

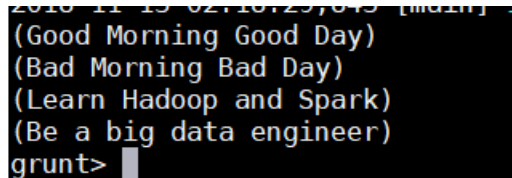
Task 1 : Write a program to implement wordcount using Pig

Load the data set

```
grunt> lines = LOAD '/home/acadgild/Desktop/Practise/PIG/wordcount.txt' USING PigStorage('\n') as  
(line:chararray);
```

Display the data set

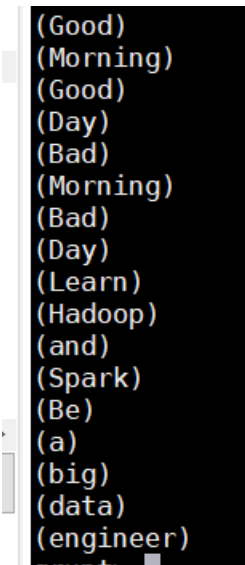
```
grunt> dump lines;
```



```
(Good Morning Good Day)  
(Bad Morning Bad Day)  
(Learn Hadoop and Spark)  
(Be a big data engineer)  
grunt>
```

Now we will read each line and we will separate every single word

```
grunt> words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line, ' ')) as word;  
grunt> dump words;
```



```
(Good)  
(Morning)  
(Good)  
(Day)  
(Bad)  
(Morning)  
(Bad)  
(Day)  
(Learn)  
(Hadoop)  
(and)  
(Spark)  
(Be)  
(a)  
(big)  
(data)  
(engineer)  
grunt>
```

Now we will group the word

```
grunt> grpwords = GROUP words BY word;  
grunt> dump grpwords;
```

```
(a,{(a)})  
(Be,{(Be)})  
(Bad,{(Bad),(Bad)})  
(Day,{(Day),(Day)})  
(and,{(and)})  
(big,{(big)})  
(Good,{(Good),(Good)})  
(data,{(data)})  
(Learn,{(Learn)})  
(Spark,{(Spark)})  
(Hadoop,{(Hadoop)})  
(Morning,{(Morning),(Morning)})  
(engineer,{(engineer)})
```

Now we will count the words in each group

```
grunt> cntwords = FOREACH grpwords GENERATE group, COUNT(words);  
grunt> dump cntwords;
```

```
(a,1)  
(Be,1)  
(Bad,2)  
(Day,2)  
(and,1)  
(big,1)  
(Good,2)  
(data,1)  
(Learn,1)  
(Spark,1)  
(Hadoop,1)  
(Morning,2)  
(engineer,1)  
grunt>
```