

HW1

Deadline: Sep. 26th 5:59 P.M. (before class)

Description

In the lab session, we have shown how to perform WordCount using Hadoop. Now in this homework, we ask you to modify the WordCount example and write a program to **count the 2nd letter of each word**.

Input

Download the Complete Works of William Shakespeare from Project Gutenberg at <http://www.gutenberg.org/cache/epub/100/pg100.txt> or download the pg100.txt from Blackboard.

TODO

In your implementation, you need to

1. transform all words to lowercase
2. ignore all non-alphabetic characters (except whitespace)
3. count nothing if there's only one alphabetic character in a word (e.g. A2)

Output must be in the following format

```
a<\tab>count of the 2nd letters "a" (e.g. bank)
b<\tab>count of the 2nd letters "b" (e.g. ****abstract, a--b--c)
...
```

Example output:

```
a      666
b      233
...
```

Submission

Upload your **java** file(s) (Mapper, Reducer, Driver) and output results to blackboard.

Note that this homework requires Java with Hadoop framework. Other language is not accepted.