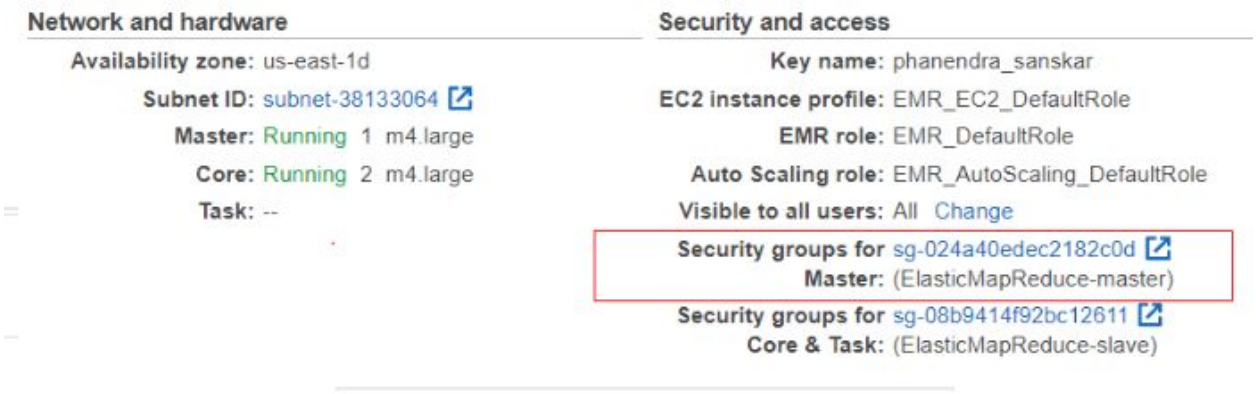
Connecting to Master Node via SSH

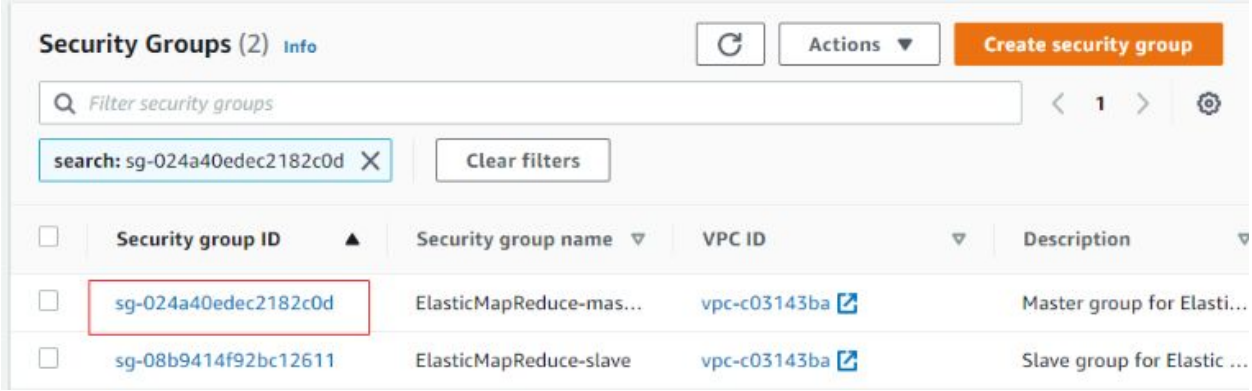
Prerequisite

After launching the EMR cluster, you need to ensure that the port 22 is open for a successful SSH connection. Follow the steps given below to open Port 22

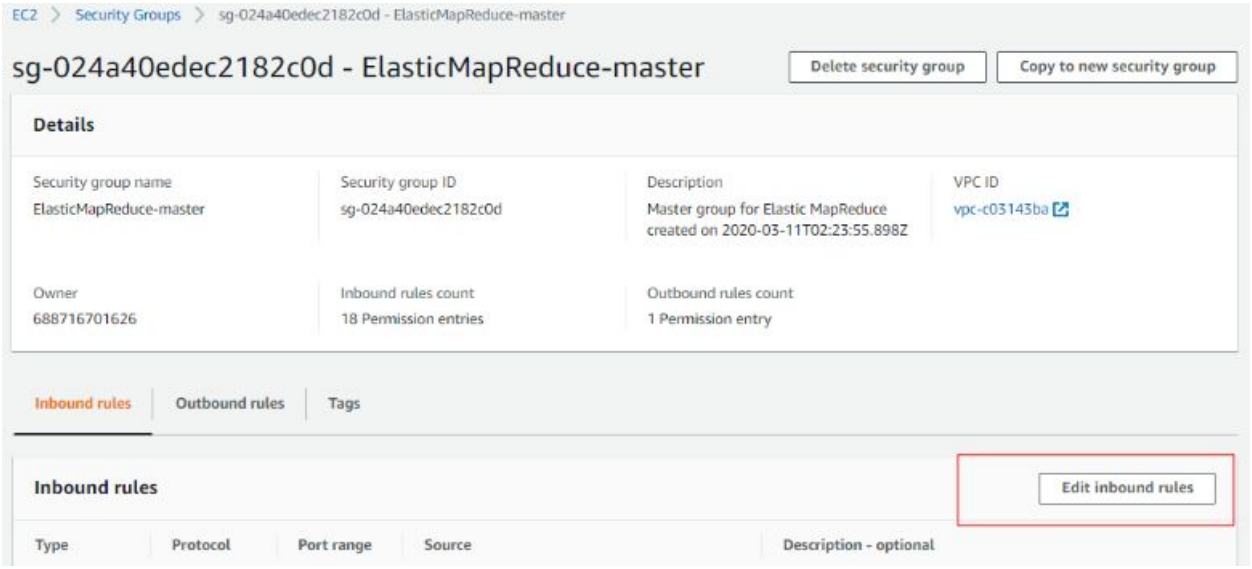
1. Under the cluster information page click on the security groups of the master node



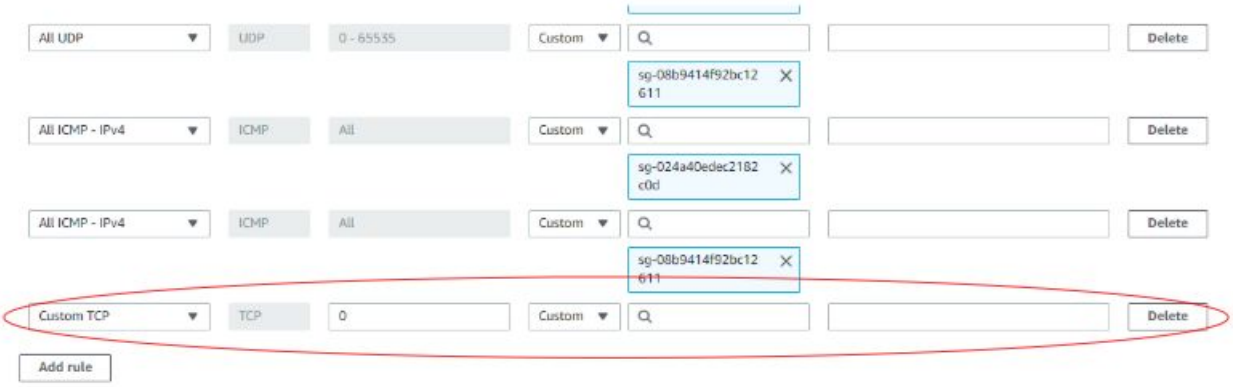
2. Clicking on the security group and you will land on a similar page. Here click on the security group of the Elastic Mapreduce-master node as highlighted in the image.



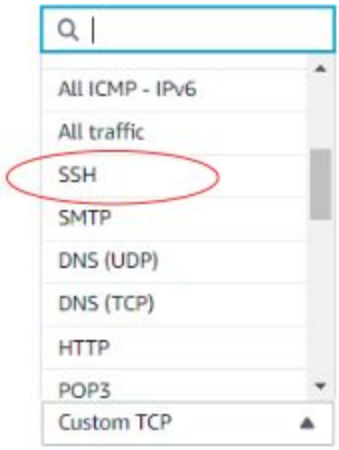
3. Clicking on the security group will land you on the corresponding security information information page. Click on edit inbound rules to add a new rule



4. This will take you list of existing rules where you have the option to delete the existing rules [Clicking on delete on the extreme right hand side] or add a new rule by clicking on **Add rule** towards the bottom of all the rules. Clicking on add rule will add a new row as shown in the figure below

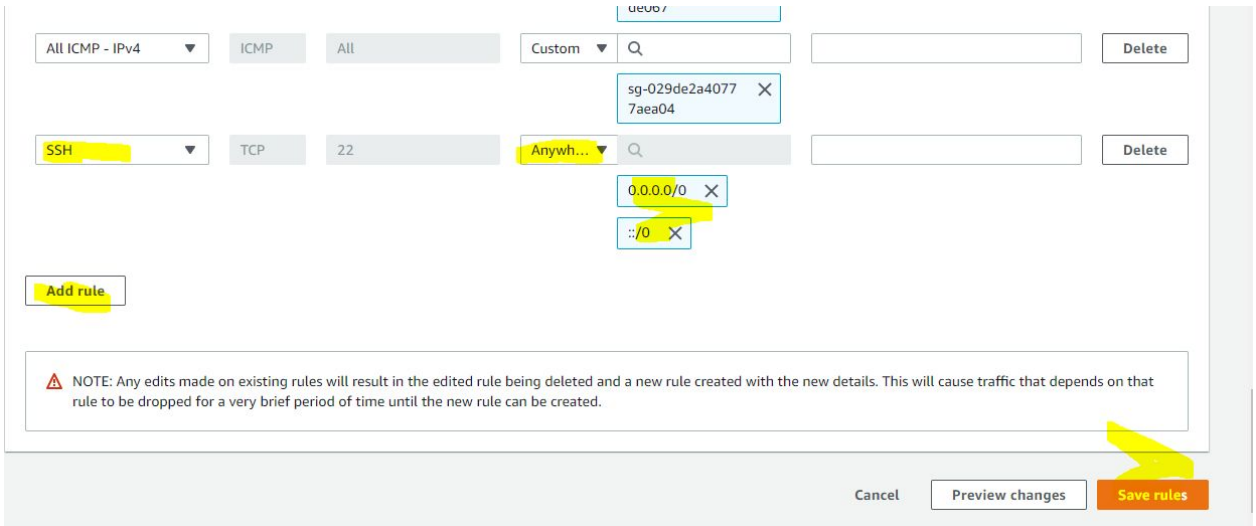


a. Under the type field of the newly added row select **SSH**



b. And choose '**Anywhere**' under the source section. This will automatically load 0.0.0.0/0 & ::/0 ip address in the adjacent blank column.

c. On addition of the rule and choosing the appropriate options as shown below, click on save rule [at the bottom of the screen] to successfully add the rule

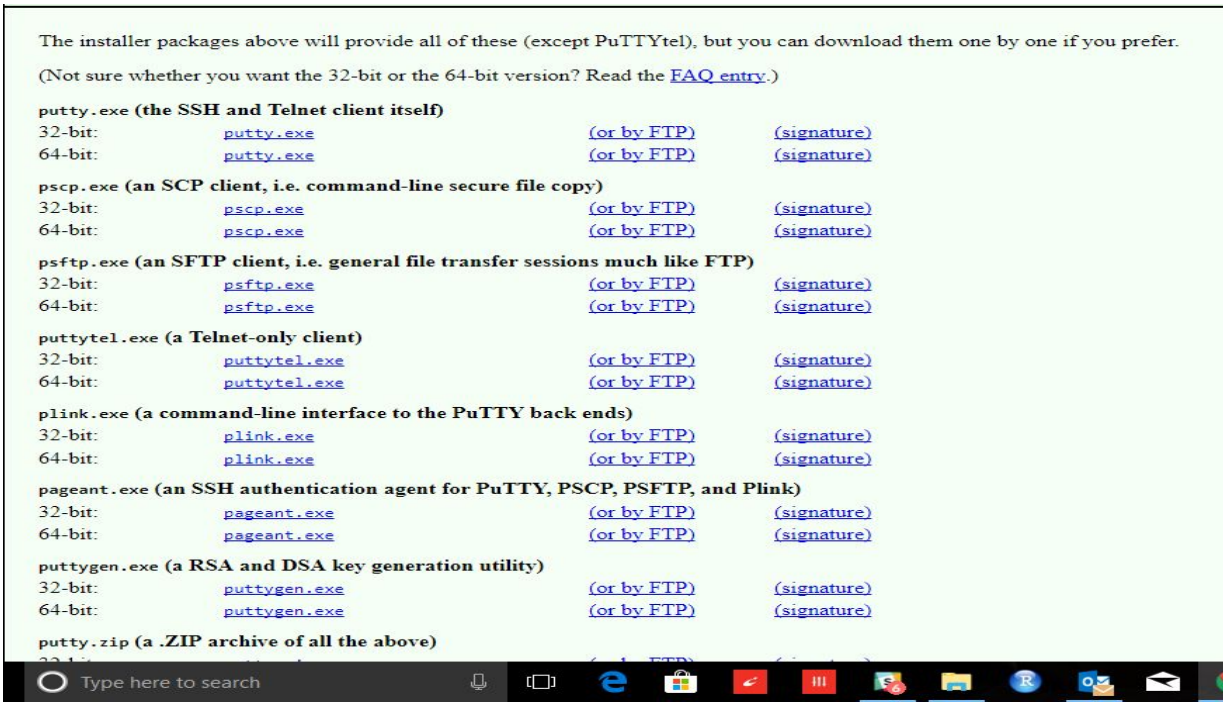


On opening the port 22 for an SSH connection, you can now connect to the master node of the cluster.

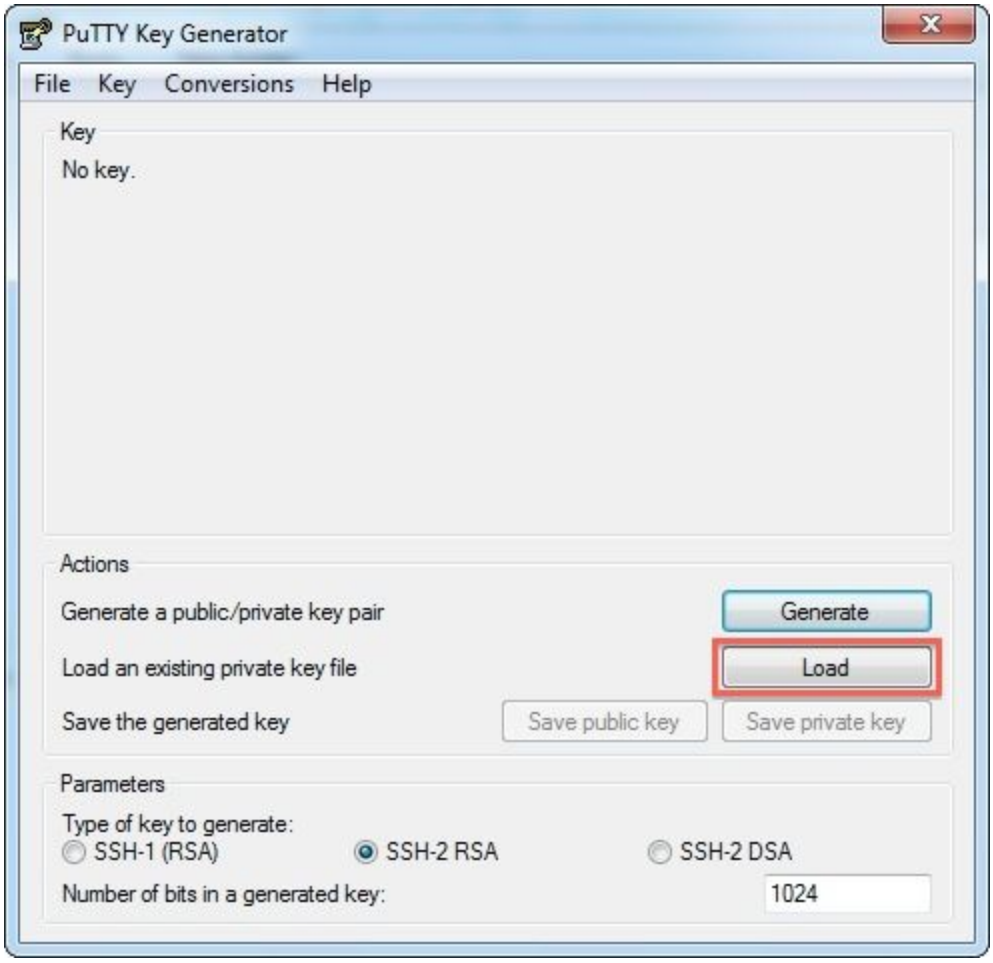
Connecting to Master Node via SSH - WINDOWS Users

1. You will need two programs for this:
    a. PuTTY
    b. PuTTYGen

2. You can download both these programs from the link given below:
   https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html
   Depending on your operating system, select either 32 bit version or 64 bit version.

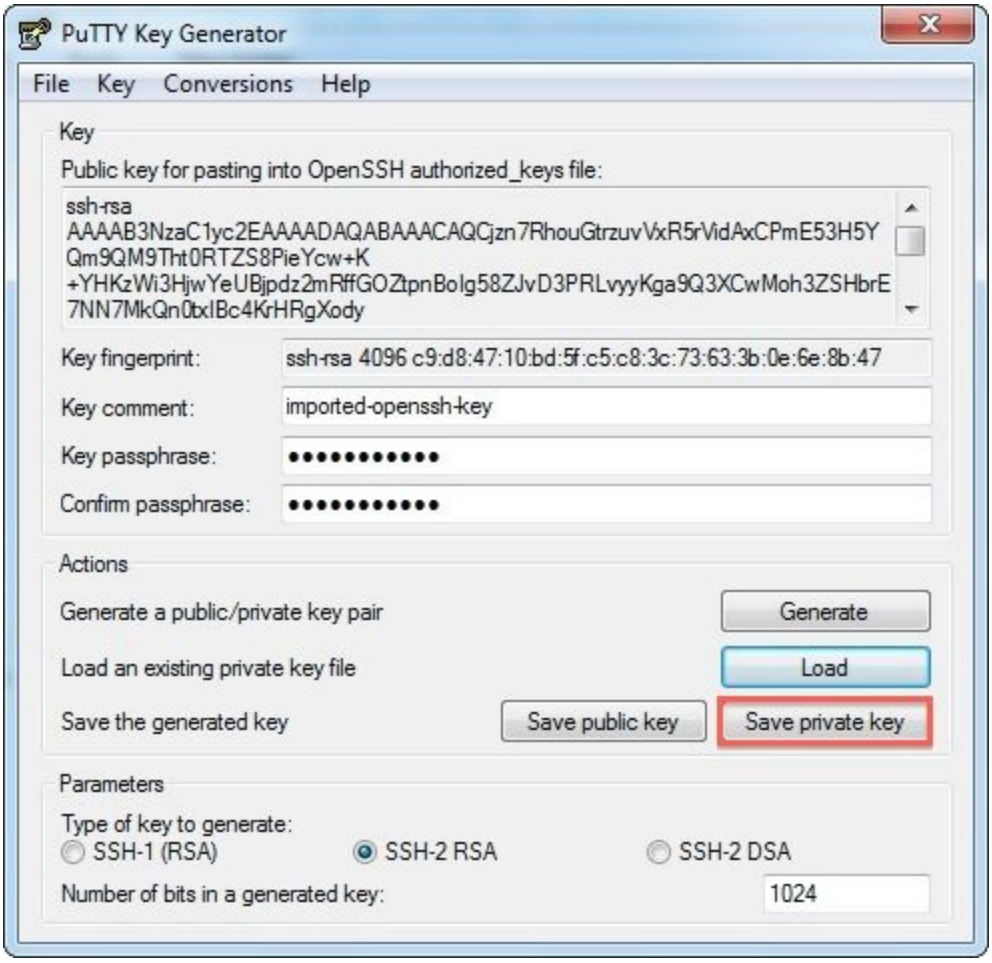


3. Create a private key pair for the downloaded .pem file
    a. Launch Puttygen
    b. Click on “Load”



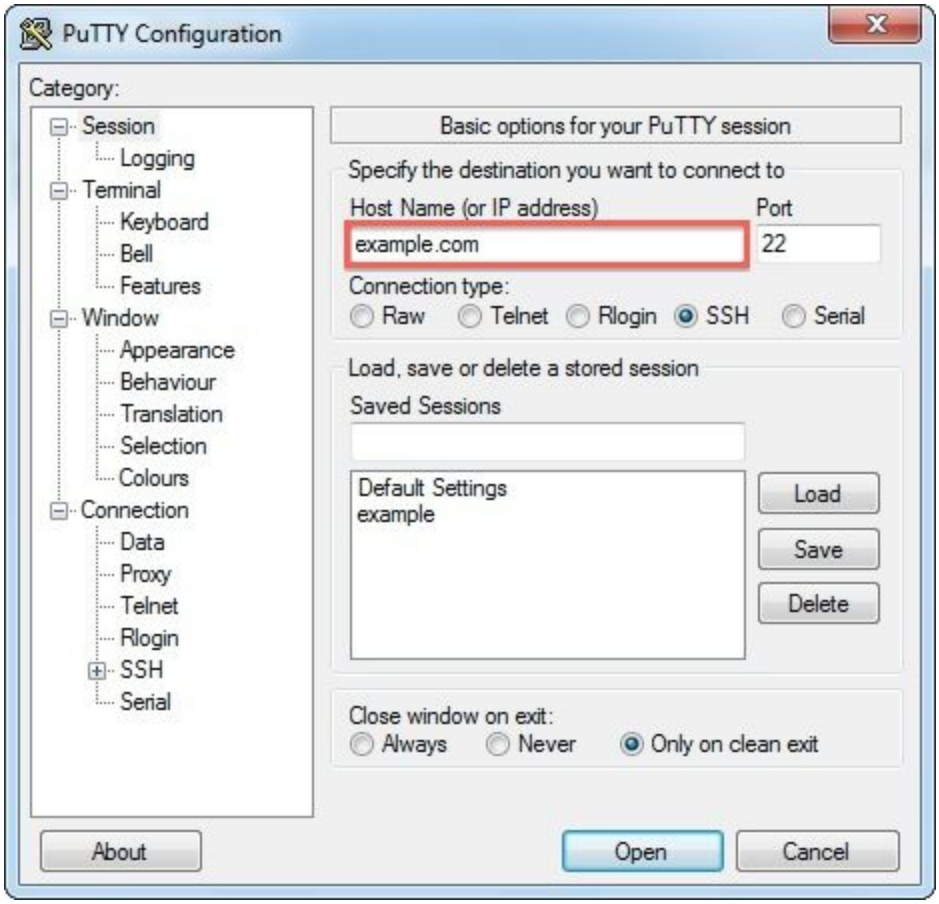
c. Locate the .pem file you downloaded on your computer

d. Click on “Save private key”

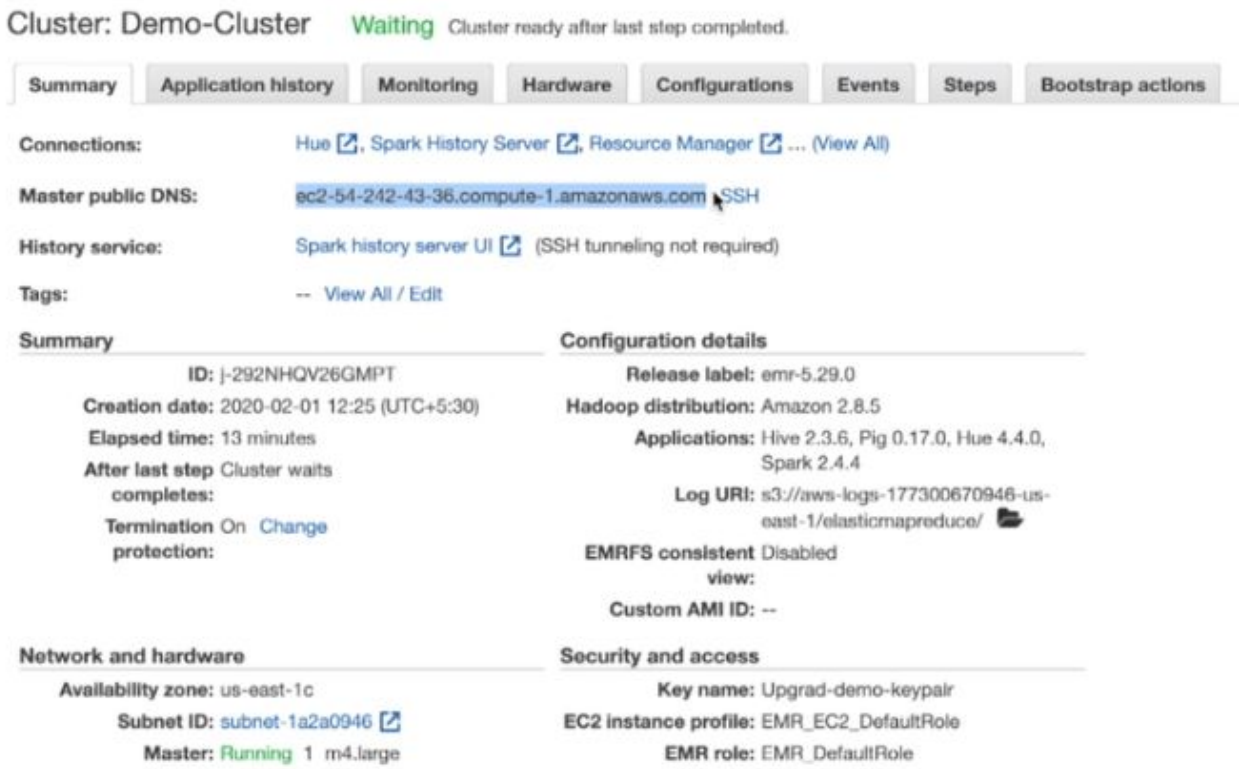


e. You will get the .ppk file, save it at a safe place.

4. Connecting to MASTER using the .ppk file
    a. Launch PuTTY
    b. In the category section on the left side, click on session.
    c. In Host name field, the format will be as follows: hadoop@<publicDNS>

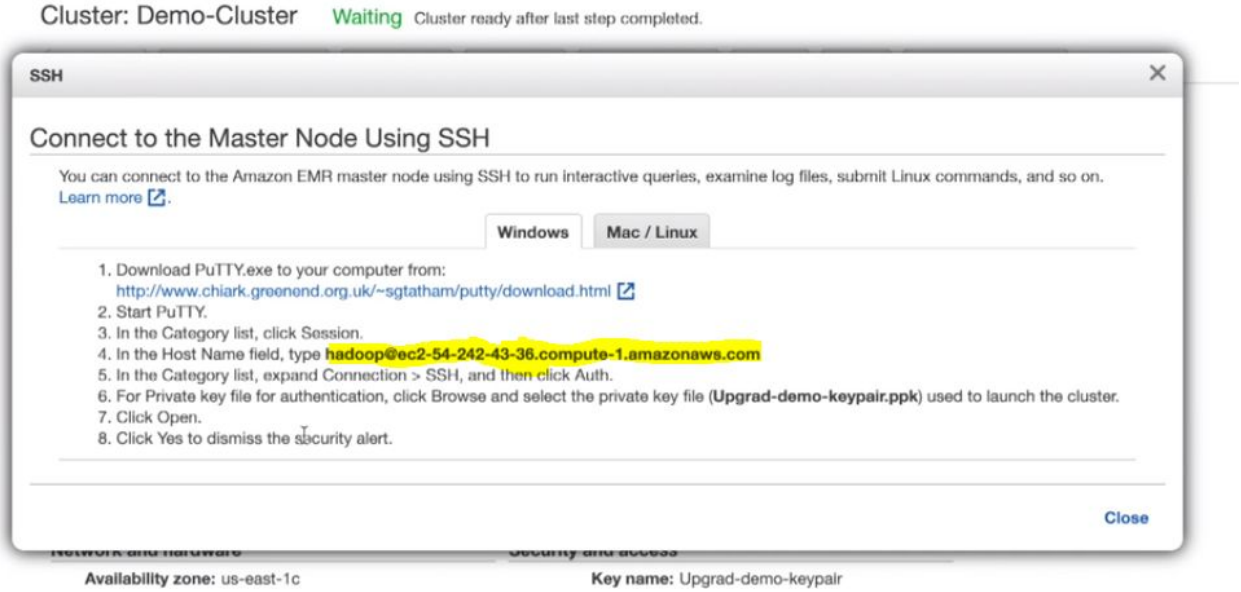


You can grab the Public DNS address from the page of your EMR cluster. Click on EMR, click on your running cluster:

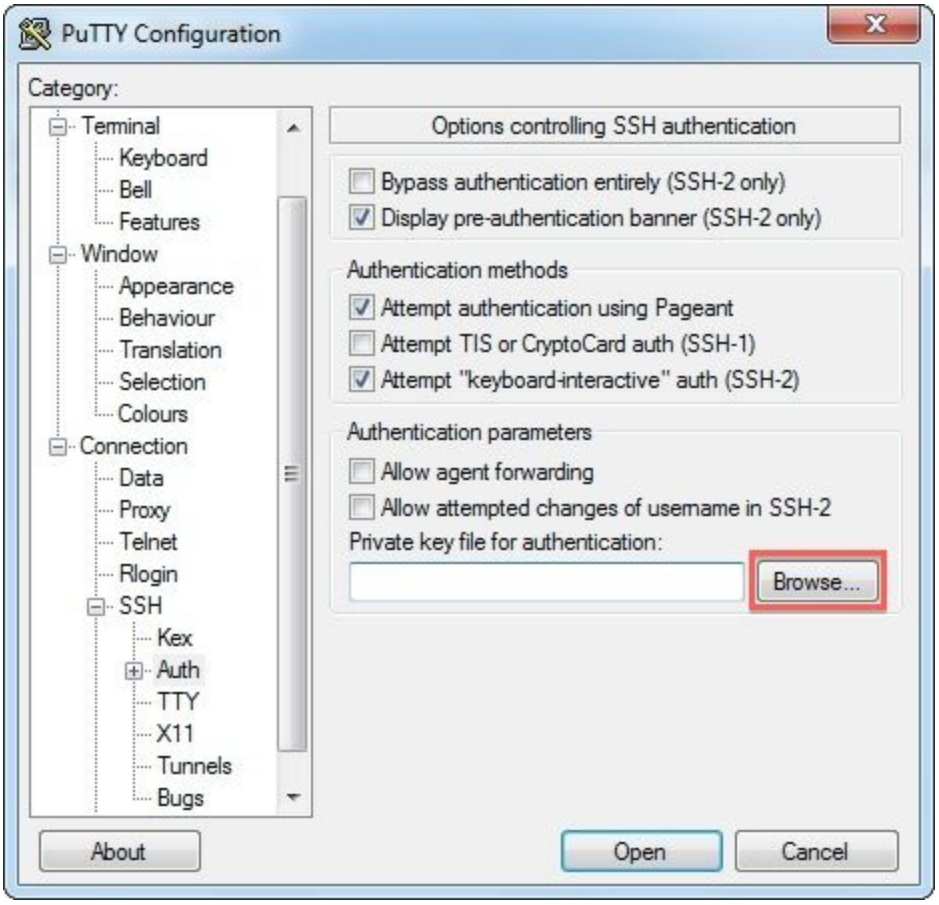


For the given example, your Host Name field will have the following text:
hadoop@ec2-54-242-43-36.compute-1.amazonaws.com

You can also copy it from the pop-up shown when you click on SSH.



d. Then, under the category section, expand connection followed by SSH. On expanding SSH, you can find Auth. Click on Auth.

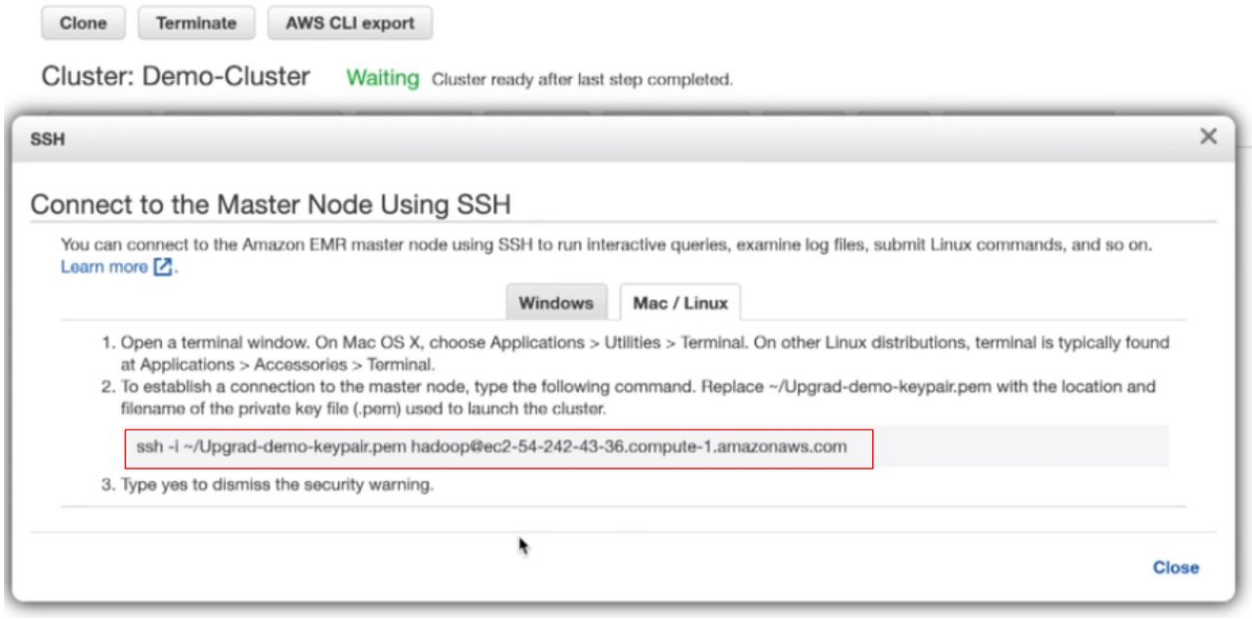


e. Now, Click on Browse and select the PPK file generated using PuttyGen for authentication.

5. Finally, click on “Open” and you will be connected to the master node and a command prompt will appear.

Connecting to Master Node via SSH - MAC Users

After opening the port 22 for an SSH connection, you can click on SSH in the cluster information page .



Now, open a new terminal and run the highlighted command. This will connect you to the master node of the cluster.