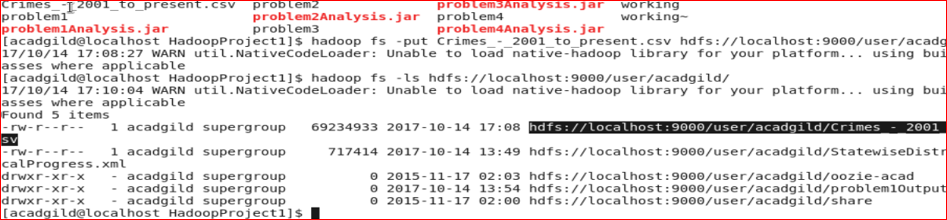
# This document explains the steps included to run map reduce jobs and screenshots of each output

First place the source file into HDFS location using hdfs put command

hadoop fs -put Crimes\_-\_2001\_to\_present.csv dfs://localhost:9000/user/acadgild/

Check file is landed into correct source location using below command

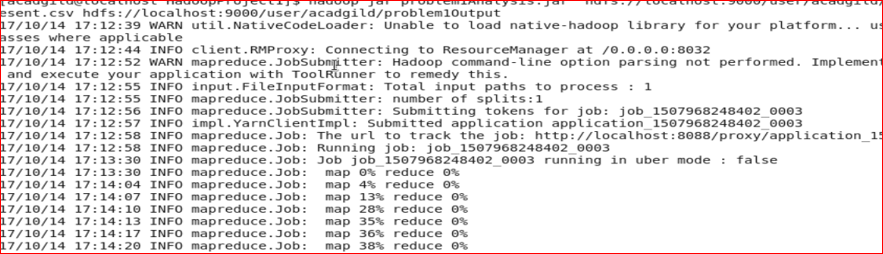
hadoop fs -ls hdfs://localhost:9000/user/acadgild/



# Write a MapReduce to calculate the number of cases investigated under each FBI code

Run the jar using command

hadoop jar problem1Analysis.jar hdfs://localhost:9000/user/acadgild/Crimes\_-\_2001\_to\_present.csv hdfs://localhost:9000/user/acadgild/problem1Output



Check output directory using below command

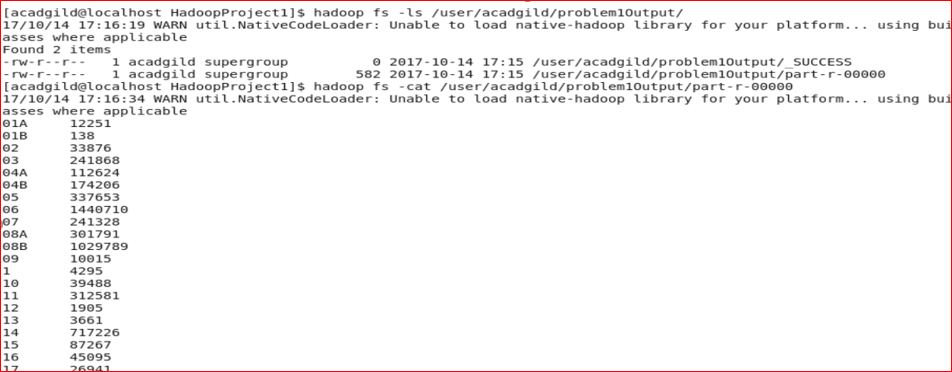
hadoop fs -ls /user/acadgild/problem1Output/

Check the output data using below command

hadoop fs -cat /user/acadgild/problem1Output/part-r-00000

The output shows the number of cases investigated by each fbi code

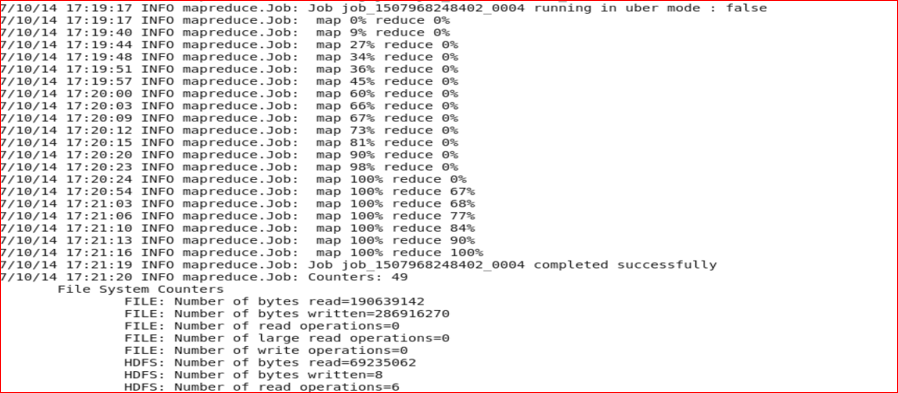
FBI code is in first columns , no of cases got investigated in in the second column



# 2. Write a MapReduce to calculate the number of cases investigated under FBIcode 32.

Run the jar using command

hadoop jar problem2Analysis.jar hdfs://localhost:9000/user/acadgild/Crimes\_-\_2001\_to\_present.csv hdfs://localhost:9000/user/acadgild/problem2Output



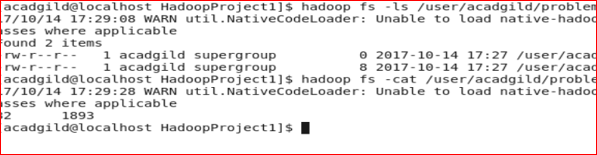
Check output directory using below command

hadoop fs -ls /user/acadgild/problem2Output/

Check the output data using below command

hadoop fs -cat /user/acadgild/problem2Output/part-r-00000

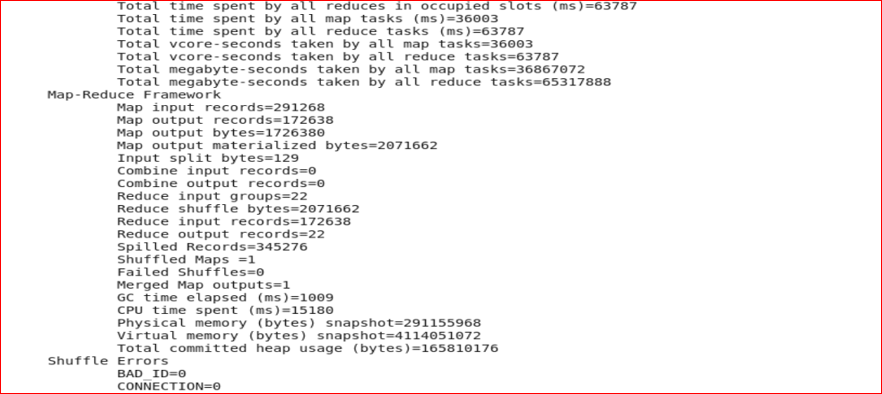
The output shows the no of cases investigated under fbi code 32



# 3. Write a MapReduce to calculate the number of arrests in theft district wise.

Run the jar using command

hadoop jar problem3Analysis.jar hdfs://localhost:9000/user/acadgild/Crimes\_-\_2001\_to\_present.csv hdfs://localhost:9000/user/acadgild/problem3Output



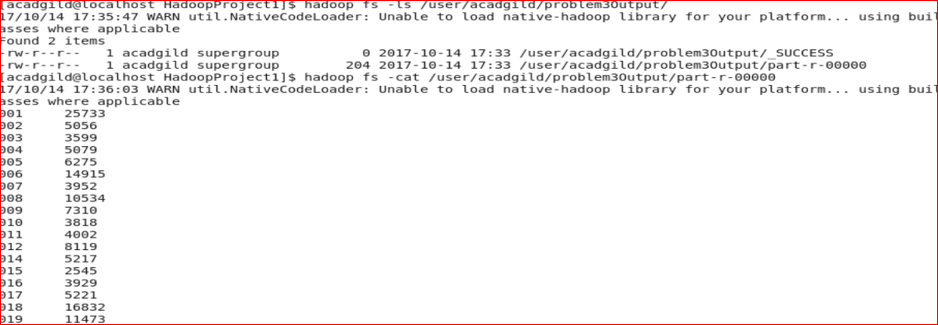
Check output directory using below command

hadoop fs -ls /user/acadgild/problem3Output/

Check the output data using below command

hadoop fs -cat /user/acadgild/problem3Output/part-r-00000

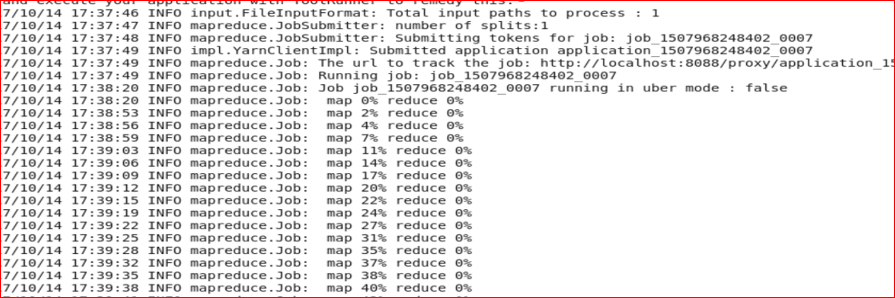
The output , the first column represents district and second column represents the number of arrests



# 4. Write a MapReduce/Pig program to calculate the number of arrests done between October 2014 and October 2015.

Run the jar using command

hadoop jar problem4Analysis.jar hdfs://localhost:9000/user/acadgild/Crimes\_-\_2001\_to\_present.csv hdfs://localhost:9000/user/acadgild/problem4Output



Check output directory using below command

hadoop fs -ls /user/acadgild/problem4Output/

Check the output data using below command

hadoop fs -cat /user/acadgild/problem4Output/part-r-00000

The output displays the no of arrests between the time duration specifid in the problem

