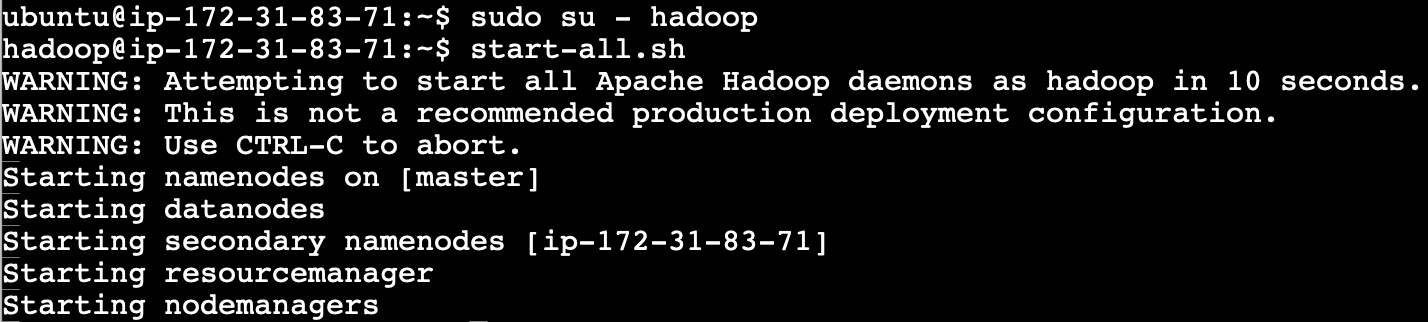
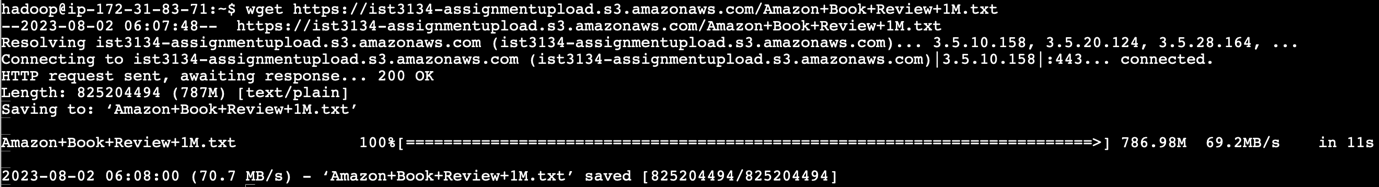
**MapReduce WordCount with Python in AWS**

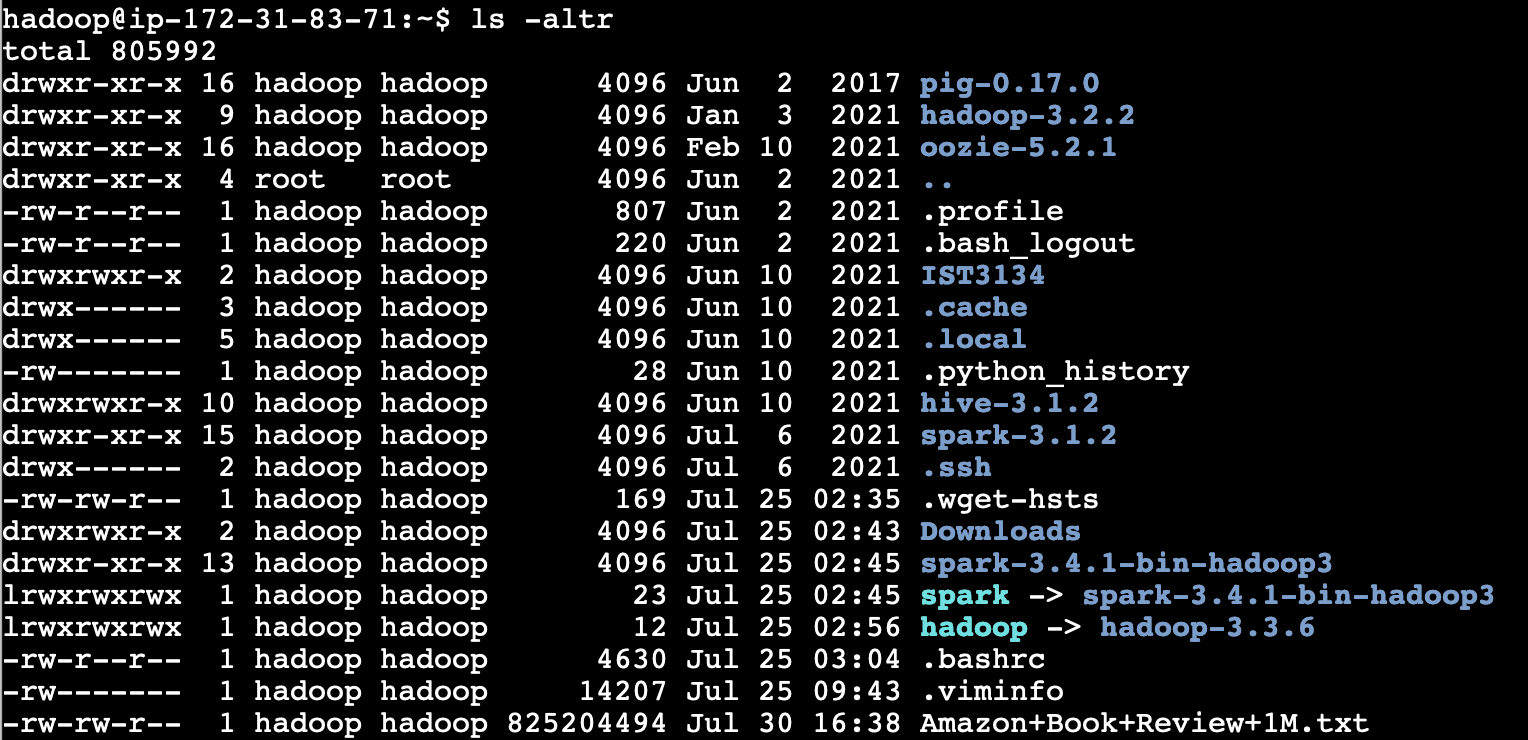
1. Enter /home/Hadoop directory and start the HDFS services.



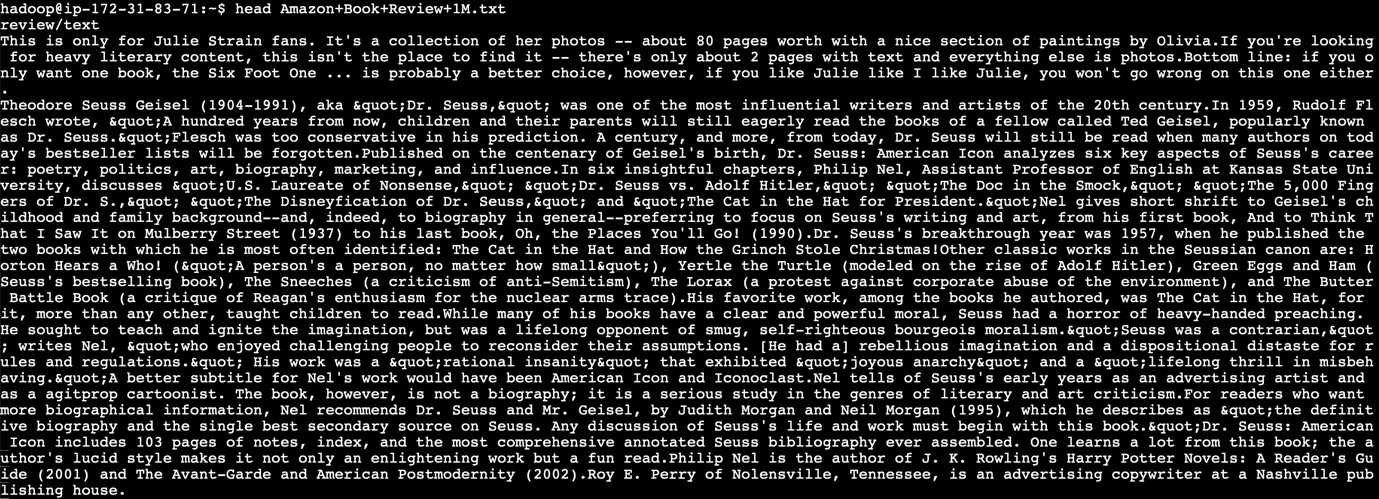
1. WGET datasets uploaded in Amazon S3 bucket



1. The Amazon Book Review file is stored in home directory



1. The contents of Amazon Book Review file

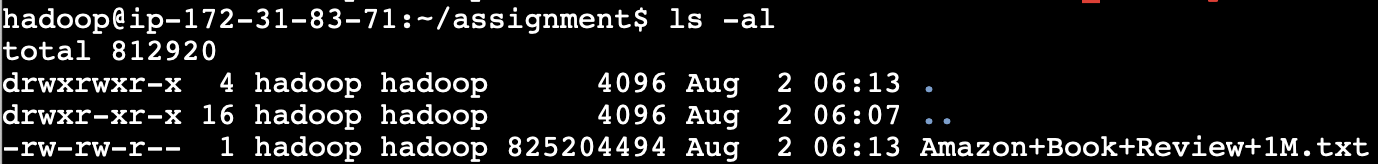


1. Make assignment directory to store the files for this assignment



1. Copy Amazon Book Review text file to assignment directory

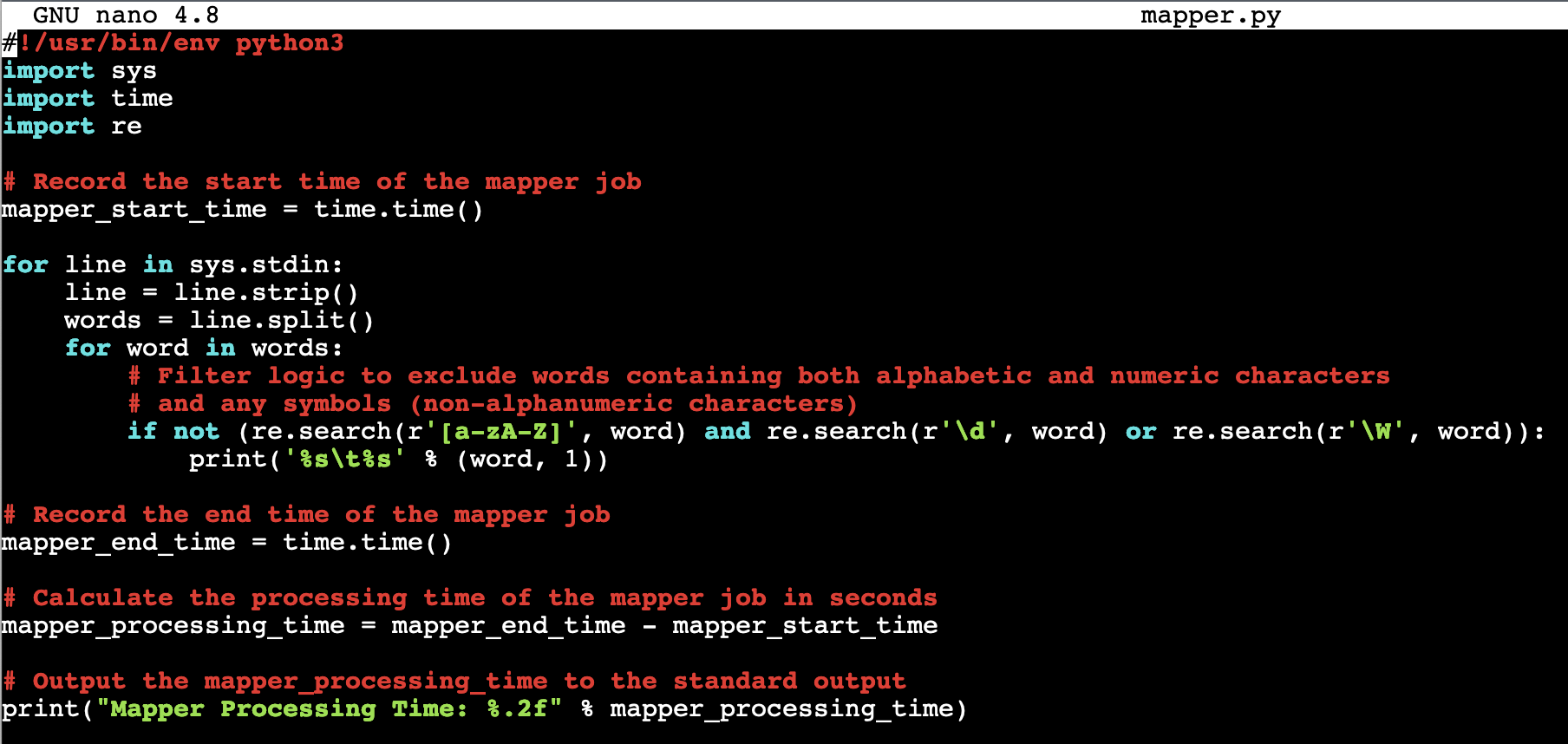




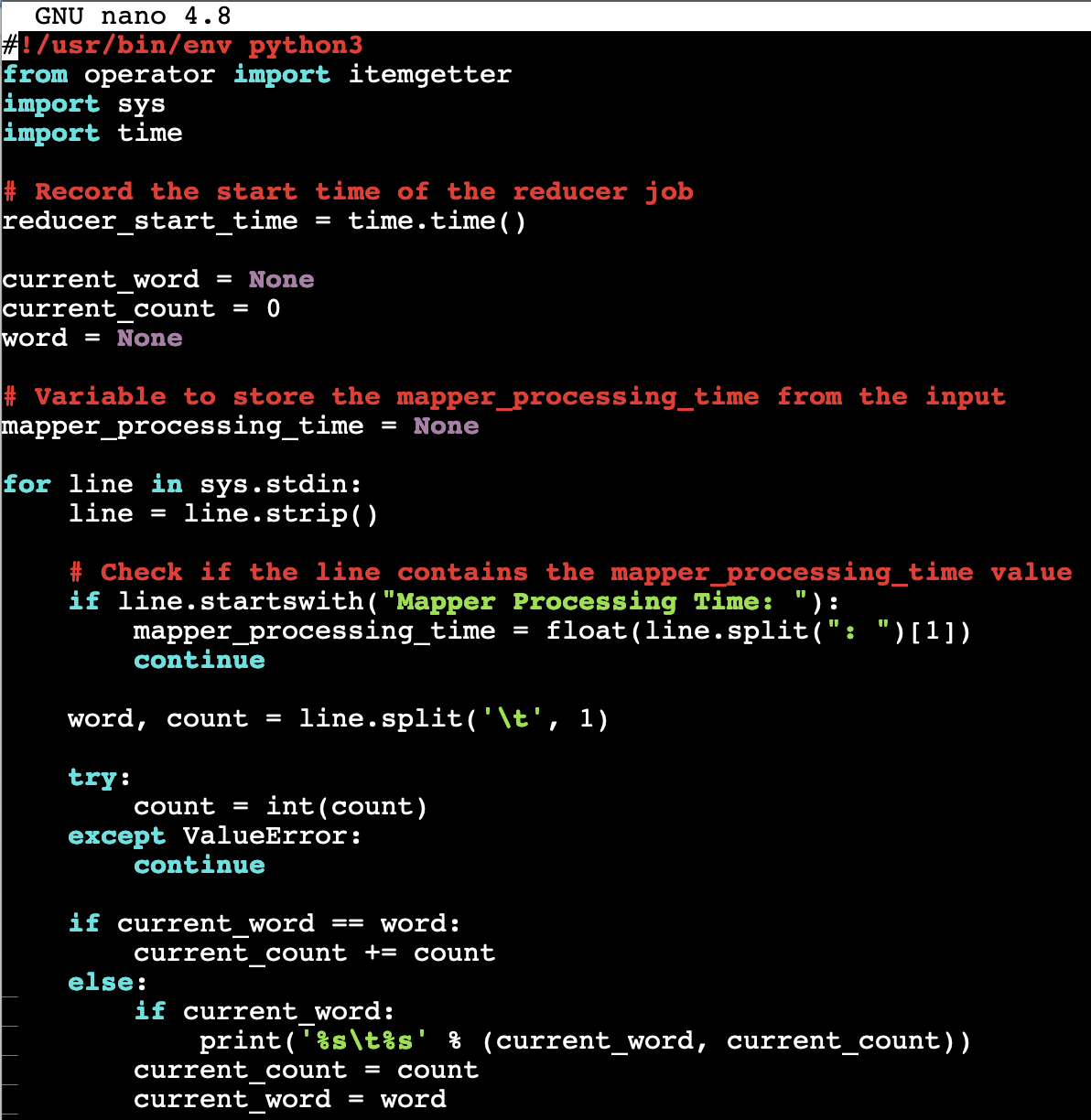
1. Create a python directory within assignment directory to store the python scripts

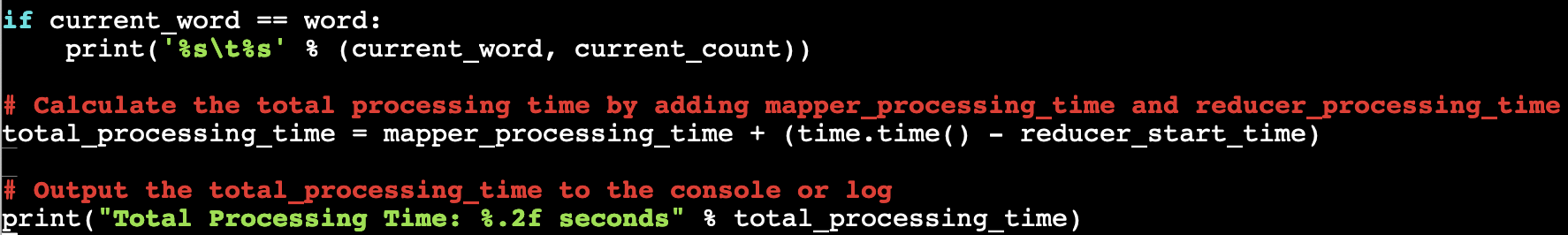


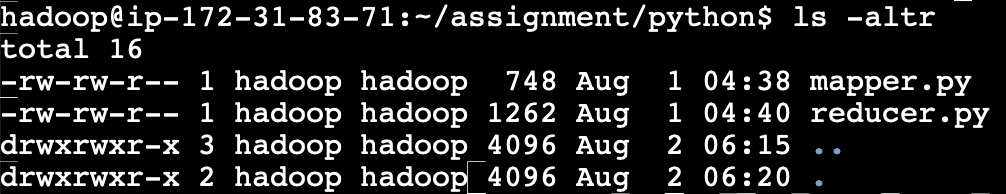
1. Write python script based on script given during tutorial class
   1. Mapper.py script



* 1. Reducer.py script







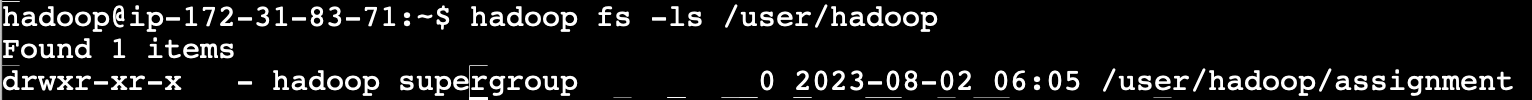
1. Create a directory in the HDFS server



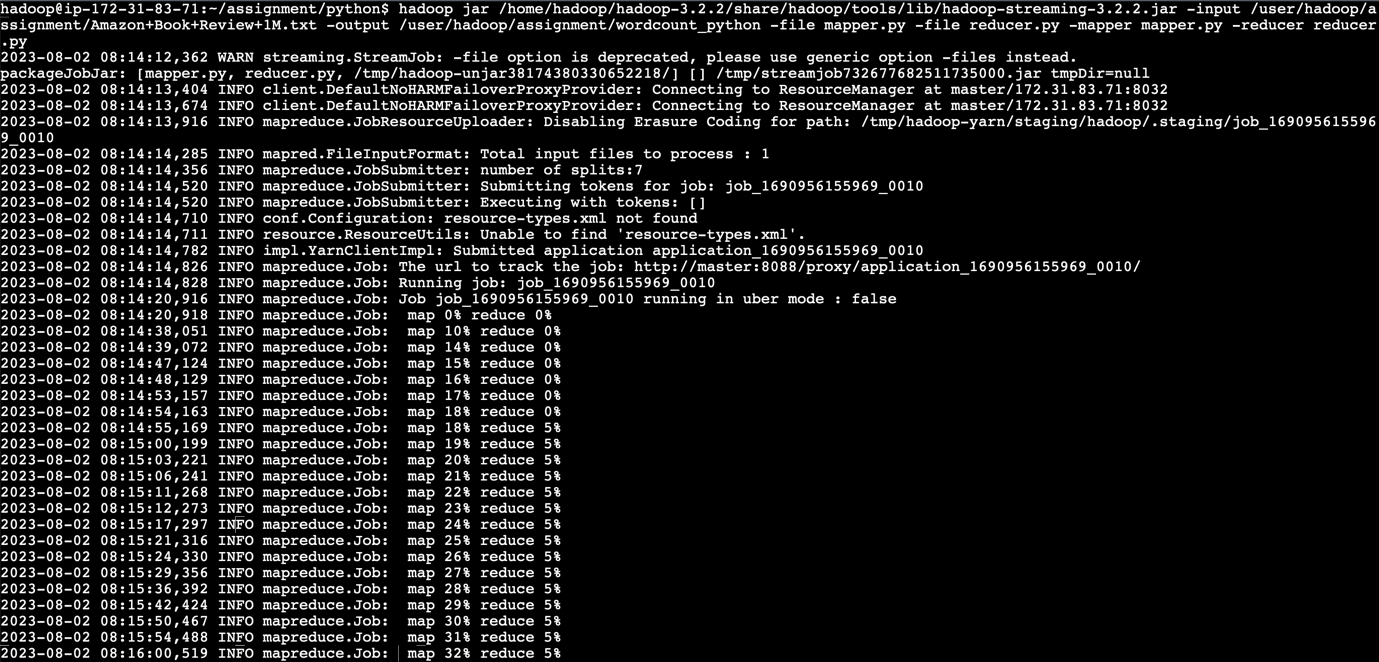


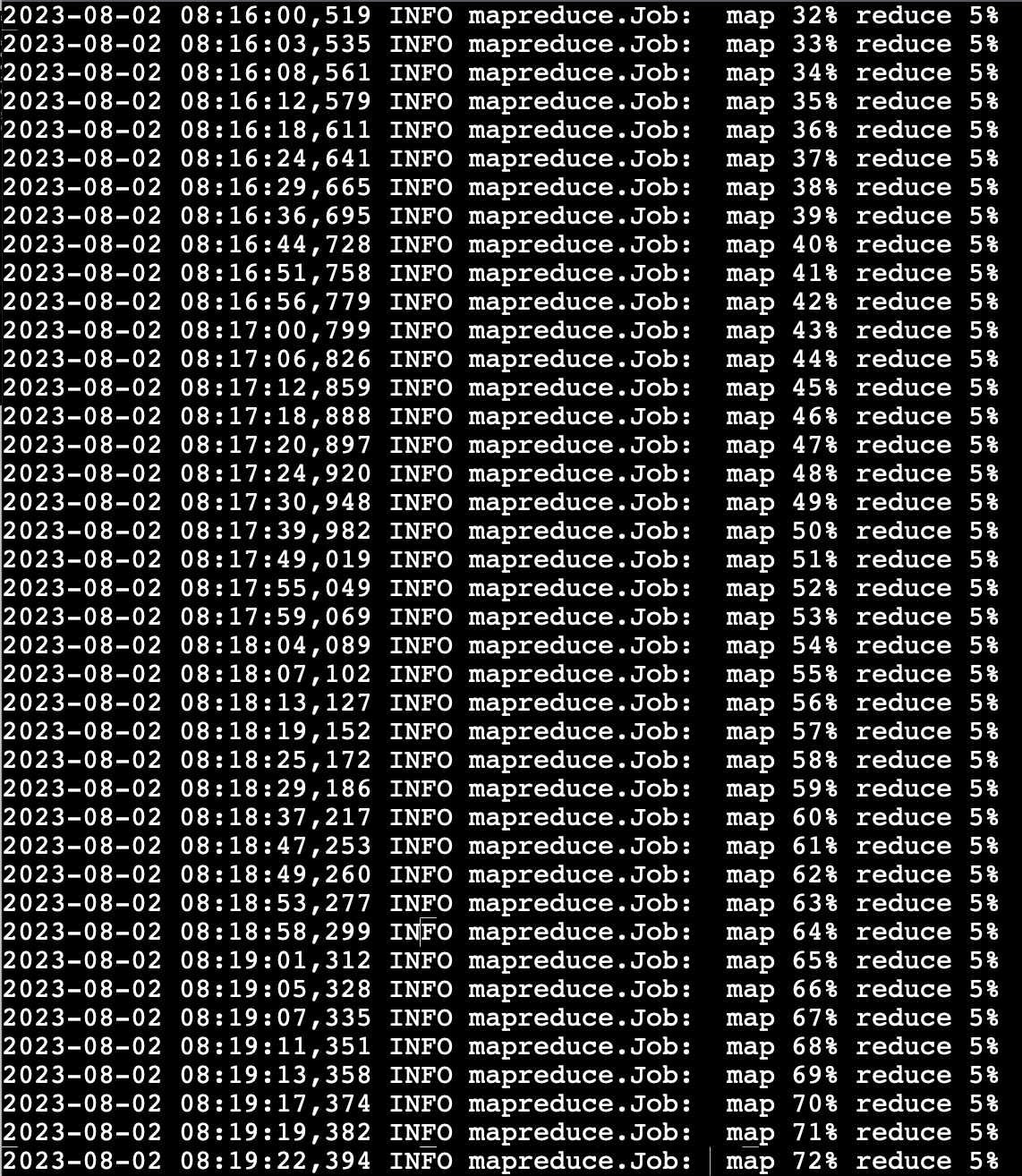
1. Put the assignment folder into the HDFS server

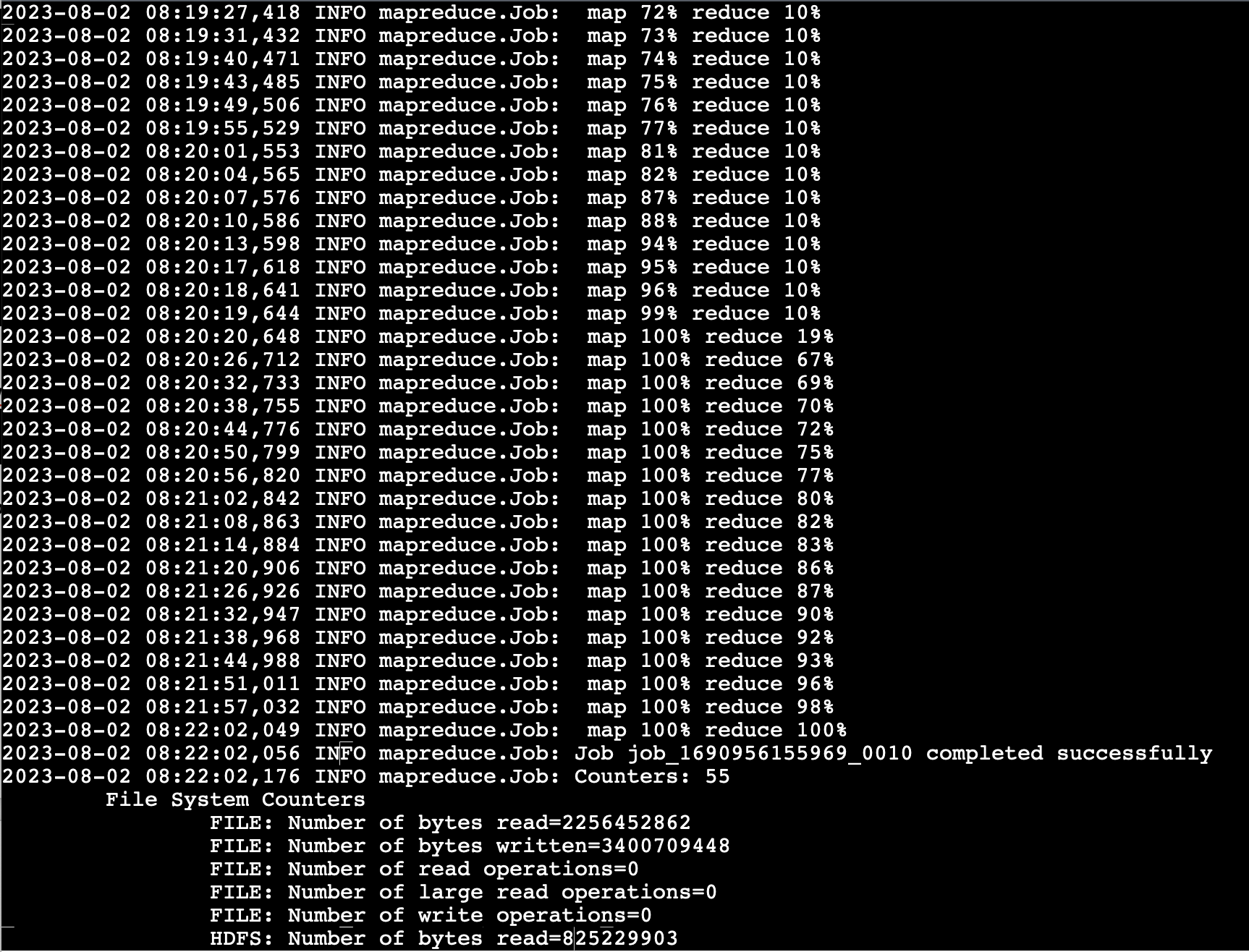


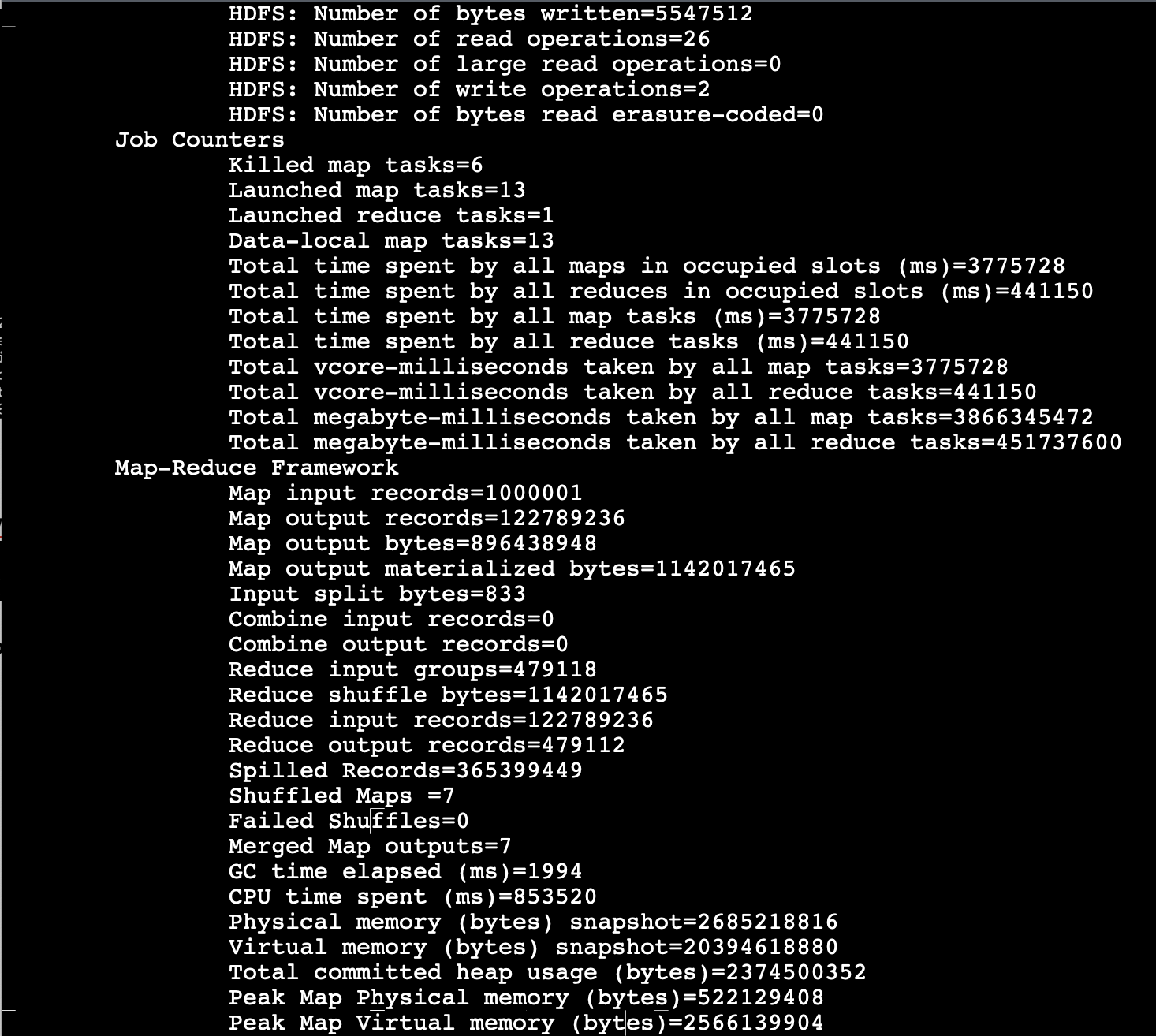


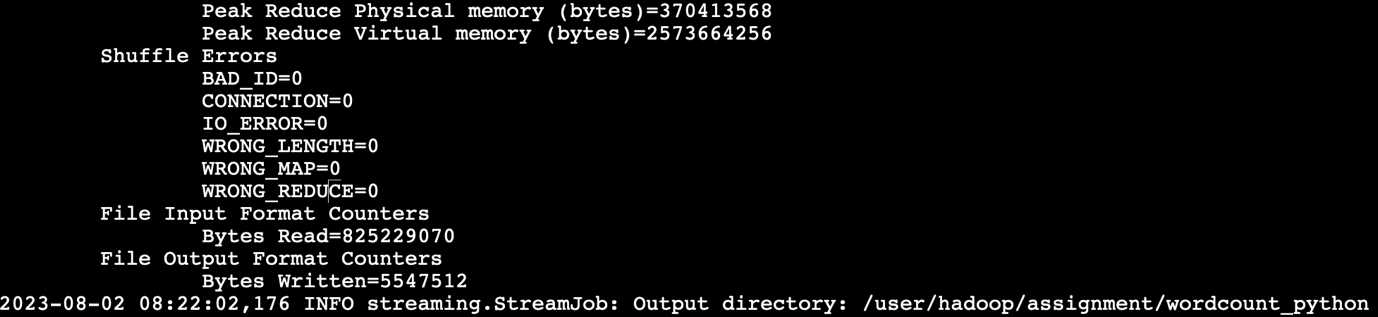
1. Command to run the MapReduce job



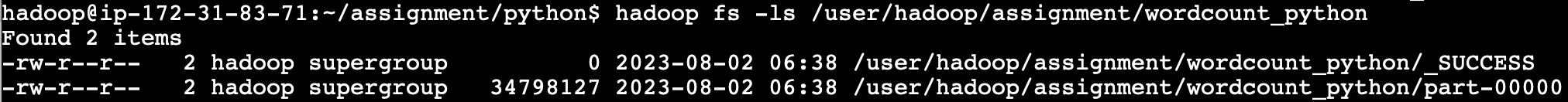




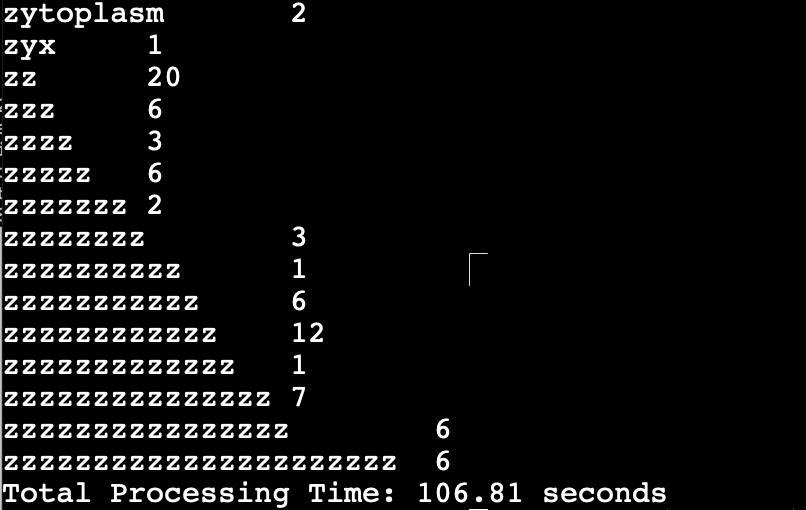




1. Show the output of the MapReduce WordCount job



1. Show the content of the WordCount job and the processing time



1. Store in home directory for download

