**USA Crime Analysis:**

1. Write a MapReduce/Pig program to calculate the number of cases investigated under each FBI code

**baseRel** = LOAD '/home/acadgild/hadoop/Crimes\_-\_2001\_to\_present.csv' USING PigStorage(',') AS (crimeID:int,CaseNumber:chararray,Date:chararray,

Block:chararray,IUCR:int,PrimaryType:chararray,Description:chararray,LocationDescription:chararray,Arrest:chararray,Domestic:chararray,

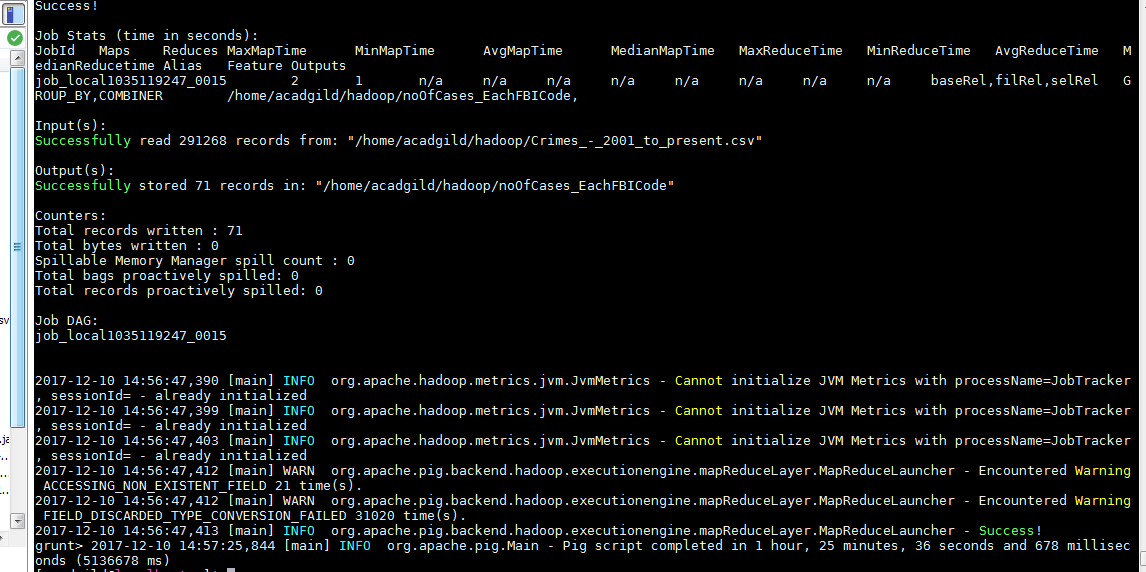
Beat:int,District:int,Ward:int,CommunityArea:int,FBICode:chararray,XCoordinate:int,YCoordinate:int,Year:chararray,UpdatedOn:chararray,

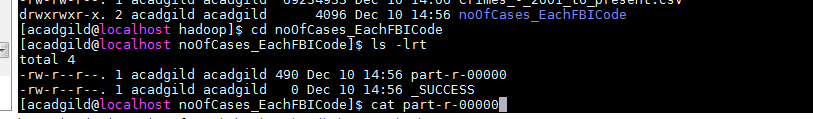
Latitude:chararray,Longitude:chararray,Location:chararray);

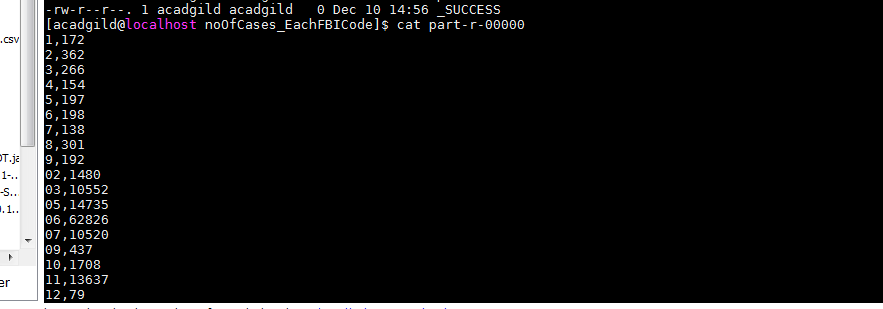
**filRel** = GROUP baseRel BY FBICode;

**selRel** = FOREACH filRel GENERATE group as FBICode,COUNT(baseRel) as NoOfCases;

**STORE** selRel INTO '/home/acadgild/hadoop/noOfCases\_EachFBICode' USING PigStorage(',');







2. Write a MapReduce/Pig program to calculate the number of cases investigated under FBI code 32.

**baseRel** = LOAD '/home/acadgild/hadoop/Crimes\_-\_2001\_to\_present.csv' USING PigStorage(',') AS (crimeID:int,CaseNumber:chararray,Date:chararray,

Block:chararray,IUCR:int,PrimaryType:chararray,Description:chararray,LocationDescription:chararray,Arrest:chararray,Domestic:chararray,

Beat:int,District:int,Ward:int,CommunityArea:int,FBICode:chararray,XCoordinate:int,YCoordinate:int,Year:chararray,UpdatedOn:chararray,

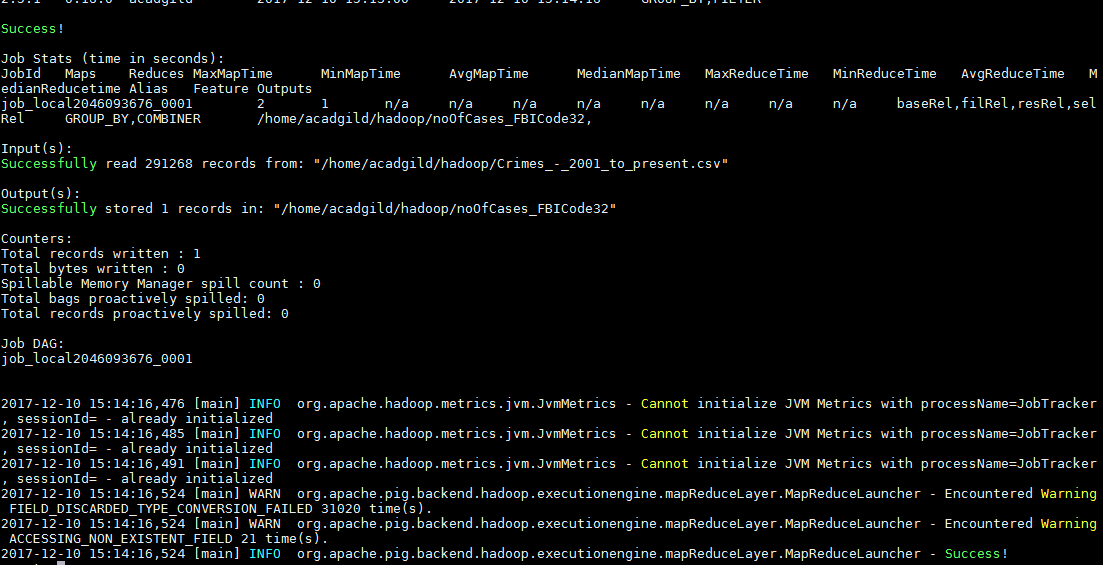
Latitude:chararray,Longitude:chararray,Location:chararray);

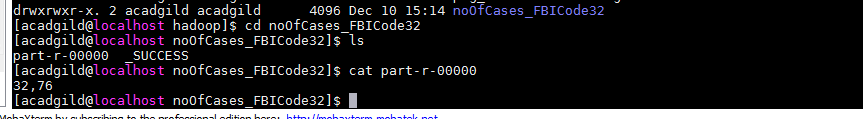
**filRel** = GROUP baseRel BY FBICode;

**selRel** = FOREACH filRel GENERATE group as FBICode,COUNT(baseRel) as NoOfCases;

**resRel** = FILTER selRel BY FBICode=='32';

STORE resRel INTO '/home/acadgild/hadoop/noOfCases\_FBICode32' USING PigStorage(',');





3. Write a MapReduce/Pig program to calculate the number of arrests in theft district wise.

baseRel = LOAD '/home/acadgild/hadoop/Crimes\_-\_2001\_to\_present.csv' USING PigStorage(',') AS (crimeID:int,CaseNumber:chararray,Date:chararray,

Block:chararray,IUCR:int,PrimaryType:chararray,Description:chararray,LocationDescription:chararray,Arrest:chararray,Domestic:chararray,

Beat:int,District:int,Ward:int,CommunityArea:int,FBICode:chararray,XCoordinate:int,YCoordinate:int,Year:chararray,UpdatedOn:chararray,

Latitude:chararray,Longitude:chararray,Location:chararray);

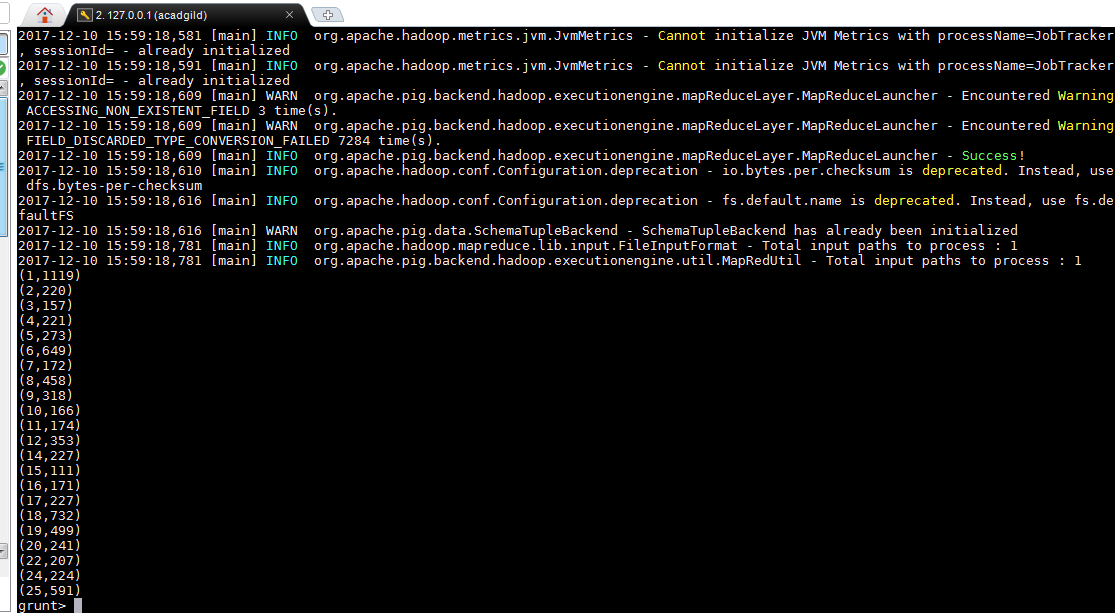
selRel = FOREACH baseRel GENERATE PrimaryType,Arrest,District;

filRel = FILTER selRel BY PrimaryType=='THEFT' and Arrest=='true';

grpRel = GROUP filRel BY District;

resRel = FOREACH grpRel GENERATE group as District,COUNT(filRel) as NumberOfDist;

dump resRel;



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

**baseRel** = LOAD '/home/acadgild/hadoop/Crimes\_-\_2001\_to\_present.csv' USING PigStorage(',') AS (crimeID:int,CaseNumber:chararray,Date:chararray,

Block:chararray,IUCR:int,PrimaryType:chararray,Description:chararray,LocationDescription:chararray,Arrest:chararray,Domestic:chararray,

Beat:int,District:int,Ward:int,CommunityArea:int,FBICode:chararray,XCoordinate:int,YCoordinate:int,Year:chararray,UpdatedOn:chararray,

Latitude:chararray,Longitude:chararray,Location:chararray);

**selRel** = FOREACH baseRel GENERATE Date, Arrest;

filRel = FILTER selRel BY Arrest=='true' and (Date >= '10/1/2014 00:00:00 AM' and Date <= '10/31/2015 23:59:59 PM') ;

**grpRel** = Group filRel BY Date;

**resRel** = FOREACH grpRel GENERATE group as Date,COUNT(filRel) as NumberOfArrests;

g = GROUP resRel BY Date;

s = FOREACH g GENERATE group, SUM(resRel.NumberOfArrests) as Sum;

DUMP s;