Program to implement wordcount using Pig.

```
input_file = LOAD '/pig_data/words.txt' using PigStorage as (line:chararray);
dump input_file;
```

```
(hadoop hive pig oozie)
(hive hbase)
(hdfs hadoop kafka)
(pig)
(spark scala hive)
(scala)
```

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

```
(hadoop)
(hive)
(pig)
(oozie)
(hive)
(hbase)
(hdfs)
(hadoop)
(kafka)
(pig)
(spark)
(scala)
(hive)
(scala)
```

```
words_grouped = GROUP words_tokenize by word;
dump words_grouped;
```

```
(pig,{(pig),(pig)})
(hdfs,{(hdfs)})
(hive,{(hive),(hive),(hive)})
(hbase,{(hbase)})
(kafka,{(kafka)})
(oozie,{(oozie)})
(scala,{(scala),(scala)})
(spark,{(spark)})
(hadoop,{(hadoop),(hadoop)})
```

```
words_count = FOREACH words_grouped GENERATE group,COUNT(words_tokenize);
dump words_count;
```

```
(pig,2)
(hdfs,1)
(hive,3)
(hbase,1)
(kafka,1)
(oozie,1)
(scala,2)
(spark,1)
(hadoop,2)
```