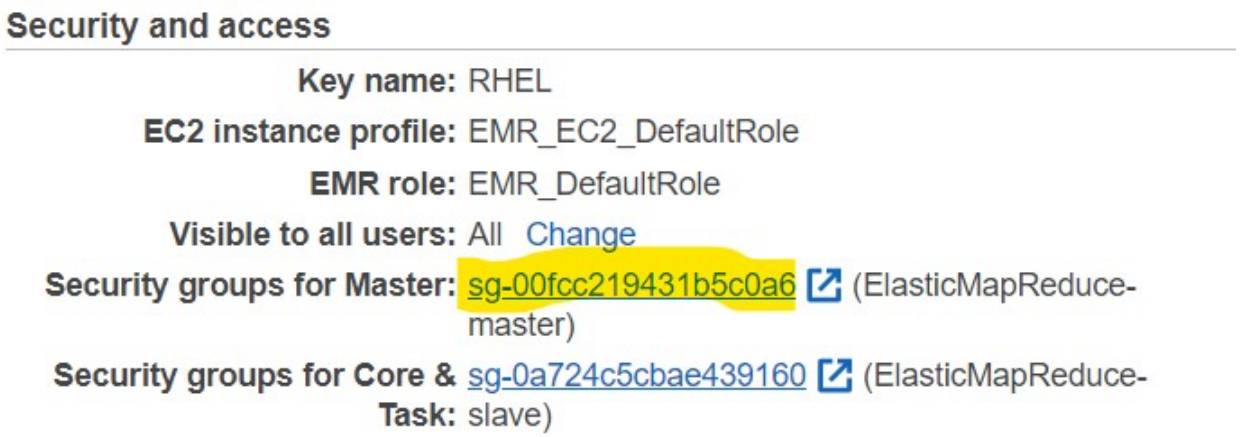
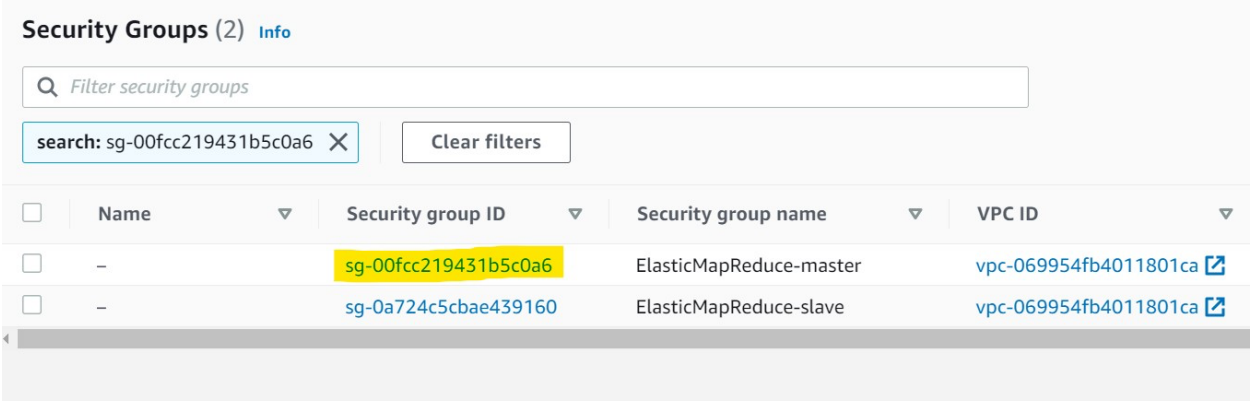
Accessing Spark History Server

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

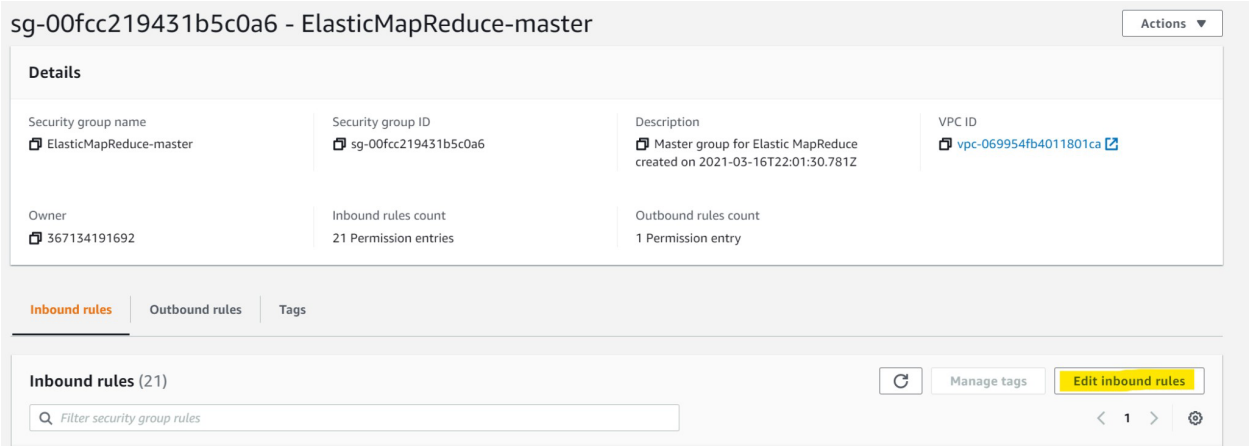
1. Go to the homepage for your EMR cluster. Then under the cluster information page, click on the **security groups of the master node**.



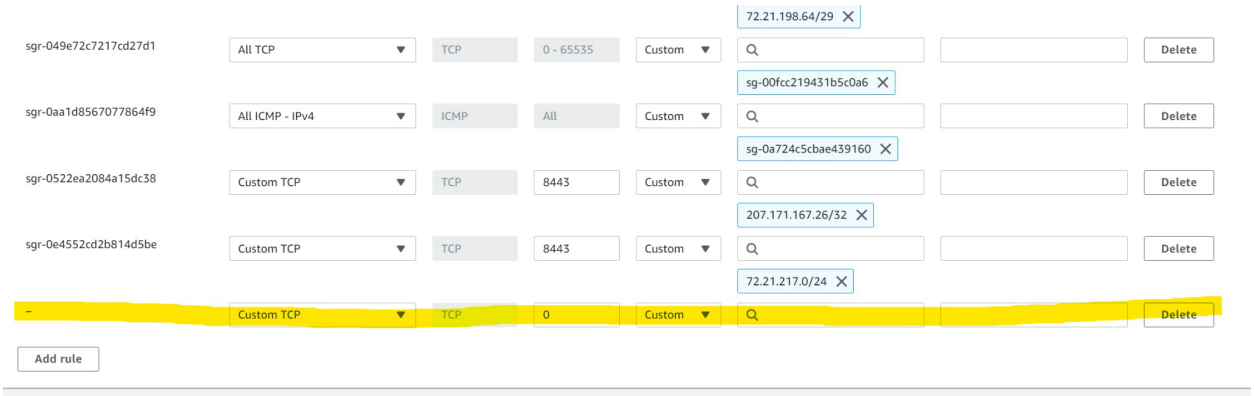
2. Click on the 'security group' and you will land on a similar page. Here, click on the security group of the **Elastic MapReduce-master node** as highlighted in the image below.



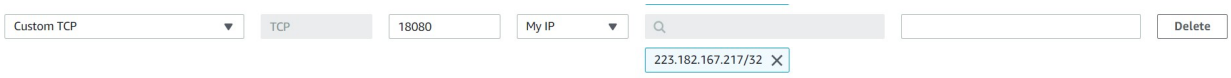
3. Clicking on the security group will land you on the corresponding security information page. Click on 'Edit inbound rules' to add a new rule.



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.

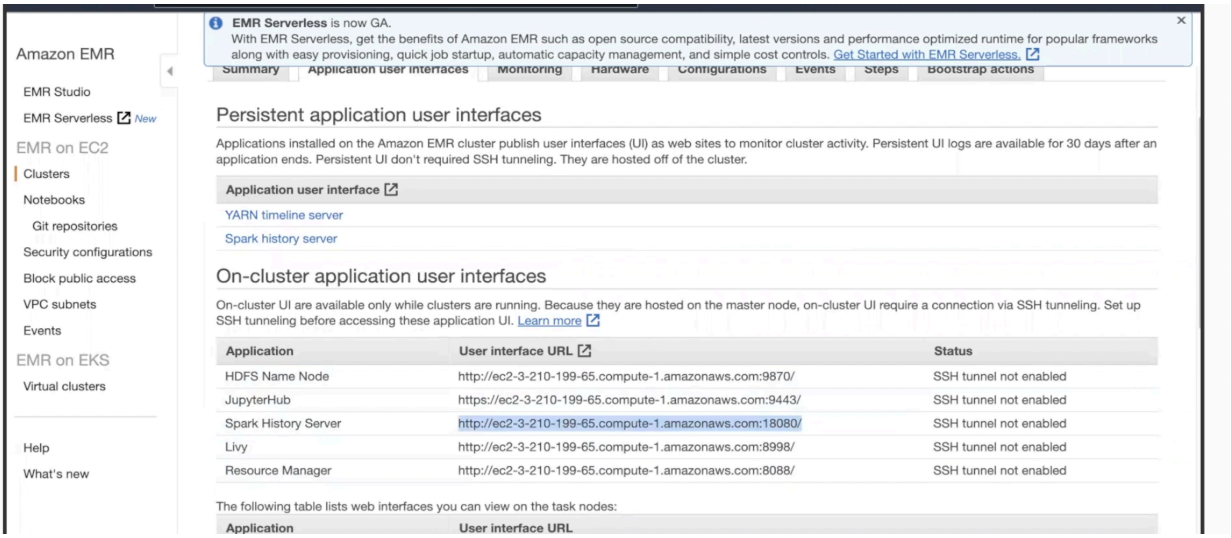


5. Now, here you need to set the rule for **Custom TCP,** and then for the port, you need to type **18080**. Next, you need to keep the Source as **'My IP'** for this rule.

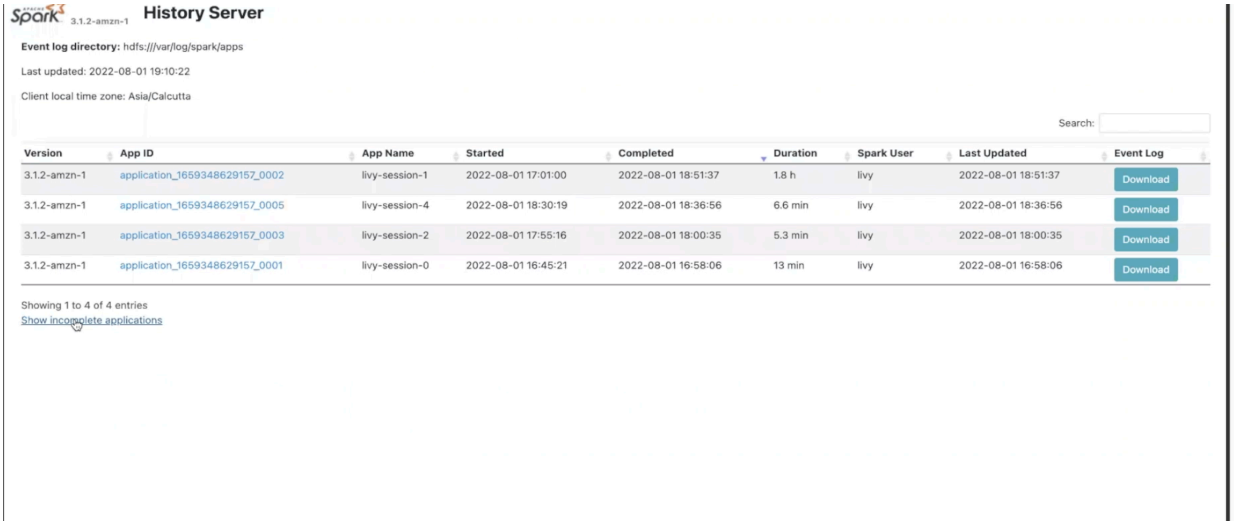


6. Now you can click on **Save rules** and then **go back to the EMR cluster homepage**.
7. Go to the **Application User Interfaces** tab for your AWS EMR cluster.

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



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.

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