



Accessing Spark History Server

This document contains the steps to access Spark History Server in AWS EMR.

1. Go to the homepage for your EMR cluster. Then under the cluster information page, click on the **security groups of the master node**.

Security and access

Key name: RHEL

EC2 instance profile: EMR_EC2_DefaultRole

EMR role: EMR_DefaultRole

Visible to all users: All Change

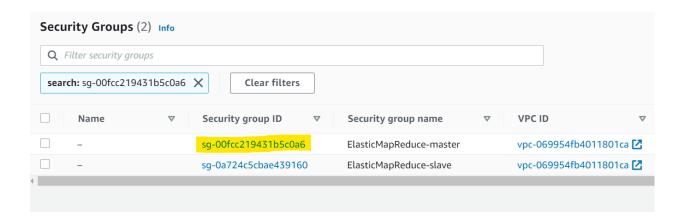
Security groups for Master: sg-00fcc219431b5c0a6 (ElasticMapReduce-

master)

Security groups for Core & sq-0a724c5cbae439160 [2] (ElasticMapReduce-

Task: slave)

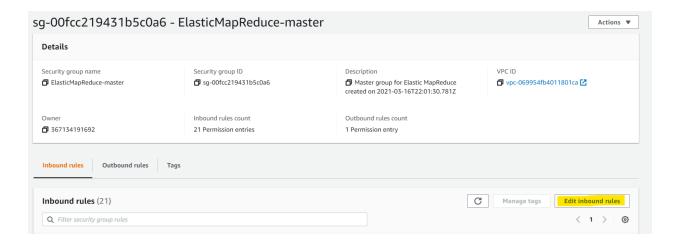
Click on the 'security group' and you will land on a similar page. Here, click on the security group of the Elastic MapReduce-master node as highlighted in the image below.



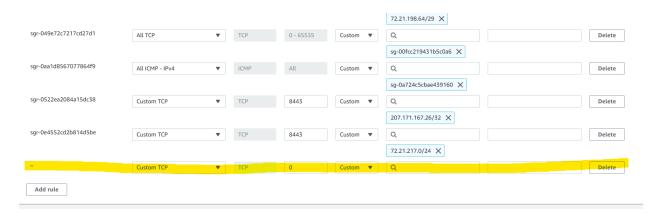




3. Clicking on the security group will land you on the corresponding security information page. Click on 'Edit inbound rules' to add a new rule.



4. Here, you need to add a new role. Click on **Add rule** towards the bottom of all the rules. Clicking on the 'add rule' will add a new row as shown in the figure below.



5. Now, here you need to set the rule for **Custom TCP**, and then for the port, you need to type **18080**. Next, you need to keep the Source as **'My IP'** for this rule.

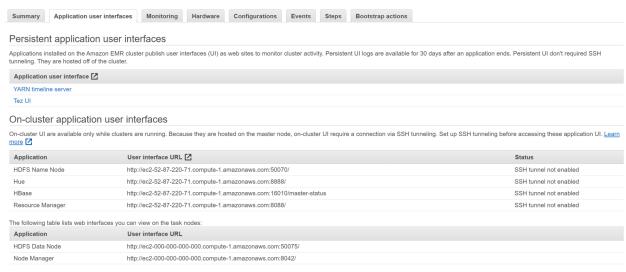


6. Now you can click on **Save rules** and then **go back to the EMR cluster homepage**.





7. Go to the Application User Interfaces tab for your AWS EMR cluster.



High-level application history

8. Copy the **User interface URL** corresponding to **Spark History Server** under the **On-cluster application user interfaces**.

On-cluster application user interfaces

On-cluster UI are available only while clusters are running. Because they are hosted on the master node, on-cluster UI

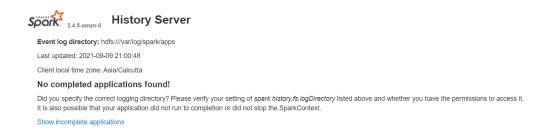
Application	User interface URL [2]
HDFS Name Node	http://ec2-34-229-60-64.compute-1.amazonaws.com:50070/
Hue	http://ec2-34-229-60-64.compute-1.amazonaws.com:8888/
HBase	http://ec2-34-229-60-64.compute-1.amazonaws.com:16010/master-status
Spark History Server	http://ec2-34-229-60-64.compute-1.amazonaws.com:18080/
JupyterHub	https://ec2-34-229-60-64.compute-1.amazonaws.com:9443/
Resource Manager	http://ec2-34-229-60-64.compute-1.amazonaws.com:8088/

The following table lists web interfered you can view on the took nedes:

9. Now you can copy the URL on another tab of your web browser and click on **Enter**. You will see the following page appear.







10. Here, you need to click on **Show incomplete applications**. All applications which haven't been finished will show here. Your job might also be present as a completed job in which case you don't need to find it in incomplete applications.



Showing 1 to 2 of 2 entries

Back to completed applications

11. From here, you can check the applications and their DAGs and access any job on the history server.