



Pig script:

lines = LOAD '/home/acadgild/wordcountdoc' AS (line:chararray);

words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;

grouped = GROUP words by word;

wordcount =FOREACH grouped GENERATE group, COUNT(words);

DUMP wordcount;