Project 1

Downloaded the dataset and uploaded the data set to the folder /user/acadgild/pig using Hadoop.

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/pig
17/12/07 11:55:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 acadgild supergroup 69234933 2017-12-06 14:20 /user/acadgild/pig/Crimes_-_2001_to_present.csv
[acadgild@localhost ~]$ ■
```

Loaded the data set in PIG using the following command:

A = LOAD '/user/acadgild/pig/Crimes_-_2001_to_present.csv' USING PigStorage(',') AS (ID:Int,Case_Number:Chararray,Date:Chararray,Block:Chararray,IUCR:Int,Primary_Type:Chararray,Description:Chararray,Location_Description:Chararray,Arrest:Chararray,Domestic:Chararray,Beat:Int,District:Int,Ward:Int,Community_Area:Int,FBICode:Chararray,X_Coordinate:Int,Y_Coordinate:Int,Y_Coordinate:Int,Y_Coordinate:Int,Updated_On:Int,Latitude:Long,Longitude:Long,Location:Long);

```
grunt> A = LOAD '/user/acadgild/pig/Crimes - 2001 to present.csv' USING PigStorage(',') AS (ID:Int,Case Number:Chararray,Date:Chararray,Block:Chararray,IUCR:Int,Primary_Type:Chararray,Description:Chararray,Location_Description:Chararray,Arrest:Chararray,Domestic:Chararray,Beat:Int,District:Int,Ward:Int,Community_Area:Int,FBICode:Chararray,X Coordinate:Int,Y Coordinate:Int,Year:Int,Updated_On:Int,Latitude:Long,Longitude:Long,Location:Long);
```

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

```
grunt> B = foreach A generate (chararray)$1 as Case_Number,(chararray)$14 as FBICode;
grunt> C = filter B by FBICode is not null and Case_Number is not null;
grunt> D = group C by FBICode;
grunt> E = foreach D Generate group, COUNT(C.FBICode);
```

- B: Generated the columns from the dataset loaded in A with column type
- C: Filtered the columns in B
- D: Grouped the filtered columns in C with the FBICode column

E: Foreach group formed in D as per the FBICode, generated the number arrest.

Output:

```
(1,172)
(2,362)
              (36,63)
              (37, 161)
(3,266)
(4,154)
              (38,117)
(5,197)
(6,198)
              (39,98)
              (40,97)
(7, 138)
(8,301)
(9,192)
             (41, 123)
             (42,87)
(02,1480)
(03,10552)
(05,14735)
             (43, 101)
             (44,35)
(06,62826)
             (45,34)
(07,10520)
(09,437)
              (46,62)
(10, 1708)
             (47, 137)
(11,13637)
(12,79)
(13,151)
              (48,61)
              (49,61)
(14,31244)
(15,3780)
             (50,40)
              (56, 15)
(16, 1949)
(17,1165)
(18,24989)
(19,590)
              (57,1)
              (58,3)
             (61,5)
(20,1435)
(21,293)
              (66,7)
(22,483)
              (68, 2)
(23,77)
(24,4114)
             (76,51)
(25, 142)
              (01A, 533)
(26,29009)
              (01B, 6)
(27,175)
(28,385)
              (04A, 4912)
(29, 196)
              (04B,7598)
(30,115)
(31,93)
              (08A, 13161)
              (08B,44935)
(32,76)
(33,105)
(34,184)
             (1923,1)
             grunt>
(35,56)
```

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

```
grunt> B = foreach A generate (chararray)$1 as Case_Number,(chararray)$14 as FBICode;
grunt> C = filter B by FBICode == '32';
grunt> D = group C by FBICode;
grunt> E = foreach D Generate group, COUNT(C.Case_Number);
grunt> Dump E;
```

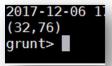
B: Generated the columns from the dataset loaded in A with column type

C: Filtered the columns in B as per the FBICode = 32 (Since we require the number of arrest happened for FBICode 32)

D: Group the filter in C

E: Foreach group in D, get the count of number of cases investigated

Output:



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

```
grunt> B1 = foreach A generate (Chararray)$5 as Primary_Type, (Chararray)$8 as Arrest,(int)$11 as District;
grunt> C1 = filter B1 by Primary_Type == 'THEFT';
grunt>
grunt> D1 = group C1 by District;
grunt>
grunt> E1 = foreach D1 generate group, COUNT(C1.Arrest);
grunt> Dump E1;
```

B1: Generate the columns from the dataset loaded in A

C1: Filter the column Primary Type by THEFT (Since we need the number of Arrest happed for THEFTS district wise)

D1: Group the filters in C1 district wise

E1: Foreach group in D1, perform a count of arrest happened.

Output:

```
(1,5893)
(9,2746)
(10, 1824)
11,2100)
(12,3915)
(14,3266)
(15, 1440)
(16, 2294)
(17,2192)
(18,5596)
(19,4663)
(20, 1297)
(22, 1996)
(24, 1801)
(25, 2996)
,1503)
grunt>
```

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

```
grunt> B = foreach A generate (Chararray)$2 as Date, (Chararray)$8 as Arrest, (Int)$17 as Year;
grunt> C = filter B by Date >= '10/1/2014' AND Date <= '10/31/2015';
grunt> D = filter B by Year >=2014 and Year <= 2015 AND Arrest is not null;
grunt> E = group D by Date;
grunt> E = foreach E generate group, COUNT(D.Arrest);
grunt> Dump F;
```

B: Filter the columns from the data loaded in A

C: Filter the Dates between October 2014 and October 2015

D: Filter the Year between 2014 and 2015 where Arrest is not null

E: Group the Year as per the Dates.

F: Foreach group in E, generate the count of number of arrest

Output

In the output, you will notice that number of arrests have been grouped as per the time stamp on daily basis, between October 2014 to October 2015.

The output is in (mm/dd/yyyy 00:00:00, count) format

(01/01/2015	01:00:00	AM. 16)	(02/12/2015	12.55.00	AM 1)	/02/00/0015	04-07-00	DW 1)
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015						(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015 (02/13/2015			(03/09/2015		
(01/01/2015						(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015 (02/13/2015			(03/09/2015 (03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015)			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2015			(02/13/2015			(03/09/2015		
(01/01/2013	02123100	, , , , ,	(02/13/2013	02.30.00	111,1)	(03/03/2013	00.09.00	111, 1

```
(06/04/2015 04:15:00 PM,1) (07/11/2015 08:20:00 AM,1) (06/04/2015 04:17:00 PM,1) (07/11/2015 08:20:00 PM,3)
  (05/09/2015 09:34:00 PM,1)
 (05/09/2015 09:35:00 AM,1)
                                                                                                    (06/04/2015 04:19:00 PM,1)
(05/09/2015 09:40:00 PM,2)
                                                                                                                                                                                                (07/11/2015 08:25:00 PM,1)
                                                                                                 (06/04/2015 04:20:00 PM,1) (07/11/2015 08:23:00 PM,1) (06/04/2015 04:30:00 AM,1) (07/11/2015 08:30:00 AM,3) (06/04/2015 04:33:00 PM,6) (07/11/2015 08:30:00 PM,3) (06/04/2015 04:33:00 PM,1) (07/11/2015 08:35:00 PM,1) (06/04/2015 04:34:00 PM,1) (07/11/2015 08:35:00 PM,1) (06/04/2015 04:40:00 AM,1) (07/11/2015 08:40:00 AM,2) (06/04/2015 04:40:00 AM,1) (07/11/2015 08:45:00 AM,2) (06/04/2015 04:45:00 PM,2) (07/11/2015 08:45:00 PM,3) (06/04/2015 04:45:00 PM,2) (07/11/2015 08:45:00 PM,3) (06/04/2015 04:45:00 PM,2) (07/11/2015 08:50:00 PM,2) (06/04/2015 04:50:00 PM,1) (07/11/2015 08:50:00 PM,1) (06/04/2015 04:51:00 PM,1) (07/11/2015 09:00:00 AM,15) (06/04/2015 04:52:00 PM,1) (07/11/2015 09:00:00 AM,15) (06/04/2015 04:52:00 PM,1) (07/11/2015 09:00:00 AM,1) (06/04/2015 05:00:00 AM,4) (07/11/2015 09:00:00 AM,1) (06/04/2015 05:00:00 PM,18) (07/11/2015 09:16:00 PM,1) (06/04/2015 05:00:00 PM,18) (07/11/2015 09:16:00 AM,1) (06/04/2015 05:00:00 PM,1) (07/11/2015 09:16:00 AM,1) (06/04/2015 05:00:00 PM,1) (07/11/2015 09:23:00 PM,1) (06/04/2015 05:15:00 AM,2) (07/11/2015 09:23:00 PM,1) (06/04/2015 05:15:00 AM,2) (07/11/2015 09:30:00 AM,4) (06/04/2015 05:15:00 PM,1) (07/11/2015 09:30:00 AM,4) (06/04/2015 05:15:00 PM,1) (07/11/2015 09:30:00 AM,4) (06/04/2015 05:15:00 AM,2) (07/11/2015 09:30:00 AM,4) (06/04/2015 05:35:00 PM,1) (07/11/2015 09:30:00 AM,4) (06/04/2015 05:35:00 PM,1) (07/11/2015 09:35:00 PM,2) (06/04/2015 05:35:00 PM,1) (07/11/2015 09:35:00 PM,2) (06/04/2015 05:35:00 PM,1) (07/11/2015 09:35:00 PM,2) (06/04/2015 05:35:00 PM,2) (07/11/2015 09:55:00 PM,2) (06/04/2015 05:37:00 AM,1) (07/11/2015 09:55:00 PM,2) (06/04/2015 05:43:00 PM,2) (07/11/2015 09:55:00 PM,2) (06/04/2015 05:44:00 PM,2) (07/11/2015 09:55:00 PM,2) (06/04/2015 05:44:00 PM,2) (07/11/2015 09:55:00 PM,1) (06/04/2015 05:40:00 PM,2) (07/11/2015 10:00:00 PM,1) (06/04/2015 05:50:00 PM,3) (07/11/2015 10:00
                                                                                                    (06/04/2015 04:20:00 PM,1)
                                                                                                                                                                                                (07/11/2015 08:28:00 AM,1)
(07/11/2015 08:30:00 AM,3)
(05/09/2015 09:44:00 AM,1)
(05/09/2015 09:45:00 AM,3)
                                                                                                    (06/04/2015 04:30:00 AM,1)
(05/09/2015 09:50:00 AM,1)
(05/09/2015 09:50:00 PM,1)
(05/09/2015 09:51:00 AM,1)
(05/09/2015 10:00:00 AM, 12)
(05/09/2015 10:00:00 PM,8)
(05/09/2015 10:01:00 PM,1)
(05/09/2015 10:03:00 PM,1)
(05/09/2015 10:05:00 PM,2)
(05/09/2015 10:06:00 PM,1)
 (05/09/2015 10:09:00 PM,2)
(05/09/2015 10:10:00 AM,4)
(05/09/2015 10:10:00 PM,1)
(05/09/2015 10:15:00 AM,3)
(05/09/2015 10:15:00 PM,1)
(05/09/2015 10:24:00 AM,1)
 (05/09/2015 10:25:00 AM,1)
(05/09/2015 10:26:00 AM,1)
(05/09/2015 10:27:00 AM,1)
(05/09/2015 10:30:00 AM,3)
(05/09/2015 10:30:00 PM,2)
(05/09/2015 10:35:00 AM,1)
 (05/09/2015 10:35:00 PM,1)
(05/09/2015 10:43:00 PM,2)
(05/09/2015 10:45:00 AM,4)
 (05/09/2015 10:45:00 PM,1)
(05/09/2015 10:55:00 AM,1)
(05/09/2015 10:55:00 PM,2)
 (05/09/2015 11:00:00 AM,13)
(05/09/2015 11:00:00 PM,7)
(05/09/2015 11:07:00 AM,1)
(05/09/2015 11:07:00 AM,1)
(05/09/2015 11:10:00 AM,4)
(05/09/2015 11:13:00 AM,1)
(05/09/2015 11:15:00 AM,1)
 (05/09/2015 11:15:00 PM,4)
(05/09/2015 11:20:00 AM,2)
(05/09/2015 11:20:00 PM,2)
                                                                                                    (06/04/2015 05:51:00 AM,1)
(06/04/2015 06:00:00 AM,3)
                                                                                                                                                                                                   (07/11/2015 10:12:00 AM,1)
                                                                                                                                                                                                   (07/11/2015 10:12:00 PM,1)
```

(08/27/2015 11:32:00 PM,1)	(09/10/2014 10:00:00	PM 15)	(10/13/2014	10:18:00	PM.1)
(08/27/2015 11:32:00 PM,1) (08/27/2015 11:33:00 PM,1)	(09/10/2014 10:07:00		(10/13/2014		
(08/27/2015 11:35:00 FM,1) (08/27/2015 11:35:00 AM,4)	(09/10/2014 10:07:00		(10/13/2014		
(08/27/2015 11:33:00 AM, 4) (08/27/2015 11:40:00 AM, 2)	(09/10/2014 10:10:00		(10/13/2014		
(08/27/2015 11:40:00 AM, 2) (08/27/2015 11:42:00 AM, 1)	(09/10/2014 10:15:00		(10/13/2014		
(08/27/2015 11:42:00 AM,1)	(09/10/2014 10:17:00		(10/13/2014		
(08/27/2015 11:50:00 AM,1)	(09/10/2014 10:20:00		(10/13/2014		
(08/27/2015 11:50:00 PM,2)	(09/10/2014 10:25:00		(10/13/2014		
(08/27/2015 11:55:00 PM,1)	(09/10/2014 10:30:00		(10/13/2014	10:35:00	PM, 1)
(08/27/2015 11:56:00 AM,1)	(09/10/2014 10:30:00		(10/13/2014		
(08/27/2015 11:58:00 AM,1)	(09/10/2014 10:37:00		(10/13/2014	10:40:00	AM, 2)
(08/27/2015 12:00:00 AM,4)	(09/10/2014 10:40:00	AM, 2)	(10/13/2014	10:41:00	AM, 1)
(08/27/2015 12:00:00 PM,10)	(09/10/2014 10:40:00	PM, 2)	(10/13/2014	10:44:00	PM, 1)
(08/27/2015 12:01:00 AM,7)	(09/10/2014 10:42:00	AM, 1)	(10/13/2014	10:45:00	AM,4)
(08/27/2015 12:05:00 PM,1)	(09/10/2014 10:43:00	PM, 1)	(10/13/2014	10:45:00	PM, 1)
(08/27/2015 12:06:00 PM,1)	(09/10/2014 10:45:00	AM, 2)	(10/13/2014		
(08/27/2015 12:09:00 PM,1)	(09/10/2014 10:45:00	PM, 1)	(10/13/2014		
(08/27/2015 12:10:00 AM,1)	(09/10/2014 10:50:00	AM,3)	(10/13/2014		
(08/27/2015 12:10:00 PM,2)	(09/10/2014 10:51:00		(10/13/2014		
(08/27/2015 12:13:00 PM,1)	(09/10/2014 10:55:00		(10/13/2014		
(08/27/2015 12:15:00 AM,4)	(09/10/2014 11:00:00		(10/13/2014		
(08/27/2015 12:15:00 PM,7)	(09/10/2014 11:00:00		(10/13/2014		
(08/27/2015 12:16:00 AM,1)	(09/10/2014 11:03:00		(10/13/2014		
(08/27/2015 12:23:00 AM,1)	(09/10/2014 11:04:00		(10/13/2014		
(08/27/2015 12:25:00 PM,1)	(09/10/2014 11:05:00		(10/13/2014		
(08/27/2015 12:28:00 AM,1)	(09/10/2014 11:08:00		(10/13/2014		
(08/27/2015 12:30:00 AM,1)	(09/10/2014 11:10:00		(10/13/2014		
(08/27/2015 12:30:00 PM,6)	(09/10/2014 11:12:00		(10/13/2014		
(08/27/2015 12:33:00 PM,1)	(09/10/2014 11:14:00		(10/13/2014 (10/13/2014		
(08/27/2015 12:35:00 AM,1)	(09/10/2014 11:15:00		(10/13/2014)		
(08/27/2015 12:35:00 PM,1)	(09/10/2014 11:15:00 (09/10/2014 11:20:00		(10/13/2014)		
(08/27/2015 12:38:00 PM,1)	(09/10/2014 11:20:00		(10/13/2014)		
(08/27/2015 12:45:00 PM,3)	(09/10/2014 11:20:00		(10/13/2014)		
(08/27/2015 12:49:00 AM,1)	(09/10/2014 11:23:00		(10/13/2014)		
(08/27/2015 12:49:00 PM,1)	(09/10/2014 11:24:00		(10/13/2014		
(08/27/2015 12:50:00 AM,1) (08/27/2015 12:50:00 PM,2)	(09/10/2014 11:25:00		(10/13/2014		
(08/27/2015 12:50:00 PM,2) (08/27/2015 12:55:00 AM,1)	(09/10/2014 11:20:00		(10/13/2014		
(08/27/2015 12:55:00 AM,1) (08/27/2015 12:55:00 PM,1)	(09/10/2014 11:30:00		(10/13/2014		
(08/28/2014 01:00:00 AM,5)	(09/10/2014 11:30:00		(10/13/2014		
(08/28/2014 01:00:00 AM, 3) (08/28/2014 01:00:00 PM, 10)	(09/10/2014 11:34:00		(10/13/2014		
(00/20/2014 01:00:00 FM,10)	, 20, 2021 2213 1100	, -,			

(11/02/2014 01:10:00 AM,3)	(12/16/2014 0	0.15.00	PM 2)	/12/21/2014	11.16.00	AM 1)
(11/02/2014 01:12:00 AM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:15:00 AM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:15:00 PM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:13:00 FM,1) (11/02/2014 01:20:00 AM,2)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:20:00 AM,2) (11/02/2014 01:20:00 PM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:20:00 PM,1) (11/02/2014 01:22:00 AM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:22:00 AH,1) (11/02/2014 01:24:00 PM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:24:00 PM,1) (11/02/2014 01:28:00 AM,1)				(12/31/2014		
	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:28:00 PM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:30:00 AM,5)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:30:00 PM,6)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:31:00 AM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:34:00 AM,1)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:35:00 AM,2)	(12/16/2014 0			(12/31/2014		
(11/02/2014 01:40:00 AM,3)	(12/16/2014 1			(12/31/2014		
(11/02/2014 01:43:00 AM,1)	(12/16/2014 1			(12/31/2014		
(11/02/2014 01:45:00 AM,1)	(12/16/2014 1			(12/31/2014		
(11/02/2014 01:45:00 PM,4)	(12/16/2014 1			(12/31/2014		
(11/02/2014 01:50:00 AM,2)	(12/16/2014 1			(12/31/2014		
(11/02/2014 01:55:00 AM,1)	(12/16/2014 1			(12/31/2014		
(11/02/2014 01:57:00 AM,2)	(12/16/2014 1			(12/31/2014		
(11/02/2014 02:00:00 AM,12)	(12/16/2014 1			(12/31/2014		
(11/02/2014 02:00:00 PM,10)	(12/16/2014 1			(12/31/2014		
(11/02/2014 02:05:00 PM,1)	(12/16/2014 1			(12/31/2014		
(11/02/2014 02:06:00 AM,1)	(12/16/2014 1			(12/31/2014		
(11/02/2014 02:07:00 AM,1)	(12/16/2014 1			(12/31/2014		
(11/02/2014 02:13:00 PM,1)	(12/16/2014 1			(12/31/2014	12:25:00	AM, 1)
(11/02/2014 02:15:00 PM,1)	(12/16/2014 1			(12/31/2014	12:27:00	PM, 1)
(11/02/2014 02:18:00 PM,1)	(12/16/2014 1			(12/31/2014	12:30:00	AM, 3)
(11/02/2014 02:21:00 AM,1)	(12/16/2014 1			(12/31/2014	12:30:00	PM, 6)
(11/02/2014 02:23:00 AM,1)	(12/16/2014 1			(12/31/2014		
(11/02/2014 02:30:00 AM,5)	(12/16/2014 1			(12/31/2014	12:40:00	AM, 1)
(11/02/2014 02:30:00 PM,1)	(12/16/2014 1			(12/31/2014	12:40:00	PM, 1)
(11/02/2014 02:34:00 PM,1)	(12/16/2014 1			(12/31/2014	12:45:00	AM, 2)
(11/02/2014 02:35:00 PM,1)	(12/16/2014 1			(12/31/2014	12:45:00	PM, 1)
(11/02/2014 02:36:00 PM,2)	(12/16/2014 1			(12/31/2014	12:48:00	PM, 1)
(11/02/2014 02:40:00 AM,1)	(12/16/2014 1			(12/31/2014	12:50:00	PM, 2)
(11/02/2014 02:45:00 AM,1)	(12/16/2014 1			(12/31/2014	12:55:00	AM, 1)
(11/02/2014 02:45:00 PM,3)	(12/16/2014 1			(12/31/2014	12:59:00	PM, 1)
(11/02/2014 02:52:00 AM,1)	(12/16/2014 1	0:50:00	AM, 1)	grunt>		