

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 & Task: **sg-0a724c5cb4e439160** [↗](#) (ElasticMapReduce-slave)

2. 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.

Security Groups (2) Info				
<input type="text" value="Filter security groups"/>				
<input type="text" value="search: sg-00fcc219431b5c0a6"/>		<input type="button" value="Clear filters"/>		
<input type="checkbox"/>	Name ▾	Security group ID ▾	Security group name ▾	VPC ID ▾
<input type="checkbox"/>	-	sg-00fcc219431b5c0a6	ElasticMapReduce-master	vpc-069954fb4011801ca ↗
<input type="checkbox"/>	-	sg-0a724c5cb4e439160	ElasticMapReduce-slave	vpc-069954fb4011801ca ↗

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

sg-00fcc219431b5c0a6 - ElasticMapReduce-master Actions ▾

Details

Security group name ElasticMapReduce-master	Security group ID sg-00fcc219431b5c0a6	Description Master group for Elastic MapReduce created on 2021-03-16T22:01:30.781Z	VPC ID vpc-069954fb4011801ca
Owner 367134191692	Inbound rules count 21 Permission entries	Outbound rules count 1 Permission entry	

Inbound rules

Outbound rules

Tags

Inbound rules (21)

Manage tags
Edit inbound rules

- Here, you need to add a new rule. 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.

sgr-049e72c7217cd27d1	All TCP	TCP	0 - 65535	Custom	72.21.198.64/29		Delete
sgr-0aa1d8567077864f9	All ICMP - IPv4	ICMP	All	Custom	sg-00fcc219431b5c0a6		Delete
sgr-0522ea2084a15dc38	Custom TCP	TCP	8443	Custom	sg-0a724c5cbac439160		Delete
sgr-0e4552cd2b814d5be	Custom TCP	TCP	8443	Custom	207.171.167.26/32		Delete
-	Custom TCP	TCP	0	Custom	72.21.217.0/24		Delete

Add rule

- 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.

Custom TCP	TCP	18080	My IP	223.182.167.217/32		Delete
------------	-----	-------	-------	--------------------	--	--------

- 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.

Summary
Application user interfaces
Monitoring
Hardware
Configurations
Events
Steps
Bootstrap actions

Persistent application user interfaces

Applications installed on the Amazon EMR cluster publish user interfaces (UI) as web sites to monitor cluster activity. Persistent UI logs are available for 30 days after an application ends. Persistent UI don't required SSH tunneling. They are hosted off of the cluster.

Application user interface
YARN timeline server
Tez UI

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 require a connection via SSH tunneling. Set up SSH tunneling before accessing these application UI. [Learn more](#)

Application	User interface URL	Status
HDFS Name Node	http://ec2-52-87-220-71.compute-1.amazonaws.com:50070/	SSH tunnel not enabled
Hue	http://ec2-52-87-220-71.compute-1.amazonaws.com:8888/	SSH tunnel not enabled
HBase	http://ec2-52-87-220-71.compute-1.amazonaws.com:16010/master-status	SSH tunnel not enabled
Resource Manager	http://ec2-52-87-220-71.compute-1.amazonaws.com:8088/	SSH tunnel not enabled

The following table lists web interfaces you can view on the task nodes:

Application	User interface URL
HDFS Data Node	http://ec2-000-000-000-000.compute-1.amazonaws.com:50075/
Node Manager	http://ec2-000-000-000-000.compute-1.amazonaws.com:8042/

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
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 interfaces you can view on the task nodes:

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.



History Server

Event log directory: hdfs:///var/log/spark/apps

Last updated: 2021-09-09 21:00:48

Client local time zone: Asia/Calcutta

No completed applications found!

Did you specify the correct logging directory? Please verify your setting of `spark.history.fs.logDirectory` listed above and whether you have the permissions to access it. It is also possible that your application did not run to completion or did not stop the SparkContext.

[Show incomplete applications](#)

- 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.



History Server

Event log directory: hdfs:///var/log/spark/apps

Last updated: 2021-09-09 21:03:29

Client local time zone: Asia/Calcutta

Search:

App ID	App Name	Started	Spark User	Last Updated	Event Log
application_1631192920774_0002	livy-session-1	2021-09-09 20:35:07	livy	2021-09-09 21:03:29	Download
application_1631192920774_0001	livy-session-0	2021-09-09 19:23:58	livy	2021-09-09 21:03:29	Download

Showing 1 to 2 of 2 entries

[Back to completed applications](#)

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