Apache Spark notebooks

Task 1. Create a notebook

1.   On the left-hand menu of your Databricks workspace, select Home.
2.   Right-click on your home folder.
3.   Select Create.
4.   Select Notebook.
5.   Name your notebook First Notebook.
6.   Set the Language to Python.
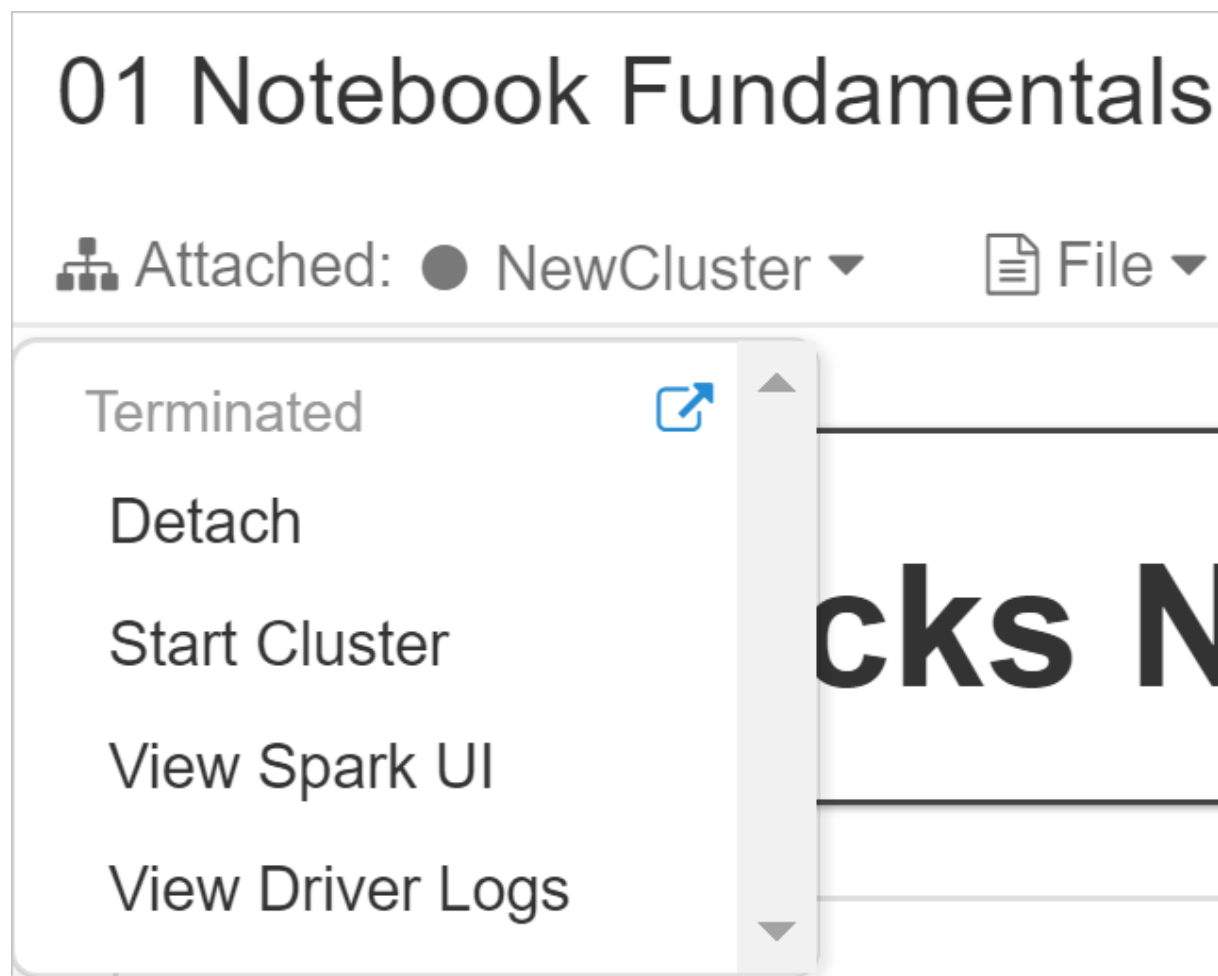7.   Select the cluster to which to attach this notebook.

Note

This option displays only when a cluster is currently running. You can still create your notebook and attach it to a cluster later.

8.   Select Create.

Task 2. Attach and detach your notebook

Now that you've created your notebook, let's use it to run some code. To use your notebook to run a code, you must attach it to a cluster. You can also detach your notebook from a cluster and attach it to another depending upon your organization's requirements.

If your notebook is attached to a cluster, you can:

1. Detach your notebook from the cluster
2. Restart the cluster
3. Attach to another cluster
4. Open the Spark UI
5. View the log files of the driver

Task 3. Clone the Databricks archive

1. In the Azure portal, navigate to your deployed Azure Databricks workspace and select Launch Workspace.
2. In the left pane, select Workspace > Users, and select your username (the entry with the house icon).
3. In the blade that appears, select the arrow next to your name, and select Import.
4. In the Import Notebooks dialog box, select the URL and paste in the following URL:
5. https://github.com/MicrosoftDocs/mslearn-azure-databricks-notebooks-fundamentals/blob/master/DBC/01-notebook-fundamentals.dbc?raw=true
6. Select Import.

Task 4. Complete the following notebooks

1. 01 Notebook Fundamentals - This notebook illustrates the fundamentals of a Databricks notebook.
2. Why Apache Spark? - In this notebook, you can practice some of the use cases for Databricks notebook.