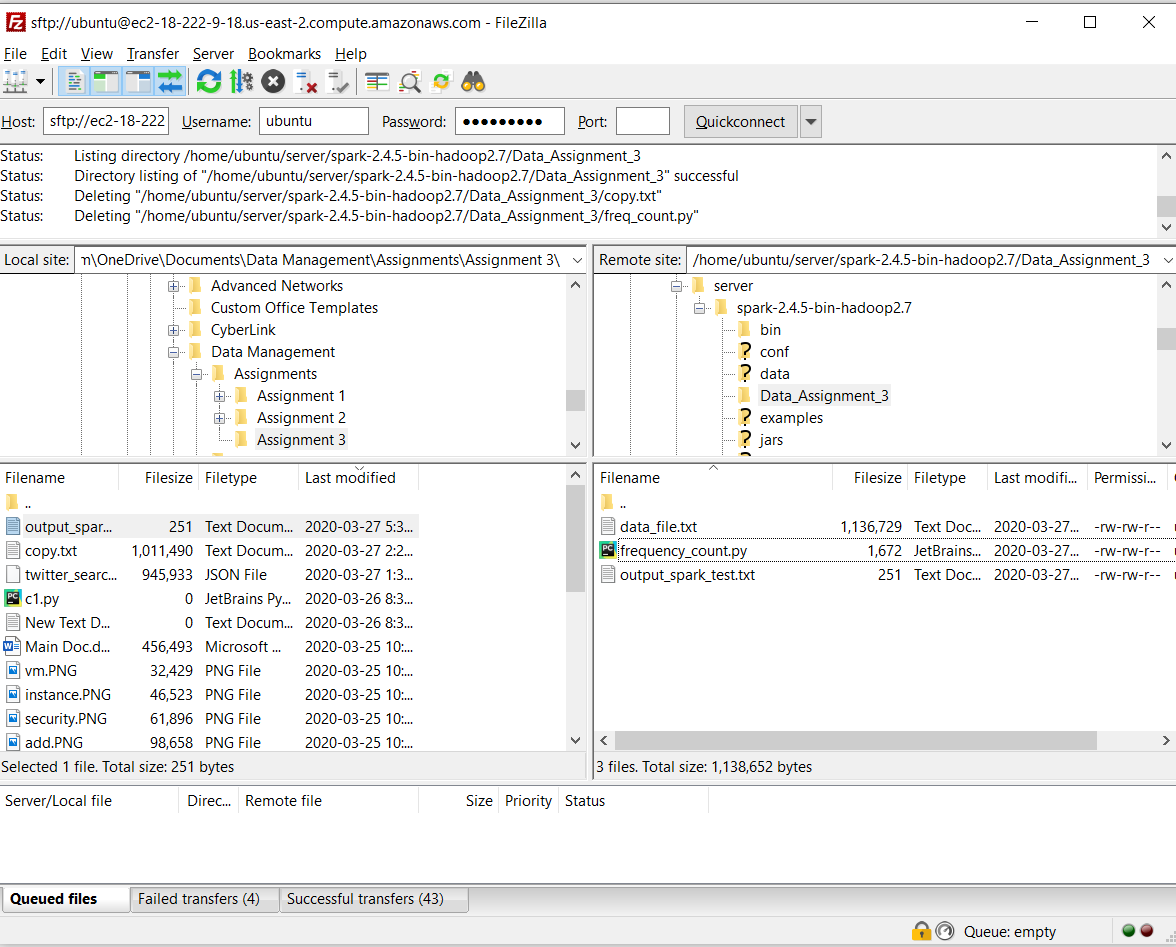
**Spark commands**

Step 1. Start the AWS instance and Ubuntu VM and start the master and slave node.

Step 2. Log in to FileZilla.

Step 3. Using fileZilla, transfer the data\_file.txt and frequency\_count.py at the location **/home/ubuntu/server/spark-2.4.5-bin-hadoop2.7/Data\_Assignment\_3/**



Step 4. cd server/spark-2.4.5-bin-hadoop2.7/

Step 5. Run the spark submit command bin/spark-submit Data\_Assignment\_3/frequency\_count.py

The output file will be generated there.