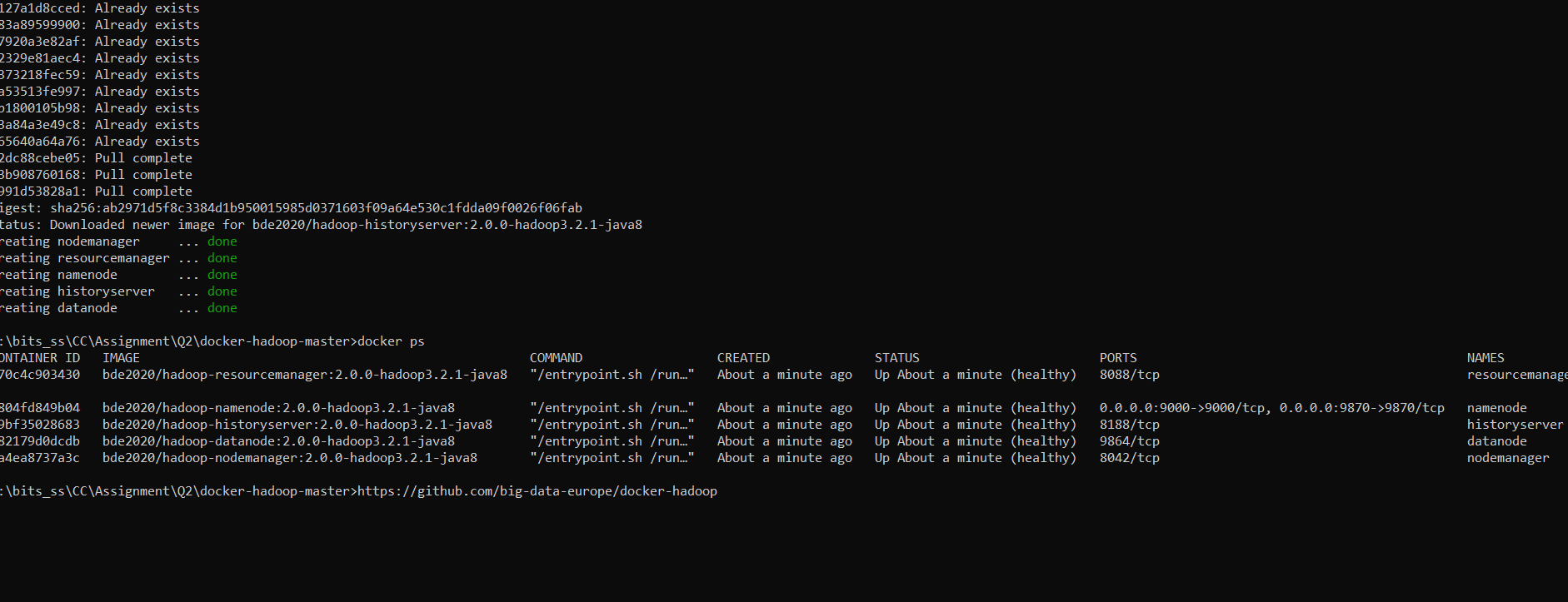
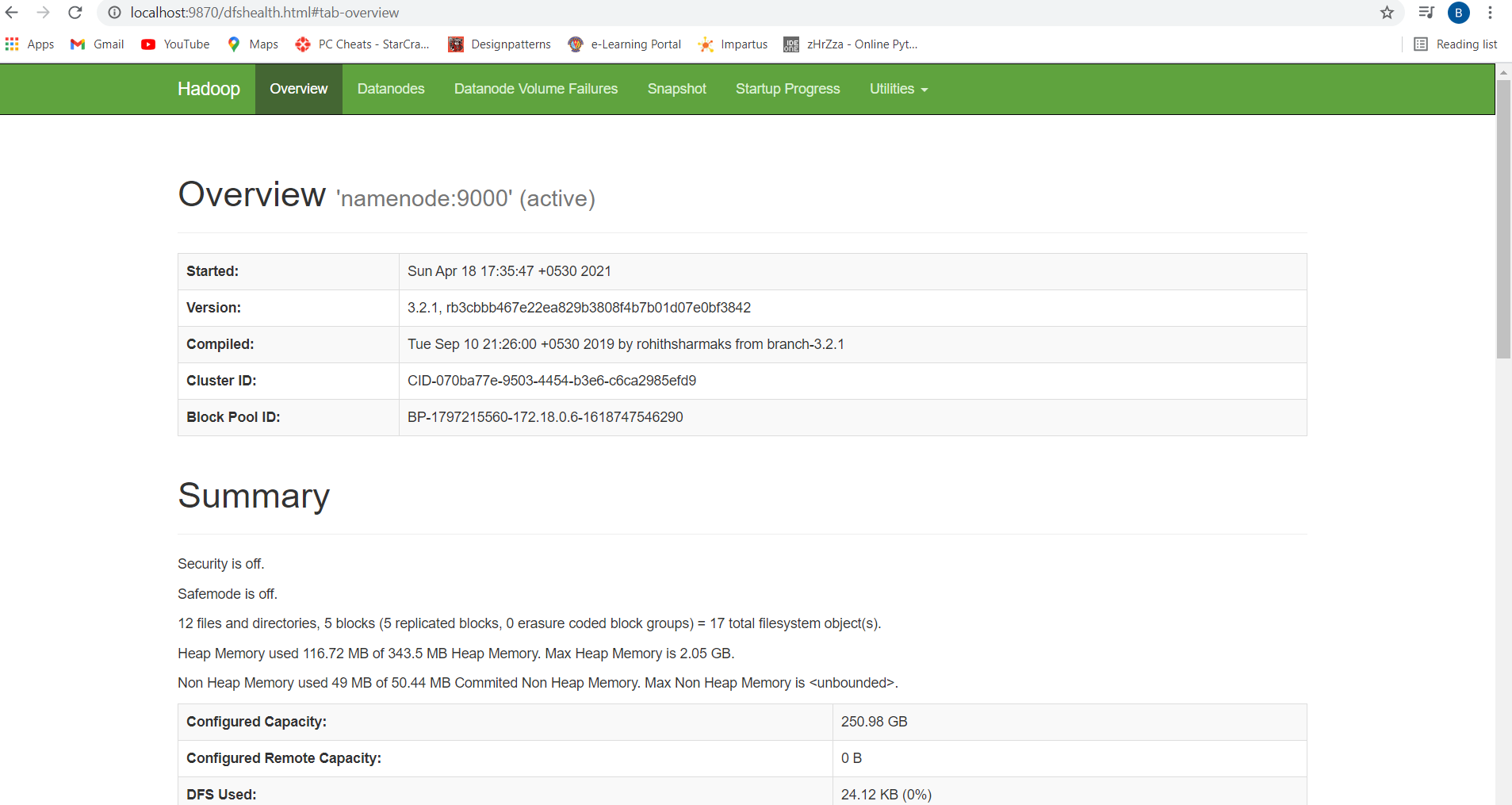
Steps and Observations:

1. Setup Hadoop locally on docker container. Use this github repo: <https://github.com/big-data-europe/docker-hadoop>. Host this on docker container.
2. Following Hadoop cluster is loaded.



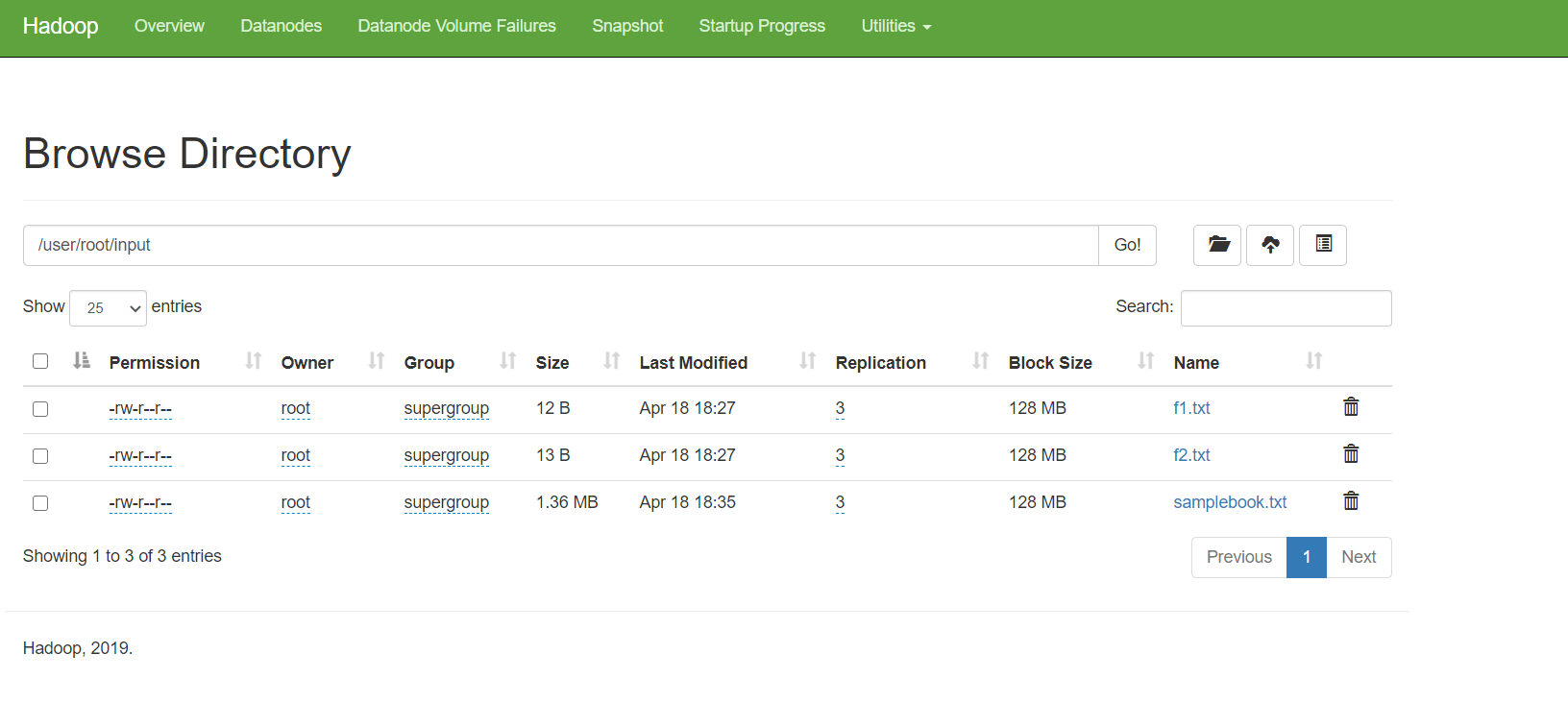
1. The following website is loaded.



1. This completes the Hadoop setup

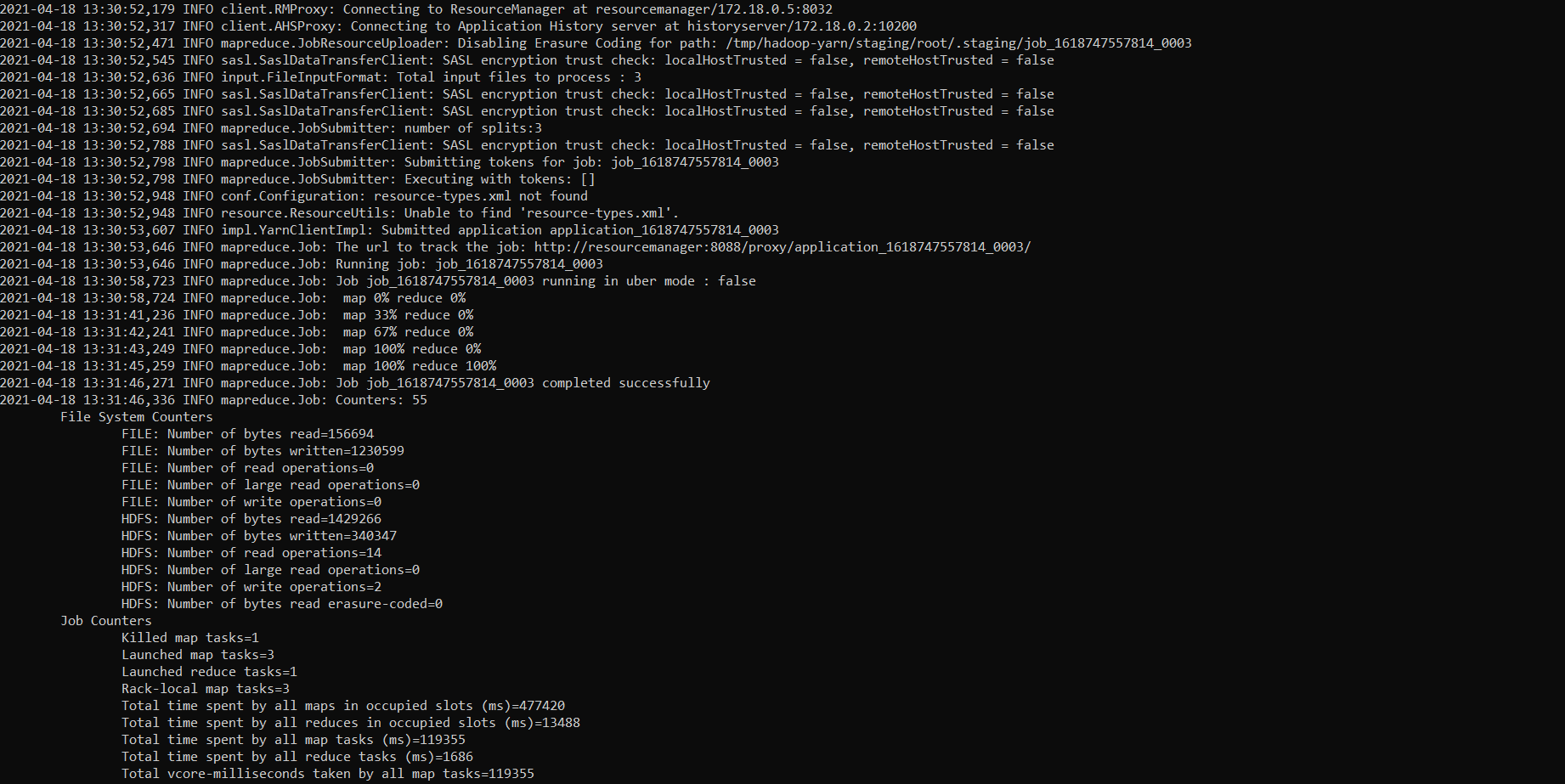
Running the word count program

1. Created 2 python files, one for mapping called mapper.py and the other to reduce called reducer.py.
2. Taken a samplebook.txt which is a long text. There are 32762 lines in the text file
3. Uploaded the mapper.py and reducer.py the docker container
4. Then logged into the docker container using **docker exec -it namenode bash**
5. Using docker cp command pushed the samplebook.txt in this container.
6. Then created a hdfs input folder using command: **hadoop fs -mkdir -p input.** Following that move the input files in the folder using the command: **hdfs dfs -put ./input/\* input**

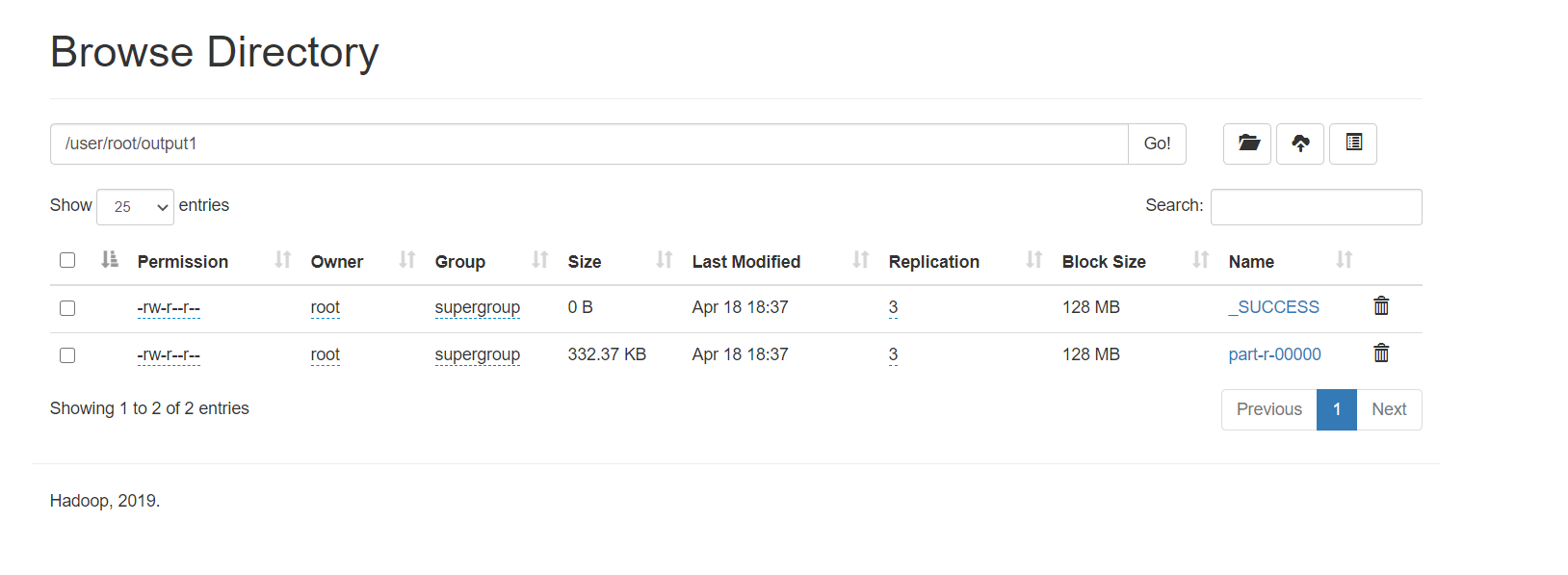


1. Run the mapper and reducer command on the input folder of the hdfs, using the below command:

**hadoop jar contrib/streaming/hadoop-0.20.0-streaming.jar -file /home/hadoop/mapper.py -mapper /home/hadoop/mapper.py -file /home/hadoop/reducer.py -reducer /home/hadoop/reducer.py input output1**



1. The command will do the word count and dump the file in output folder.

. 

1. To print out the WordCount program result: hdfs dfs -cat output/part-r-00000
2. I have copied the output file out of the hdfs into local folder as **output.txt. Please check the output.txt file in the current folder.**