

Assignment

Crime Data analysis using Pig

1. To find the total number of crimes which occurred in all the states in the year 2006.

```
L = LOAD '/user/training/vaibhav/crimeyear1.csv' using PigStorage(',')
AS
```

```
(State:chararray,crimehead:chararray,y1:double,y2:double,y3:double,y4:double,y5:double,y6:double,y7:double,y8:double,y9:double,y10:double,y11:double,y12:double);
```

```
grunt> L = LOAD '/user/training/vaibhav/crimeyear1.csv' using PigStorage(',')
AS (State:chararray,crimehead:chararray,y1:double,y2:double,y3:double,y4:double,y5:double,y6:double,y7:double,y8:double,y9:double,y10:double,y11:double,y12:double);
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.permissions.supergroup is deprecated. Instead, use dfs.permissions.superusergroup
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.max.objects is deprecated. Instead, use dfs.namenode.max.objects
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.interval is deprecated. Instead, use dfs.namenode.replication.interval
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.data.dir is deprecated. Instead, use dfs.datanode.data.dir
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.access.time.precision is deprecated. Instead, use dfs.namenode.access.time.precision
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.min is deprecated. Instead, use dfs.namenode.replication.min
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - fs.checkpoint.dir is deprecated. Instead, use dfs.namenode.checkpoint.dir
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.http.address is deprecated. Instead, use dfs.namenode.http.address
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.considerLoad is deprecated. Instead, use dfs.namenode.replication.considerLoad
```

```
grunt> G = Group L all;
grunt> S = FOREACH G GENERATE SUM(L.y6);
grunt> describe S;
S: {double}
```

```
grunt> G = Group L all;
grunt> S = FOREACH G GENERATE SUM(L.y6);
grunt> describe S;
S: {double}
grunt> dump S;
2022-09-08 11:03:47,895 [main] INFO org.apache.
```

```
2022-09-08 11:04:01,082 [mai
2022-09-08 11:04:01,082 [mai
(559646.0)
```

2. Calculate the number of females (age between 18-30 years) who were victims in different crimes in different states.

```
L = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',') AS
(state:chararray,crime:chararray,M18:int,F18:int,m1830:int,f1830:int,m3045:int,f3045:int,m4560:i
nt,f4560:int,m60:int,f60:int,tm:int,tf:int,t:int);
```

```
grunt> L = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',') AS (state:chararray,
crime:chararray,M18:int,F18:int,m1830:int,f1830:int,m3045:int,f3045:int,m4560:int,f4560:int,m60:i
nt,f60:int,tm:int,tf:int,t:int);
2022-09-08 11:43:59,953 [main] WARN org.apache.hadoop.conf.Configuration - io.bytes.per.checksum
is deprecated. Instead, use dfs.bytes-per-checksum
```

```
C = FOREACH L generate crime,f1830;
P = GROUP C ALL;
R = FOREACH P GENERATE SUM(C.$1);
```

```
grunt> C = FOREACH L generate crime,f1830;
grunt> P = GROUP C ALL;
grunt> describe C;
C: {crime: chararray,f1830: int}
grunt> describe P;
P: {group: chararray,C: {(crime: chararray,f1830: int)}}
grunt> R = FOREACH P GENERATE SUM(C.$1);
grunt> describe R;
R: {long}
```

```
2022-09-08 11:45:52,
2022-09-08 11:45:52,
2022-09-08 11:45:52,
(86071)
grunt>
```

3. Which state have highest rape cases across all years.

```
YC = FILTER D BY (crime matches '.*(RAPE).*');
```

```
dump YC;
```

```
CV = FOREACH YC generate state,crime,t;
```

```
dump CV;
```

```
GC = GROUP CV BY state;
```

```
grunt> D = LOAD '/user/training/kirti/crimeagel.csv' using PigStorage(',') AS
0:int,tm:int,tf:int,t:int);
grunt> CC = FOREACH D generate state,crime,t;
grunt> dump CC;
```

```
2022-09-08 12:00:41,270 [main] INFO org.apache.p
(ANDHRA PRADESH,RAPE (SECTION 376 IPC),1664)
(ARUNACHAL PRADESH,RAPE (SECTION 376 IPC),47)
(ASSAM,RAPE (SECTION 376 IPC),1626)
(BIHAR,RAPE (SECTION 376 IPC),1327)
(CHHATTISGARH,RAPE (SECTION 376 IPC),1214)
(GOA,RAPE (SECTION 376 IPC),61)
(GUJARAT,RAPE (SECTION 376 IPC),647)
(HARYANA,RAPE (SECTION 376 IPC),940)
(HIMACHAL PRADESH,RAPE (SECTION 376 IPC),259)
(JAMMU & KASHMIR,RAPE (SECTION 376 IPC),388)
(JHARKHAND,RAPE (SECTION 376 IPC),780)
(KARNATAKA,RAPE (SECTION 376 IPC),842)
(KERALA,RAPE (SECTION 376 IPC),1259)
(MADHYA PRADESH,RAPE (SECTION 376 IPC),4822)
(MAHARASHTRA,RAPE (SECTION 376 IPC),2591)
(MANIPUR,RAPE (SECTION 376 IPC),46)
(MEGHALAYA,RAPE (SECTION 376 IPC),182)
(MIZORAM,RAPE (SECTION 376 IPC),122)
(NAGALAND,RAPE (SECTION 376 IPC),26)
(ODISHA,RAPE (SECTION 376 IPC),1666)
(PUNJAB,RAPE (SECTION 376 IPC),895)
(RAJASTHAN,RAPE (SECTION 376 IPC),1807)
(SIKKIM,RAPE (SECTION 376 IPC),29)
(TAMIL NADU,RAPE (SECTION 376 IPC),962)
(TRIPURA,RAPE (SECTION 376 IPC),202)
(UTTAR PRADESH,RAPE (SECTION 376 IPC),3593)
(UTTARAKHAND,RAPE (SECTION 376 IPC),184)
```

```

(YC DOCUMENT, TOTAL CRIMES AGAINST WOMEN, 110)
grunt> YC = FILTER CB BY (crime matches '.*(RAPE).*');
grunt> dump YC;

grunt> GC = GROUP CV BY state;
grunt> dump GC;
2022-09-08 12:08:19,863 [main] INFO org.apache.pig.t
2022-09-08 12:08:19,877 [main] INFO org.apache.pig.n
2022-09-08 12:08:19,892 [main] INFO org.apache.pig.b
2022-09-08 12:08:19,897 [main] INFO org.apache.pig.b
2022-09-08 12:08:19,897 [main] INFO org.apache.pig.b

2022-09-08 12:08:32,727 [main] INFO org.apache.pig.backend
(GOA,{(GOA,RAPE (SECTION 376 IPC),61)})
(ASSAM,{(ASSAM,RAPE (SECTION 376 IPC),1626)})
(BIHAR,{(BIHAR,RAPE (SECTION 376 IPC),1327)})
(DELHI,{(DELHI,RAPE (SECTION 376 IPC),892)})
(KERALA,{(KERALA,RAPE (SECTION 376 IPC),1259)})
(ODISHA,{(ODISHA,RAPE (SECTION 376 IPC),1666)})
(PUNJAB,{(PUNJAB,RAPE (SECTION 376 IPC),895)})
(SIKKIM,{(SIKKIM,RAPE (SECTION 376 IPC),29)})
(GUJARAT,{(GUJARAT,RAPE (SECTION 376 IPC),647)})
(HARYANA,{(HARYANA,RAPE (SECTION 376 IPC),940)})
(MANIPUR,{(MANIPUR,RAPE (SECTION 376 IPC),46)})
(MIZORAM,{(MIZORAM,RAPE (SECTION 376 IPC),122)})
(TRIPURA,{(TRIPURA,RAPE (SECTION 376 IPC),202)})
(NAGALAND,{(NAGALAND,RAPE (SECTION 376 IPC),26)})
(JHARKHAND,{(JHARKHAND,RAPE (SECTION 376 IPC),780)})
(KARNATAKA,{(KARNATAKA,RAPE (SECTION 376 IPC),842)})
(MEGHALAYA,{(MEGHALAYA,RAPE (SECTION 376 IPC),182)})
(RAJASTHAN,{(RAJASTHAN,RAPE (SECTION 376 IPC),1807)})
(CHANDIGARH,{(CHANDIGARH,RAPE (SECTION 376 IPC),34)})
(PUDUCHERRY,{(PUDUCHERRY,RAPE (SECTION 376 IPC),15)})
(TAMIL NADU,{(TAMIL NADU,RAPE (SECTION 376 IPC),962)})
(DAMAN & DIU,{(DAMAN & DIU,RAPE (SECTION 376 IPC),10)})

```

Output:

```

(UTTAR PRADESH,{(UTTAR PRADESH,RAPE (SECTION 376 IPC),3593)})
(ANDHRA PRADESH,{(ANDHRA PRADESH,RAPE (SECTION 376 IPC),1664)})
(MADHYA PRADESH,{(MADHYA PRADESH,RAPE (SECTION 376 IPC),4822)})
(JAMMU & KASHMIR,{(JAMMU & KASHMIR,RAPE (SECTION 376 IPC),388)})
(HIMACHAL PRADESH,{(HIMACHAL PRADESH,RAPE (SECTION 376 IPC),259)})

```

6. Calculate the number of males (age between 45-60 years) who were victims in “Kidnapping and abduction” in different states.

```
L1 = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',')
AS
```

```
AS
```

```
(State, Crimehead, MB18, FB18, MB1830, FB1830, MB3035, FB3045, MB4560, FB4560, MA60, FA60,
TOTALMALE, TOTALFEMALE);
```

```
grunt> L1 = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',')
AS (State, Crimehead, MB18, FB18, MB1830, FB1830, MB3035, FB3045, MB4560, FB4560, MA60, FA60, TOTALMALE, TOTALFEMALE);
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - io.bytes.per.checksum is deprecated
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.permissions.supergroup is deprecated
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.max.objects is deprecated. Instance
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.interval is deprecated
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.data.dir is deprecated. Instead
```

```
Y = FILTER L1 BY (Crimehead matches '("KIDNAPPING AND ABDUCTION").');
```

```
grunt> Y = FILTER L1 BY (Crimehead matches '("KIDNAPPING AND ABDUCTION").');
2022-09-08 11:18:27,195 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CH
```

```
C = FOREACH Y generate Crimehead, MB4560;
```

```
PA = GROUP C ALL;
```

```
QA = FOREACH PA GENERATE SUM(C.$1);
```

```
grunt> C = FOREACH Y generate Crimehead, MB4560;
2022-09-08 11:19:04,329 [main] WARN org.apache.pig.PigServer
grunt> PA = GROUP C ALL;
2022-09-08 11:19:18,094 [main] WARN org.apache.pig.PigServer
grunt> QA = FOREACH PA GENERATE SUM(C.$1);
```

```
(1022)
```

```
grunt>
```