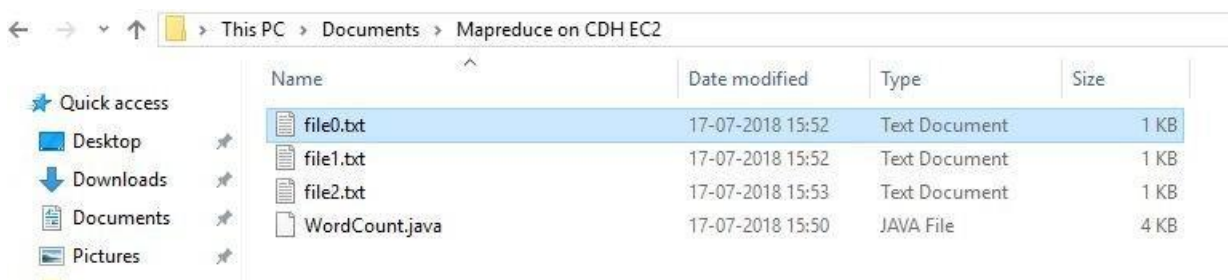


PREREQUISITES

- Please ensure that you have installed the following on your Windows machine:
 - [WinSCP](#) tool
 - [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.



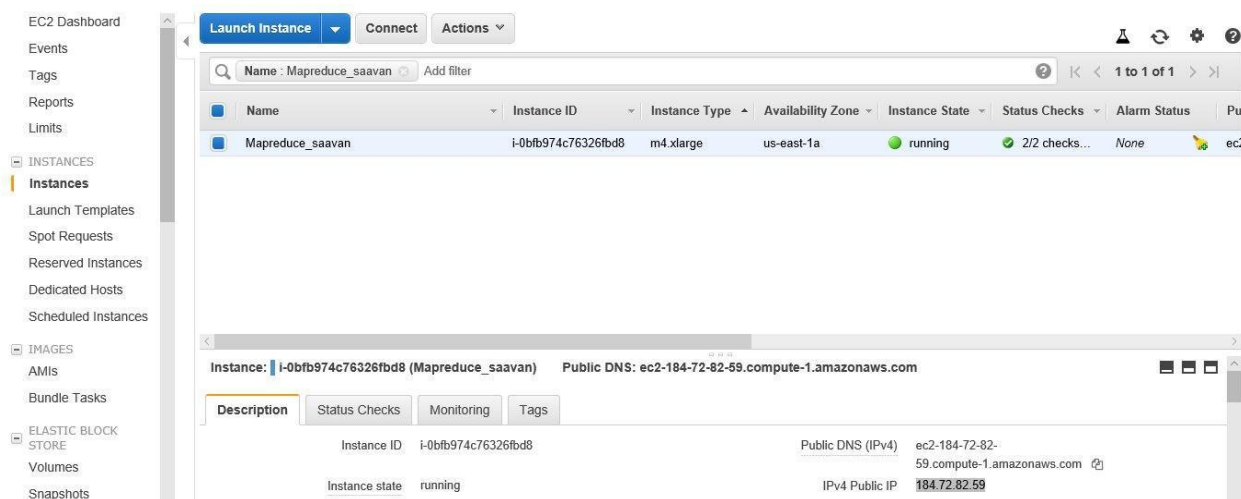
- Write the files with the following Data.
 - file0.txt** - "Hadoop is an elephant"
 - file1.txt** - "Hadoop is as yellow as can be"
 - file2.txt** - "Oh what a yellow fellow is Hadoop"
- WinSCP is a tool to transfer a file from a Windows machine to a Linux machine (EC2 instance) and vice versa.

Before moving forward you need to install WinSCP on your machine. You can download WinSCP from [here](#).

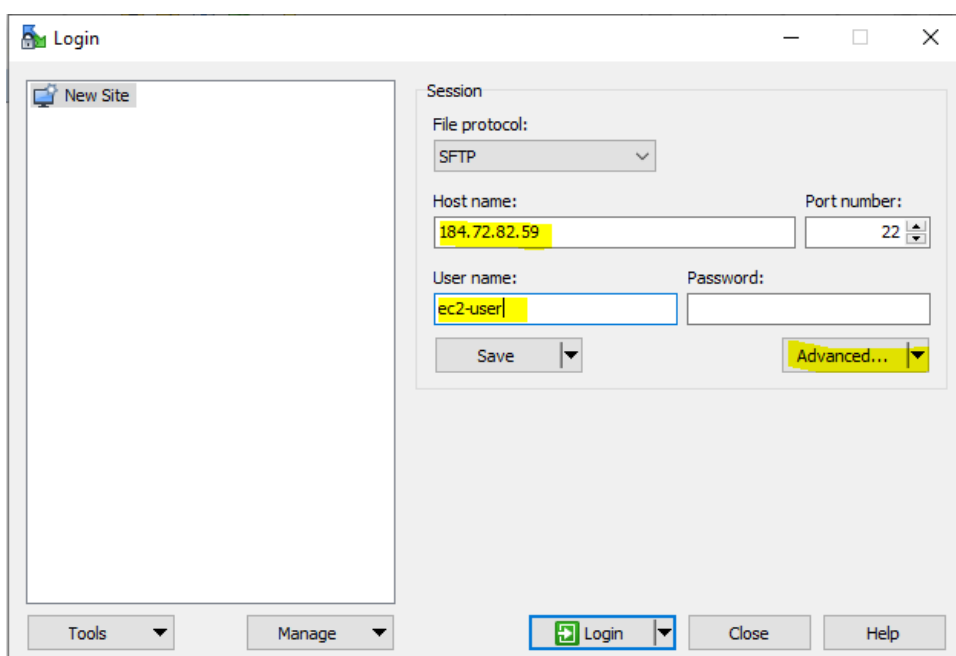
- Open WinSCP.**
- Enter the following credentials**

Hostname: Provide the public IP from the EC2 dashboard.

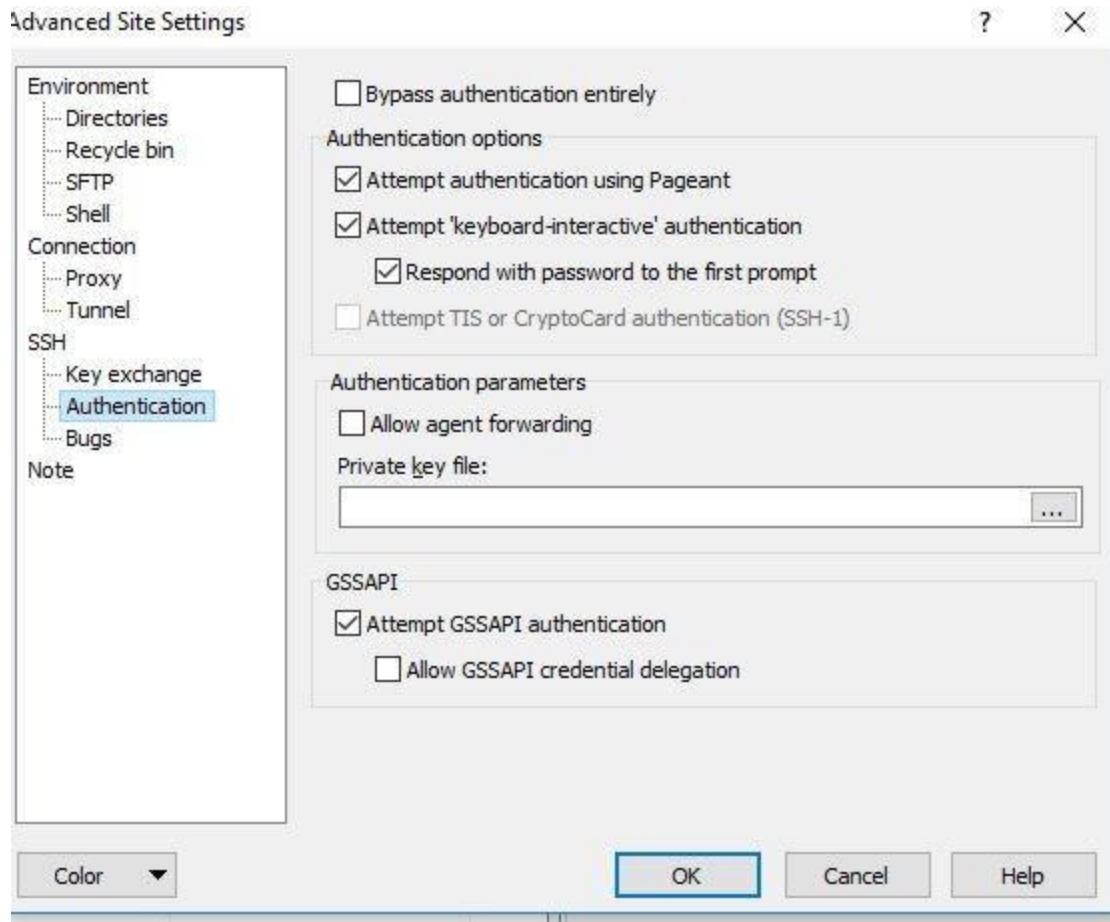
Enter Username: **ec2-user**
Then, click on '**Advanced**'.



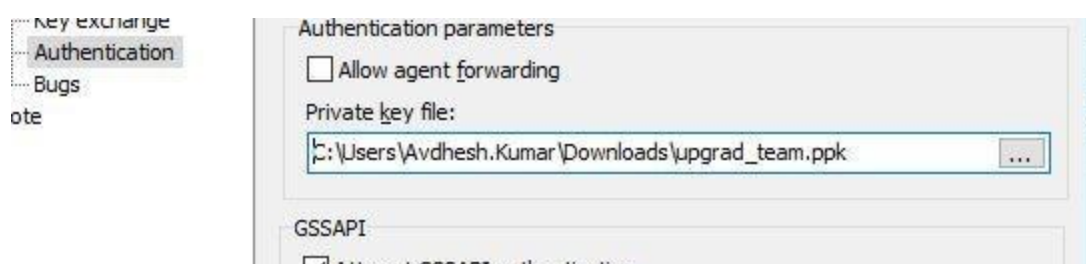
The screenshot shows the AWS Management Console interface. On the left, the navigation menu includes EC2 Dashboard, Events, Tags, Reports, Limits, INSTANCES, Launch Templates, Spot Requests, Reserved Instances, Dedicated Hosts, Scheduled Instances, IMAGES, AMIS, Bundle Tasks, ELASTIC BLOCK STORE, Volumes, and Snapshots. The main content area displays a table of EC2 instances. The table has columns for Name, Instance ID, Instance Type, Availability Zone, Instance State, Status Checks, Alarm Status, and Public IP. One instance, 'Mapreduce_saavan', is listed with Instance ID 'i-0bfb974c76326fbd8', Instance Type 'm4.xlarge', Availability Zone 'us-east-1a', Instance State 'running', Status Checks '2/2 checks...', Alarm Status 'None', and Public IP '184.72.82.59'. Below the table, the details for the selected instance are shown, including the Instance ID, Instance state (running), Public DNS (IPv4), and IPv4 Public IP.



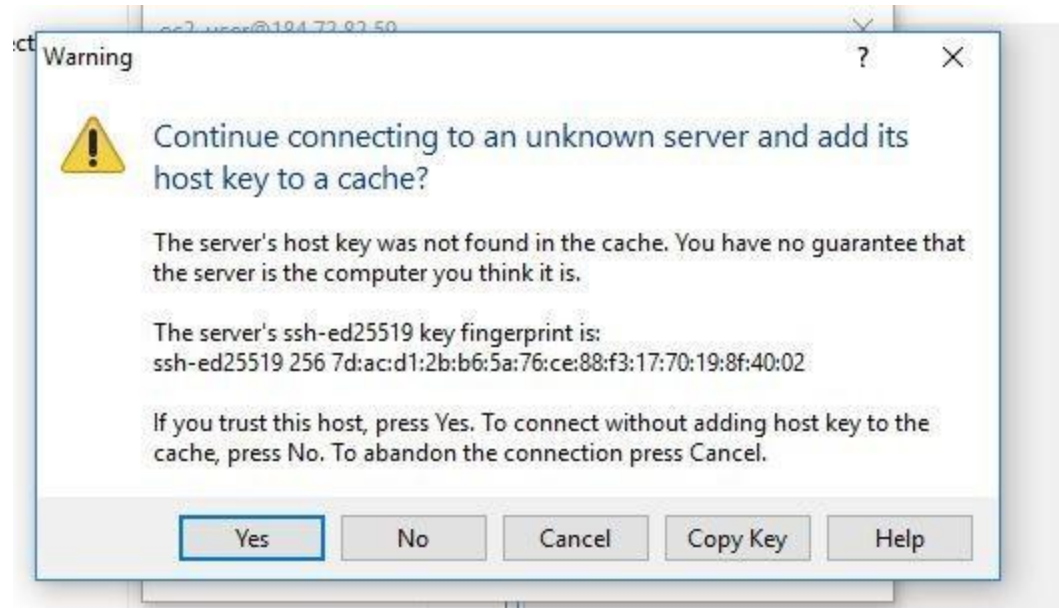
The screenshot shows a 'Login' dialog box with a 'New Site' tab. The dialog box contains a 'Session' section with the following fields: 'File protocol' (SFTP), 'Host name' (184.72.82.59), 'Port number' (22), 'User name' (ec2-user), and 'Password'. There are 'Save' and 'Advanced...' buttons. At the bottom, there are 'Tools', 'Manage', 'Login', 'Close', and 'Help' buttons.



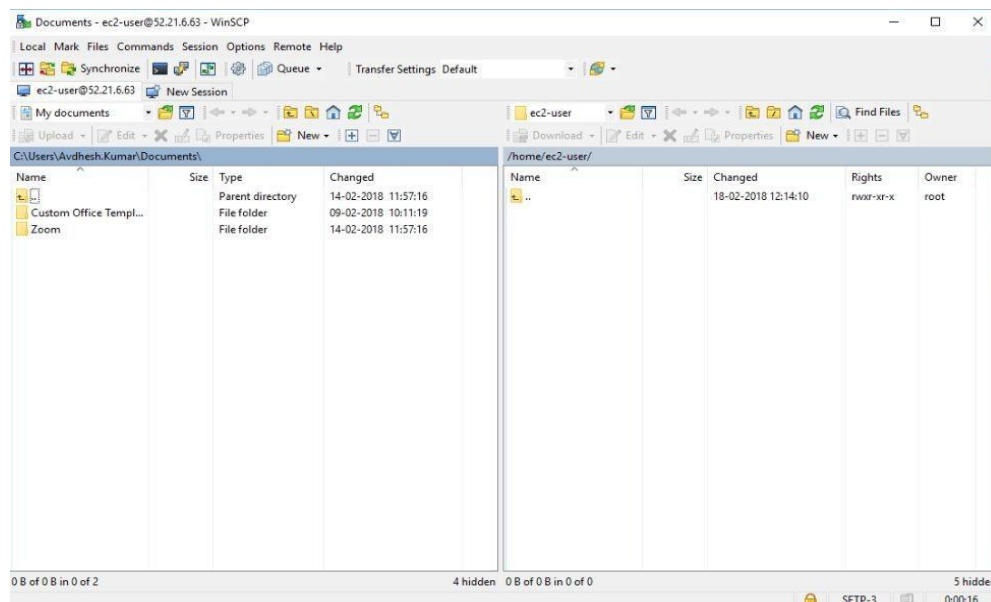
- After clicking on '**Authentication**', enter the path of your PPK file.



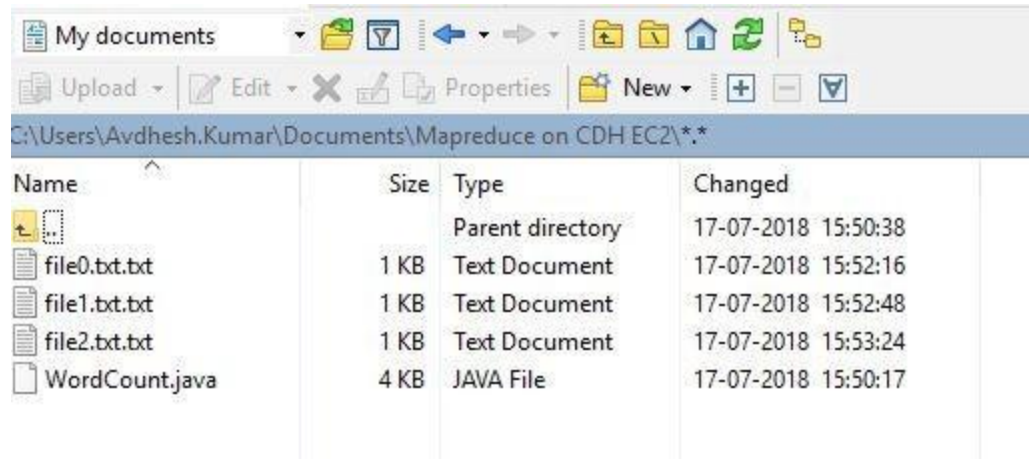
- Click on 'OK' followed by 'Login' after which a pop-up will appear. Click on 'Yes'.



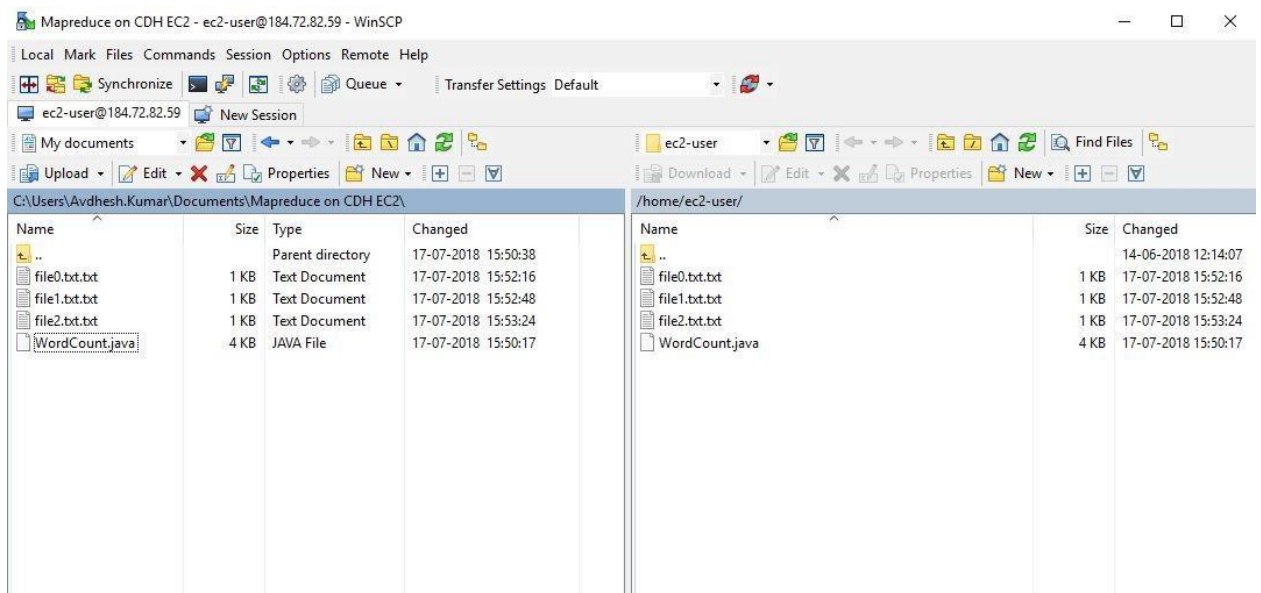
- The following screen appears.
 - Left side screen: your local machine (Windows, in our case)
 - Right side screen: your Linux machine (AWS EC2 instance)



- 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
```

```
[ec2-user@ip-172-31-84-196 ~]$ ls  
file0.txt.txt  file1.txt.txt  file2.txt.txt  WordCount.java  
[ec2-user@ip-172-31-84-196 ~]$
```

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

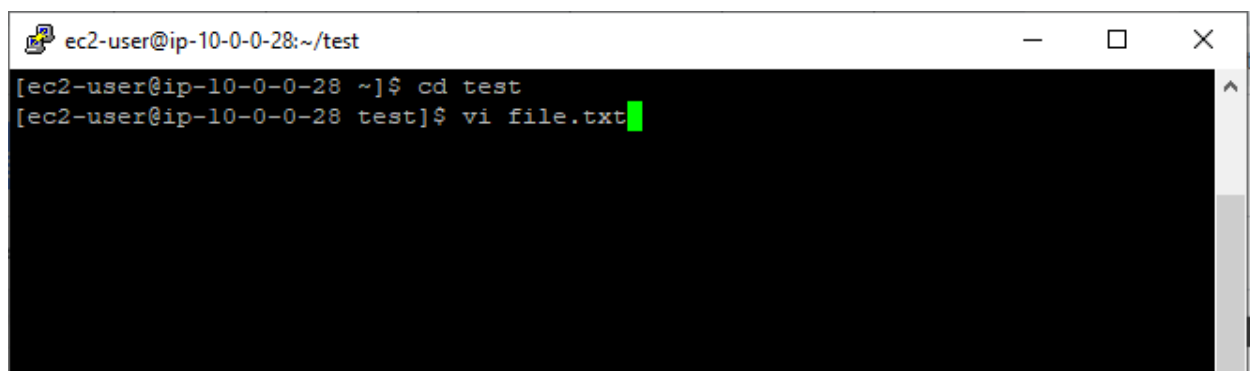
- Now let us transfer file from the Linux (EC2 instance to the local system i.e. Windows).
- Firstly let's create a file in the EC2 instance.
- Log into your EC2 instance using PuTTY. Now go the desired folder in which you want to create the file.
- Let's say we create a directory named "test" using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



```
ec2-user@ip-10-0-0-28:~  
[ec2-user@ip-10-0-0-28 ~]$ mkdir test  
[ec2-user@ip-10-0-0-28 ~]$ ls  
test  
[ec2-user@ip-10-0-0-28 ~]$
```

- 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
```



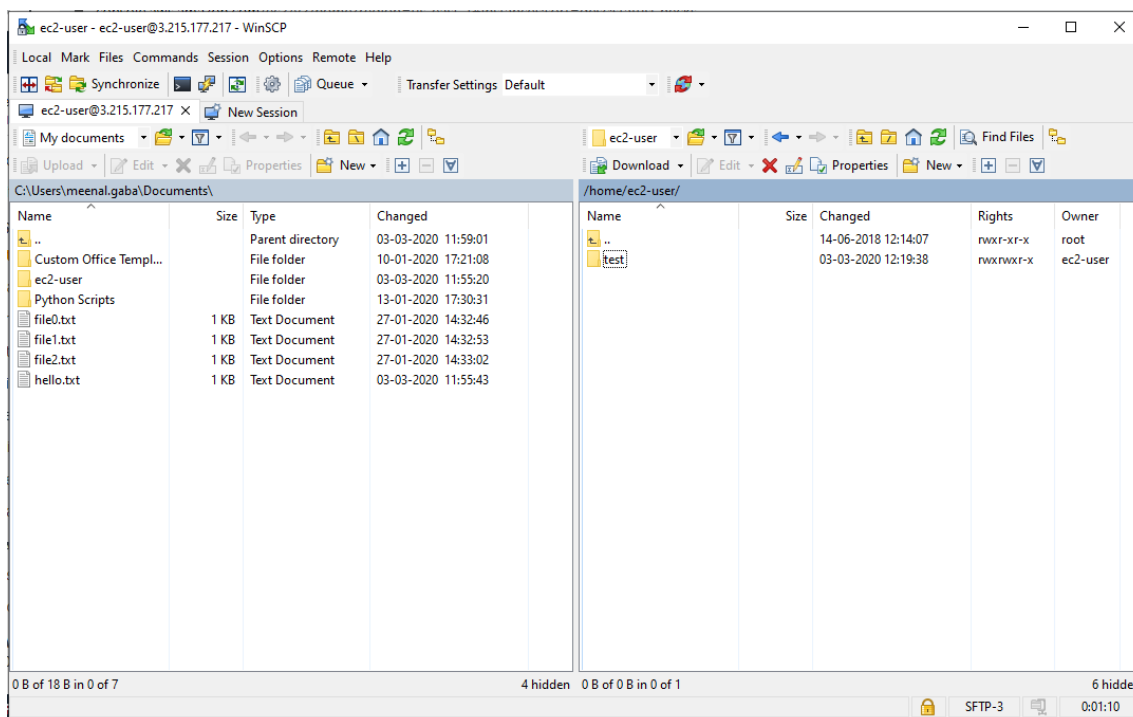
```
ec2-user@ip-10-0-0-28:~/test  
[ec2-user@ip-10-0-0-28 ~]$ cd test  
[ec2-user@ip-10-0-0-28 test]$ vi file.txt
```

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.

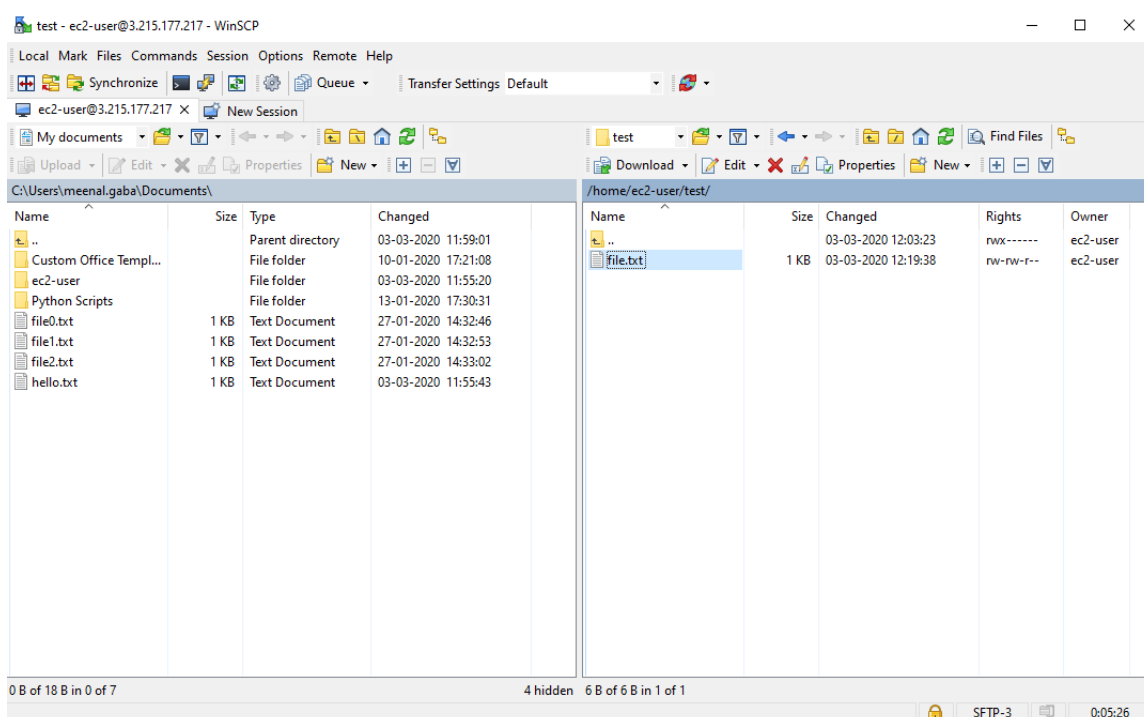


A terminal window titled "ec2-user@ip-10-0-0-28: ~/test". The window has a black background. The word "Hello" is written in white text at the top left. Below it, there are several lines of blue tilde characters (~). At the bottom left, there is a green prompt character followed by "wq!".

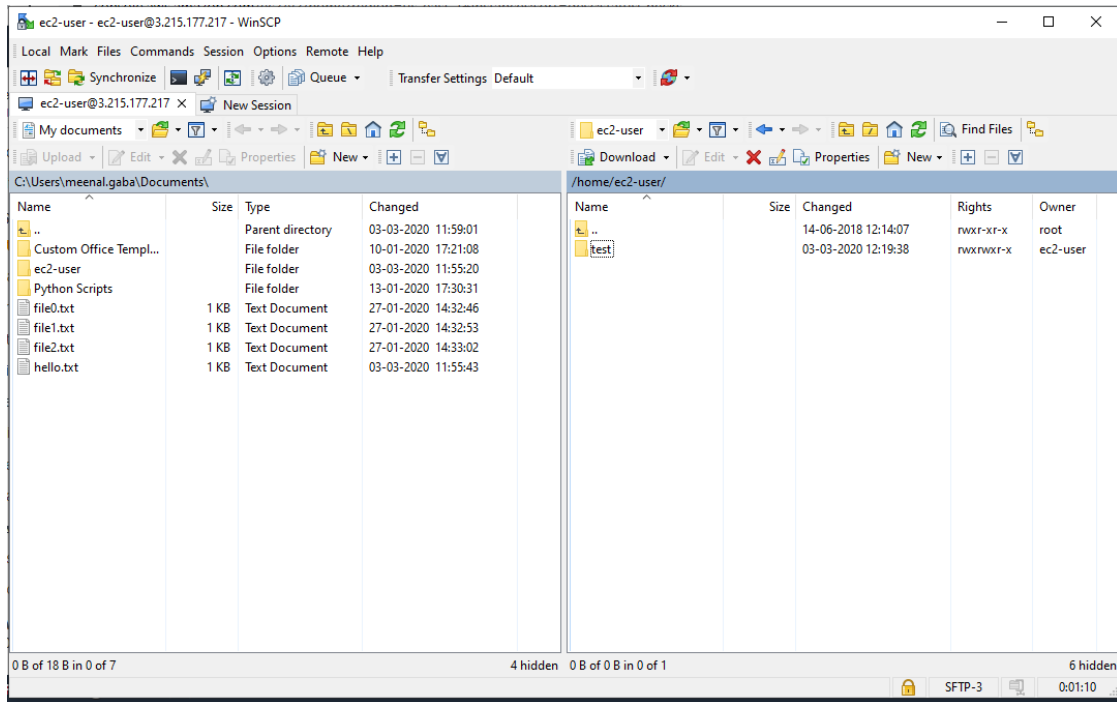
- Now start WinSCP as stated in the above steps.
- The following screen appears.
 - **Left side screen: your local machine (Windows, in our case)**
 - **Right side screen: your Linux machine (AWS EC2 instance)**
- You will see the following screen



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.

