**Road Accident Dataset:**

The columns in road\_accident.csv STATE/UT,YEAR,JANUARY,FEBRUARY,MARCH,APRIL,MAY,JUNE,JULY,AUGUST,SEPTEMBER,OCTOBER,NOVEMBER,DECEMBER,TOTAL

1.Remove the headers

2.Rename to '.txt'

3.Copy the local file in hdfs

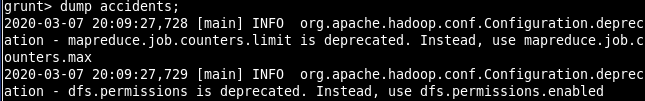
#hadoop dfs -put 'file' /

**LOAD Dataset:**

accidents = LOAD '/road\_accidents.txt' using PigStorage (',') as (state:chararray,year:int,jan:int,feb:int,mar:int,apr:int,may:int,jun:int,jul:int,aug:int,sep:int,oct:int,nov:int,dec:int,total:int);

**View Dataset:**

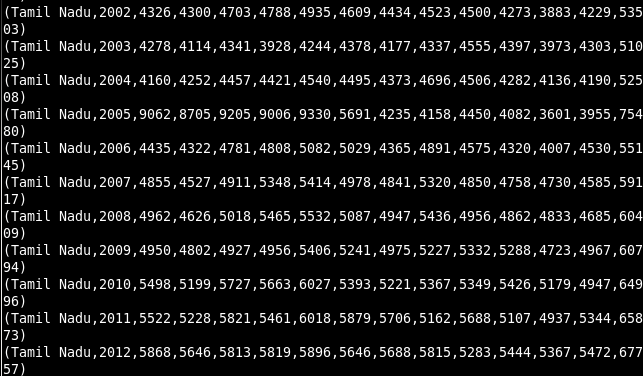
dump accidents;



**Analysing Dataset:**

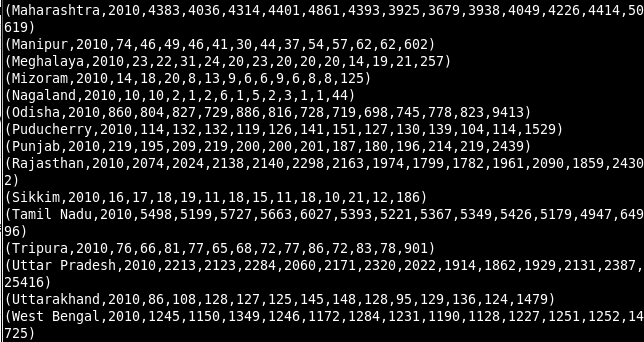
**1)Accindents Records in TamilNadu:**

filter\_data = FILTER accidents BY state == 'Tamil Nadu';dump filter\_data;



**2)Accident taken place in 2010:**

filter\_data = FILTER accidents BY year == 2010;dump filter\_data;

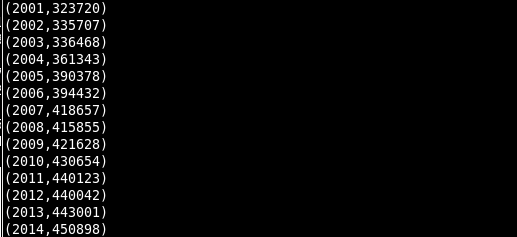


**3)Accident count by year:**

GroupByYear = GROUP accidents BY year;

dump GroupByYear;

accidents\_year = foreach GroupByYear Generate group,SUM(accidents.total);

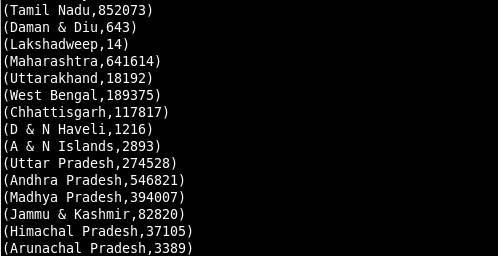
 dump accidents\_year

**4)Accident count by state:**

GroupByState = GROUP accidents BY state;

dump GroupByState ;

accidents\_state = foreach GroupByState Generate group,SUM(accidents.total);

 dump accidents\_state;

**5) Average accidents in January By Year:**

GroupByYear = GROUP accidents BY year;

dump GroupByYear;

accidents\_year\_avg = foreach GroupByYear Generate group,AVG(accidents.jan);

dump accidents\_year\_avg;

