

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.

Amazon EMR

Clusters

Notebooks


Git repositories

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.

Notebooks

Use EMR notebooks based on Jupyter to analyze data interactively with live code, narrative text, visualizations, and more. Create independently of clusters. Standard billing for clusters and Amazon S3 apply. [Learn more](#)

[Create notebook](#) [View details](#) [Open in JupyterLab](#) [Open in Jupyter](#) [Start](#) [Stop](#) [Delete](#)

Filter: All notebooks 2 notebooks (all loaded) 

	Name
<input type="radio"/>	SparkNotebook
<input type="radio"/>	MyNewNoteBook

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

Create notebook

Name and configure your notebook

Name your notebook, choose a cluster or create one, and customize configuration options if desired. [Learn more](#)

Notebook name*

Names may only contain alphanumeric characters, hyphens (-), or underscores (_).

Description

256 characters max.

Cluster*

☒ Choose an existing cluster

Choose

☐ Create a cluster ⓘ

Security groups

☒ Use default security groups ⓘ

☐ Choose security groups

AWS service role*

EMR_Notebooks_DefaultRole ▼ ⓘ

Notebook location*

Choose an S3 location where files for this notebook are saved.

☒ Use the default S3 location
s3://aws-emr-resources-864328032829-us-east-1/notebooks/

☐ Choose an existing S3 location in us-east-1

▶ Git repository

Link to a Git repository

▶ Tags ⓘ

* Required

Cancel

Create notebook

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.

Notebook name*

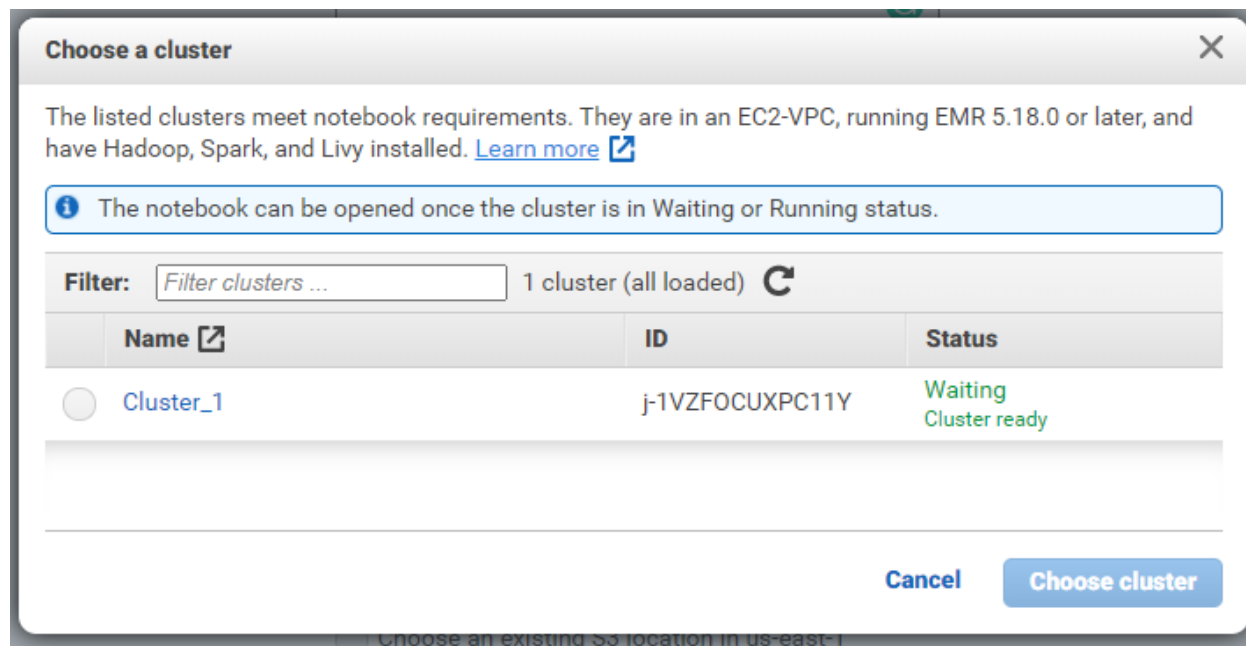
Notebook1_test

Names may only contain alphanumeric characters, hyphens (-), or underscores (_).

Description

This is a test notebook.

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.

Notebook: Notebook1_test Starting Starting notebook for cluster j-1VZF0CUXPC11Y.

[Open in JupyterLab](#) [Open in Jupyter](#) [Stop](#) [Delete](#)

Notebook

Notebook ID: e-6CPAK9W6ZUVC0VW0DQ41N6Z4P
Description: This is a test notebook.
Last modified: 0 seconds ago ⓘ
Last modified by: ...federated-user/rishav.talwar@upgrad.com ⓘ
Created on: 2020-09-04 00:01 (UTC+5:30)
Created by: ...federated-user/rishav.talwar@upgrad.com ⓘ
Service IAM role: EMR_Notebooks_DefaultRole ⓘ
Notebook tags: creatorUserId = 864328032829:rishav.talwar@upgrad.com [View All / Edit](#)
Notebook location: s3://aws-emr-resources-864328032829-us-east-1/notebooks/ ⓘ

Cluster

Cluster: Cluster_1
Cluster id: j-1VZF0CUXPC11Y
Cluster status: Waiting Cluster ready after last step completed.
Cluster tags: --
Step logs: --

Git repositories

The repository can be linked to a notebook once the notebook is ready. Make sure your cluster, service role and security groups have the required settings. [Learn more](#) ⓘ

[Link new repository](#) [Unlink repository](#)

Repository name	URL	Branch	Link status	Failure reason
-----------------	-----	--------	-------------	----------------

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

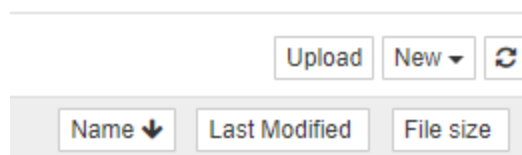
Step 7: Finally, after the status of the Jupyter Notebook shows 'Ready', you can launch the notebook. To do this, simply click on the '**Open in Jupyter**' button.

Notebook: Notebook1_test Ready Notebook is ready to run jobs on cluster j-1VZFOCUXPC11Y.



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.

			Name	Last Modified	File size
<input type="checkbox"/>		Notebook1_test.ipynb		10 minutes ago	72 B
<input type="checkbox"/>		Project1.ipynb		in a few seconds	14.1 kB

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

Notebook: Notebook1_test **Ready** Notebook is ready to run jobs on cluster j-1VZF0CUXPC11Y.

Open in JupyterLab

Open in Jupyter

Stop

Delete

Notebook

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.

Notebooks

Use EMR notebooks based on Jupyter to analyze data interactively with live code, narrative text, visualizations, and more. Create and independently of clusters. Standard billing for clusters and Amazon S3 apply. [Learn more](#)

Create notebook

View details


Open in JupyterLab

Open in Jupyter

Start

Stop

Delete

Filter: All notebooks 3 notebooks (all loaded) 

Name



Notebook1_test



SparkNotebook



MyNewNoteBook