In this activity, you will utilize the word count program from [Mini-Lesson 18.4](https://classroom.emeritus.org/courses/10605/pages/mini-lesson-18-dot-4-executing-the-word-count-program-in-hadoop-60-00) on a larger text file that you will download from [Project Gutenberg](https://gutenberg.org/) .

An alternate site for downloading the required .zip file is the [Internet Archive](https://archive.org/details/mobydick00015gut) . If you have difficulty accessing Project Gutenberg, please download the .zip file from the [Internet Archive](https://archive.org/details/mobydick00015gut)

instead.

**Reference**

[Project Gutenberg. “Welcome to Project Gutenberg.” *Project Gutenberg*. Accessed December 21, 2021. https://gutenberg.org/](https://gutenberg.org/)

[Links to an external site.](https://gutenberg.org/)

Prior to beginning this activity, review the submission instructions below to ensure that you collect the required screenshots as you progress through the activity.

1. Make sure that you have installed Hadoop in Docker and that all of the *containers* are running.  
   Provide a screenshot of your Docker desktop to show all of the Hadoop *containers* running.
2. Open your browser to [Project Gutenberg](https://gutenberg.org/)
3. and look for *Moby Dick* by Herman Melville. Download the .zip file containing the book as a text (.txt) file to your local machine. Unzip the file to your local machine. Provide a screenshot of your local machine to show that you successfully downloaded the Moby Dick .zip file and unzipped its contents to your local machine. (If you have any difficulty accessing Project Gutenberg, please try downloading the .zip file from the [Internet Archive](https://archive.org/details/mobydick00015gut)
4. instead.)
5. From your Docker desktop, select <CLI> from the namenode *container*. From the bash window, create a folder in the namenode *container* and call it input. Provide a screenshot to show that you successfully created the input folder in the namenode *container*.
6. From your local machine, run a Docker cp command to copy the .txt file you downloaded from [Project Gutenberg](https://gutenberg.org/)
7. or the [Internet Archive](https://archive.org/details/mobydick00015gut)
8. to the namenode input *container*. Provide a screenshot to show that you successfully copied the .txt file to the namenode *container*.
9. From the bash window, run a HDFS command to create a folder named input. Provide a screenshot to show that you successfully created an input folder.
10. From the bash window, run a HDFS command to copy the contents of the local input folder to the HDFS input Provide a screenshot to show that you successfully ran the HDFS command to copy the contents of the local input folder to the HDFS input folder.
11. From the bash window, run the curl command to download the hadoop-mapreduce-examples-2.7.1-sources.jar jar file to the current *directory*. Provide a screenshot to show that you successfully ran the curl command to download the jar file.
12. From the bash window, run the word count program. Provide a screenshot to show that you successfully ran the word count program. For this step, ensure that the output *directory* does not exist already. If it does, you can delete it before completing this step by running the command rm -r output.
13. From the bash window, run the cat command on the HDFS output *directory* to display the contents of the file. Provide a screenshot showing that you successfully executed the cat command to display the contents of the file.

In this activity, you ran a word count program to illustrate Hadoop’s MapReduce architecture. You also manipulated the HDFS so that Hadoop could process files.

**Submission Instructions:**

Your submission for this activity should be a Word document that includes the following screenshots, each labeled for the step that the screenshot represents:

1. Provide a screenshot of your Docker desktop to show the Hadoop *containers* running.
2. Provide a screenshot of your local machine to show that you successfully downloaded the Moby Dick .zip file and unzipped its contents to your local machine.
3. Provide a screenshot to show that you successfully created the input folder in the namenode *container*.
4. Provide a screenshot to show that you successfully copied the .txt file to the namenode *container*.
5. Provide a screenshot to show that you successfully created an input folder.
6. Provide a screenshot to show that you successfully ran the HDFS command to copy the contents of the local input folder to the HDFS input folder.
7. Provide a screenshot to show that you successfully ran the curl command to download the jar file.
8. Provide a screenshot to show that you successfully ran the word count program.
9. Provide a screenshot to show that you successfully executed the cat command to display the contents of the file.