**PROJECT I**

\*\*We have data file: usa\_crime\_data(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)

>pig -x local

grunt>crime\_data = LOAD ' usa\_c rime\_data.csv ' USING PigStorage(',') AS (ID:int, CaseNumber:chararray, Date:datetime, Block: chararray, IUCR: chararray, PrimaryType: chararray, Description:chararray, LocationDescription: chararray, Arrest: chararray, Domestic: chararray, Beat:, District: chararray, Ward: chararray, CommunityArea: chararray, FBICode:int, XCoordinate:float, YCoordinate:float, Year:, UpdatedOn: datetime, Latitude:double, Longitude:double, Location: chararray);

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

>cases\_with\_fbicode = GROUP crime\_data BY FBICode;

>result = FOREACH cases\_with\_fbicode GENERATE group, COUNT(crime\_data) ;

>DUMP result;

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

>filtered\_cases = FILTER crime\_data BY FBICode==32;

>cases\_with\_fbicode32 = GROUP filtered\_cases BY FBICode;

>result = FOREACH cases\_with\_fbicode32 GENERATE group, COUNT(filtered\_cases.ID) ;

>DUMP result;

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

>input1 = FILTER crime\_data BY PrimaryType=="THEFT" && Arrest =="TRUE"

>input2 = GROUP input1 BY District;

>Result = FOREACH input2 GENERATE group, COUNT(input1.ID);

>DUMP Result;

**4.Write a mapreduce/pig program to calculate the number of arrests done between october 2014 and october 2015.**

>data\_by\_monthyear = FOREACH crime\_data GENERATE ID, Date.getMonth() AS Month , Date.getYear() AS Year, Arrest;

>filtered\_data = FILTER data\_by\_monthyear BY Arrest=="TRUE" && Month == "October" && Year ==2014 OR Year==2015;

>group\_data = GROUP filtered\_data BY All;

>Result = FOREACH group\_data GENERATE group, COUNT(filtered\_data.ID);

>DUMP Result;