The goal for this assignment is to get a Map-Reduce library up and running on your machine and to do a few exercises to gain familiarity with the Map-Reduce paradigm.

I recommend MrJob as a Map-Reduce library. Here is a helpful tutorial on MrJob.
https://pythonhosted.org/mrjob/guides/quickstart.html

Additionally, there is useful information in the Python and Hadoop book. (If you haven't already, you should download a copy of this book. It is free from the publisher's website.)
http://www.oreilly.com/programming/free/hadoop-with-python.csp

On Canvas, I have placed a pair of example text files that you can use as input for your MapReduce programs. (I have removed punctuation from the files, and so you don't need to worry about that for this first assignment.)

Please attempt the following:

1. Write a MapReduce program that counts the total number of words in a file.

2. Write a MapReduce program that counts the number of times "the" appears in the file.

3. Write a MapReduce program that counts the number of 3-letter words in a file.

4. Write a MapReduce program that counts the number of words that begin with each letter. (That is, how many words start with 'a'? How many words start with 'b'? … )

5. Write a MapReduce program that for each possible word length computes the average number of vowels in the word. (That is, words with 4 letters typically have how many vowels? Words with 5 letters typically have how many vowels?)