

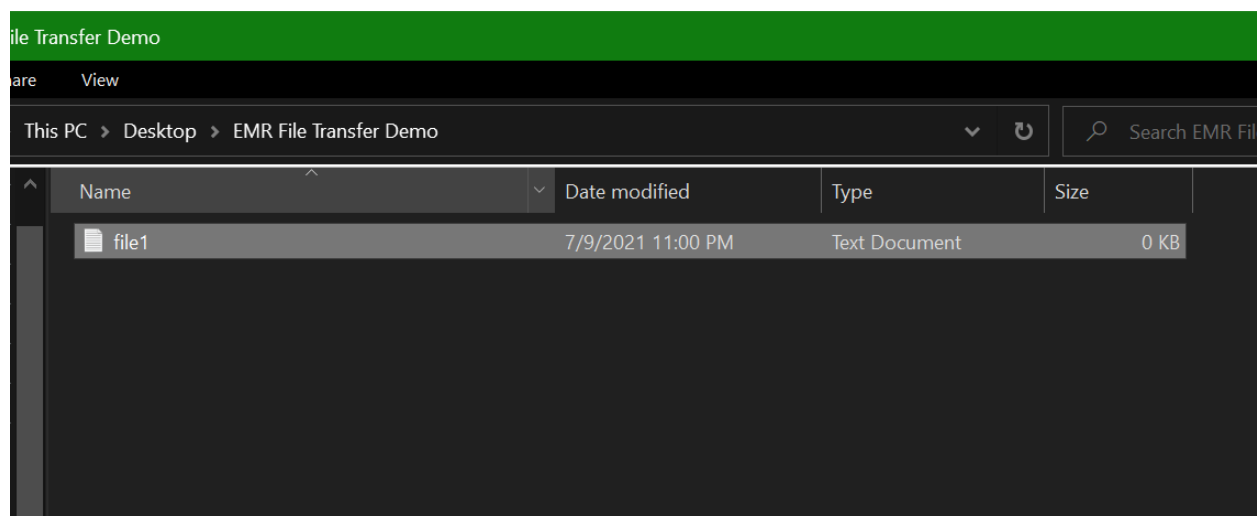
File Transfer – Windows

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**'.

Summary

ID: j-2G35TPE3669YZ

Creation date: 2021-07-09 20:42 (UTC+5:30)

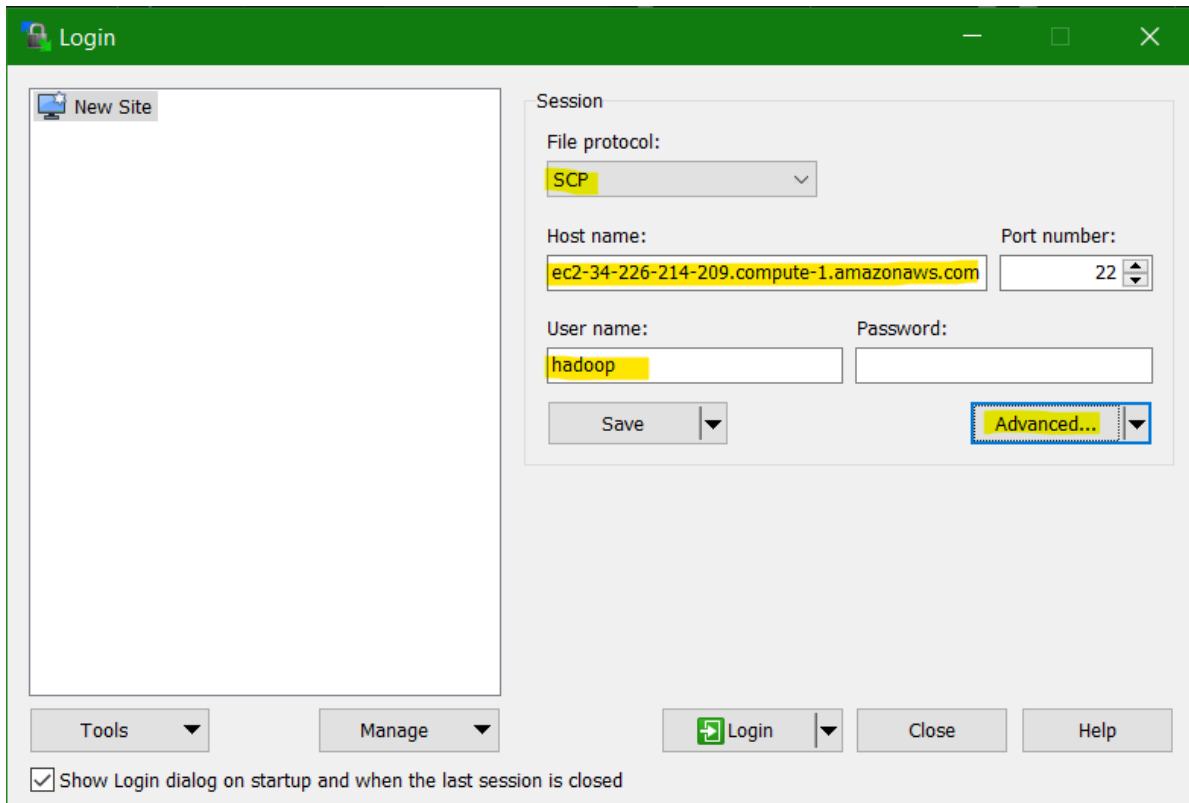
Elapsed time: 2 hours, 19 minutes

After last step completes: Cluster waits

Termination protection: Off [Change](#)

Tags: -- [View All / Edit](#)

Master public DNS: ec2-54-196-22-126.compute-1.amazonaws.com 
[Connect to the Master Node Using SSH](#)



Login

New Site

Session

File protocol:
SCP

Host name:
ec2-34-226-214-209.compute-1.amazonaws.com

Port number:
22

User name:
hadoop

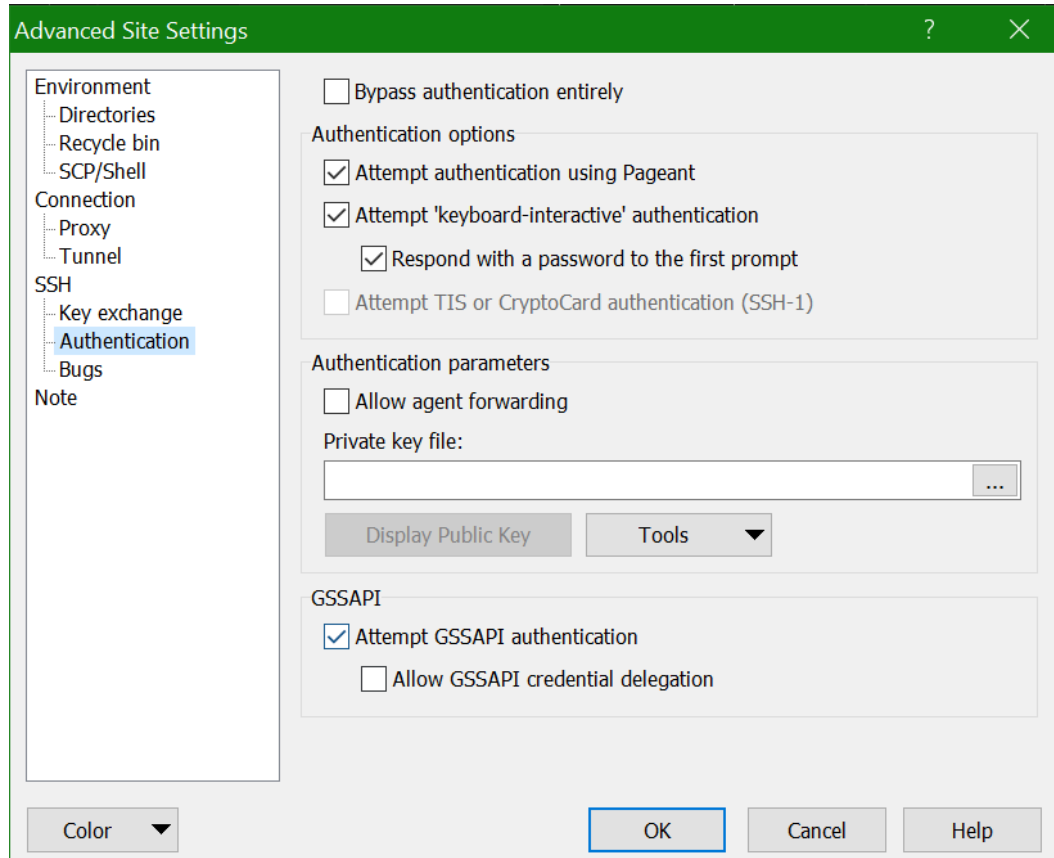
Password:

Save Advanced...

Tools Manage Login Close Help

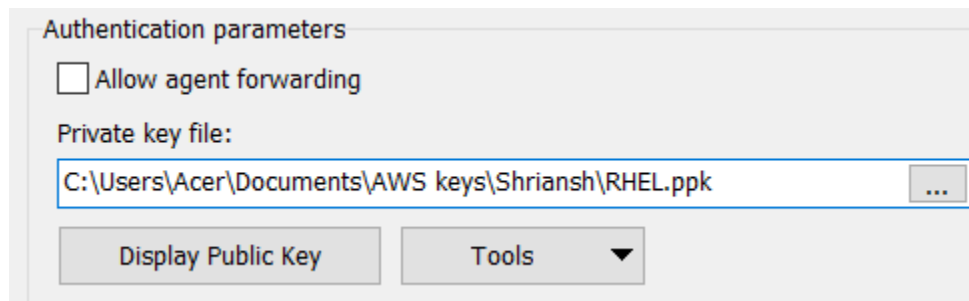
☒ Show Login dialog on startup and when the last session is closed

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



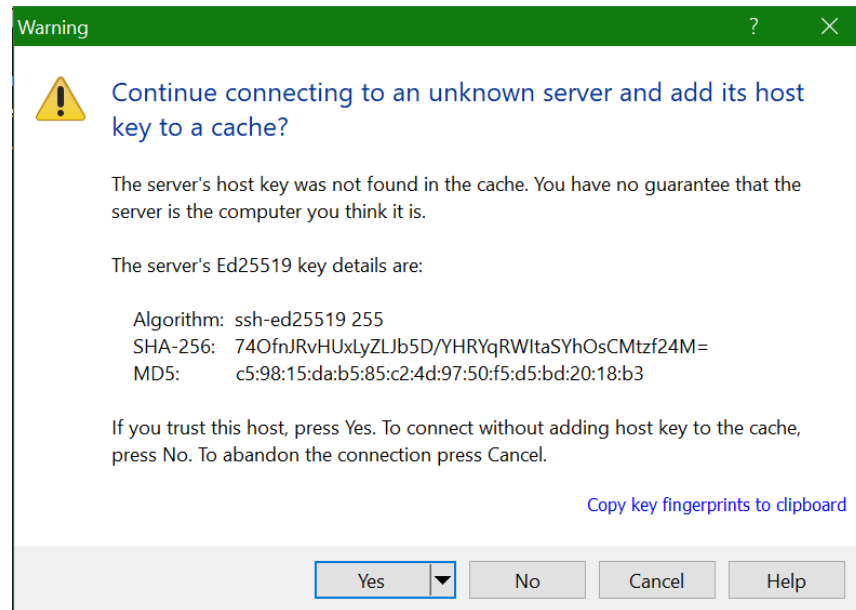
The image shows the 'Advanced Site Settings' dialog box. On the left, a tree view shows the following categories: Environment, Directories, Recycle bin, SCP/Shell, Connection, Proxy, Tunnel, SSH, Key exchange, Authentication (highlighted), Bugs, and Note. The main area contains several sections: 'Bypass authentication entirely' with an unchecked checkbox; 'Authentication options' with three checked checkboxes: 'Attempt authentication using Pageant', 'Attempt 'keyboard-interactive' authentication', and 'Respond with a password to the first prompt'; 'Attempt TIS or CryptoCard authentication (SSH-1)' with an unchecked checkbox; 'Authentication parameters' with an unchecked checkbox for 'Allow agent forwarding' and a 'Private key file:' field with a three-dotted button; 'Display Public Key' button; 'Tools' dropdown; and 'GSSAPI' with a checked checkbox for 'Attempt GSSAPI authentication' and an unchecked checkbox for 'Allow GSSAPI credential delegation'. At the bottom are 'Color' dropdown, 'OK', 'Cancel', and 'Help' buttons.

- 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.

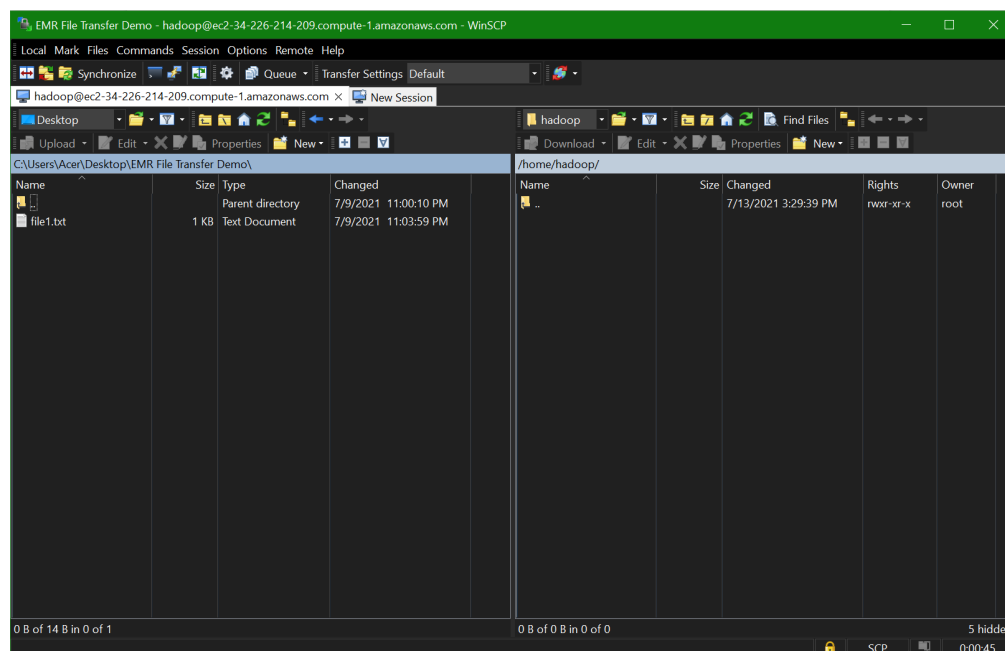


The image shows a close-up of the 'Authentication parameters' section. It contains an unchecked checkbox for 'Allow agent forwarding', a 'Private key file:' label, and a text field containing the path 'C:\Users\Acer\Documents\AWS keys\Shriansh\RHEL.ppk'. To the right of the text field is a three-dotted button. Below the text field are 'Display Public Key' and 'Tools' buttons.

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



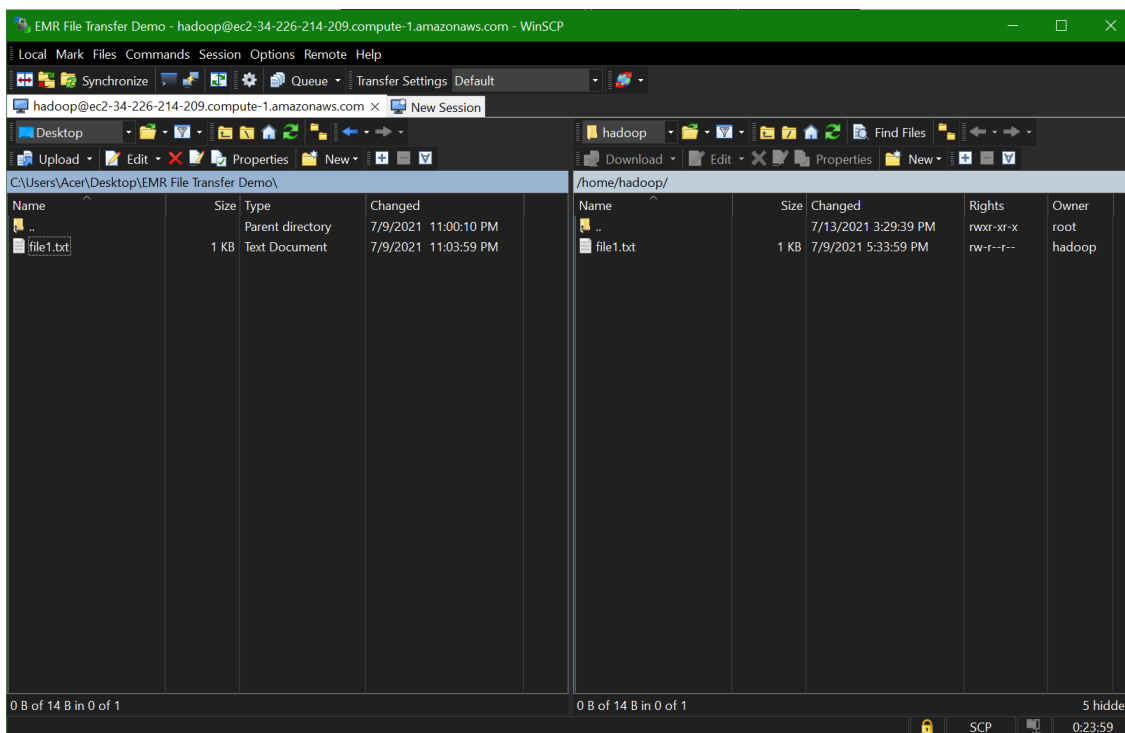
- The following screen will appear.
 - Left side screen: Your local machine (Windows, in our case)
 - Right side screen: Your linux machine (AWS EMR instance)



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

Name	Size	Type	Changed
Parent directory			7/9/2021 11:00:10 PM
file1.txt	1 KB	Text Document	7/9/2021 11:03:59 PM

- 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
```

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

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.

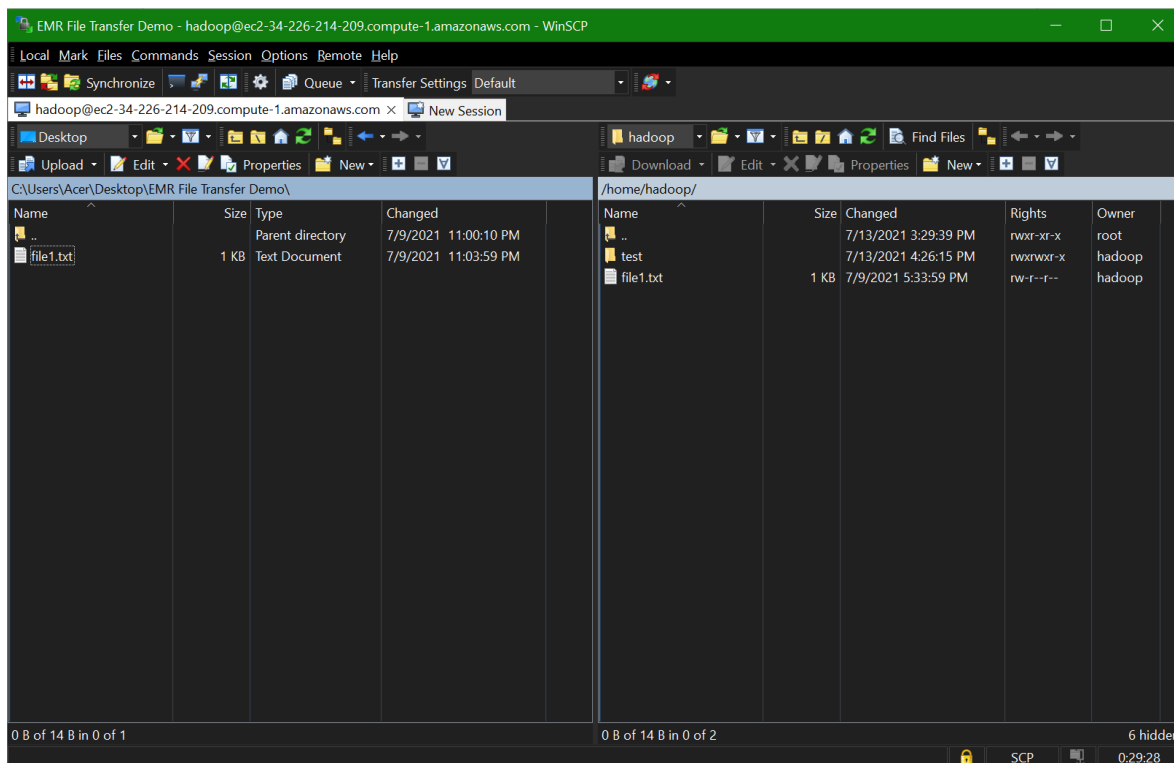
Press 'esc', and then type :wq! to save and exit.

```

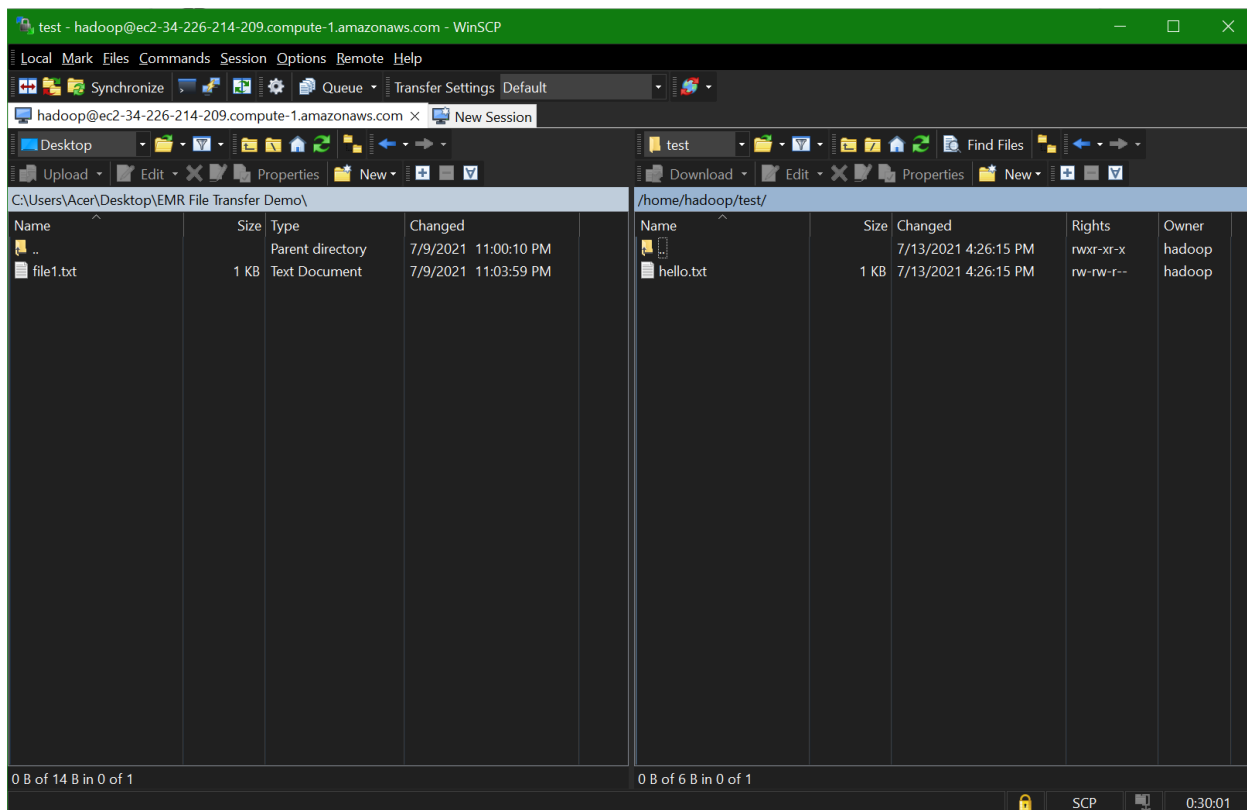
Hello
-- INSERT --
1,6
ALL

```

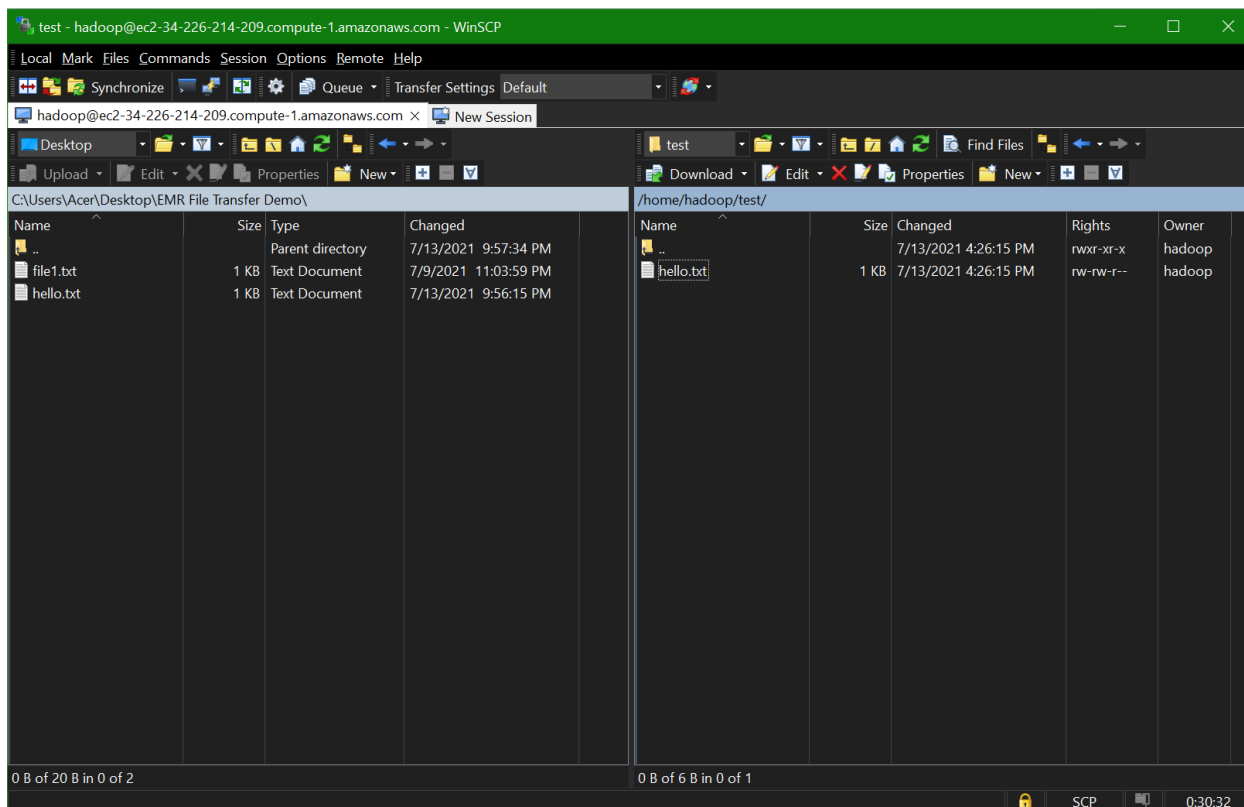
- Now start WinSCP as stated in the above steps.
- The following screen will appear:
 - **Left-side screen: Your local machine (Windows, in our case).**
 - **Right-side screen: Your linux machine (AWS EMR instance).**
- You will see the following screen.



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.

