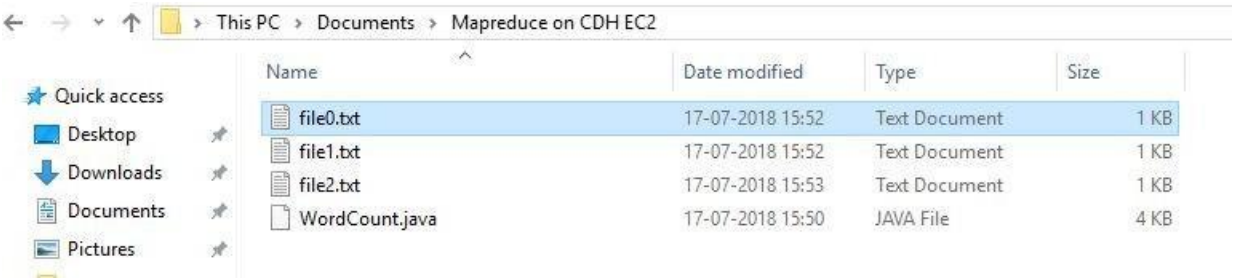
PREREQUISITES

- Please ensure that you have installed the following on your Windows machine:
    1. WinSCP tool
    2. Notepad++

Uploading the data from local file system to the CDH Instance on AWS

- We will now upload the java and input files on EC2.
- Download the WordCount.java given on the platform and create 3 text files naming them file0.txt, file1.txt and file2.txt.
- Store these files in a folder.
- We stored the files in a folder inside the Documents directory.
- You can store these files anywhere but then you will have to change the steps below accordingly.



- Write the files with the following Data.
    - **file0.txt** - “Hadoop is an elephant”
    - **file1.txt** - “Hadoop is as yellow as can be”
    - **file2.txt** - “Oh what a yellow fellow is Hadoop”

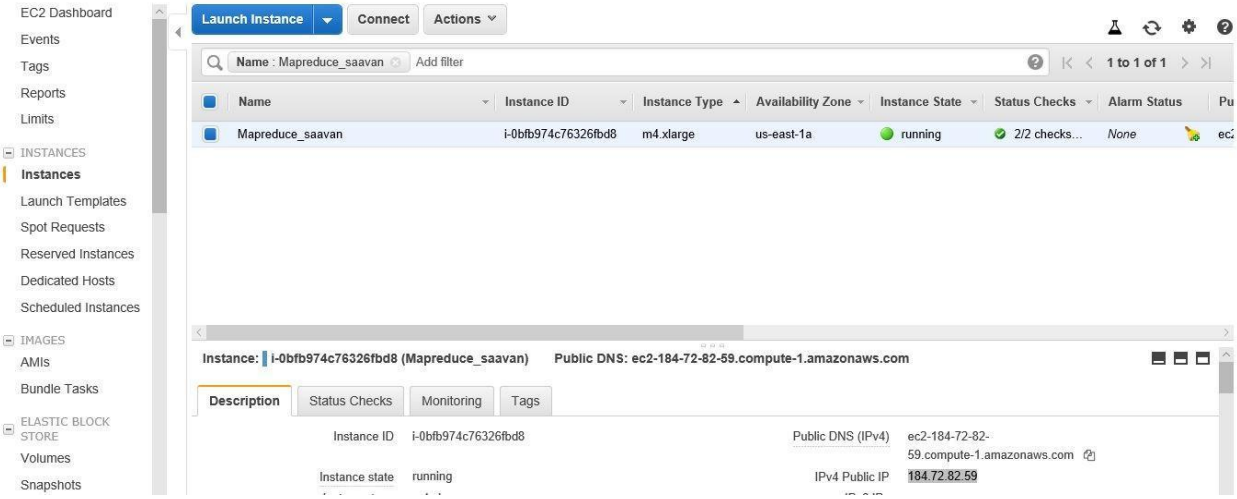
- WinSCP is a tool to transfer a file from a Windows machine to a Linux machine (EC2 instance) and vice versa.

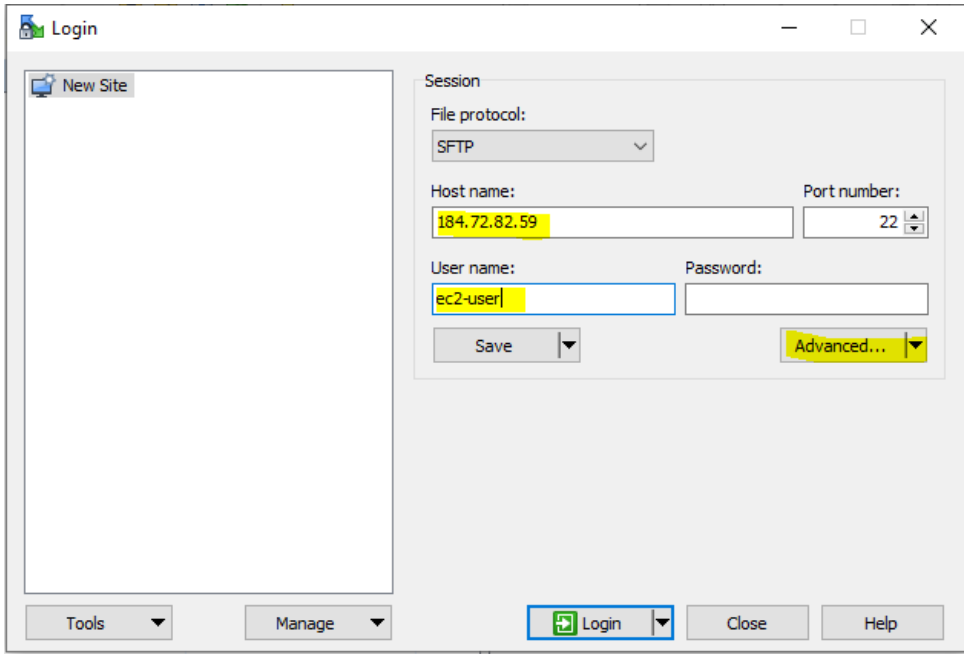
  Before moving forward you need to install WinSCP on your machine. You can download WinSCP from here.

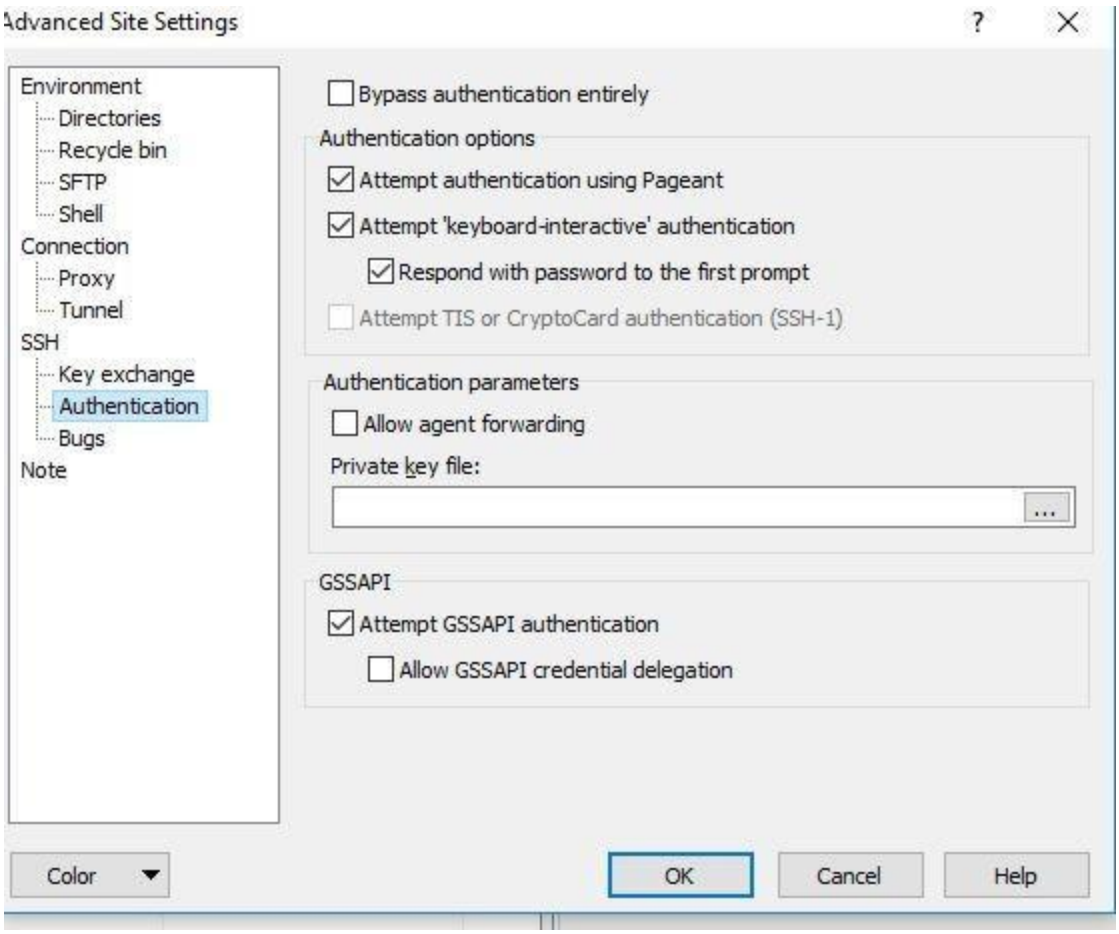
    - **Open WinSCP.**
    - **Enter the following credentials**

      Hostname: Provide the public IP from the EC2 dashboard.

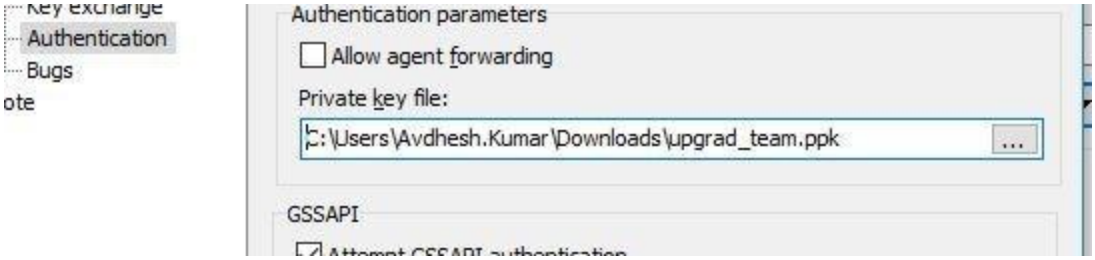
Enter Username: **ec2-user**
Then, click on '**Advanced**'.



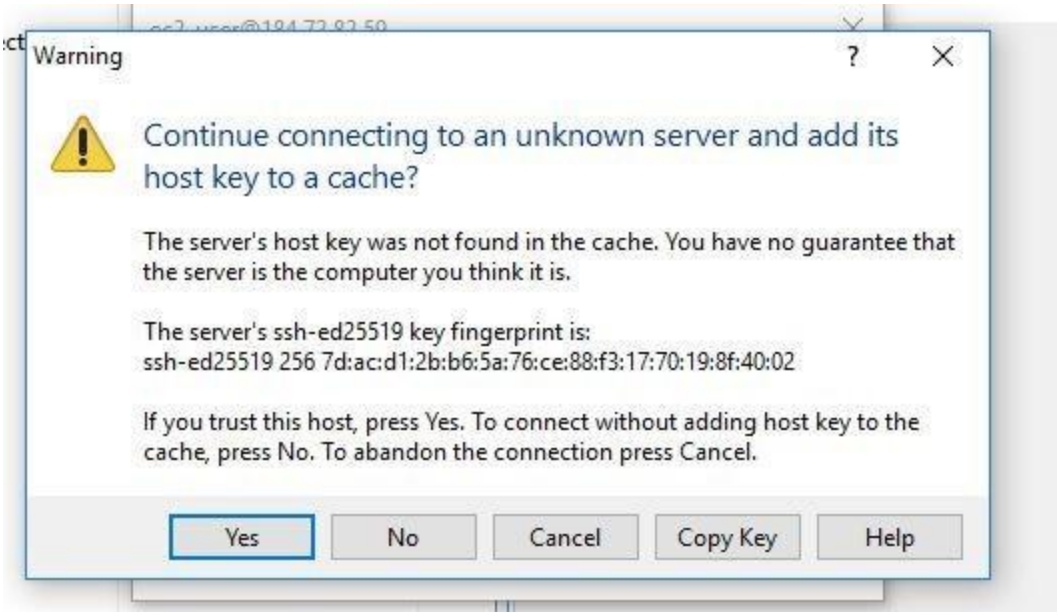




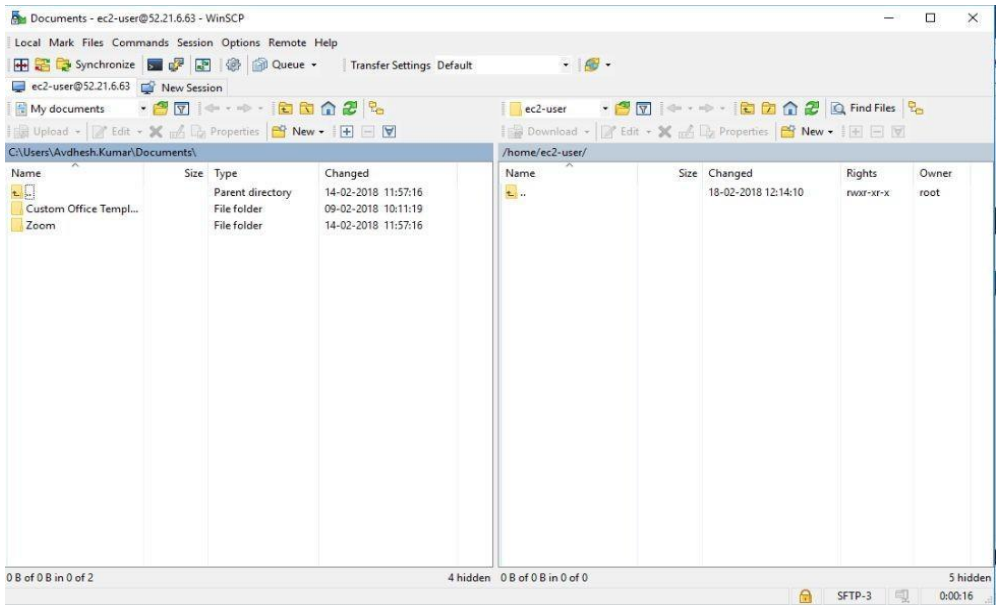
- After clicking on '**Authentication**', enter the path of your PPK file.



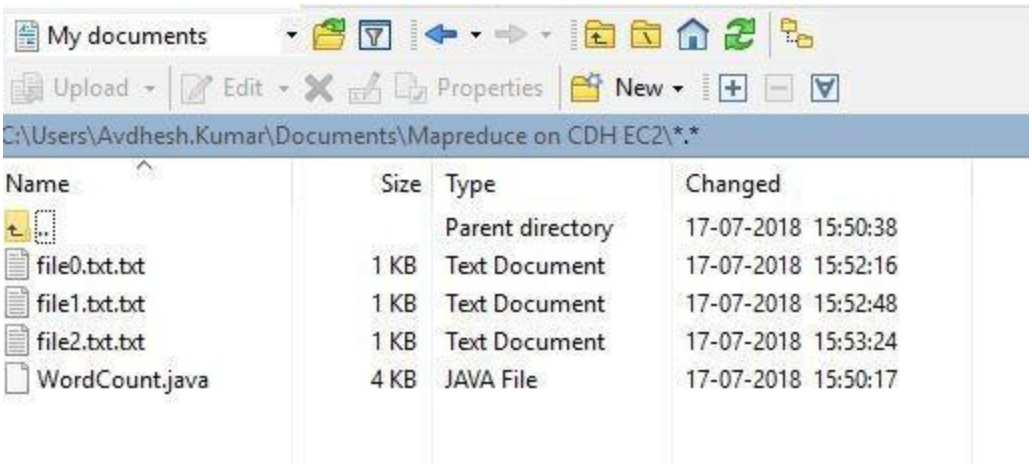
- Click on '**OK**' followed by '**Login**' after which a pop-up will appear. Click on '**Yes**'.



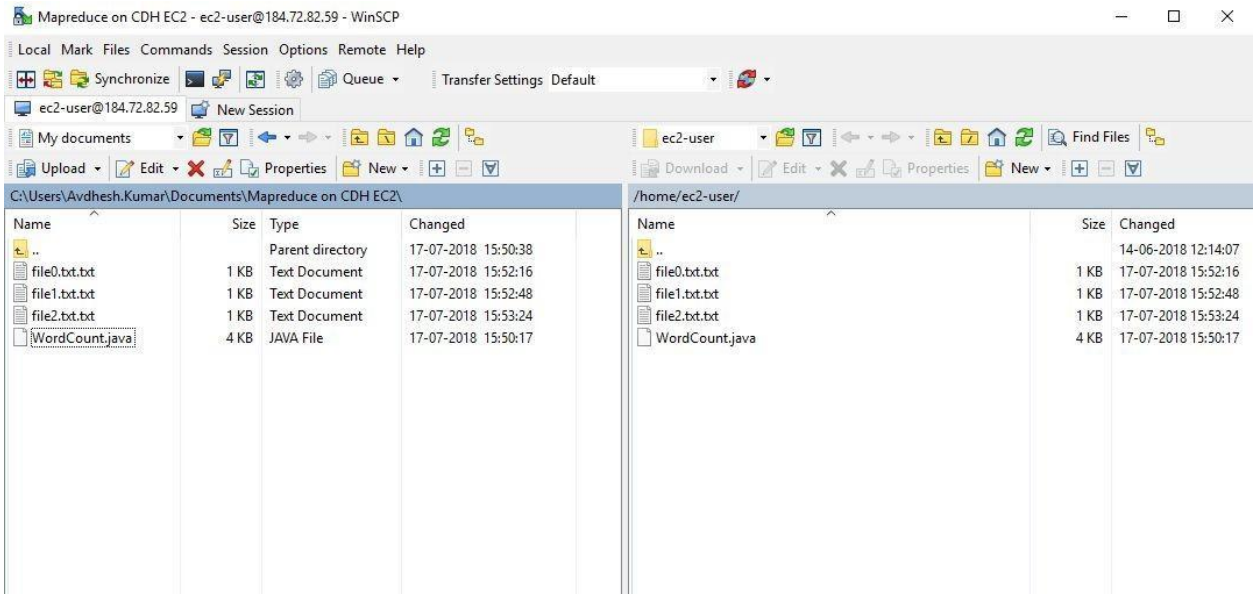
- The following screen appears.
  - **Left side screen: your local machine (Windows, in our case)**
  - **Right side screen: your Linux machine (AWS EC2 instance)**



- On the left side, browse to the folder containing the 'WordCount.java' and input text files.



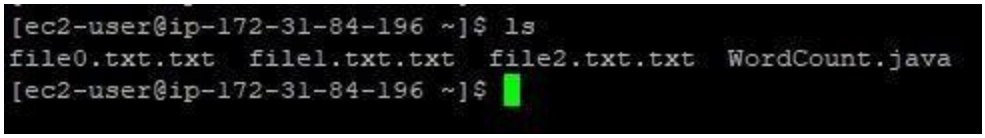
- Now drag the 'WordCount.java' and input text files on the left side and drop them to the right. Click on '**OK**' on the prompt which appears.



- We have now successfully copied the 'WordCount.java' and the input files - file0,file1,file2 from our local machine to our EC2 instance.

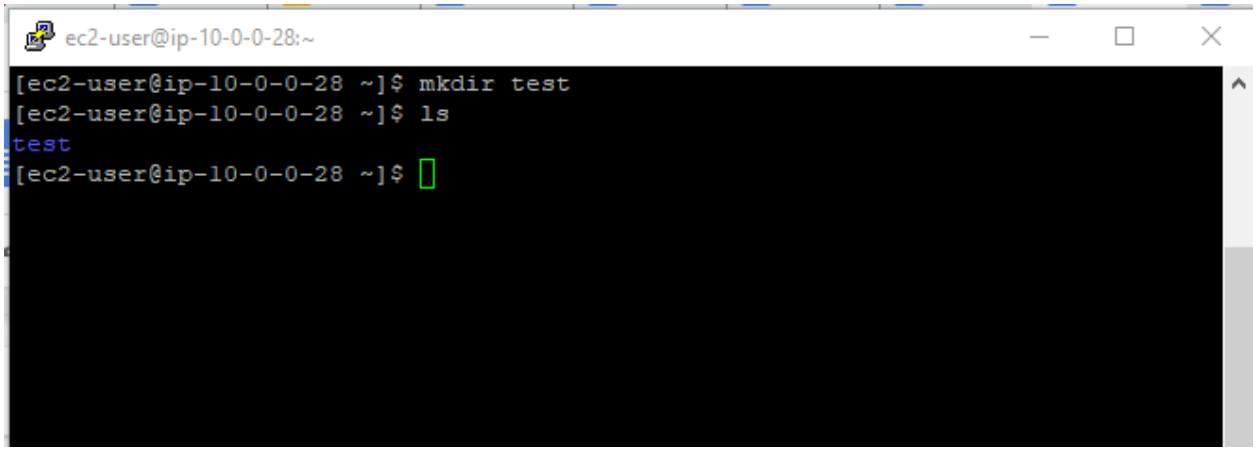
- Now, go back to AWS EC2 instance and verify if the files are uploaded or not using the 'ls' command.

```
ls
```



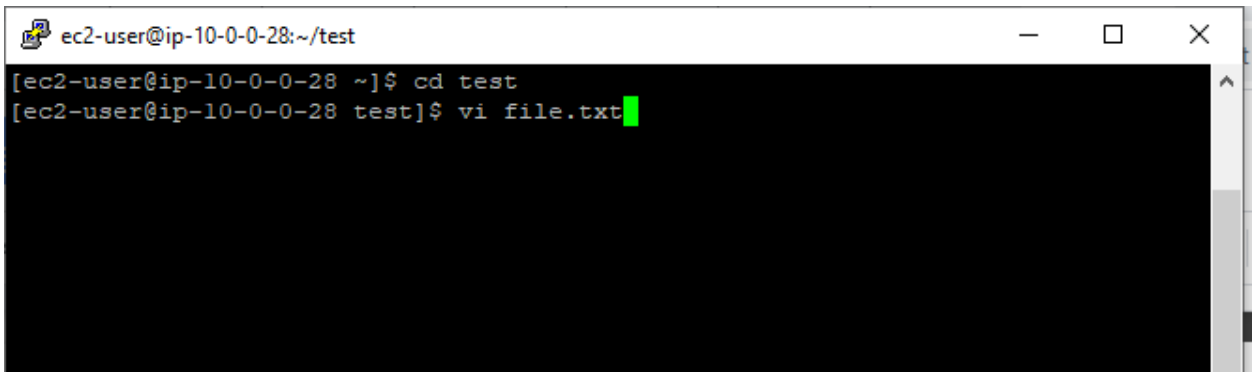
Downloading the data from CDH instance to the local file system.

- Now let us transfer file from the Linux (EC2 instance to the local system i.e. Windows).
- Firstly let's create a file in the EC2 instance.
- Log into your EC2 instance using PuTTy. Now go the desired folder in which you want to create the file.
- Let's say we create a directory named "test" using the command " `mkdir test` ". Use `ls` command to check it.

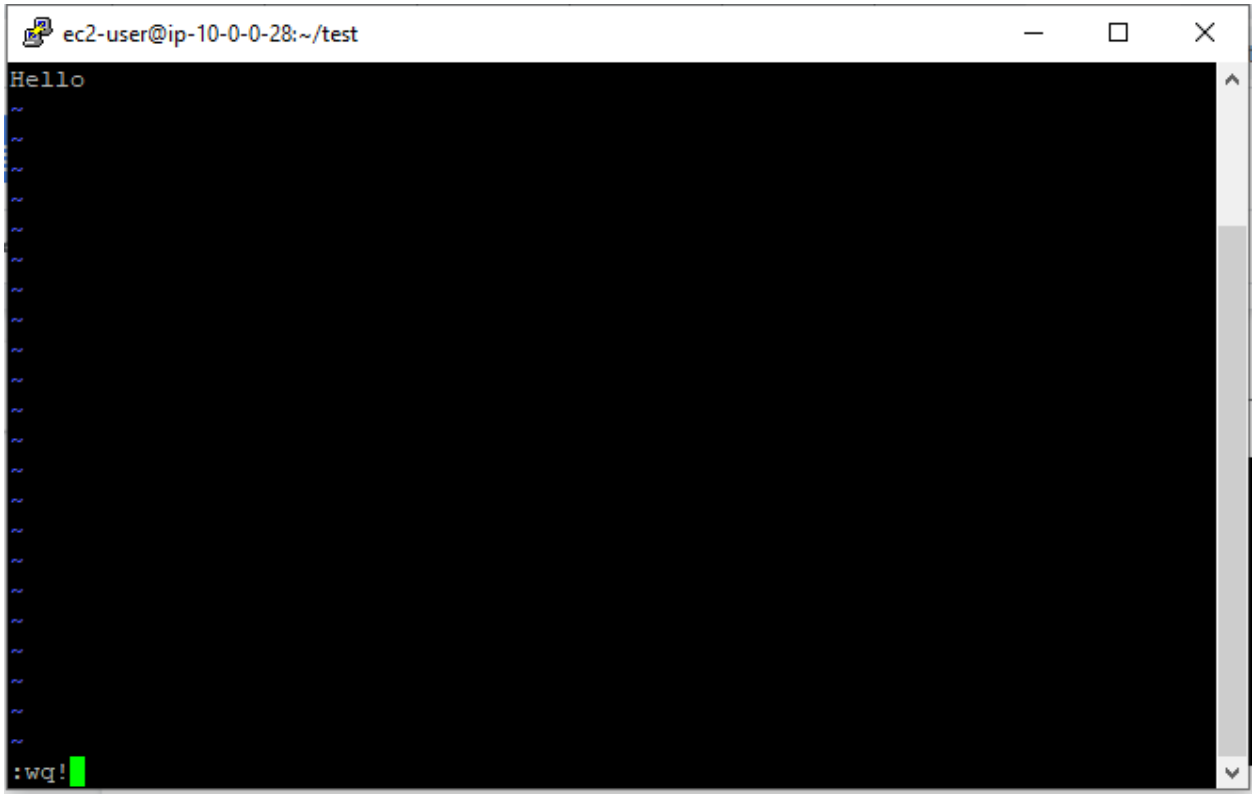


- Now let's create a file named "file.txt" in this directory. For this use the following commands one by one:

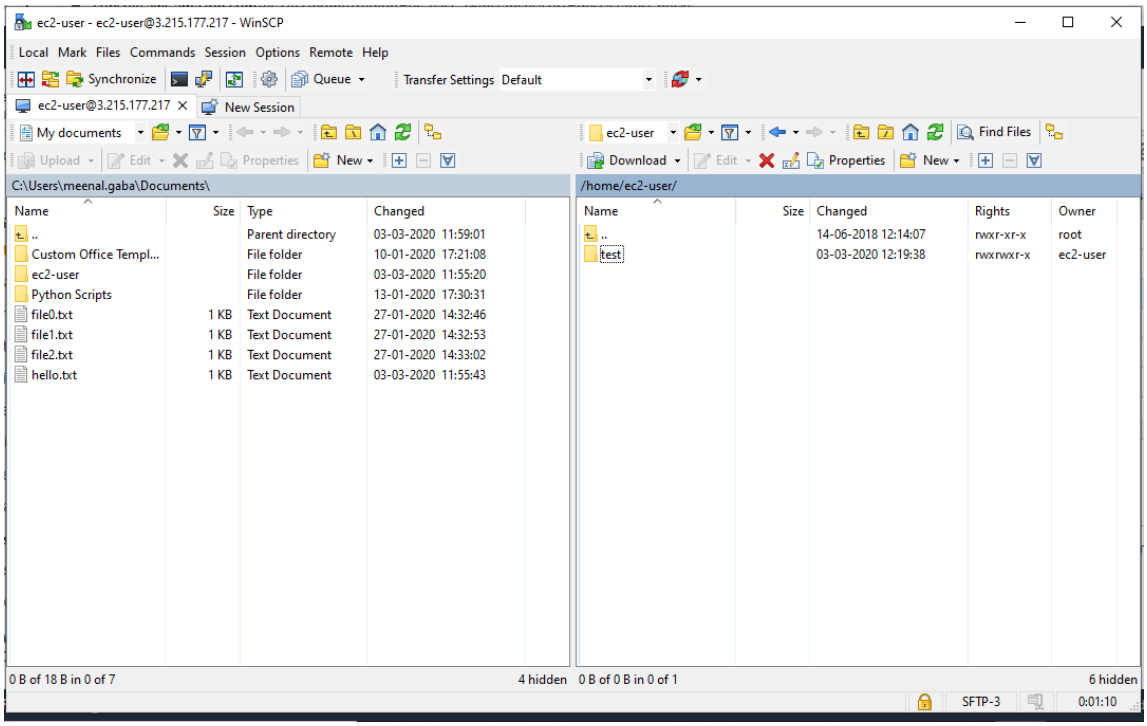
```
cd test
vi file.txt
```



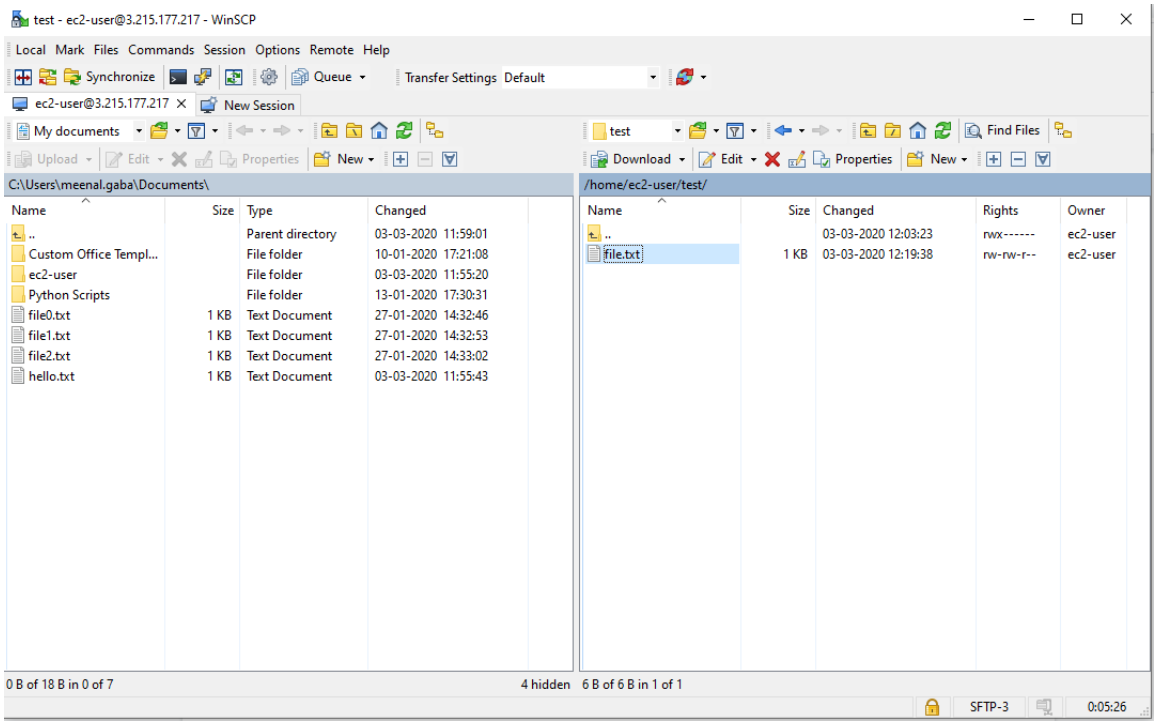
This will open up the file. Now press "i" to insert the text in the file created.
Add "Hello" to file.
Press esc and then type :wq! to save and exit.



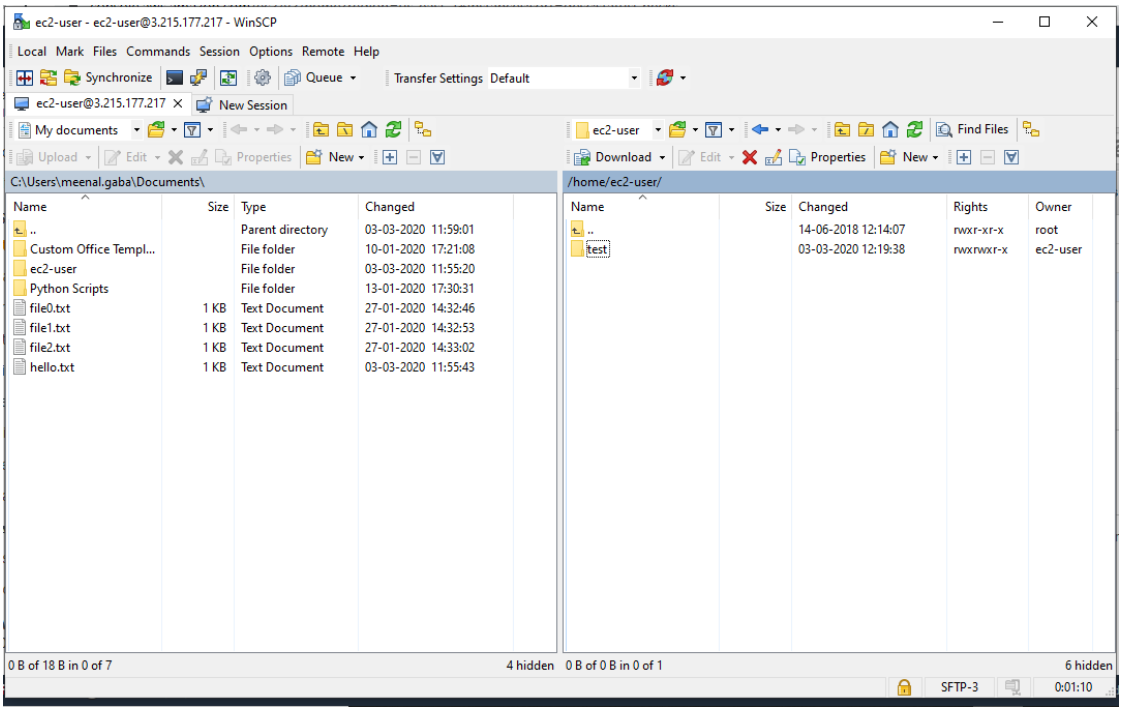
- Now start WinSCP as stated in the above steps.
- The following screen appears.
    - **Left side screen: your local machine (Windows, in our case)**
    - **Right side screen: your Linux machine (AWS EC2 instance)**
- You will see the following screen



You can see the test folder created by you on the right-hand side of your screen. Now open the folder and you will see your file.



Now drag and drop it on the left-hand side of your screen and you will see the file on the left-hand side as well.



- Now to check, go to the folder in which you have transferred the file in your windows system. You will see your file in that folder.