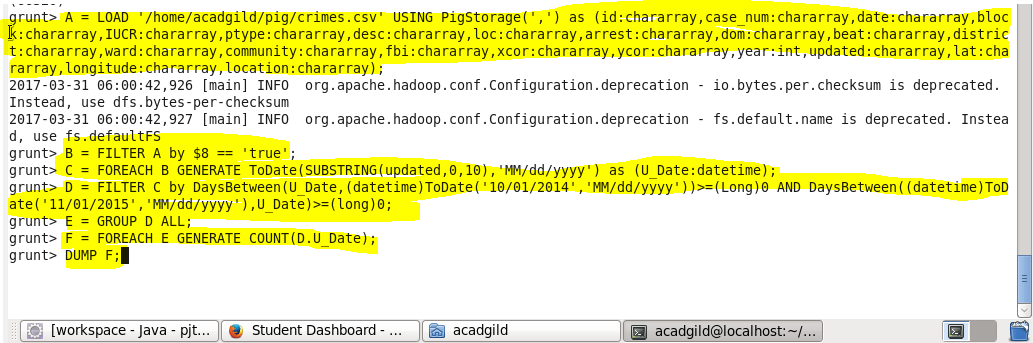


Above screenshots shows the output of our mapreduce program that shows the district in first column and corresponding number of arrests in second column.

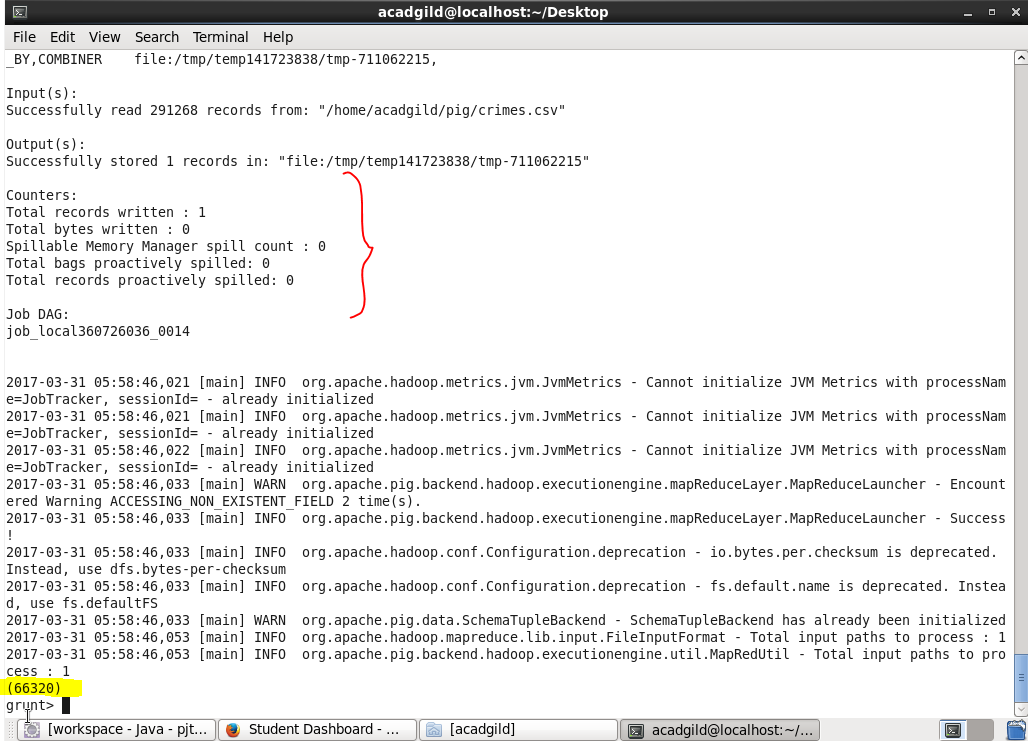
**USING PIG PROGRAM :**

**Pig script :**

* First we will load our file using LOAD command then we will filter our data filter our data with arrest equals ‘true’. Then foreach arrest data we will generate substring that is date of updated date time then we will filter our data and will find if date lies in between given date range then we will group our date and will generate the count of arrest between that date. Then using DUMP command we will show the result.



**Output :**

****

**So,we got the same result from mapreduce and pig**