

Water Quality Dataset:

The columns in Waterquality.csv are State Name,District Name,Block Name,Panchayat Name,Village Name,Habitation Name,Quality Parameter,Year

1.Remove the headers

2.Copy the local file in hdfs

```
#hadoop dfs -put 'file' /
```

LOAD Dataset:

```
water_record = LOAD '/Waterquality.csv' using PigStorage(',') as  
(state:chararray,district:chararray,block:chararray,panchayat:chararray,village:chararray,habit:chararray,quality:chararray,year:chararray);
```

```
grunt> water_record = LOAD '/Waterquality.csv' using PigStorage(',') as (state:chararray,district:chararray,block:chararray,panchayat:chararray,village:chararray,habit:chararray,quality:chararray,year:chararray);
```

View Dataset:

```
dump water_record;
```

```
(ORISSA,Nayagarh,Nuagaon,SINGARAPALLI,Sundhijhola,NUASAH,Iron,1/4/2009)  
(CHHATTISGARH,BILASPUR,BELHA,DAGORY,SATTIGHAT,SATIGHAT,Salinity,1/4/2012)  
(ORISSA,PURI,KRUSHNAPRASAD,SIANDI,GOPALPUR,RATAMATI,Salinity,1/4/2012)  
(MADHYA PRADESH,SEONI,KEOLARI,KHAMI,KHAMI,ZIDDI TOLA,Iron,1/4/2009)  
(BIHAR,KATIHAR,AZAMNAGAR,MAHESHPUR,MAHESHPUR,MAHESHPUR,Iron,1/4/2009)  
(ORISSA,JAJAPUR,JAJPUR,CHAINPUR,GOPALPUR,GOUDA SAHI,Iron,1/4/2009)  
(MAHARASHTRA,JALGAON,PACHORA,KURHAD KH.,KURHAD KH.,KURHAD KH.,Fluoride,1/4/2010)  
(CHHATTISGARH,KORBA,KORBA,KARUMOHA,KARUMAUHA,TIKRAPARA,Iron,1/4/2012)  
(ORISSA,BALESHWAR,BALIPAL,KUNDULI,INDIPADA,INDIPADA,Salinity,1/4/2012)
```

Analysing Dataset:

1)Records containing Salinity:

```
filter_data = FILTER water_record by quality=='Salinity';  
dump filter_data;
```

```
(MADHYA PRADESH,REWA,NAIGARHI,SHAHPUR,SHAHPUR,HARIZAN TOLA,Salinity,1/4/2011)  
(RAJASTHAN,AJMER,JAWAJA,SURADIYA,SANGARWAS,SANGARWAS,Salinity,1/4/2012)  
(RAJASTHAN,JODHPUR,BALESAR,BELWA,BELWA,LAKHAMA RAM NAI KD,Salinity,1/4/2009)  
(KARNATAKA,DAVANGERE,CHANNAGIRI,DURVIGERE,DURVIGERE,DURVIGERE,Salinity,1/4/2009)  
(CHHATTISGARH,BILASPUR,BELHA,DAGORY,SATTIGHAT,SATIGHAT,Salinity,1/4/2012)  
(ORISSA,PURI,KRUSHNAPRASAD,SIANDI,GOPALPUR,RATAMATI,Salinity,1/4/2012)  
(ORISSA,BALESHWAR,BALIPAL,KUNDULI,INDIPADA,INDIPADA,Salinity,1/4/2012)
```

2)Records Containing Fluoride or Iron:

```
filter_data = FILTER water_record by quality in ('Flouride','Iron');  
dump filter_data;
```

```
(ORISSA,Nayagarh,Nuagaon,SINGARAPALLI,Sundhijhola,NUASAH,Iron,1/4/2009)  
(MADHYA PRADESH,SEONI,KEOLARI,KHAMI,KHAMI,ZIDDI TOLA,Iron,1/4/2009)  
(BIHAR,KATI HAR,AZAMNAGAR,MAHESHPUR,MAHESHPUR,MAHESHPUR,Iron,1/4/2009)  
(ORISSA,JAAPUR,JAJPUR,CHAINPUR,GOPALPUR,GOUDA SAHI,Iron,1/4/2009)  
(CHHATTISGARH,KORBA,KORBA,KARUMOH, KARUMAUHA,TIKRAPARA,Iron,1/4/2012)
```

3)Records containing Salinity or Iron in Orissa:

```
filter_data = FILTER water_record by quality in ('Salinity','Iron') and state == 'ORISSA';  
dump filter_data;
```

```
(ORISSA,PURI,KRUSHNAPRASAD,PIRIJIPUR,SIPAKUDAPATNA,NUA SAHI,Salinity,1/4/2009)  
(ORISSA,ANUGUL,TALCHER,GOPALPRASAD,BAGHABASPUR,BAGHABASAPUR,Iron,1/4/2009)  
(ORISSA,Nayagarh,Nuagaon,SINGARAPALLI,Sundhijhola,NUASAH,Iron,1/4/2009)  
(ORISSA,PURI,KRUSHNAPRASAD,SIANDI,GOPALPUR,RATAMATI,Salinity,1/4/2012)  
(ORISSA,JAAPUR,JAJPUR,CHAINPUR,GOPALPUR,GOUDA SAHI,Iron,1/4/2009)  
(ORISSA,BALESHWAR,BALIPAL,KUNDULI,INDIPADA,INDIPADA,Salinity,1/4/2012)
```

4)Count of Salinity in each state:

```
filter_data = FILTER water_record by quality == 'Salinity';  
dump filter_data;  
GroupByState = GROUP filter_data by state;  
dump GroupByState;  
countbystate = foreach GroupByState GENERATE group,COUNT(filter_data.state);  
dump countbystate;
```

```
(TAMIL NADU,13)  
(CHATTISGARH,8)  
(MAHARASHTRA,36)  
(WEST BENGAL,42)  
(CHHATTISGARH,5)  
(UTTAR PRADESH,42)  
(ANDHRA PRADESH,9)
```

5)Count of quality in each state:

```
GroupByqualitystate = GROUP water_record BY (quality,state);  
dump GroupByqualitystate;  
countbyqualitystate = foreach GroupByqualitystate GENERATE  
group,COUNT(water_record.state);  
dump countbyqualitystate;
```

```
((Iron,ASSAM),1334)  
((Iron,BIHAR),1290)  
((Iron,KERALA),58)  
((Iron,ORISSA),1070)  
((Iron,PUNJAB),3)  
((Iron,TRIPURA),450)  
((Iron,NAGALAND),13)  
((Iron,JHARKHAND),49)
```