

Connect to the EMR Instance from your local machine

(Windows Machine)

PuTTY Installation:

- For Windows users, the required software are as follows:
 - PuTTY
 - PuTTYgen
1. Download and install PuTTY and PuTTYgen from the link below.
<https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html>
 or
Click on the first link:
[PuTTY - Secure Download | SSH.COM - SSH Communications Security](#)

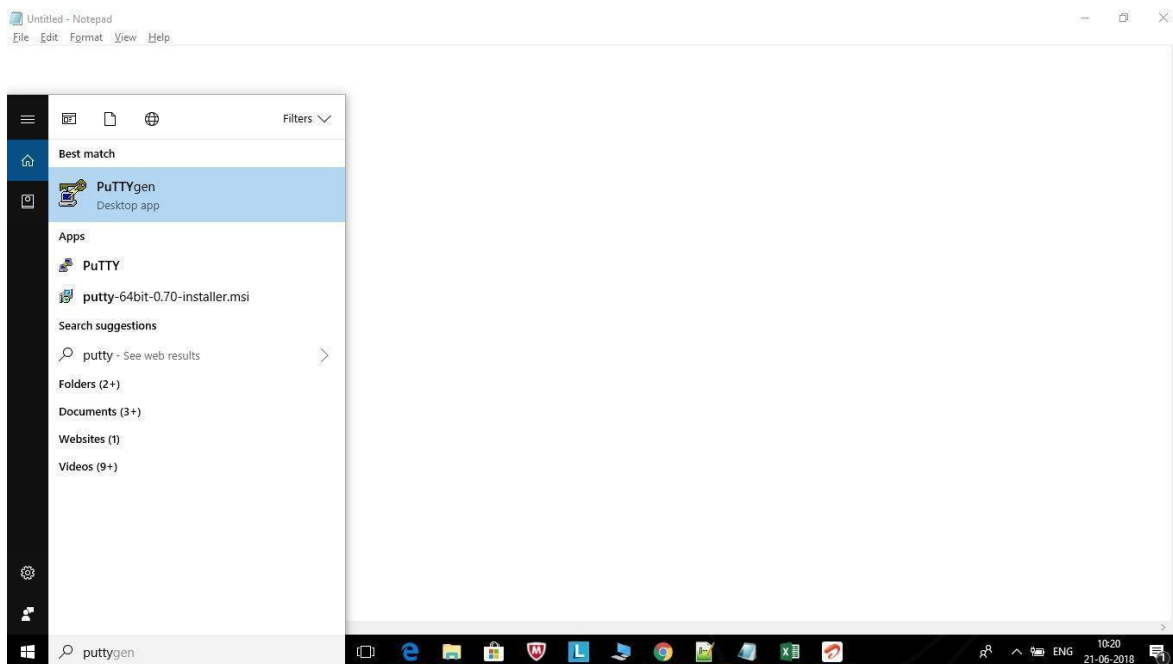
DOWNLOAD PUTTY INSTALLATION PACKAGE FOR WINDOWS

Binary	Platform	Signature	Date
putty-0.70-installer.msi	Windows (any)	GPG signature	2017-07-08
putty-64bit-0.70-installer	Windows (64-bit)	GPG signature	2017-07-08

- If you have a 32-bit OS, then you need to install putty-0.70-installer.msi.
- If you have a 64-bit OS, then choose the latest 64-bit installer file.
- Select the link and it will download PuTTY automatically in your machine and run the software.

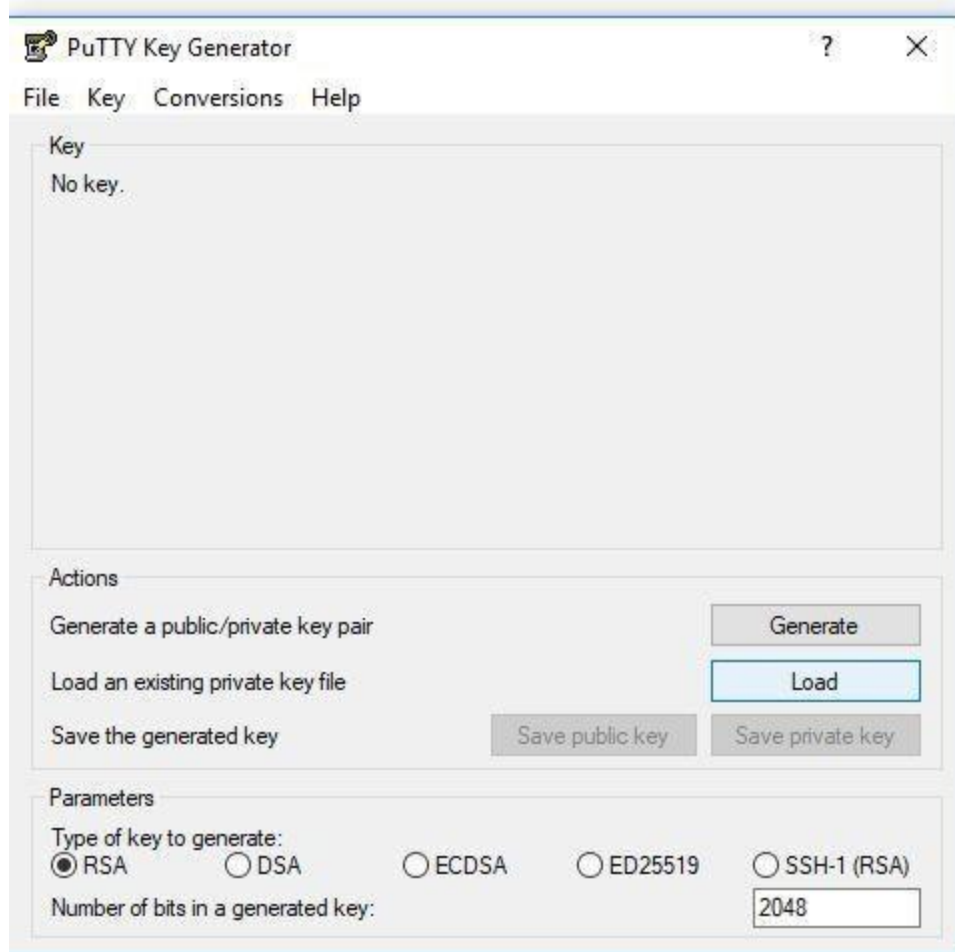
Note: We have successfully installed both PuTTY and PuTTYgen.

- Now, go to the 'Search' tab on your OS and type 'putty'; the results will show both PuTTY and PuTTYgen.

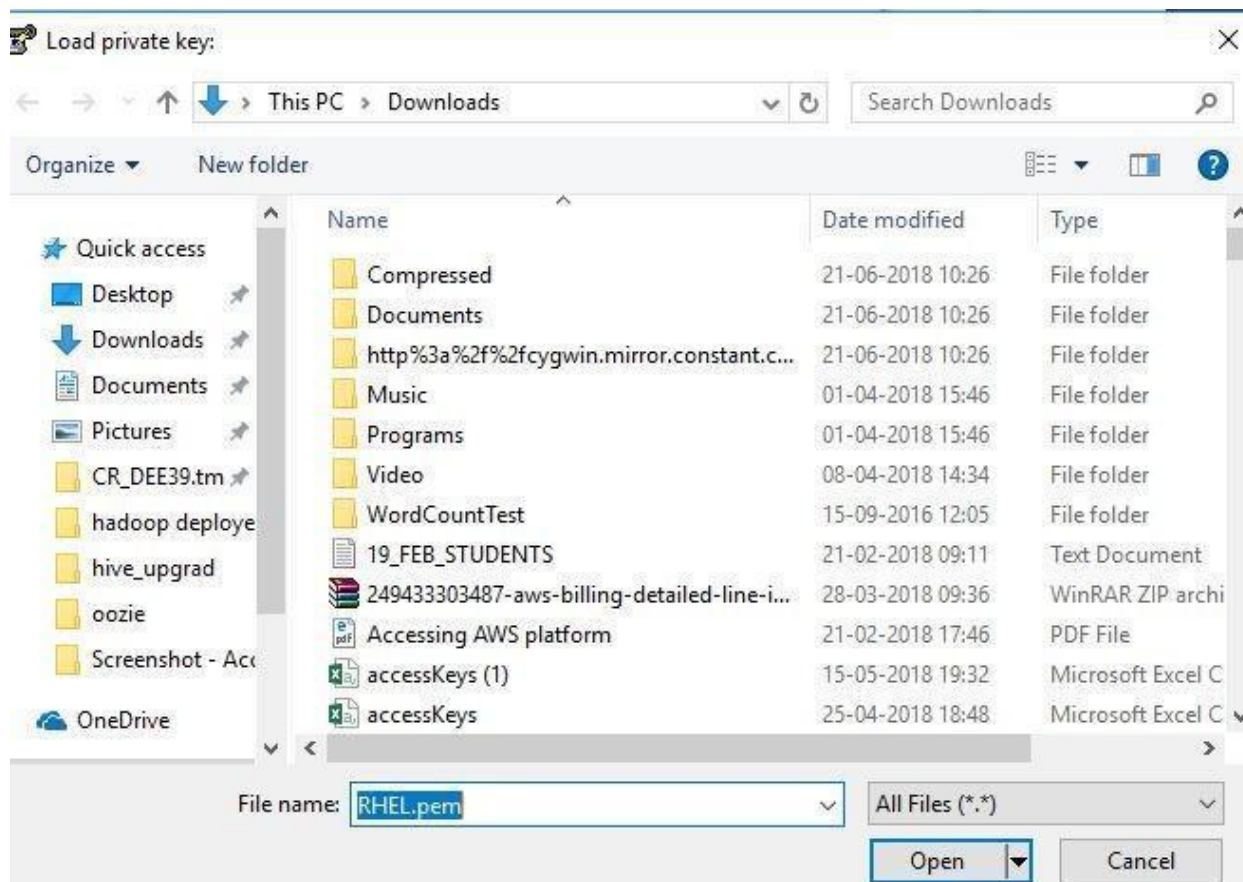


Accessing EMR instance using PuTTY:

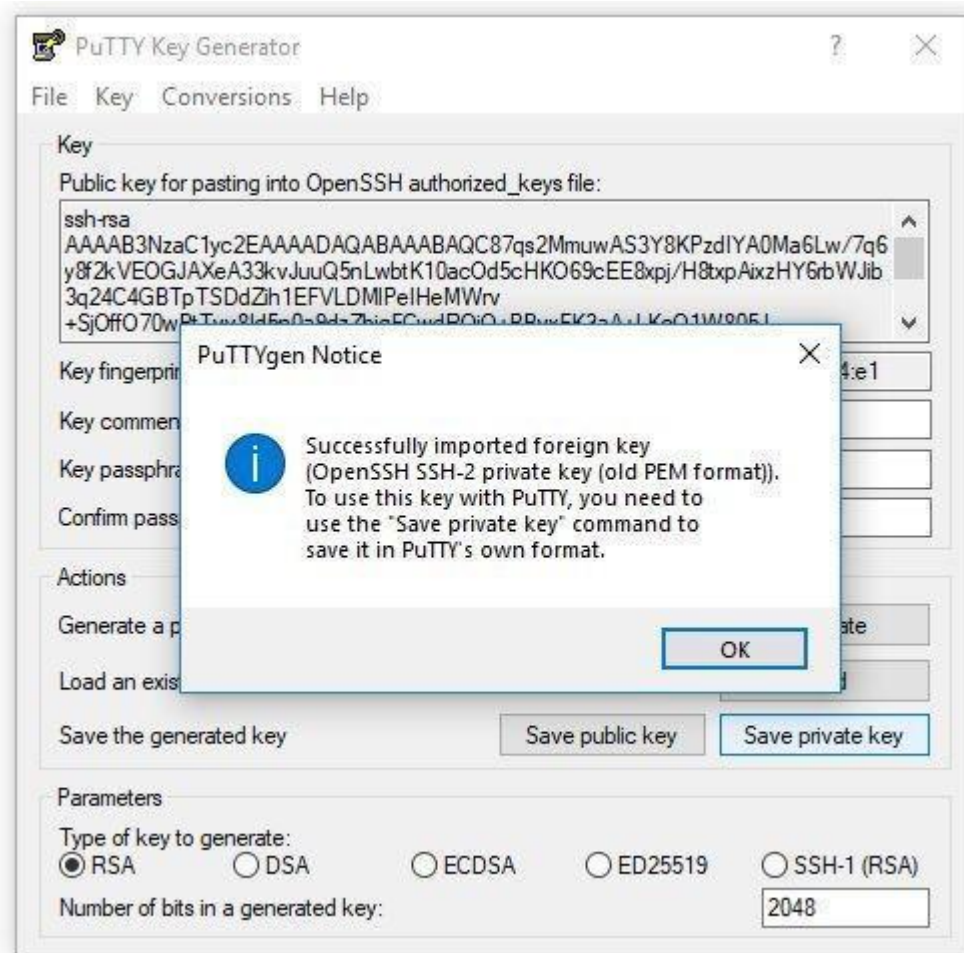
1. For Windows systems, you need to first convert your .pem file to a .ppk file using PuTTYgen. To do this, **open PuTTYgen** and click on '**Load**'.



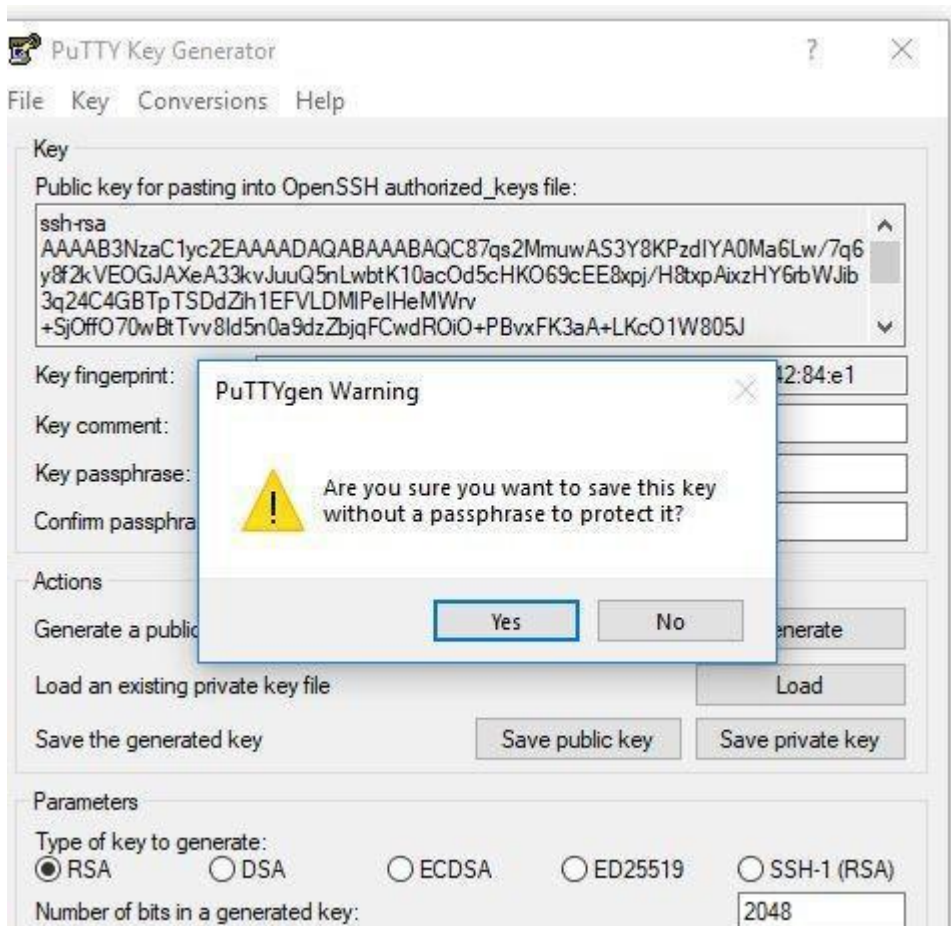
2. Locate the .pem file that you downloaded on your computer and select it. Do not forget to change the file type from .ppk to '**All files**' to locate your .pem file.



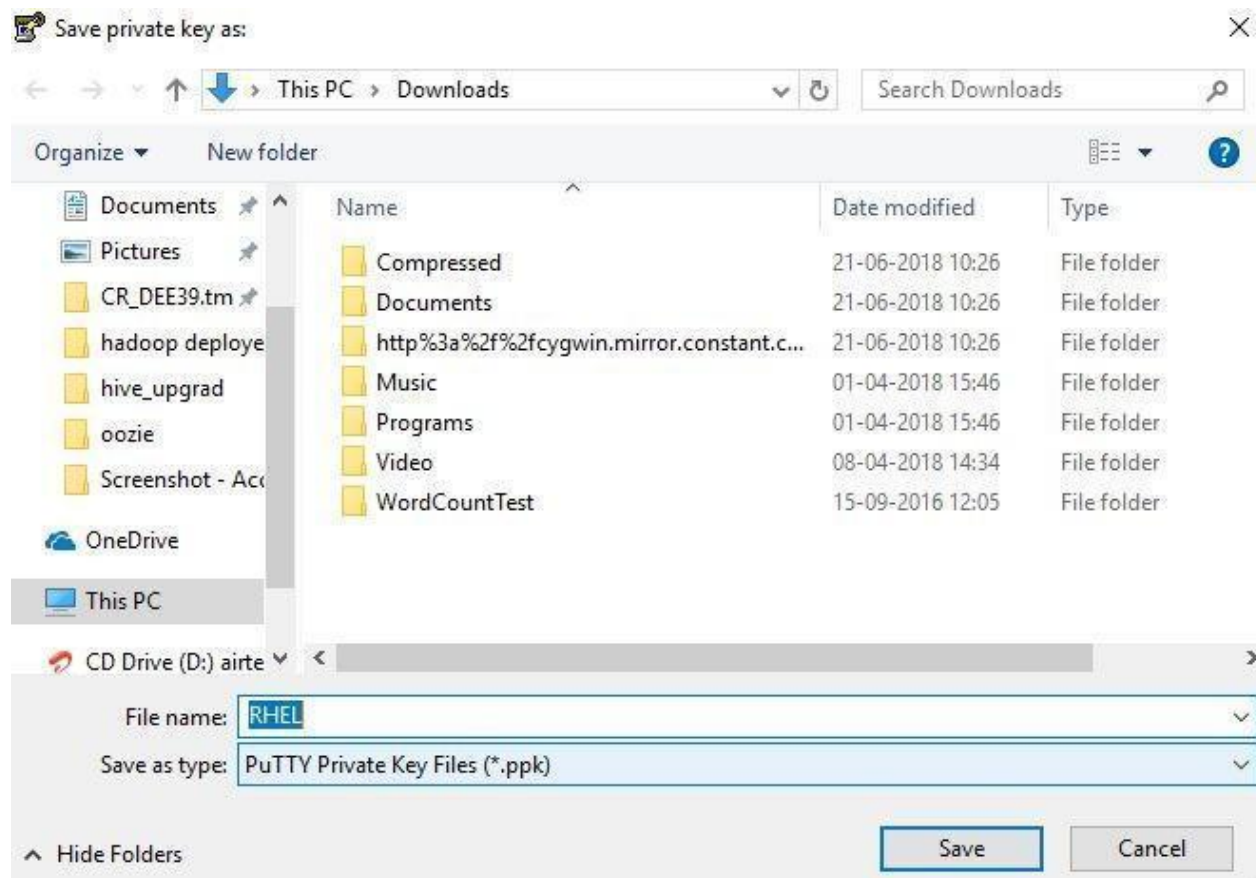
- Click on '**Open**', and then click on '**Ok**' on the pop-up that appears on the screen.



4. The '**Key Passphrase**' is entirely optional. If you want to set a Key Passphrase, then remember to store it in a safe place. This Key Passphrase will be required to connect the local machine to the EMR instance. Click on '**Save private key**', and then click on '**Yes**'.



5. Save your .ppk file (**RHEL** in our case).



Then, click on close (X).

Note:

- Make sure you have set the Port 22(SSH) as open as mentioned in the EMR Setup document before using Putty to login to your EMR cluster otherwise you won't be able to SSH to your EMR cluster

Rule ID	Protocol	Port	Source	Destination	Action
sgr-0522ea2084a15dc38	Custom TCP	TCP	8443	Custom	Q
sgr-0e4552cd2b814d5be	Custom TCP	TCP	8443	Custom	Q
-	SSH	TCP	22	Anywhere...	Q

207.171.167.26/32 X

72.21.217.0/24 X

0.0.0.0/0 X

Add rule

Cancel Preview changes Save rules

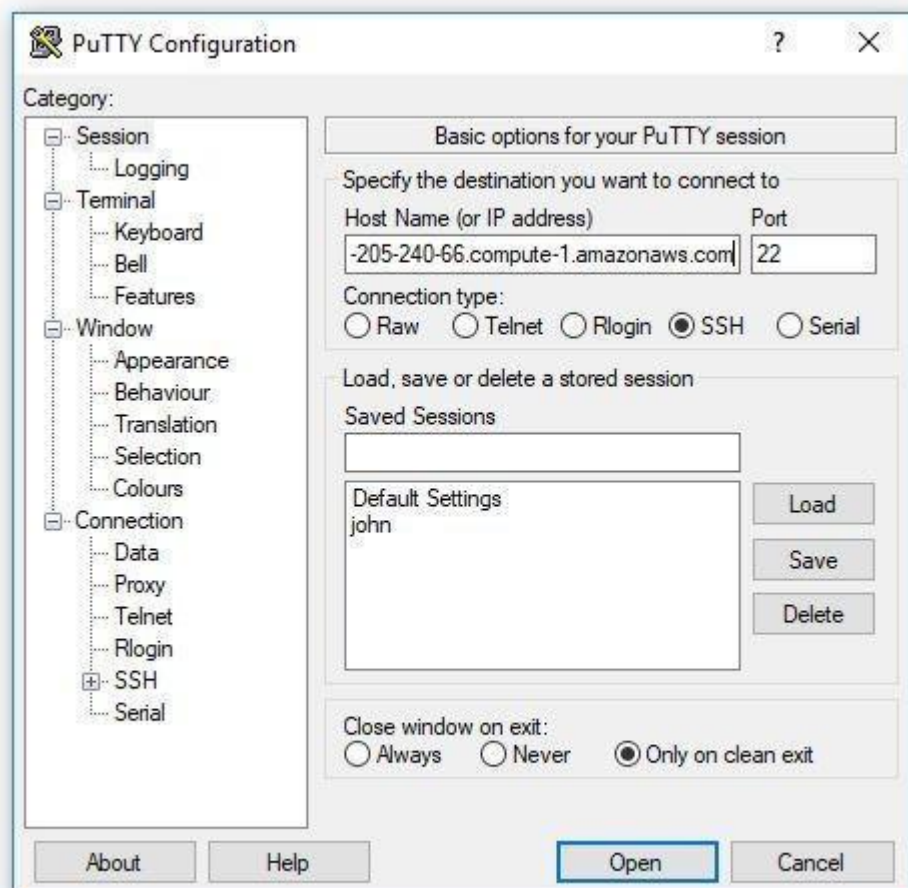
Now, open your EMR dashboard and select your instance. Copy your '**Master public DNS**' information, as shown in the screenshot below.

Cluster: upGrad_EMR Waiting Cluster ready after last step completed.

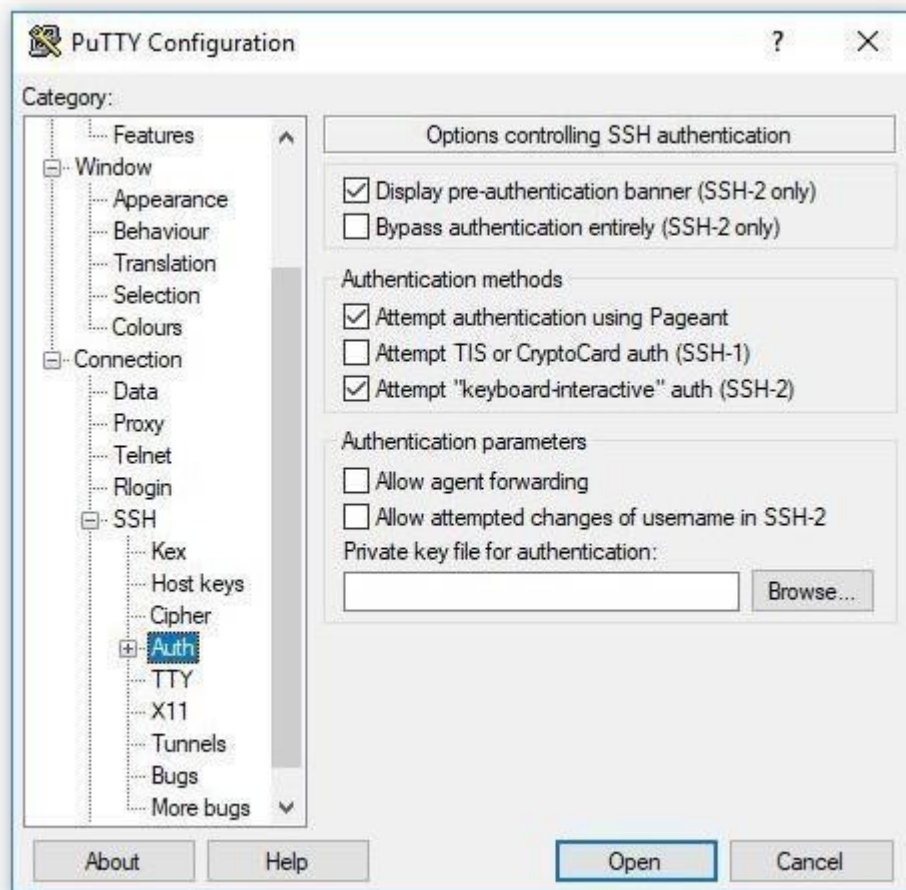
Summary	Application user interfaces	Monitoring	Hardware	Configurations	Events	Steps	Bootstrap actions
<p>Summary</p> <p>ID: j-2G35TPE3669YZ</p> <p>Creation date: 2021-07-09 20:42 (UTC+5:30)</p> <p>Elapsed time: 42 minutes</p> <p>After last step completes: Cluster waits</p> <p>Termination protection: Off Change</p> <p>Tags: -- View All / Edit</p> <p>Master public DNS: pc2-54-196-22-126.compute-1.amazonaws.com Copy</p> <p>Connect to the Master Node Using SSH</p>	<p>Configuration details</p> <p>Release label: emr-5.30.1</p> <p>Hadoop distribution: Amazon 2.8.5</p> <p>Applications: Hive 2.3.6, Hue 4.6.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2</p> <p>Log URI: s3://aws-logs-367134191692-us-east-1/elasticmapreduce/ View</p> <p>EMRFS consistent view: Disabled</p> <p>Custom AMI ID: --</p>	<p>Application user interfaces</p> <p>Persistent user interfaces YARN timeline server, Tez UI</p> <p>On-cluster user interfaces Not Enabled Enable an SSH Connection</p>	<p>Network and hardware</p> <p>Availability zone: us-east-1a</p> <p>Subnet ID: subnet-0085edc222d4dad91 Copy</p> <p>Master: Running 1 m4.xlarge</p> <p>Core: --</p> <p>Task: --</p> <p>Cluster scaling: Not enabled</p>				

6. Open PuTTY:

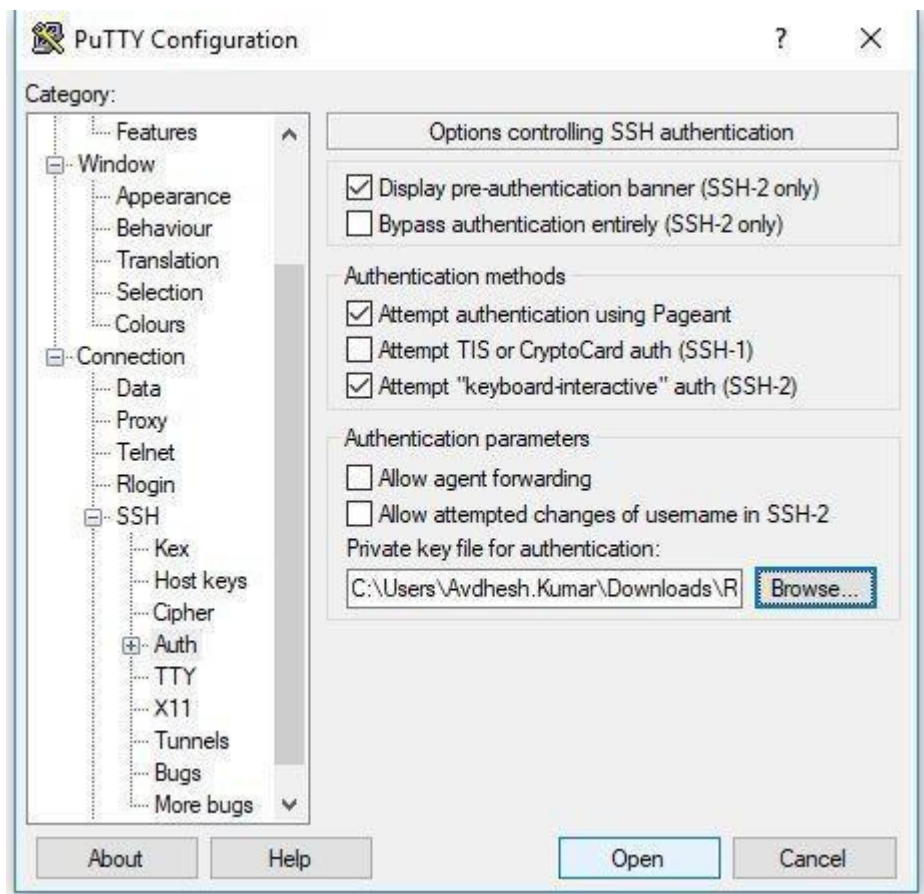
Under the '**Host Name**' section, paste the public DNS information of your instance that you just copied.



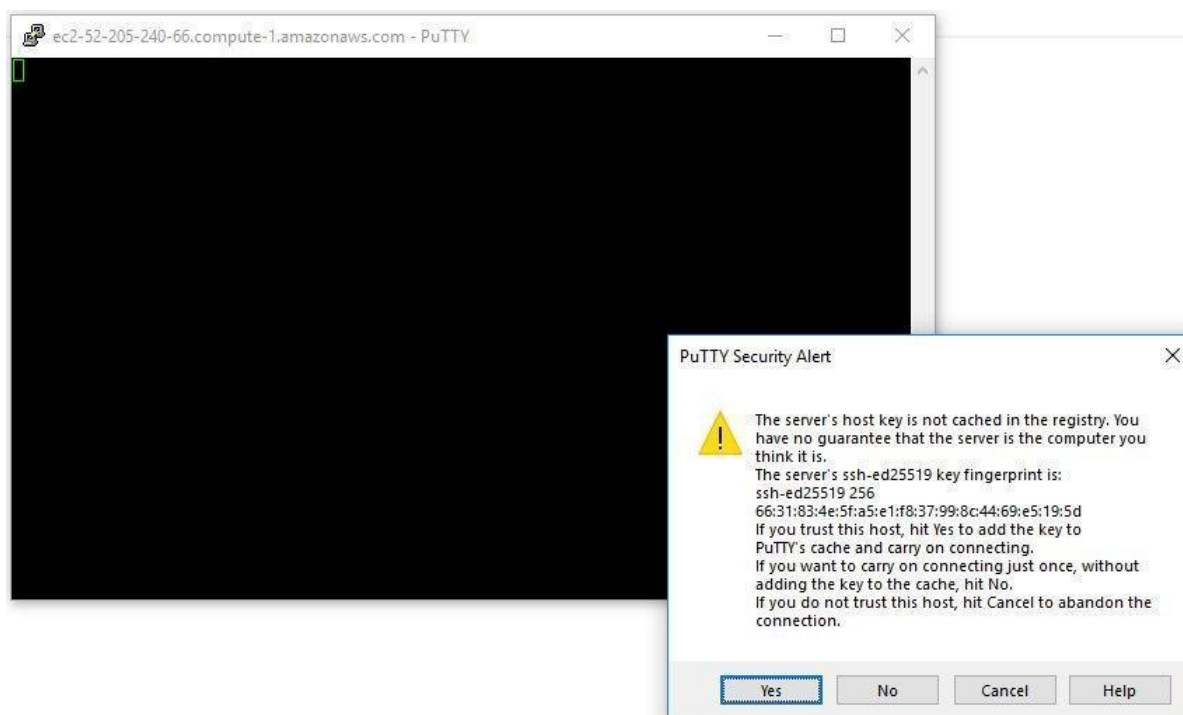
7. On the left-hand side panel, click on '**Connection**'. Then click on '**SSH**' followed by '**Auth**'. In the private key field, click on '**Browse**'.



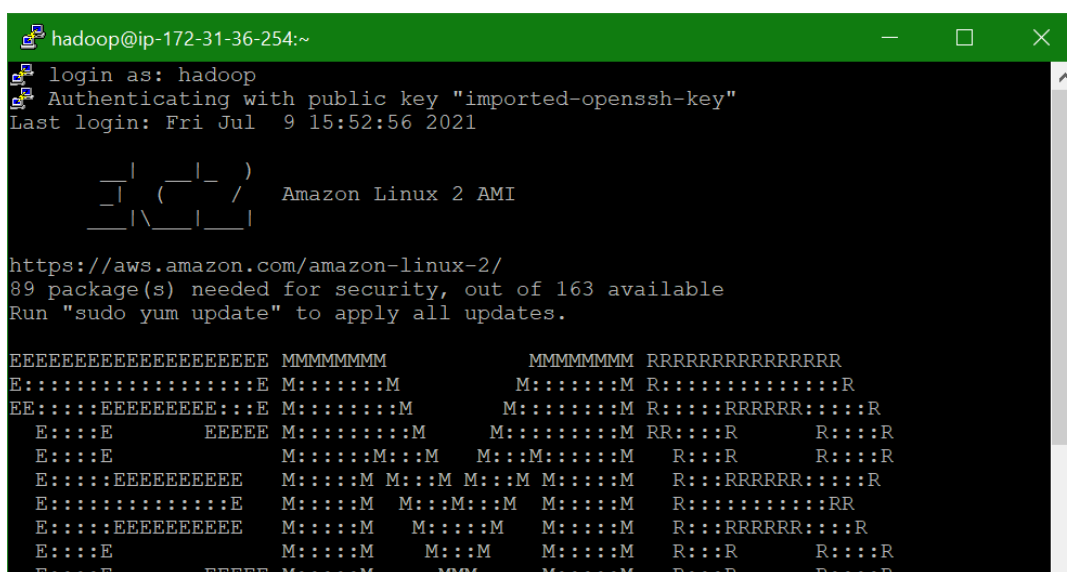
8. Select the .ppk file(**RHEL.ppk**) you generated using PuTTYgen and click on 'Open'.



9. Click on 'Yes' and login with **hadoop**.



Login as **hadoop**.



10. Now, your local machine has successfully established a connection with the EMR Instance.

11. After this, run the following command to check whether S3 is accessible from your instance.

```
aws s3 ls
```

12. If the above command is not working, then you will need to follow the S3 access document given on the platform.

```
[hadoop@ip-172-31-46-44 ~]$ aws s3 ls
2020-06-28 04:56:43 atmetloutput5files
2020-12-14 17:59:20 aws-emr-resources-367134191692-us-east-1
2020-12-14 17:47:40 aws-logs-367134191692-us-east-1
2021-03-07 14:56:29 demobucket00123
2020-09-11 06:49:25 rawatatestbucket123
2020-06-24 15:33:12 shrianshs3testbucket
2020-12-04 11:36:10 testbucketupgrad
[hadoop@ip-172-31-46-44 ~]$ |
```

13. Everytime after login , make sure that you have run the following command to update the linux packages.

```
sudo yum update
```

```
[hadoop@ip-172-31-32-207 ~]$ sudo yum update
Loaded plugins: extras_suggestions, langpacks, priorities, update-motd
amzn2-core | 3.7 kB 00:00:00
10 packages excluded due to repository priority protections
Resolving Dependencies
--> Running transaction check
--> Package amazon-linux-extras.noarch 0:1.6.10-1.amzn2 will be updated
--> Package amazon-linux-extras.noarch 0:2.0.0-1.amzn2 will be an update
--> Package amazon-linux-extras-yum-plugin.noarch 0:1.6.10-1.amzn2 will be updated
--> Package amazon-linux-extras-yum-plugin.noarch 0:2.0.0-1.amzn2 will be an update
--> Package amazon-ssm-agent.x86_64 0:2.3.714.0-1.amzn2 will be updated
--> Package amazon-ssm-agent.x86_64 0:3.0.1124.0-1.amzn2 will be an update
--> Package avahi-libs.x86_64 0:0.6.31-19.amzn2.0.1 will be updated
--> Package avahi-libs.x86_64 0:0.6.31-20.amzn2 will be an update
--> Package aws-cfn-bootstrap.noarch 0:1.4-31.amzn2 will be updated
--> Package aws-cfn-bootstrap.noarch 0:2.0-6.amzn2 will be an update
--> Processing Dependency: python3-pystache for package: aws-cfn-bootstrap-2.0-6.amzn2.noarch
--> Processing Dependency: python3-daemon for package: aws-cfn-bootstrap-2.0-6.amzn2.noarch
--> Package awscli.noarch 0:1.16.300-1.amzn2.0.1 will be updated
--> Package awscli.noarch 0:1.18.147-1.amzn2.0.1 will be an update
--> Package bash.x86_64 0:4.2.46-33.amzn2 will be updated
--> Package bash.x86_64 0:4.2.46-34.amzn2 will be an update
```

When you get a prompt to verify if the update can be done, type 'y' and press Enter. This will successfully update your EMR instance.

```
Transaction Summary
=====
===Install   10 Packages (+6 Dependent packages)
Upgrade  161 Packages

Total download size: 308 M
Is this ok [y/d/N]: y|
```

Wait a few minutes until the update has been completed. You will eventually be able to see the following screen.

```
Replaced:
  ec2-net-utils.noarch 0:1.3-1.amzn2
86_64 1:2.02-35.amzn2.0.4
  python3-tools.x86_64 0:3.7.6-1.amzn2.0.1

Complete!
[hadoop@ip-172-31-32-207 ~]$ |
```

You have now successfully logged into your EMR instance.