

Assignment 5: Word Count using Map Reduce

Problem Description:

The purpose of this assignment is to get familiar and gain practical experience with the MapReduce programming model. MapReduce is used by Google for processing large data sets (terabytes of data). You will build your assignment on top of the Hadoop software platform. Hadoop is an open-source version of Map Reduce written in Java. For your assignment you are required to use Hadoop as a local node on your machine and solve the task below.

Now you:

1. Write a simple map-reduce program using Hadoop to count the number of words that appear in the any dataset (the dataset consists of multiple documents that you create). Use the stop word list to avoid stop words:
<http://www.textfixer.com/resources/common-englishwords-with-contracti-ns.txt>
2. Create a short documentation in which you briefly describe your implementation, such that somebody who has not seen your code can understand what you did. (Not more than 1 page).

Deadline 14 April 2021 Midnight