

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**

The screenshot shows the 'Network and hardware' and 'Security and access' tabs for an EMR cluster. Under 'Network and hardware', the 'Availability zone' is 'us-east-1d', 'Subnet ID' is 'subnet-38133064', and the 'Master' node is 'Running' with '1 m4.large' instance. Under 'Security and access', the 'Key name' is 'phanendra_sanskar', 'EC2 instance profile' is 'EMR_EC2_DefaultRole', 'EMR role' is 'EMR_DefaultRole', 'Auto Scaling role' is 'EMR_AutoScaling_DefaultRole', and 'Visible to all users' is 'All'. The 'Security groups for sg-024a40edec2182c0d' are highlighted with a red box, showing 'Master: (ElasticMapReduce-master)'. The 'Security groups for sg-08b9414f92bc12611' are also listed, showing 'Core & Task: (ElasticMapReduce-slave)'.

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.

The screenshot shows the 'Security Groups (2)' page in the AWS console. The search bar contains 'sg-024a40edec2182c0d'. The table lists two security groups:

	Security group ID	Security group name	VPC ID	Description
<input type="checkbox"/>	sg-024a40edec2182c0d	ElasticMapReduce-mas...	vpc-c03143ba	Master group for Elasti...
<input type="checkbox"/>	sg-08b9414f92bc12611	ElasticMapReduce-slave	vpc-c03143ba	Slave group for Elastic ...

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

EC2 > Security Groups > sg-024a40edec2182c0d - ElasticMapReduce-master

sg-024a40edec2182c0d - ElasticMapReduce-master Delete security group Copy to new security group

Details

Security group name ElasticMapReduce-master	Security group ID sg-024a40edec2182c0d	Description Master group for Elastic MapReduce created on 2020-03-11T02:23:55.898Z	VPC ID vpc-c03143ba
Owner 688716701626	Inbound rules count 18 Permission entries	Outbound rules count 1 Permission entry	

Inbound rules Edit inbound rules

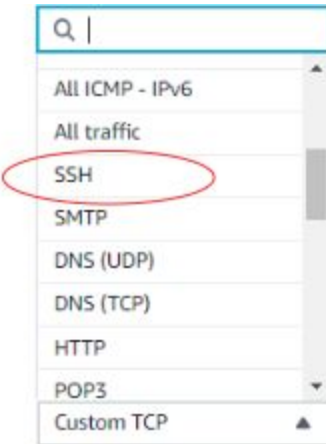
Type	Protocol	Port range	Source	Description - optional
------	----------	------------	--------	------------------------

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

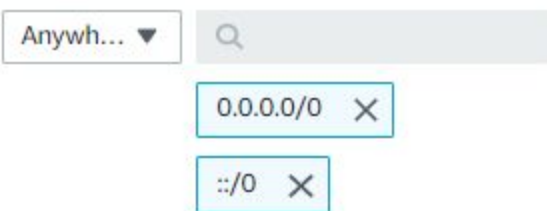
All UDP	UDP	0 - 65535	Custom	Q		Delete
All ICMP - IPv4	ICMP	All	Custom	Q		Delete
All ICMP - IPv4	ICMP	All	Custom	Q		Delete
Custom TCP	TCP	0	Custom	Q		Delete

Add rule

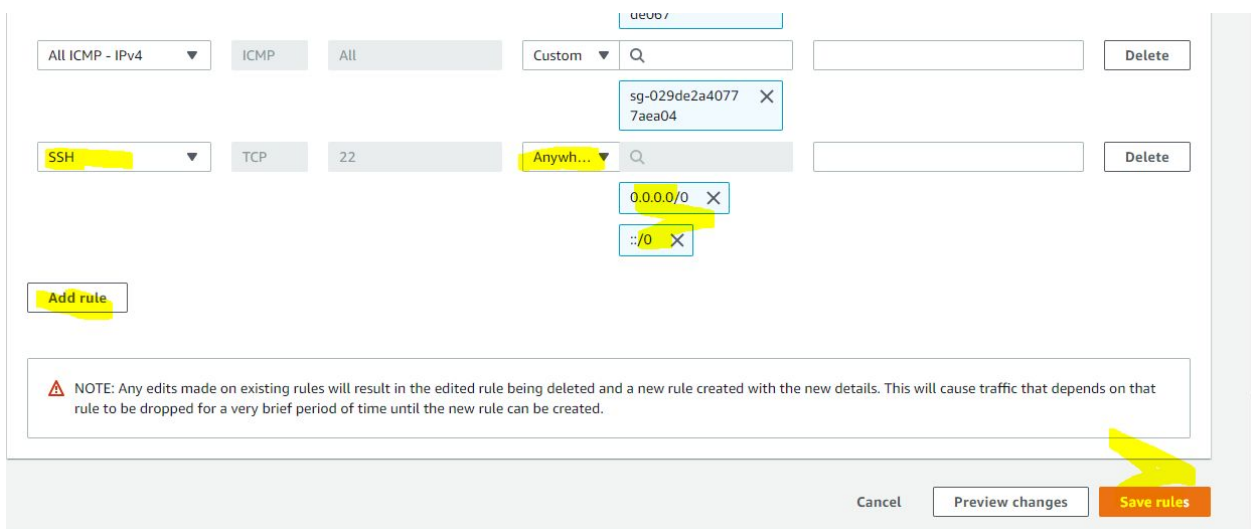
- a. Under the type field of the newly added row select **SSH**



- 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.



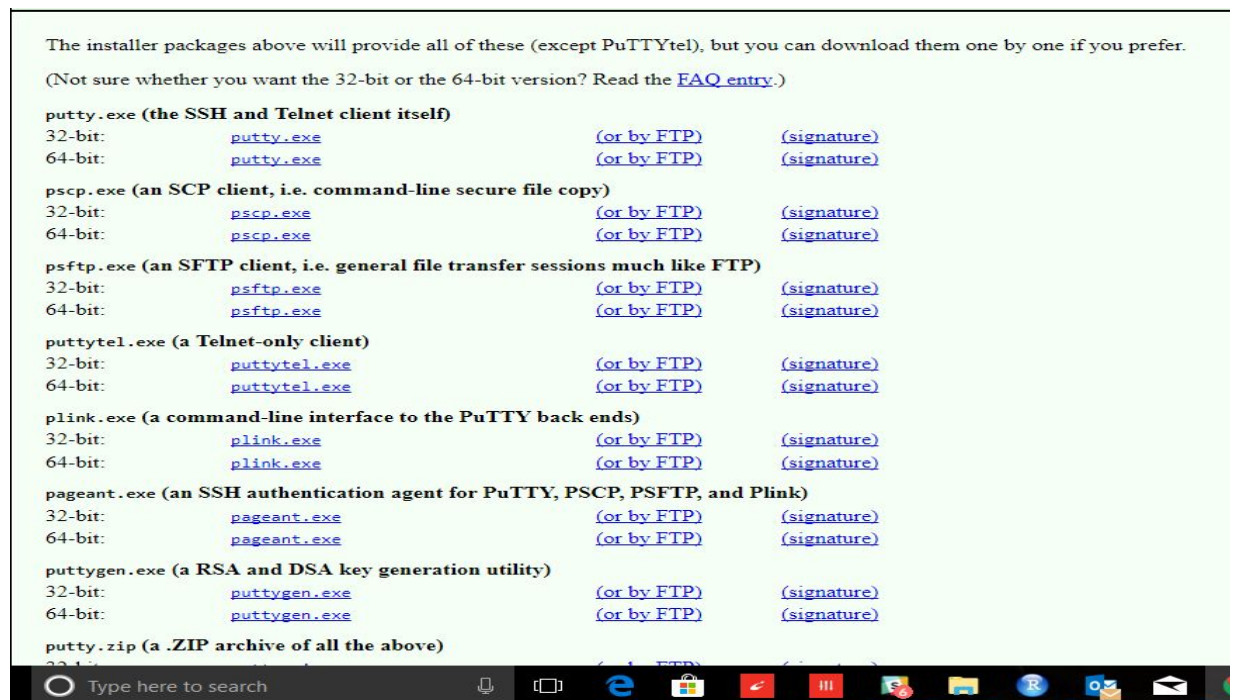
- c. On addition of the rule and choosing the appropriate options as shown below, click on save rule [at the bottom of the screen] to successfully add the rule



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.

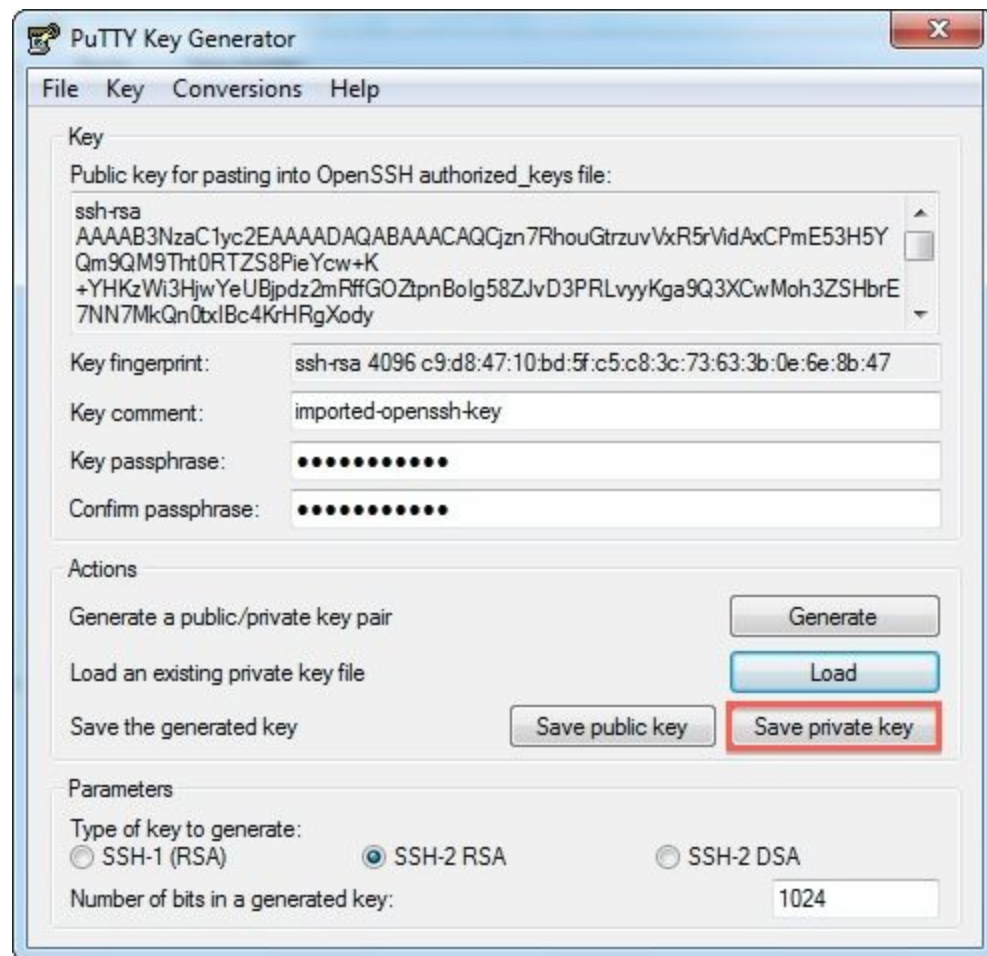


3. Create a private key pair for the downloaded .pem file
 - a. Launch Puttygen
 - b. Click on "Load"



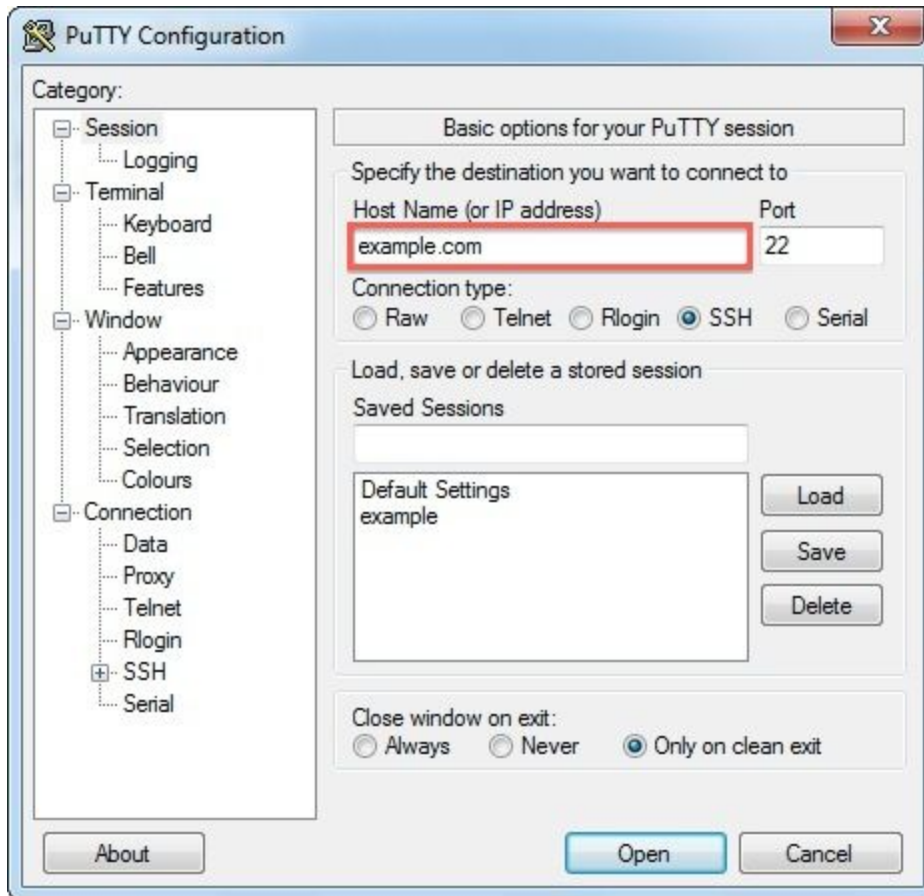
- c. Locate the .pem file you downloaded on your computer

- 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
- Launch PuTTY
 - In the category section on the left side, click on session.
 - 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:

Cluster: Demo-Cluster **Waiting** Cluster ready after last step completed.

Summary Application history Monitoring Hardware Configurations Events Steps Bootstrap actions

Connections: Hue [\[link\]](#), Spark History Server [\[link\]](#), Resource Manager [\[link\]](#) ... (View All)

Master public DNS: [ec2-54-242-43-36.compute-1.amazonaws.com](#) SSH

History service: Spark history server UI [\[link\]](#) (SSH tunneling not required)

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

Summary

ID: j-292NHQV26GMPT
Creation date: 2020-02-01 12:25 (UTC+5:30)
Elapsed time: 13 minutes
After last step Cluster waits completes:
Termination On [Change](#) protection:

Configuration details

Release label: emr-5.29.0
Hadoop distribution: Amazon 2.8.5
Applications: Hive 2.3.6, Pig 0.17.0, Hue 4.4.0, Spark 2.4.4
Log URI: [s3://aws-logs-177300670946-us-east-1/elasticmapreduce/](#) [\[link\]](#)
EMRFS consistent view: Disabled
Custom AMI ID: --

Network and hardware

Availability zone: us-east-1c
Subnet ID: [subnet-1a2a0946](#) [\[link\]](#)
Master: **Running** 1 m4.large

Security and access

Key name: Upgrad-demo-keypair
EC2 instance profile: EMR_EC2_DefaultRole
EMR role: EMR_DefaultRole

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.

Cluster: Demo-Cluster **Waiting** Cluster ready after last step completed.

SSH

Connect to the Master Node Using SSH

You can connect to the Amazon EMR master node using SSH to run interactive queries, examine log files, submit Linux commands, and so on. [Learn more](#) [\[link\]](#).

Windows

Mac / Linux

1. Download PuTTY.exe to your computer from:
[http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html](#) [\[link\]](#)

2. Start PuTTY.

3. In the Category list, click Session.

4. In the Host Name field, type [hadoop@ec2-54-242-43-36.compute-1.amazonaws.com](#)

5. In the Category list, expand Connection > SSH, and then click Auth.

6. For Private key file for authentication, click Browse and select the private key file (Upgrad-demo-keypair.ppk) used to launch the cluster.

7. Click Open.

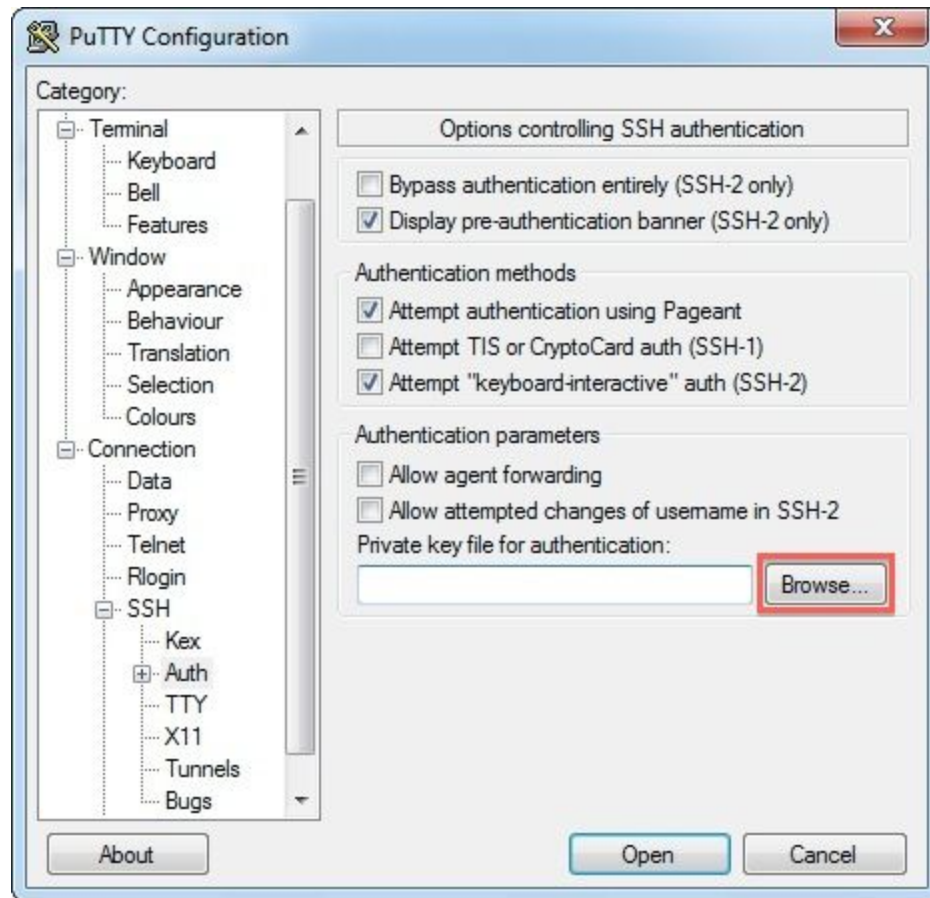
8. Click Yes to dismiss the security alert.

Close

Availability zone: us-east-1c

Key name: Upgrad-demo-keypair

- 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 .

Clone

Terminate

AWS CLI export

Cluster: Demo-Cluster Waiting Cluster ready after last step completed.

SSH

Connect to the Master Node Using SSH

You can connect to the Amazon EMR master node using SSH to run interactive queries, examine log files, submit Linux commands, and so on.
[Learn more](#)

Windows

Mac / Linux

1. Open a terminal window. On Mac OS X, choose Applications > Utilities > Terminal. On other Linux distributions, terminal is typically found at Applications > Accessories > Terminal.

2. To establish a connection to the master node, type the following command. Replace ~/Upgrad-demo-keypair.pem with the location and filename of the private key file (.pem) used to launch the cluster.

ssh -i ~/Upgrad-demo-keypair.pem hadoop@ec2-54-242-43-36.compute-1.amazonaws.com

3. Type yes to dismiss the security warning.

Close

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