

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;
```

```
grunt> dump accidents;  
2020-03-07 20:09:27,728 [main] INFO org.apache.hadoop.conf.Configuration.deprec  
ation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.c  
ounters.max  
2020-03-07 20:09:27,729 [main] INFO org.apache.hadoop.conf.Configuration.deprec  
ation - dfs.permissions is deprecated. Instead, use dfs.permissions.enabled
```

Analysing Dataset:

1)Accidents Records in TamilNadu:

```
filter_data = FILTER accidents BY state == 'Tamil Nadu';dump filter_data;
```

```
(Tamil Nadu,2002,4326,4300,4703,4788,4935,4609,4434,4523,4500,4273,3883,4229,535  
03)  
(Tamil Nadu,2003,4278,4114,4341,3928,4244,4378,4177,4337,4555,4397,3973,4303,510  
25)  
(Tamil Nadu,2004,4160,4252,4457,4421,4540,4495,4373,4696,4506,4282,4136,4190,525  
08)  
(Tamil Nadu,2005,9062,8705,9205,9006,9330,5691,4235,4158,4450,4082,3601,3955,754  
80)  
(Tamil Nadu,2006,4435,4322,4781,4808,5082,5029,4365,4891,4575,4320,4007,4530,551  
45)  
(Tamil Nadu,2007,4855,4527,4911,5348,5414,4978,4841,5320,4850,4758,4730,4585,591  
17)  
(Tamil Nadu,2008,4962,4626,5018,5465,5532,5087,4947,5436,4956,4862,4833,4685,604  
09)  
(Tamil Nadu,2009,4950,4802,4927,4956,5406,5241,4975,5227,5332,5288,4723,4967,607  
94)  
(Tamil Nadu,2010,5498,5199,5727,5663,6027,5393,5221,5367,5349,5426,5179,4947,649  
96)  
(Tamil Nadu,2011,5522,5228,5821,5461,6018,5879,5706,5162,5688,5107,4937,5344,658  
73)  
(Tamil Nadu,2012,5868,5646,5813,5819,5896,5646,5688,5815,5283,5444,5367,5472,677  
57)
```

2)Accident taken place in 2010:

```
filter_data = FILTER accidents BY year == 2010;dump filter_data;
```

```
(Maharashtra,2010,4383,4036,4314,4401,4861,4393,3925,3679,3938,4049,4226,4414,50619)
(Manipur,2010,74,46,49,46,41,30,44,37,54,57,62,62,602)
(Meghalaya,2010,23,22,31,24,20,23,20,20,20,14,19,21,257)
(Mizoram,2010,14,18,20,8,13,9,6,6,9,6,8,8,125)
(Nagaland,2010,10,10,2,1,2,6,1,5,2,3,1,1,44)
(Odisha,2010,860,804,827,729,886,816,728,719,698,745,778,823,9413)
(Puducherry,2010,114,132,132,119,126,141,151,127,130,139,104,114,1529)
(Punjab,2010,219,195,209,219,200,200,201,187,180,196,214,219,2439)
(Rajasthan,2010,2074,2024,2138,2140,2298,2163,1974,1799,1782,1961,2090,1859,24302)
(Sikkim,2010,16,17,18,19,11,18,15,11,18,10,21,12,186)
(Tamil Nadu,2010,5498,5199,5727,5663,6027,5393,5221,5367,5349,5426,5179,4947,64996)
(Tripura,2010,76,66,81,77,65,68,72,77,86,72,83,78,901)
(Uttar Pradesh,2010,2213,2123,2284,2060,2171,2320,2022,1914,1862,1929,2131,2387,25416)
(Uttarakhand,2010,86,108,128,127,125,145,148,128,95,129,136,124,1479)
(West Bengal,2010,1245,1150,1349,1246,1172,1284,1231,1190,1128,1227,1251,1252,14725)
```

3)Accident count by year:

```
GroupByYear = GROUP accidents BY year;
```

```
dump GroupByYear;
```

```
accidents_year = foreach GroupByYear Generate group,SUM(accidents.total);
```

```
dump accidents_year
```

```
(2001,323720)
(2002,335707)
(2003,336468)
(2004,361343)
(2005,390378)
(2006,394432)
(2007,418657)
(2008,415855)
(2009,421628)
(2010,430654)
(2011,440123)
(2012,440042)
(2013,443001)
(2014,450898)
```

4)Accident count by state:

```
GroupByState = GROUP accidents BY state;  
dump GroupByState ;  
accidents_state = foreach GroupByState Generate group,SUM(accidents.total);  
dump accidents_state;
```

```
(Tamil Nadu,852073)  
(Daman & Diu,643)  
(Lakshadweep,14)  
(Maharashtra,641614)  
(Uttarakhand,18192)  
(West Bengal,189375)  
(Chhattisgarh,117817)  
(D & N Haveli,1216)  
(A & N Islands,2893)  
(Uttar Pradesh,274528)  
(Andhra Pradesh,546821)  
(Madhya Pradesh,394007)  
(Jammu & Kashmir,82820)  
(Himachal Pradesh,37105)  
(Arunachal Pradesh,3389)
```

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;
```

```
(2001,772.5428571428571)  
(2002,783.6)  
(2003,805.9714285714285)  
(2004,877.7714285714286)  
(2005,1021.1428571428571)  
(2006,958.1428571428571)  
(2007,1040.9142857142858)  
(2008,1028.9142857142858)  
(2009,1009.8857142857142)  
(2010,1047.8285714285714)  
(2011,1094.4285714285713)  
(2012,1097.0)  
(2013,1119.5714285714287)  
(2014,1134.2571428571428)
```