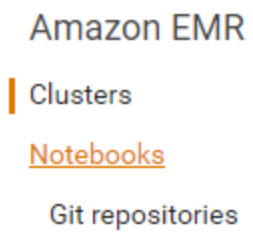
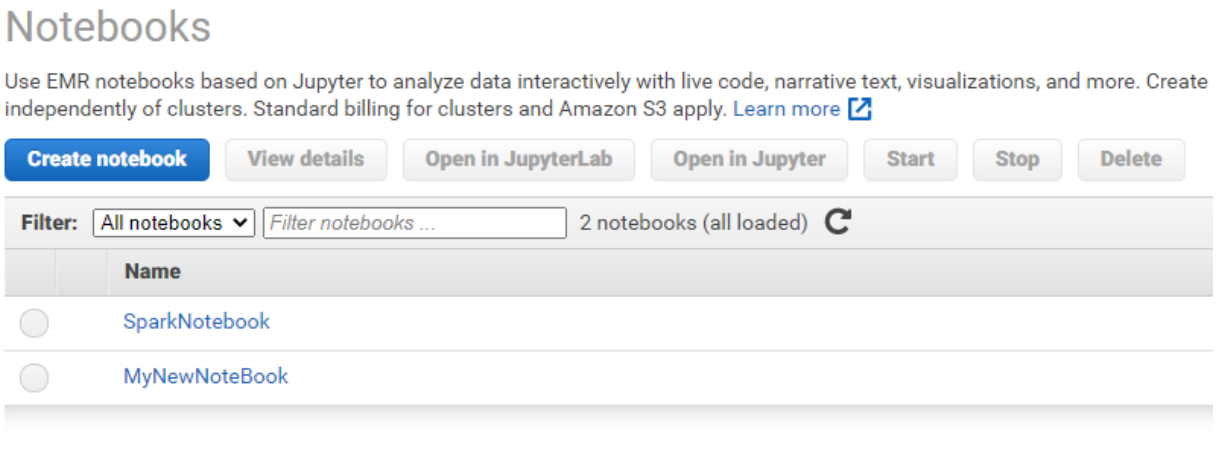
Creating a Jupyter Notebook on an EMR Cluster

Once you have created and started your EMR cluster, you can then create a new Jupyter Notebook wherein you can write your Spark jobs. Note that these notebooks are persisted on S3 even after you terminate the EMR cluster, so you don't have to worry about creating a Jupyter Notebook again from scratch. You can follow the steps below to create a Jupyter Notebook:

Step 1: First, you need to go to the 'Notebooks' link in the left navigation pane under Amazon EMR.

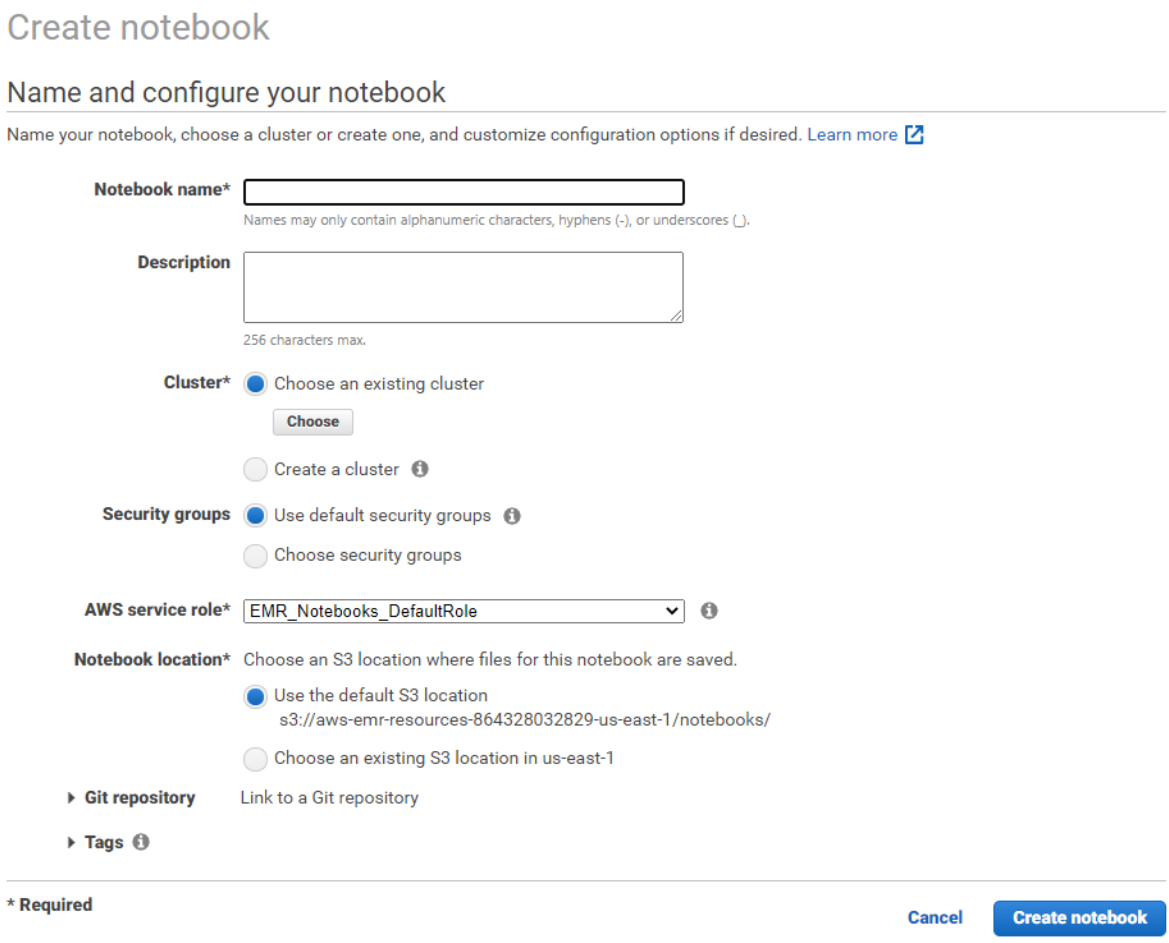


Step 2: On clicking the link, your screen will appear as shown below. Now, to create a new Jupyter Notebook, simply click on the 'Create notebook' button.

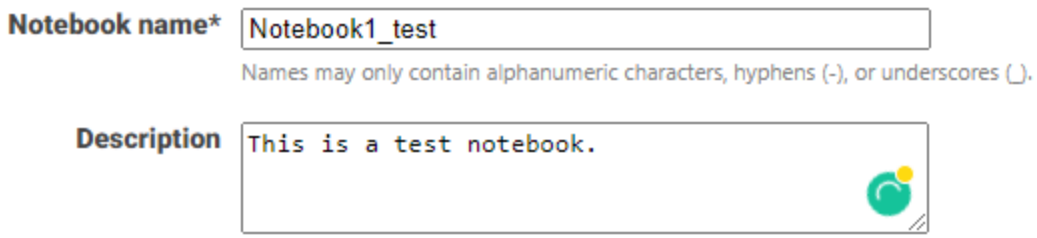


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

Once you click on the button, the following page will open on your screen.

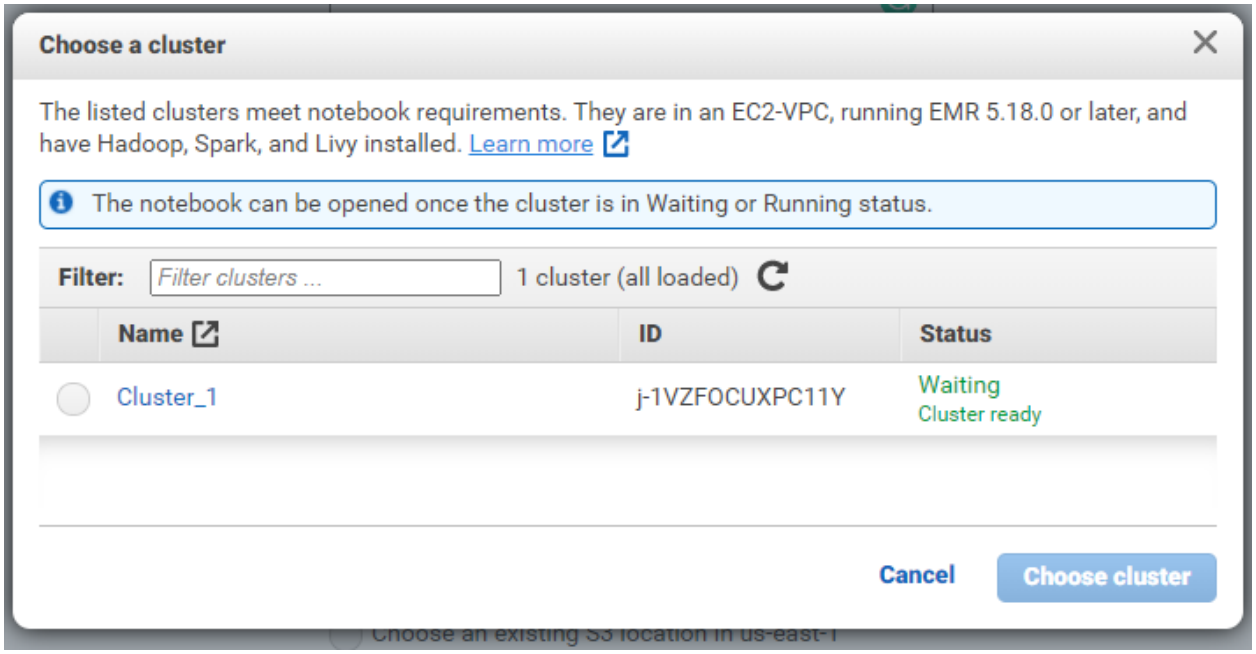


Step 3: Here, you can write a 'Notebook name' for your Jupyter Notebook. Under 'Description', you can write a few lines to describe the notebook that you are creating.



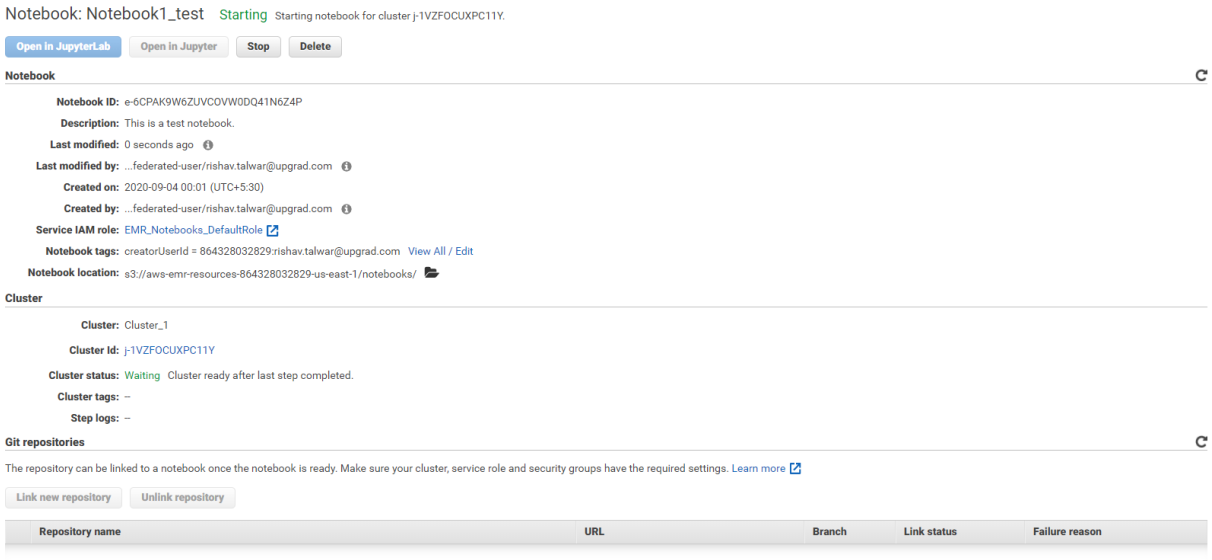
© Copyright. upGrad Education Pvt. Ltd. All rights reserved

Step 4: Now, under ‘Cluster’, click the radio button for ‘Choose an existing cluster’ and then click on the ‘Choose’ button. As soon as you do that, a pop-up will appear, showing you the list of all currently running EMR clusters.



Step 5: In this step, just select the EMR cluster that you have created and then click on the ‘Choose cluster’ button at the bottom right corner of the page.

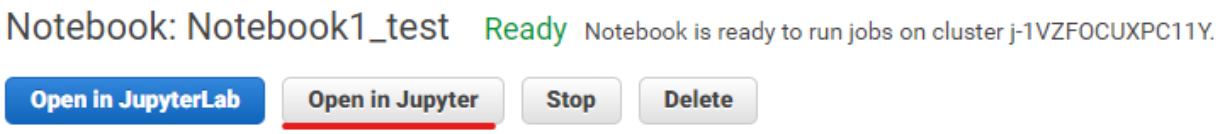
Step 6: After this step, you can keep the other settings as default and click on the ‘Create notebook’ button at the bottom right of the page. As soon as you do this, the following page will open on your screen.



© Copyright. upGrad Education Pvt. Ltd. All rights reserved

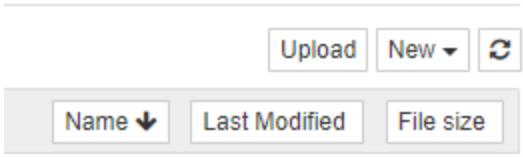
This means the Notebook has been created and is now starting. Shortly after, the notebook will show at the top with status ‘Ready’.

Step 7: Finally, after the status of the Jupyter Notebook shows ‘Ready’, you can launch the notebook. To do this, simply click on the ‘**Open in Jupyter**’ button.



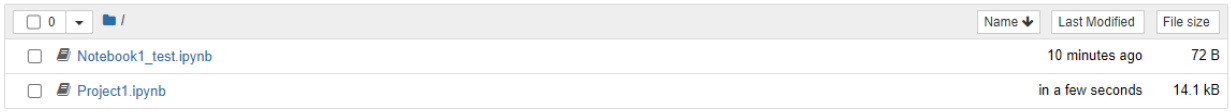
This will open the familiar Jupyter UI. Here, you can start creating your own Jupyter Notebooks as you have done in the previous module on Spark.

You can also upload any Jupyter Notebook that you want easily. For this, simply click on the ‘Upload’ button to the top right on the Jupyter UI.



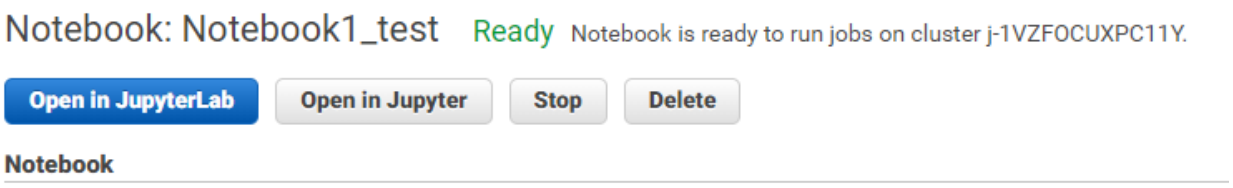
A Windows ‘Open’ dialogue box will appear. From here, you can simply find the location of your Jupyter Notebook and then click on the ‘Open’ button. After this, you will see that the name has been appended to the list of notebooks. You now need to click on the ‘Upload’ button next to your notebook file.

This will upload the Jupyter Notebook file to your EMR notebooks folder.

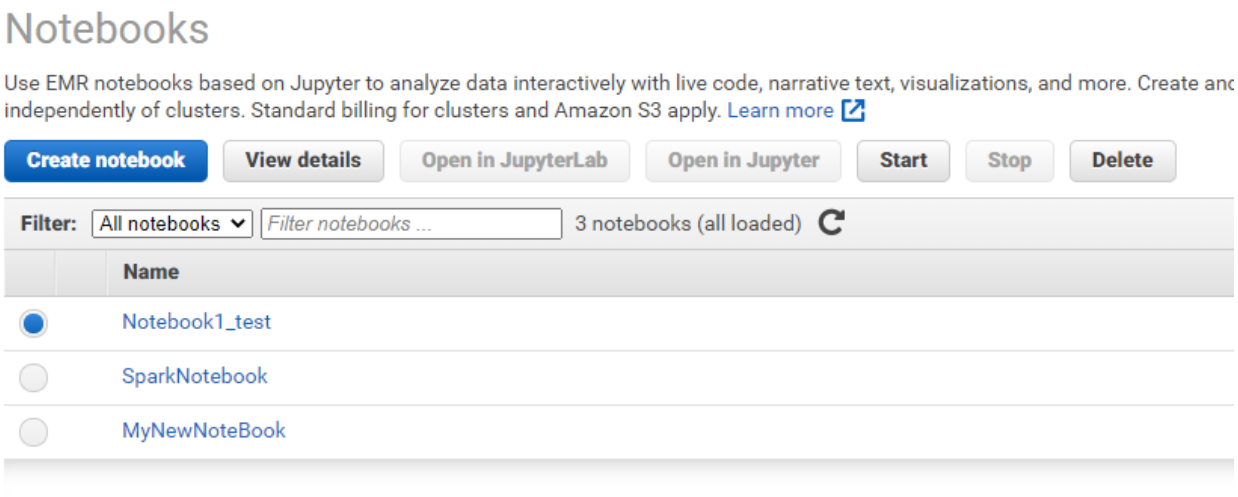


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

You can also stop the Notebook whenever you want by simply clicking on the ‘Stop’ button in the Notebook UI, as shown below:



And, if you need to resume your Jupyter Notebook, then you can do so by going to the notebooks list, selecting your notebook and clicking on the ‘Start’ button.



© Copyright. upGrad Education Pvt. Ltd. All rights reserved