Tracy Michaels Assignment 1

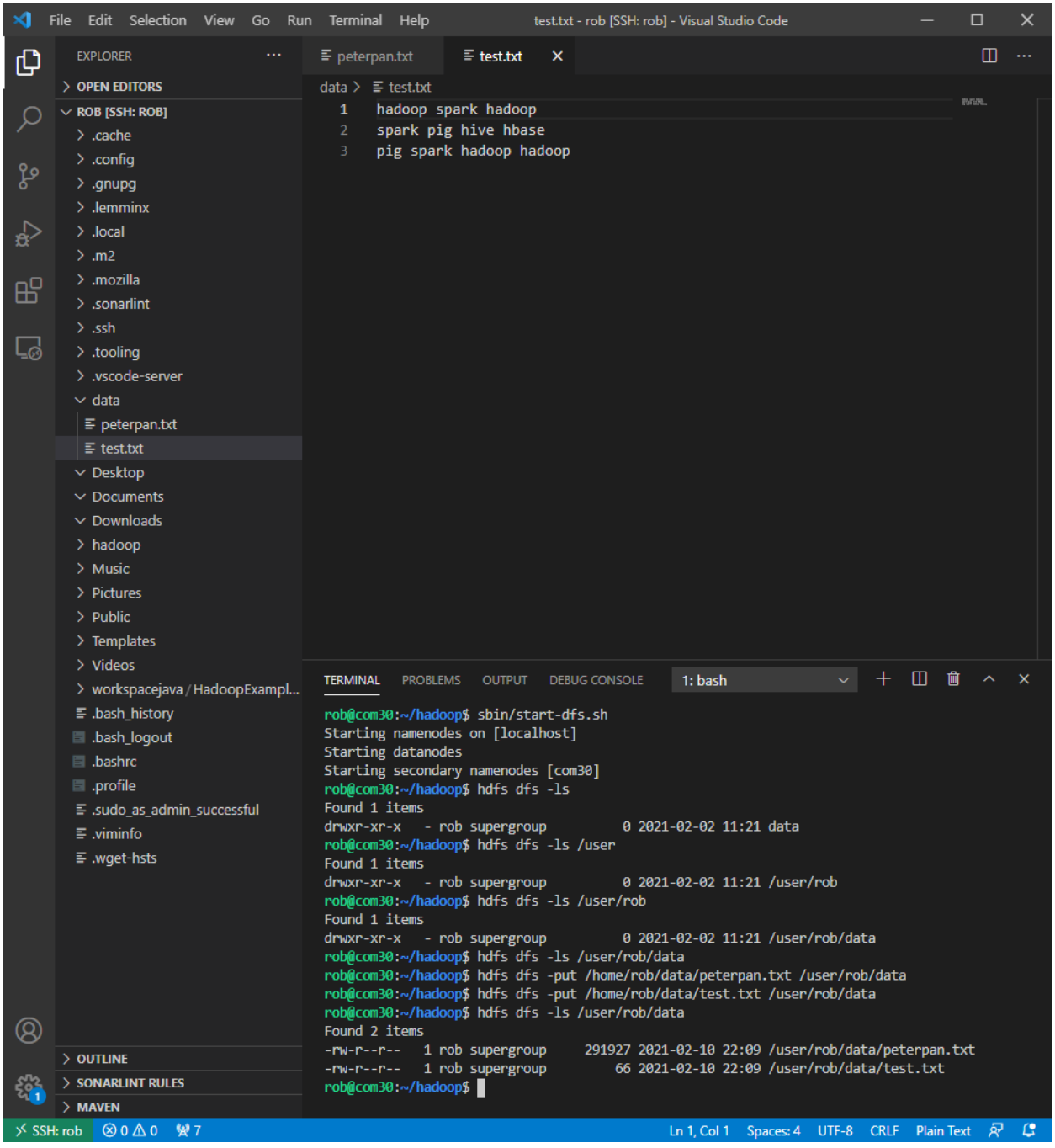
1) Setting up Hadoop.

- Install java JDK and JRE
- Download Hadoop .tar file
- Uncompress file
- Update Hadoop-env.sh file to include 'export JAVA_HOME=/usr/lib.jvm/java-11-openjdk-amd64/'
- Create tmp folder in hadoop
- Update core-site.xml in etc directory to include necessary properties
- Do the same for hdfs-site.xml
- Update .bashrc for Hadoop to append following lines to file
    - Export HADOOP_HOME=/home/rob/Hadoop
    - Export PATH=$HADOOP_HOME/bin:$PATH

(I do not have screen shots of this part as I had set it up before this assignment was assigned)

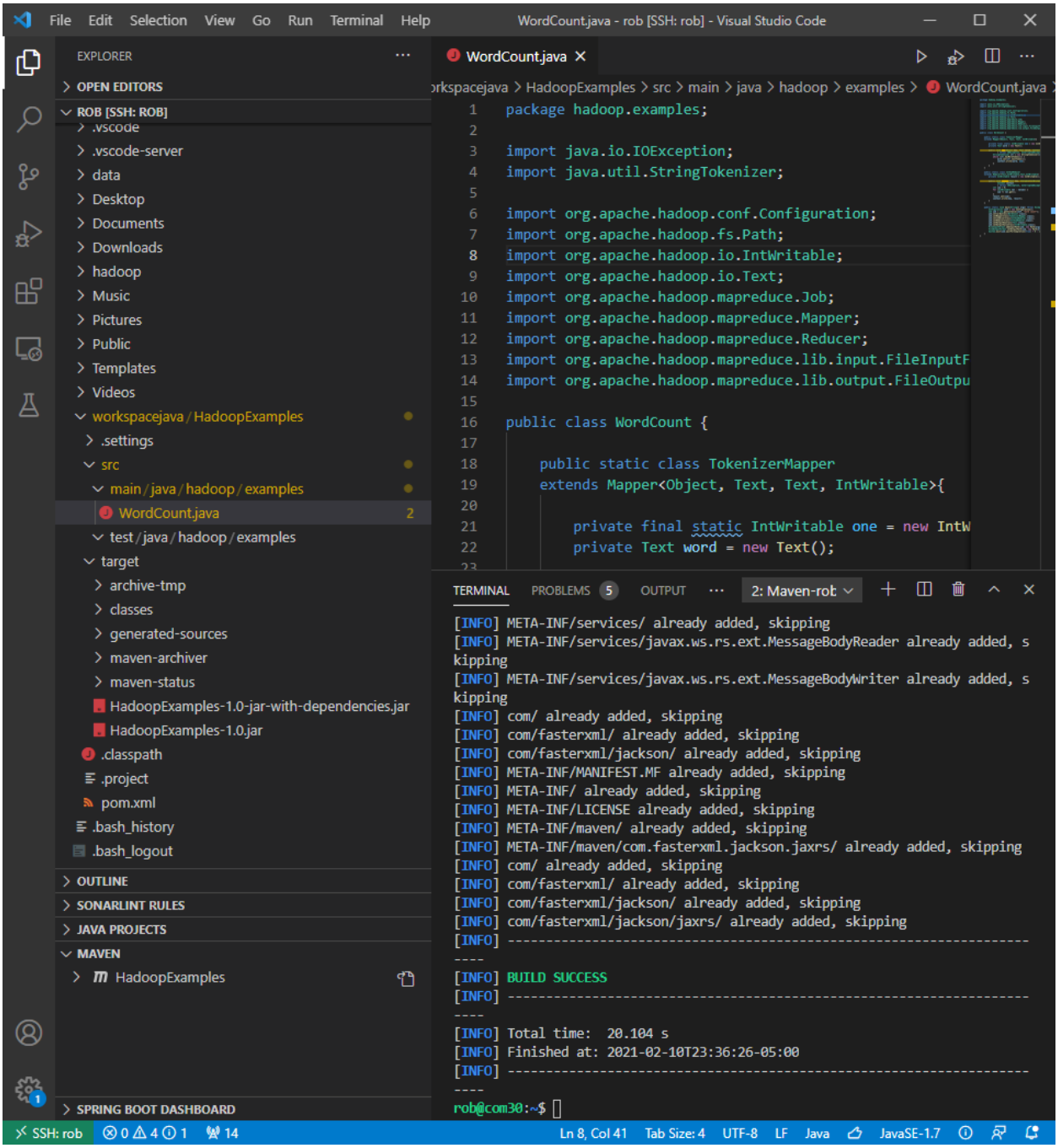
2) Setup HDFS and upload datasets ‘test.txt’ and ‘peterpan.txt’ into HDFS

- Install openssh-client and server
- Create an rsa encrypted public key and store in ‘authorized_keys’ file
- This allows ssh to local host without need for a password
- Use command ‘hsfs namenode -format’ to prepare hdfs
- Start hdfs with command ‘sbin/start-dfs.sh’ from inside the Hadoop folder
- Use command ‘hdfs dfs -mkdir /path’ to create necessary folders
- Use command ‘hdfs dfs -put [local source] [hdfs destination] to copy files from local directory to hdfs



3) Create a maven project in VS Code, import the WordCount.java file

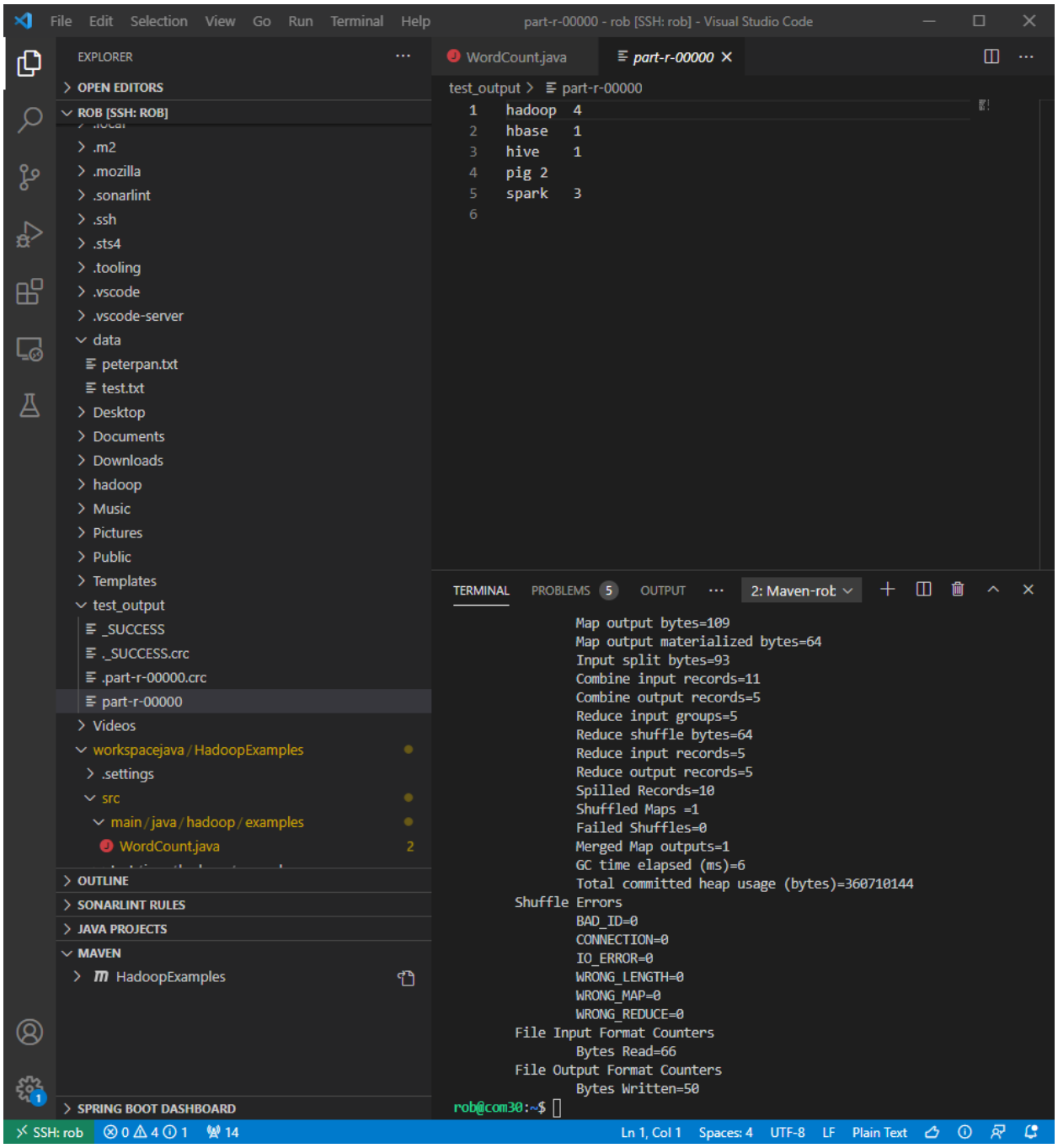
- Add dependencies and plug-ins to pom.xml file
- Create .jar file using maven



4) Open a terminal, and run the 'WordCount.jar" file on "test.txt" and "peterpan.txt"

On test.txt:

hadoop jar /home/rob/workspacejava/HadoopExamples/target/HadoopExamples-1.0-jar-with-dependencies.jar file:///home/rob/data/test.txt file:///home/rob/test_output



On peterpan.txt:

hadoop jar /home/rob/workspacejava/HadoopExamples/target/HadoopExamples-1.0-jar-with-dependencies.jar file:///home/rob/data/peterpan.txt file:///home/rob/peterpan_output