download

```
wget https://archive.apache.org/dist/pig/pig-0.16.0/pig-0.16.0.tar.gz
tar xvf pig-0.16.0.tar.gz
```

modify profile

```
vim ~/.bashrc
```

```
export PIG_HOME=${HADOOP_BASE_PATH}/pig-0.16.0
export PATH=$PATH:${PIG_HOME}/bin

# hadoop config setting
export PIG_CLASSPATH=${HADOOP_BASE_PATH}/hadoop-2.7.3/etc/hadoop
```

start as localhost session

```
$PIG_HOME/bin/pig -x local
# $HADOOP_HOME/bin/hdfs dfs -ls /input
grunt> ls /input
```

start as service

```
$PIG_HOME/bin/pig
# $HADOOP_HOME/bin/hdfs dfs -ls /input
grunt> ls /input
```

WordCount

```
// 加载输入文件，并按行分隔
grunt> a = LOAD '/input/immortals.txt' as (line:chararray);
// 将每行分割成单词
grunt> words = FOREACH a GENERATE flatten(TOKENIZE(line)) as w;
// 按单词分组
grunt> g = GROUP words by w;
// 单词记数
grunt> wordcount = FOREACH g GENERATE group,COUNT(words);
```

WordCount2

带词频倒排序

```
a = LOAD '/input/immortals.txt' as (line:chararray);
words = FOREACH a GENERATE flatten(TOKENIZE(line)) as w;
g = GROUP words by w;
// 给单词数所在列加一个别名count
wordcount = FOREACH g GENERATE group,COUNT(words) as count;
// 将结果列交换，将变成{count，word}这种结构
r = foreach wordcount generate count,group;
// 按count分组
g2 = group r by count;
// 去掉无用的列
x = foreach g2 generate group,r.group;
// 按count倒排
y = order x by group desc;
```