Programming Assignment 1

1 Assignments

1. If applicable, install Oracle virtual box and Linux on your Windows OS or Mac OS.

   - Follow instructions in virtualbox_Ubuntu_install.docx
   - Skip this step if your OS is Linux.

2. Follow the instructions in HadoopSetup_3.2.0.doc to install Java and Hadoop on your Linux system.

3. Test the given WordCount program.

4. Change the contents in "input" folder using the provided two files as follows.

   - file0: contains ∼100 lines of texts.
   - file1: contains ∼150 lines of texts.

5. Revise the code so that it not only performs the word count, but also figures out how many times `map` method is invoked in the `map` stage.

   - The number of times the `map` method is invoked needs to be written to one output file.

2 Submission

- In answer.txt, (1) explain how to figure out how many time the `map` method is invoked in the `map` stage, and (2) provide your solution given the input.

- tar -cvf pa1_<your last name>.tar input output WordCount.java answer.txt

- Upload the tar file to blackboard before the deadline

  – Deadline: January 29 (Tuesday), 2019 @ 11:59AM