Project 1

1. Write a MapReduce/Pig program to calculate the number of cases investigated under each FBI code

load

```
grunt> A = LOAD 'Crimes.csv' using PigStorage(',')as(id:int,casse_number:chararr
ay,date:chararray,block:chararray,IUCR:int,primary_type:chararray,description:ch
ararray,location:chararray,Arrest:boolean,domestic:boolean,beat:int,district:int
,ward:int,community_area:int,fbicode:chararray,x_coordinate:int,y_coordinate:int
,year:int,updated_on:chararray,latitude:float,longitude:float,location_coordinat
es:tuple(lat:float,long:float));
```

```
grunt> b = foreach A generate id,fbicode;
grunt> c = filter b by fbicode is not null;
grunt> d = group c by fbicode;
```

```
grunt> describe d;
d: {group: chararray,c: {(id: int,fbicode: chararray)}}
grunt> e = foreach d generate group,COUNT(c.id);
grunt> dump e;
```

Output

```
2017-11-20 23:23:52,774 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(1,172)
(2,362)
(3,266)
(4,154)
(5,197)
(6,198)
(7,138)
(8,301)
(9,192)
(02,1480)
(03,10552)
(05,14735)
(06,62826)
(07,10520)
(09,437)
(10,1708)
(11,13637)
(12,79)
(13,151)
(14,31244)
(15,3780)
(16,1949)
```

```
(17,1165)
(18,24989)
(19,590)
(20,1435)
(21,293)
(22,483)
(23,77)
(24,4114)
(25,142)
(26,29009)
(27,175)
(28,385)
(29,196)
(30,115)
(31,93)
(32,76)
(33,105)
(34,184)
(35,56)
(36,63)
(37,161)
(38,117)
(39,98)
(40,97)
```

```
(41,123)
(42,87)
(43,101)
(44,35)
(45,34)
(46,62)
(47,137)
(48,61)
(49,61)
(50,40)
(56,15)
(57,1)
(58,3)
(61,5)
(66,7)
(68,2)
(76,51)
(01A,533)
(01B,6)
(04A,4912)
(04B,7598)
(08A,13161)
(08B,44935)
(1923,1)
```

2. Write a MapReduce/Pig program to calculate the number of cases investigated under FBI code 32.

Load

```
grunt> A = LOAD 'Crimes.csv' using PigStorage(',')as(id:int,casse_number:chararr
ay,date:chararray,block:chararray,IUCR:int,primary_type:chararray,description:ch
ararray,location:chararray,Arrest:boolean,domestic:boolean,beat:int,district:int
,ward:int,community_area:int,fbicode:chararray,x_coordinate:int,y_coordinate:int
,year:int,updated_on:chararray,latitude:float,longitude:float,location_coordinat
es:tuple(lat:float,long:float));
```

```
grunt> b = foreach A generate id,fbicode;
grunt> c = filter b by fbicode is not null;
```

```
grunt> d = filter c by fbicode =='32' ;
grunt> e = group d by fbicode;
grunt> describe e;
e: group: chararray,d: {(id: int,fbicode: chararray)}
grunt> f = foreach e generate group,COUNT(d.id);
grunt> dump f;
```

OUTPUT:

```
ne.util.MapRedUtil - Total input paths to process : 1
(32,76)
grunt>
```

3. Write a MapReduce/Pig program to calculate the number of arrests in theft district wise.

```
grunt> A = LOAD 'Crimes.csv' using PigStorage(',')as(id:int,casse_number:chararr
ay,date:chararray,block:chararray,IUCR:int,primary_type:chararray,description:ch
ararray,location:chararray,Arrest:boolean,domestic:boolean,beat:int,district:int
,ward:int,community_area:int,fbicode:chararray,x_coordinate:int,y_coordinate:int
,year:int,updated_on:chararray,latitude:float,longitude:float,location_coordinat
es:tuple(lat:float,long:float));
```

```
grunt> b = foreach A generate primary_type,Arrest,district;
grunt> c = filter b by Arrest == TRUE;
grunt> d = filter c by primary_type == 'THEFT';
grunt> e = group d by district;
grunt> f = foreach e generate group,COUNT(d.Arrest);
grunt> dump f;
```

OUTPUT

```
(1,1119)
(2,220)
(3,157)
(4,221)
(5,273)
(6,649)
(7,172)
(8,458)
(9,318)
(10,166)
(11,174)
(12,353)
(14,227)
(15,111)
(16,171)
(17,227)
(18,732)
(19,499)
(20,241)
(22,207)
(24,224)
(25,591)
grunt>
```

4. Write a MapReduce/Pig program to calculate the number of arrests done between October 2014 and October 2015.

```
A = LOAD 'Crimes.csv' using PigStorage(',')as
(id:int,casse_number:chararray,date:chararray,block:chararray,IUCR:int,primary
(lat:float,long:float));

b = foreach A generate GetMonth(ToDate(date,'MM/dd/yyyy HH:mm:ss aa')) as
month,GetYear(ToDate(date,'MM/dd/yyyy HH:mm:ss aa')) as yr,Arrest as arrest;

c = filter b by arrest == TRUE AND month is not null AND yr is not null;

d = filter c by (month  >9 AND yr == 2014) OR (month <11 AND  yr == 2015);

e = group d by arrest;

f = foreach e generate group,COUNT(d.arrest);
```

OUTPUT:

```
(true,63173)
grunt>
```