**Crime Dataset:**

The columns in Crime.csv are Area\_Name,Year,Group\_Name,Sub\_Group\_Name, Cases\_Property\_Recovered,Cases\_Property\_Stolen,Value\_of\_Property\_Recovered,Value\_of\_Property\_Stolen

1.Remove the headers

2.Copy the local file in hdfs

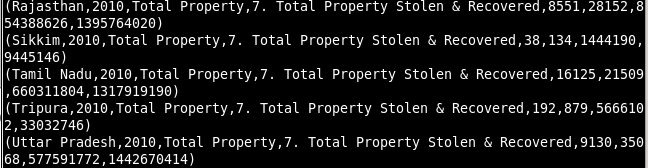
#hadoop dfs -put 'file' /

**LOAD Dataset:**

crime\_record = LOAD '/Crime.csv' using PigStorage(',') as (area:chararray,year:int,groups:chararray,subgroups:chararray,casesrecovered:int,casesstolen:int,valuerecovered:int,valuestolen:int);

**View Dataset:**

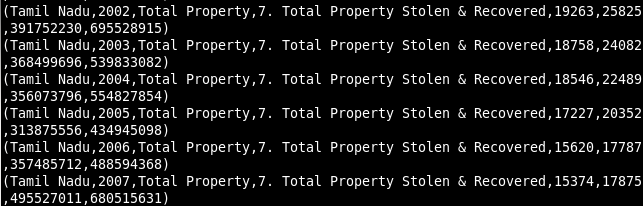
dump crime\_record;



**Analysing Dataset:**

**1)Crimes accured in TamilNadu:**

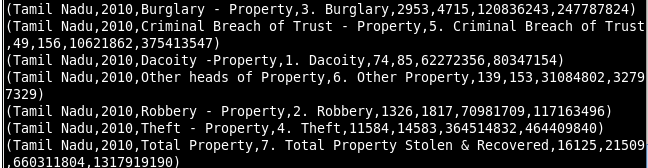
filter\_data = FILTER crime\_record by area == 'Tamil Nadu';

 dump filter\_data;

**2)Crimes in TamilNadu at 2010:**

filter\_data = FILTER crime\_record by area == 'Tamil Nadu' and year==2010;

dump filter\_data;

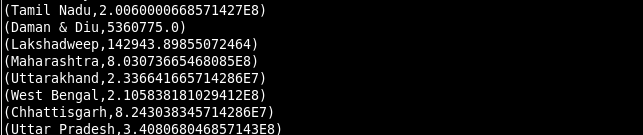


**3)Crimes Average Value Stolen by Area:**

GroupByArea = GROUP crime\_record by area;

dump GroupByArea;

area\_avg\_value = foreach GroupByArea GENERATE group,AVG(crime\_record.valuestolen);

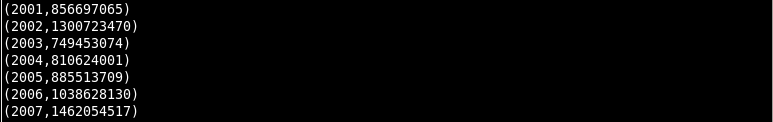
 dump area\_avg\_value;

**4)Maximum Value Recovered by year:**

GroupByyear = GROUP crime\_record by year;

dump GroupByyear;

max\_val\_rec = FOREACH GroupByyear GENERATE group,MAX(crime\_record.valuerecovered);

 dump max\_val\_rec;

**5)Average ValueStolen of cases by area and year:**

GroupByareayear = GROUP crime\_record by (area,year);

dump GroupByareayear;

avg\_area\_year = FOREACH GroupByareayear GENERATE group, AVG(crime\_record.valuestolen);

dump avg\_area\_year;

