**PROJECT 1**

Dataset description

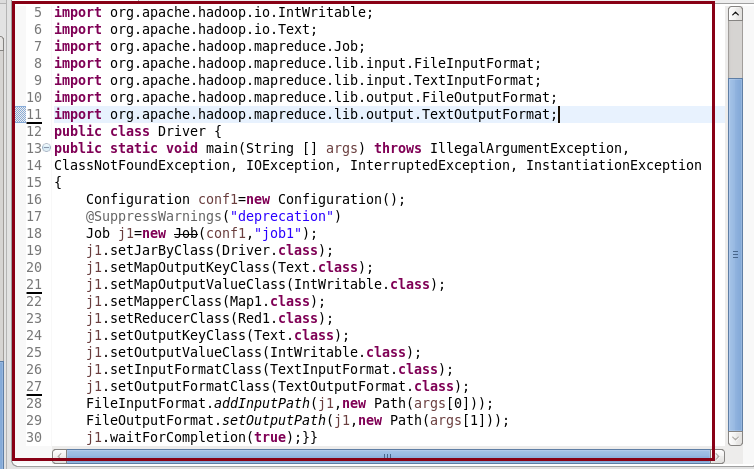
ID,Case Number,Date,Block,IUCR,Primary Type,Description,Location Description,Arrest,Domestic,Beat,District,Ward,Community Area,FBICode,X Coordinate,Y Coordinate,Year,Updated On,Latitude,Longitude,Location

Problem Statement :

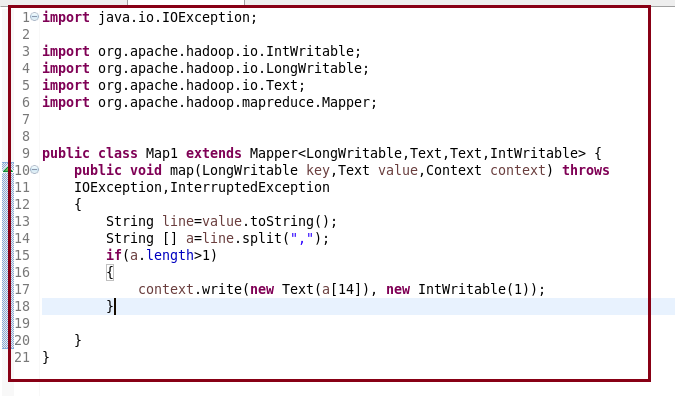
1. Write a mapreduce and pig program to calculate the number of cases investigated under each FBI code.

Mapreduce program:

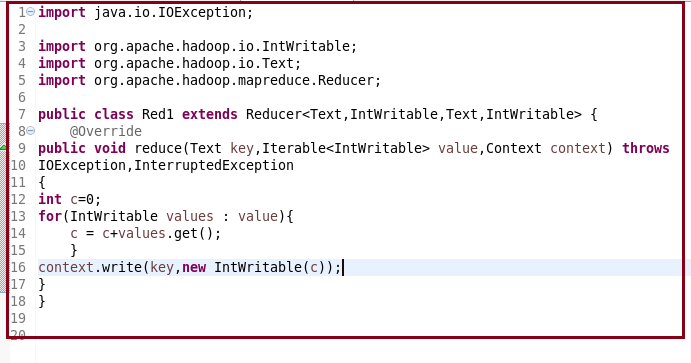
Driver class



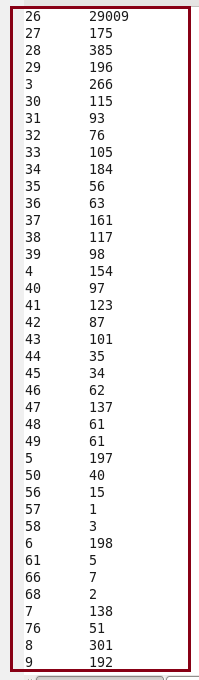
Mapper class



Reducer class



Output :



PIG CODE:

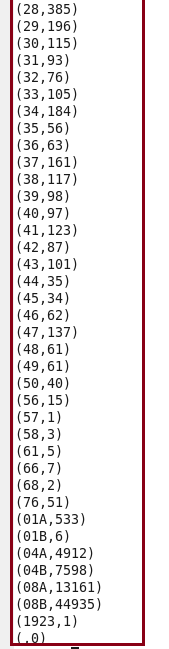
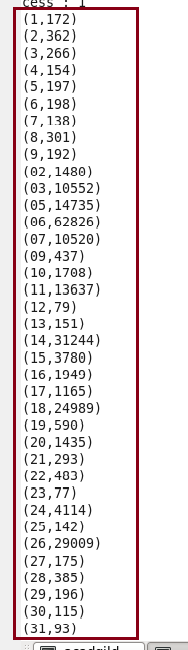
A = LOAD ‘/home/acadgild/crimes.csv’ USING PigStorage (‘,’) AS (ID:chararray,CaseNumber:chararray,Date:chararray, Block:chararray, IUCR:chararray ,PrimaryType:chararray, Description:chararray, Location Description:chararray, Arrest:chararray, Domestic:chararray, Beat:chararray, District:chararray, Ward:chararray, CommunityArea:chararray, FBICode:chararray, XCoordinate:chararray, YCoordinate:chararray, Year:chararray, UpdatedOn:chararray, Latitude:chararray, Longitude:chararray, Location:chararray);

B = GROUP A by $14;

C = FOREACH B generate group,COUNT(A.FBICODE);

DUMP C:

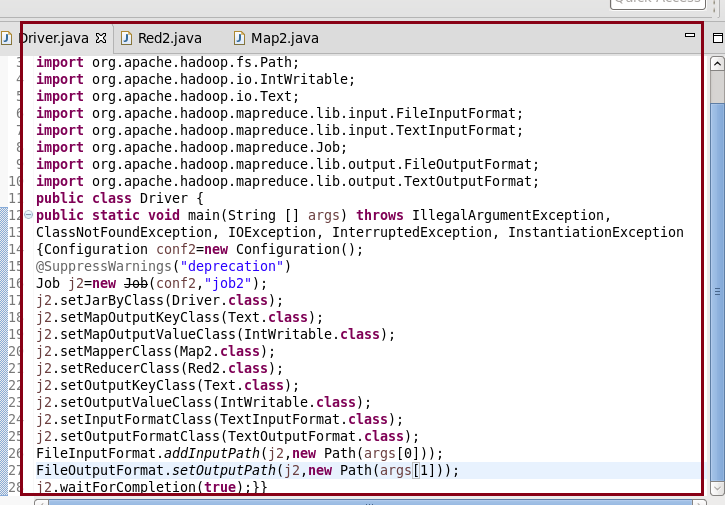
Output :



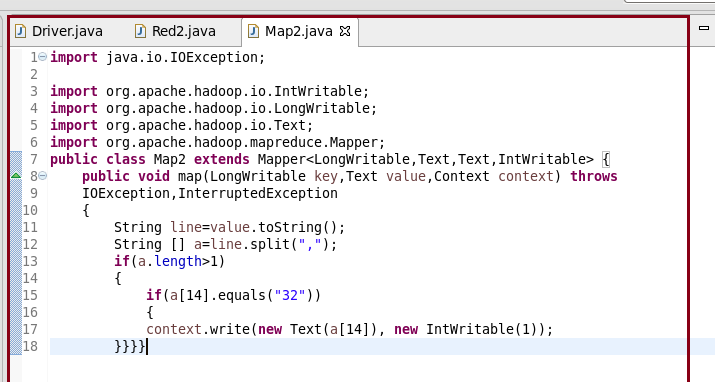
1. Write a mapreduce and pig program to calculate the number of cases investigated under FBI code 32.

Mapreduce program:

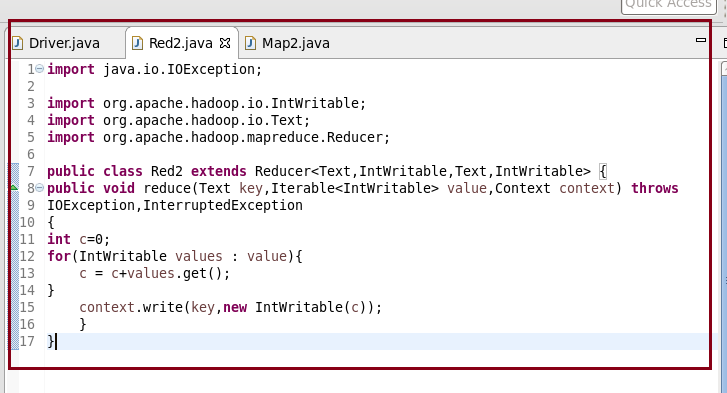
Driver class:



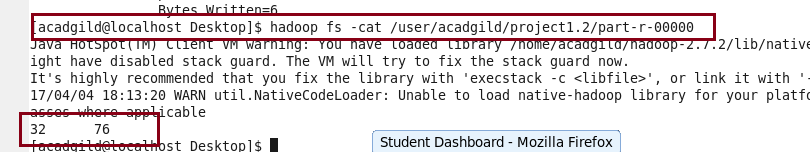
Mapper Class:



Reducer class :



Output :



PIG CODE :

A = LOAD ‘/home/acadgild/crimes.csv’ USING PigStorage (‘,’) AS (ID:chararray,CaseNumber:chararray,Date:chararray, Block:chararray, IUCR:chararray ,PrimaryType:chararray, Description:chararray, Location Description:chararray, Arrest:chararray, Domestic:chararray, Beat:chararray, District:chararray, Ward:chararray, CommunityArea:chararray, FBICode:chararray, XCoordinate:chararray, YCoordinate:chararray, Year:chararray, UpdatedOn:chararray, Latitude:chararray, Longitude:chararray, Location:chararray);

B = FILTER A by FBICODE == ‘32’;

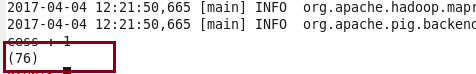
C = GROUP B by FBICODE;

D =FOREACH C generate group,COUNT(B.FBICODE);

E = FOREACH D generate $1;

DUMP E;

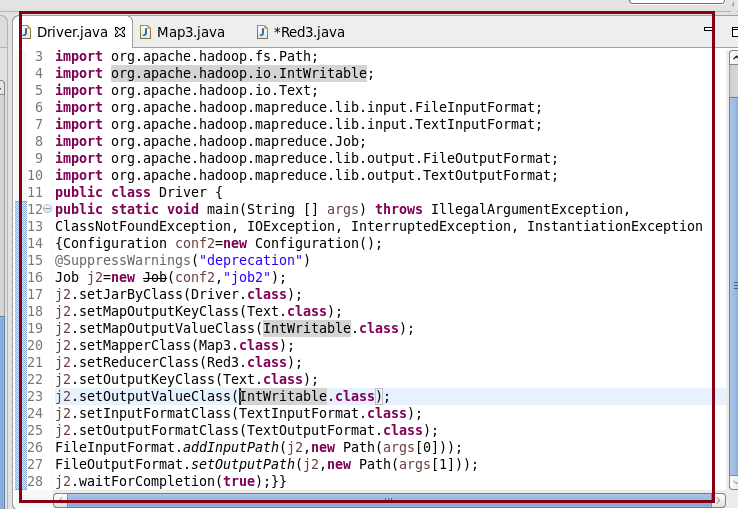
Output :



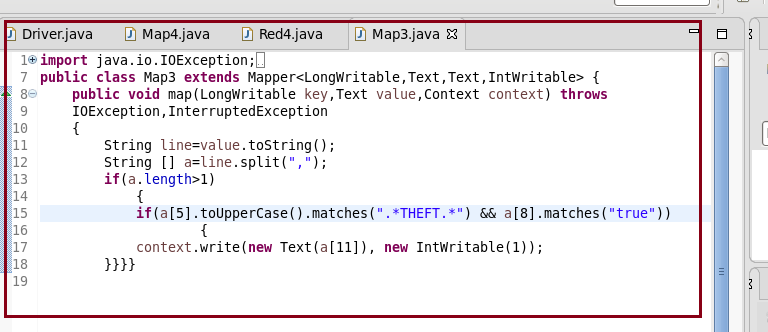
1. Write a mapreduce and pig program to calculate the number of arrests in theft district wise.

Mapreduce Program:

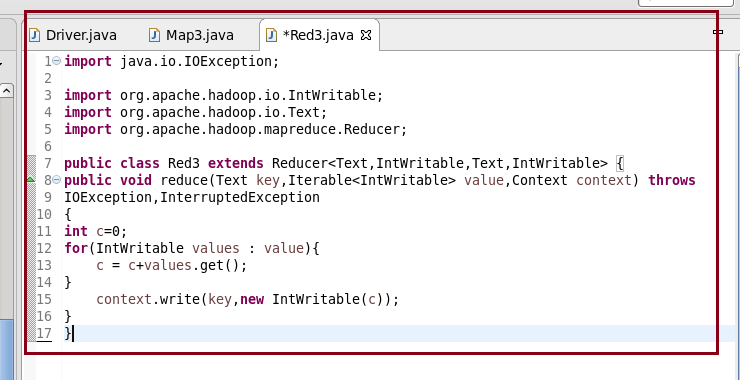
Driver class :



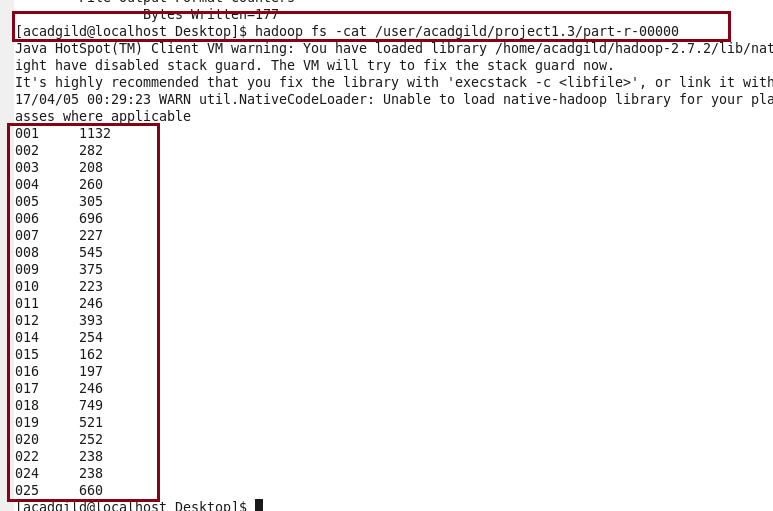
Mapper class :



Reducer class:



Output :



PIG CODE :

A = LOAD ‘/home/acadgild/crimes.csv’ USING PigStorage (‘,’) AS (ID:chararray,CaseNumber:chararray,Date:chararray, Block:chararray, IUCR:chararray ,PrimaryType:chararray, Description:chararray, Location Description:chararray, Arrest:chararray, Domestic:chararray, Beat:chararray, District:chararray, Ward:chararray, CommunityArea:chararray, FBICode:chararray, XCoordinate:chararray, YCoordinate:chararray, Year:chararray, UpdatedOn:chararray, Latitude:chararray, Longitude:chararray, Location:chararray);

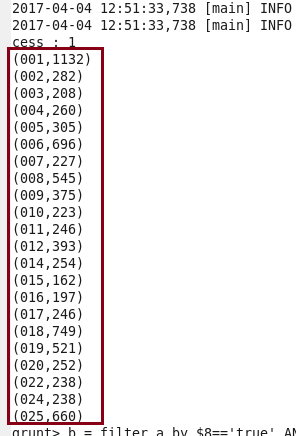
B = FILTER A by $8==’true’ && ($5 matches ‘.\*THEFT.\*’);

C = GROUP B by $11;

D = FOREACH C generate group,COUNT(B.District);

DUMP D;

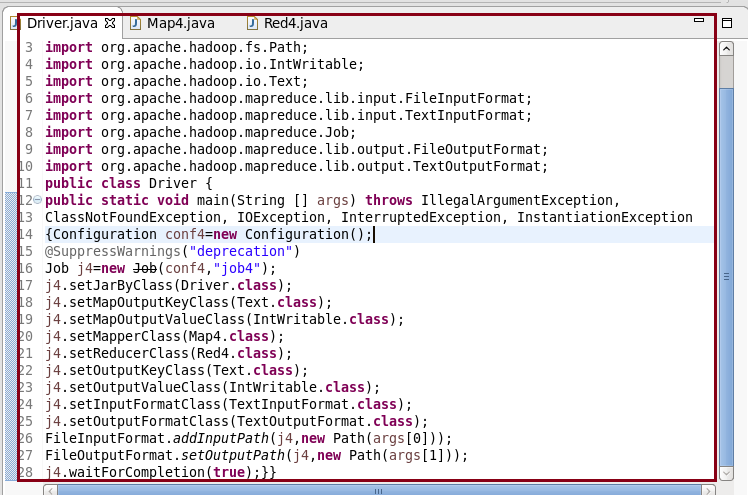
Output :



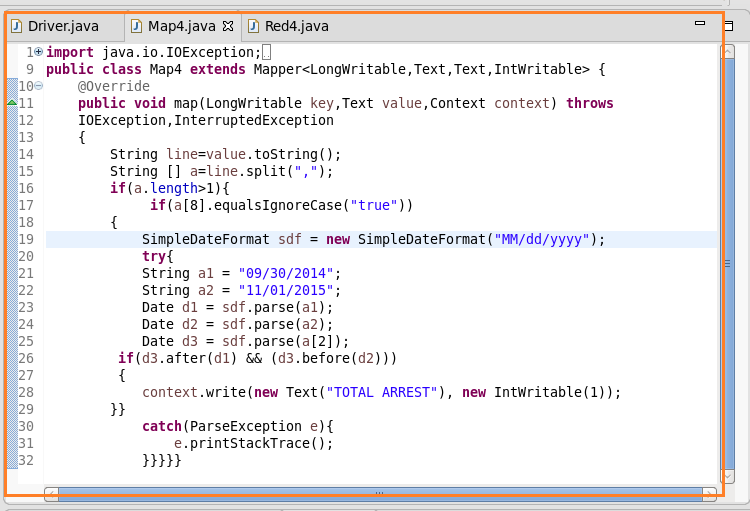
1. Write a mapreduce and pig program to calculate the number of arrests done between October 2014 and October 2015.

Mapreduce program:

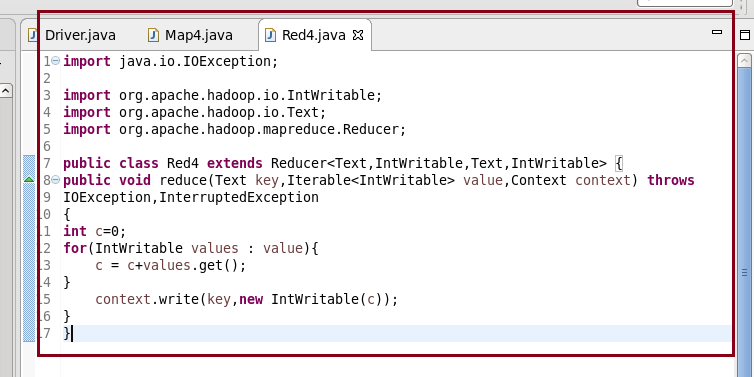
Driver class:



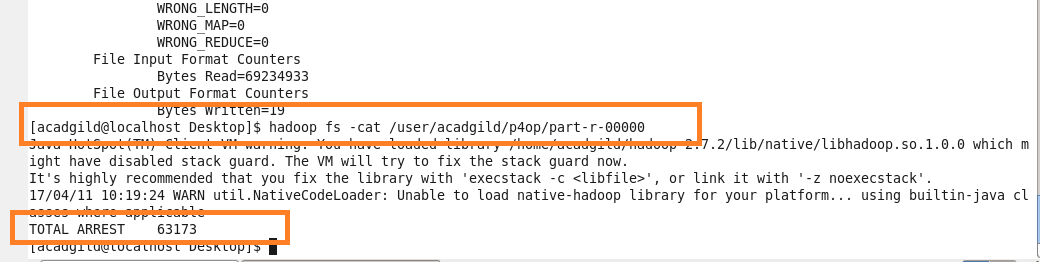
Mapper class:



Reducer class:



Output :



PIG CODE :

A = LOAD ‘/home/acadgild/crimes.csv’ USING PigStorage (‘,’) AS (ID:chararray,CaseNumber:chararray,Date:chararray, Block:chararray, IUCR:chararray ,PrimaryType:chararray, Description:chararray, Location Description:chararray, Arrest:chararray, Domestic:chararray, Beat:chararray, District:chararray, Ward:chararray, CommunityArea:chararray, FBICode:chararray, XCoordinate:chararray, YCoordinate:chararray, Year:chararray, UpdatedOn:chararray, Latitude:chararray, Longitude:chararray, Location:chararray);

B = FOREACH A generate ToDate(date,’MM/dd/yyyy hh:mm:ss aa’) as (DT:DateTime),Arrest;

C = FILTER B by DT >= (10/01/2014 12:00:00 AM,’MM/dd/yyyy hh:mm:ss aa’) AND DT < (11/01/2015 12:00:00 AM,’MM/dd/yyyy hh:mm:ss aa’) AND Arrest == ‘true’;

D = GROUP C by Arrest;

E = FOREACH D generate COUNT(C.Arrest);

DUMP E;

Output:

