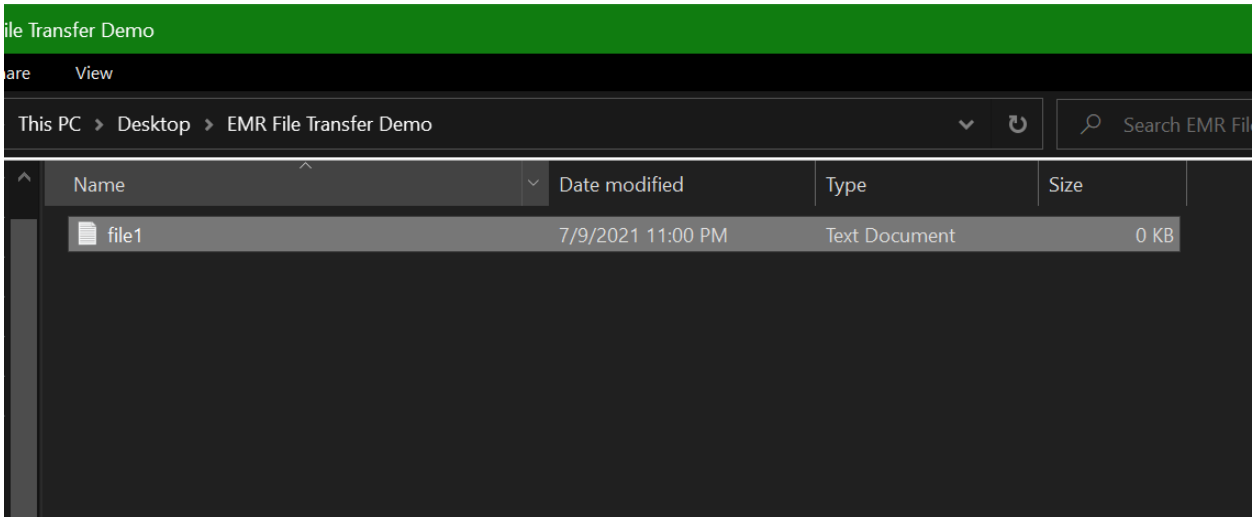
File Transfer – Windows

Prerequisites:

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

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

- We will now upload the Java and input files on EMR.
- Create a text file and name it **file1.txt**.
- Store this file in a folder.
- You can store the file anywhere, but you will have to change the below steps accordingly.

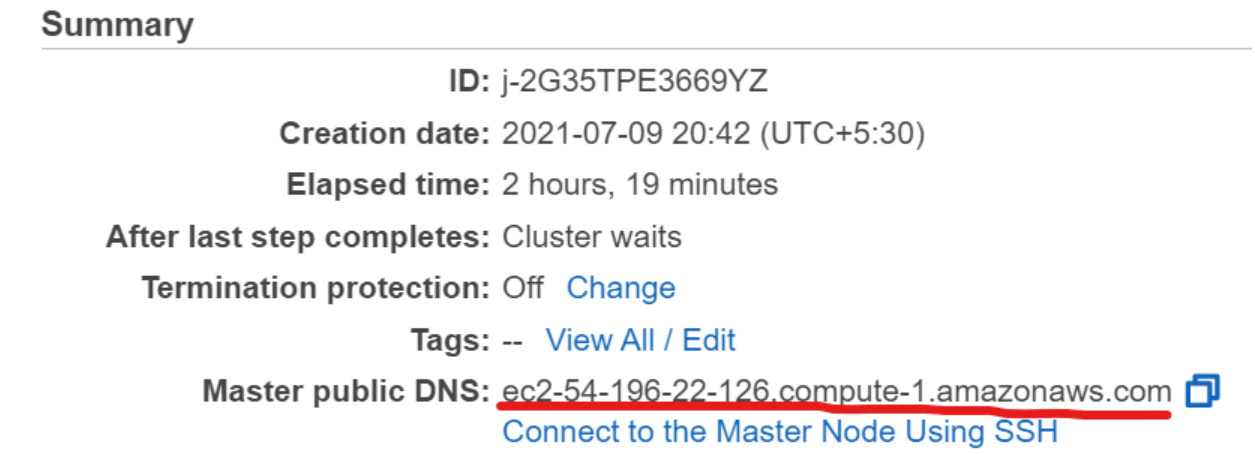


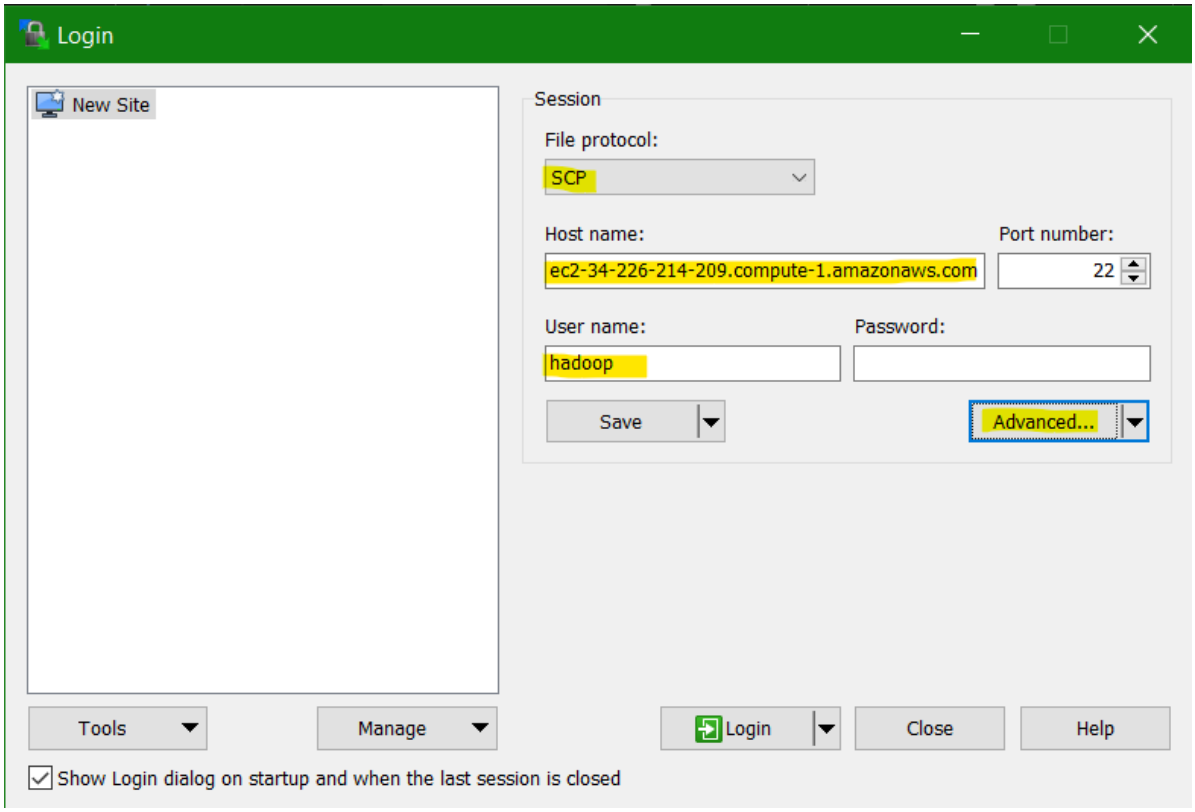
- Write the files with the following data:
  - **file1.txt** - “This is a test”
- WinSCP is a tool to transfer a file from a Windows machine to a Linux machine (EMR instance) and vice versa.

  Before moving forward, you will need to install WinSCP on your machine. You can download WinSCP from here.
  - **Open WinSCP.**
  - **Change the file protocol to SCP.**

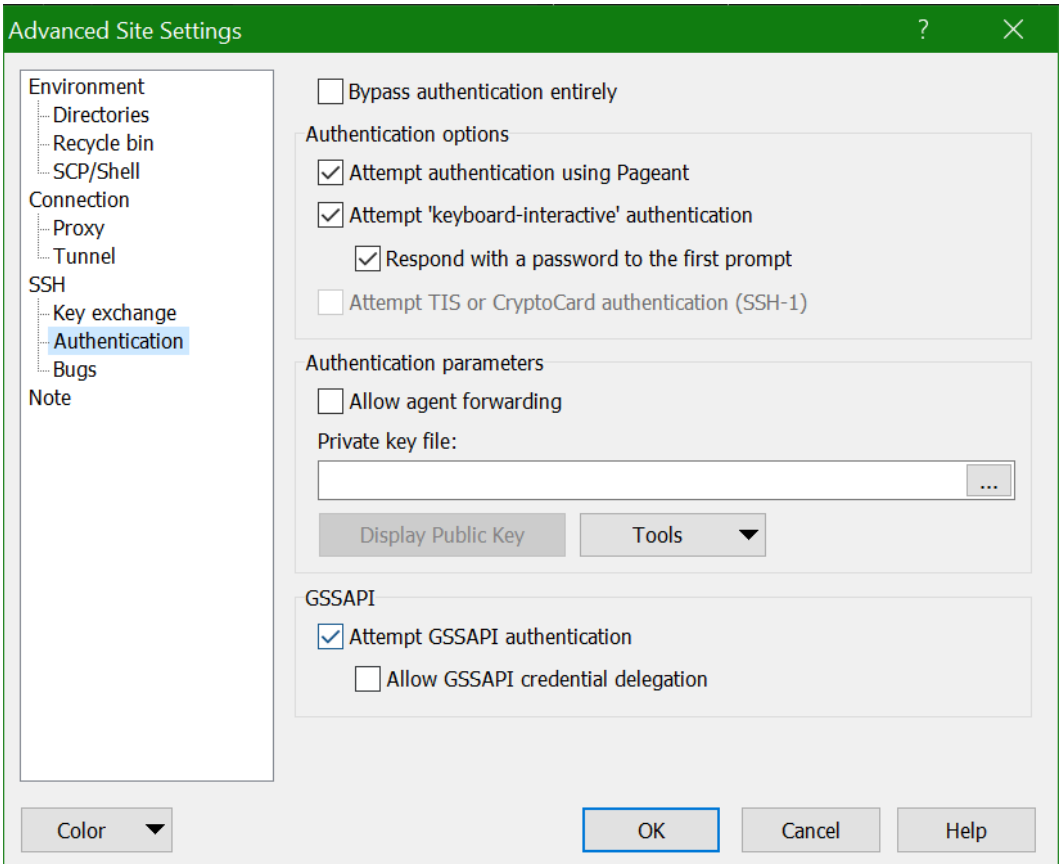
- **Enter the following credentials**:

  Hostname: Provide the public DNS from the EMR dashboard.
  Enter Username: **hadoop**
  Then, click on '**Advanced**'.

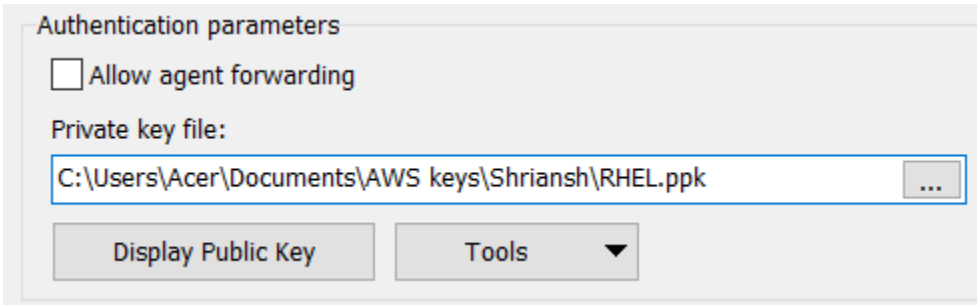




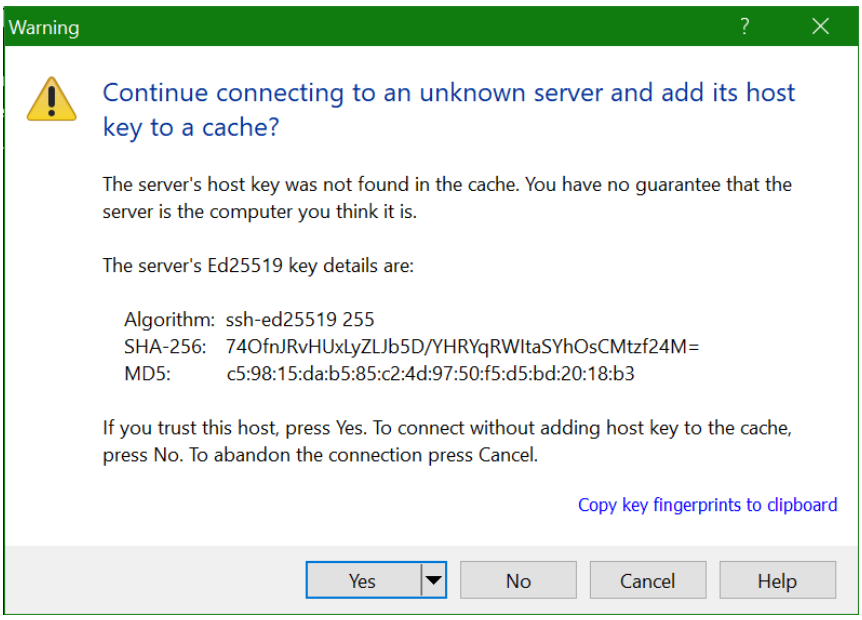
- Click on '**Authentication'** under **SSH**, in the drop-down menu to the left.



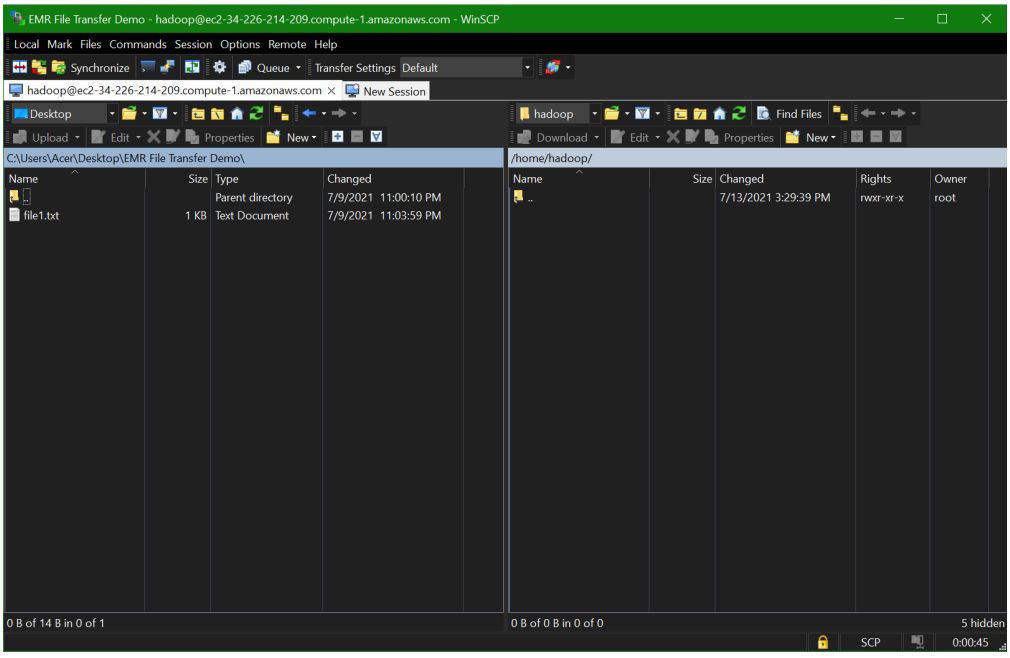
- After clicking on '**Authentication**', enter the path of your PPK file or press on the three-dotted button on the right to navigate to your PPK file, and then select the PPK file.



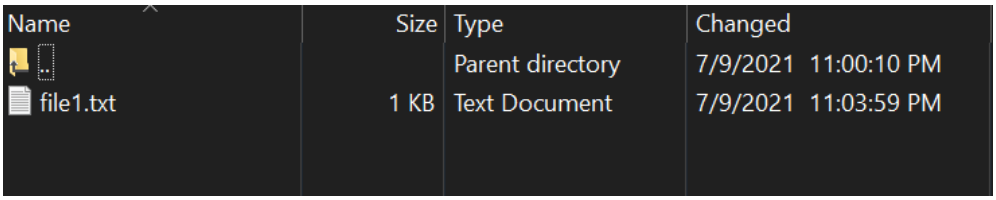
- Click on '**OK**' followed by '**Login**' after which a pop-up will appear. Click on '**Yes**'.



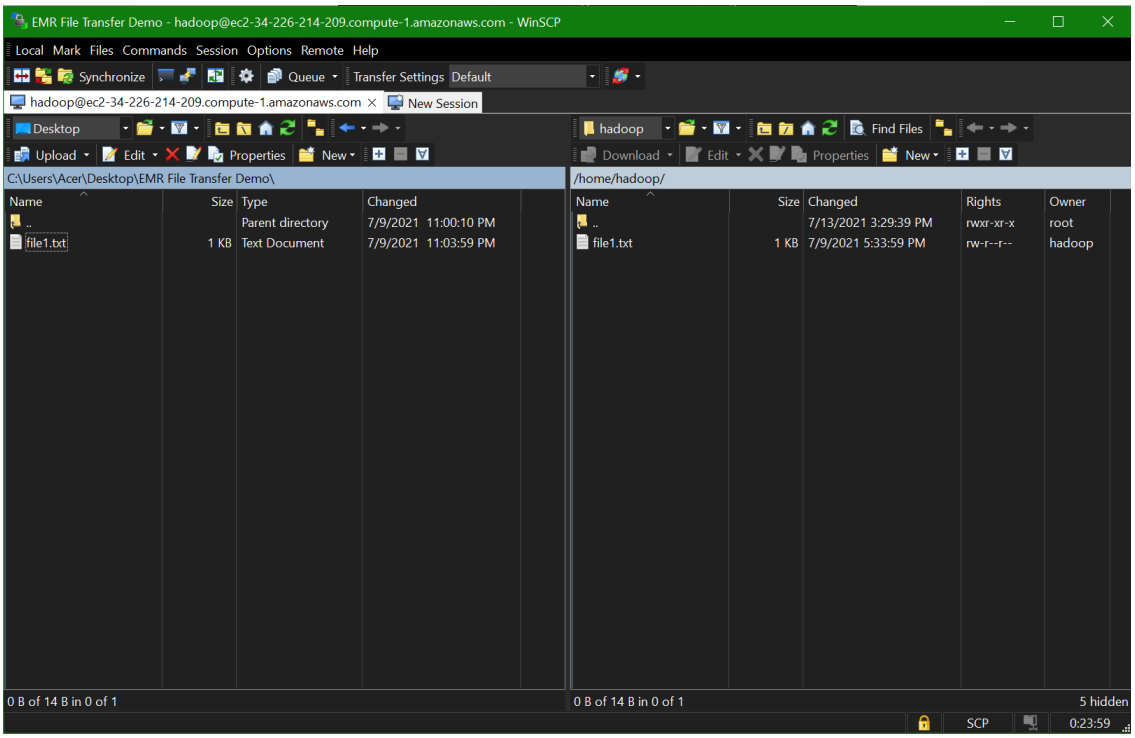
- The following screen will appear.
  - **Left side screen: Your local machine (Windows, in our case)**
  - **Right side screen: Your linux machine (AWS EMR instance)**



- On the left side, browse to the folder containing the **file1.txt**.



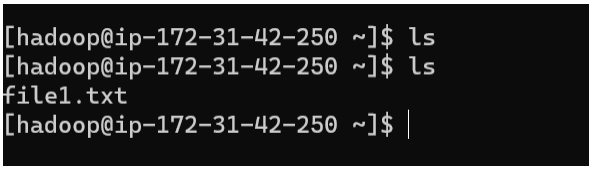
- Now drag the **file1.txt** on the left side and drop them to the right. Click on '**OK**' when the prompt appears.



- We have now successfully copied the **file1.txt** file from our local machine to our EMR instance.

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

```
ls
```



Downloading the data from the EMR instance to the local file system

- Let's transfer the files from the Linux (EMR instance to the local system, i.e., Windows).
- Firstly, let's create a file in the EMR instance.
- Log into your EMR instance using PuTTy. Now go to 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 the `ls` command to check the same.

```
[hadoop@ip-172-31-42-250 ~]$ mkdir test
[hadoop@ip-172-31-42-250 ~]$ ls
file1.txt  test
[hadoop@ip-172-31-42-250 ~]$
```

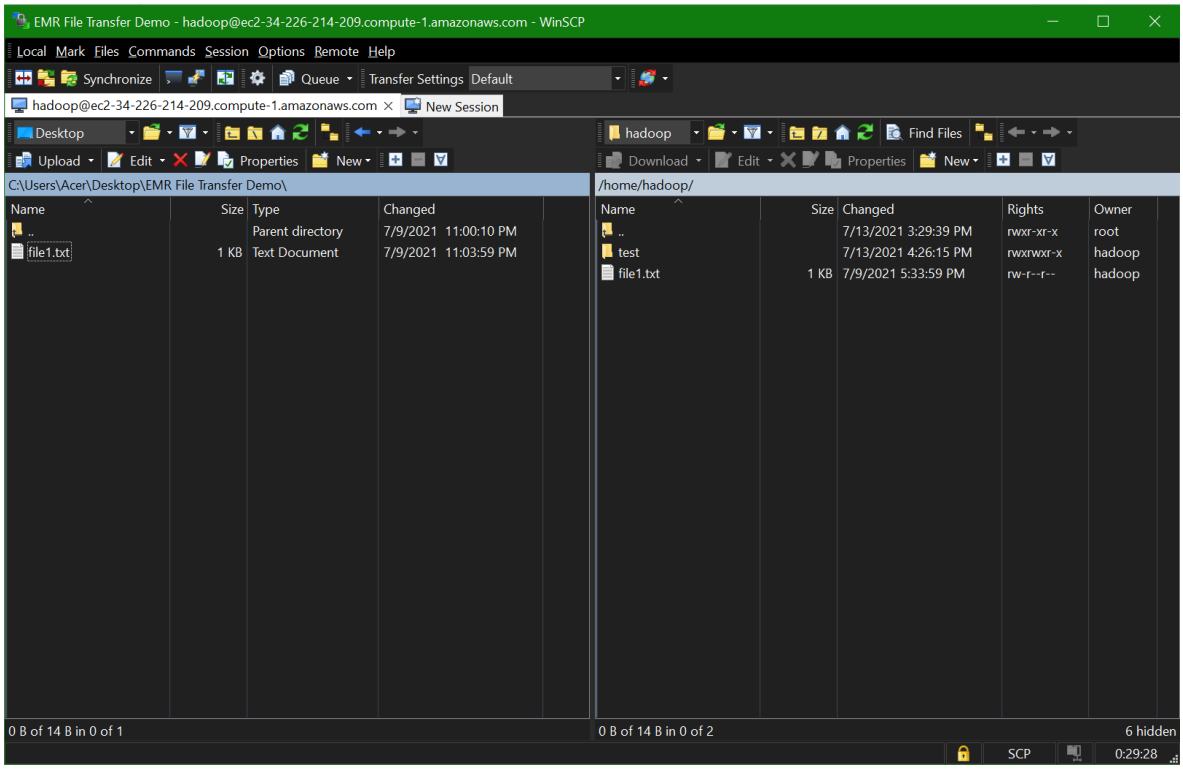
- Now, let's create a file named "hello.txt" in this directory. For this, use the following commands one after the other:

```
cd test
vi hello.txt
```

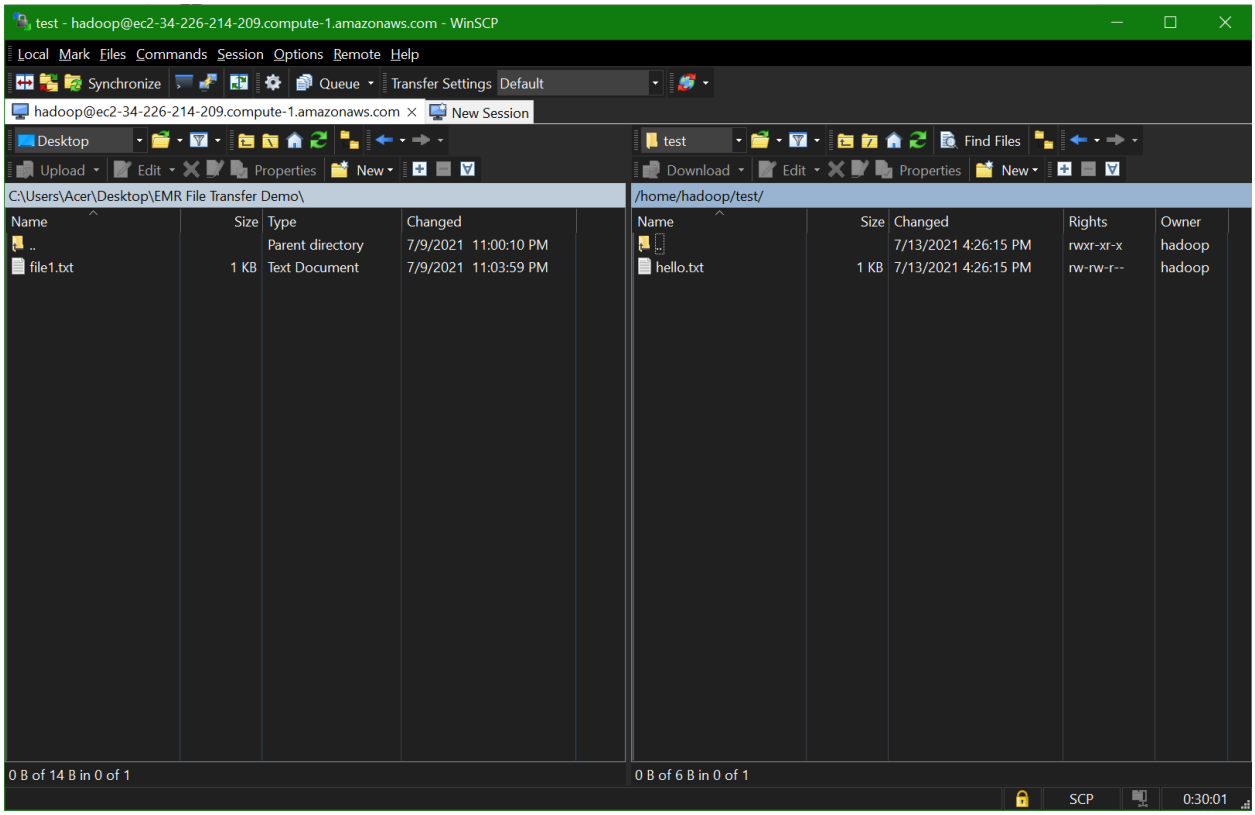
```
[hadoop@ip-172-31-42-250 ~]$ cd test/
[hadoop@ip-172-31-42-250 test]$ vi hello.txt
```

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

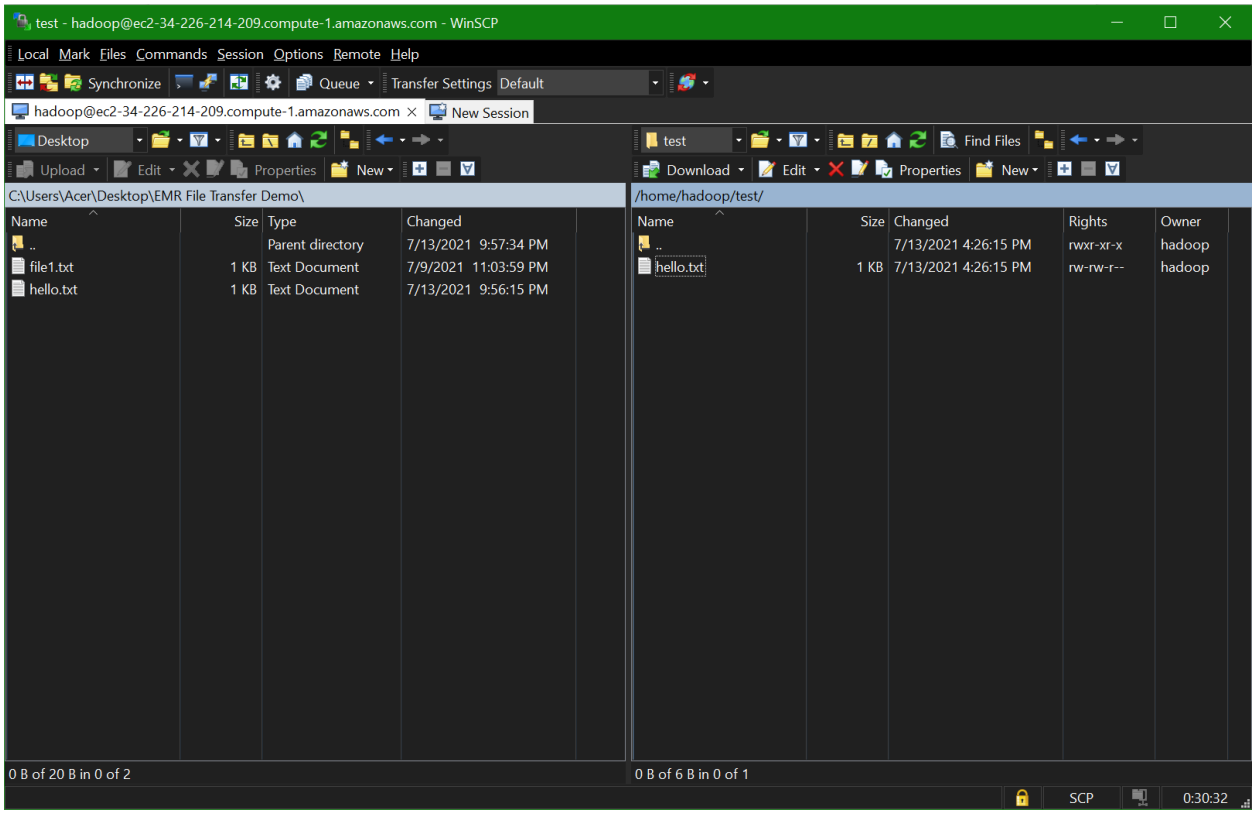
- Now start WinSCP as stated in the above steps.
- The following screen will appear:
    - **Left-side screen: Your local machine (Windows, in our case).**
    - **Right-side screen: Your linux machine (AWS EMR instance).**
- You will see the following screen.



You can view the test folder that you created on the right-hand side of your screen. Now you can see your file by opening the folder.



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 can view your file in that particular folder.