Assignment 2.3

Enter grunt shell with command:

```
[acadgild@localhost ~]$ pig -x LOCAL
```

Load text file

```
grunt> text_input = LOAD '/home/acadgild/imput.txt' AS (line:chararray);
```

Separate words by using TOKENIZE.
Since foreach will generate bags of token, hence use FLATTEN to get them flattened.

```
grunt> the_words = foreach text_input GENERATE FLATTEN(TOKENIZE(line)) as word;
```

Grouping by words and using COUNT to get the total count of each word.

```
grunt> words_grouped = GROUP the_words BY word;
grunt> word_count = FOREACH words_grouped GENERATE group,COUNT(the_words);
grunt> dump word_count;
```

OUTPUT:

```
(14,1)
(15,1)
(Bax,1)
(Cay,1)
(Dal,1)
(Del,1)
(Male,11)
(Saur,1)
(Tome,1)
(Webb,1)
(Ilise,1)
(Jasun,1)
(Porte,1)
(Anstee,1)
(Brolan,1)
(Carlye,1)
(Female,4)
(Keefer,1)
(Liddel,1)
(Seeley,1)
(Trever,1)
(Whyard,1)
(Burgess,1)
(Cameron,1)
(Joselin,1)
(Larkcum,1)
(Mallory,1)
(Martine,1)
(Moreinu,1)
(Quilter,1)
(Annmaria,1)
(Terrence,1)
(Geroldini,1)
(Whitington,1)
(34.9.187.58,1)
(157.87.97.11,1)
(22.28.81.135,1)
(103.12.209.61,1)
(166.181.35.71,1)
(41.197.110.62,1)
(76.104.44.154,1)
(kcayc@mit.edu,1)
(180.125.107.20,1)
(209.226.45.138,1)
(38.184.231.111,1)
(100.243.145.200,1)
(126.161.252.141,1)
(188.225.247.186,1)
(218.250.204.131,1)
(253.240.112.101,1)
(tbrolan4@bbb.org,1)
(twhyardd@nih.gov,1)
(dporte5@patch.com,1)
(isaurb@disqus.com,1)
(tjoselin3@mlb.com,1)
(dquilter9@noaa.gov,1)
(clarkcum0@youku.com,1)
(janstee2@latimes.com,1)
(wgeroldinia@cnet.com,1)
(mmoreinu7@pcworld.com,1)
(bliddel8@over-blog.com,1)
(bmartinee@nydailynews.com,1)
(aseeley6@independent.co.uk,1)
```