

# Azure Databricks lab

---

In this lab, you will use notebooks within an Azure Databricks workspace that cover the following topics:

- Notebook fundamentals
- Introduction to Azure Databricks
- Using Apache Spark
- Reading and writing data
- Basic ETL (extract-transform-load)
- Advanced ETL
- Delta Lake
- Stream processing

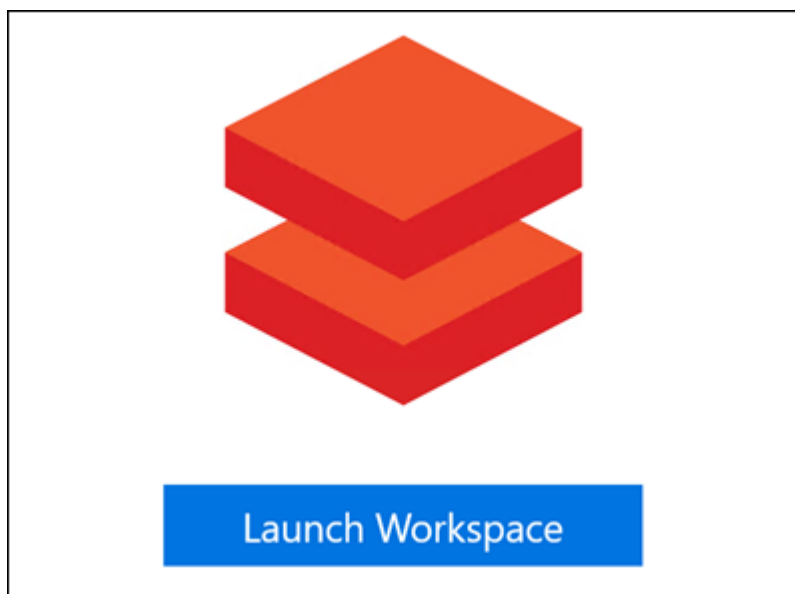
The topics have been ordered in such a way that the concepts become more advanced as you progress. The typical time to complete all the notebooks is around 2 hours.

Some of the later notebooks include challenges on the bottom that you must complete by applying concepts learned in earlier steps. If you get stuck, you can reference the notebooks within the **Solutions** sub-folder to find the code used to complete the challenges.

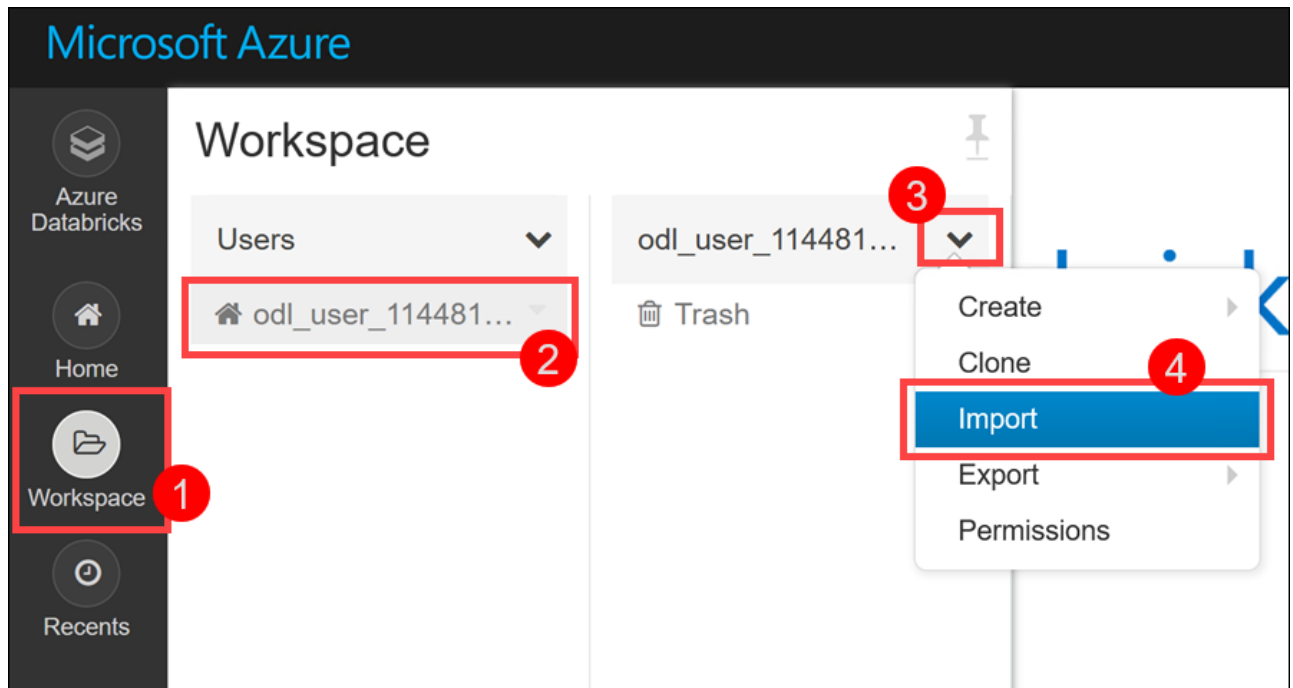
## Lab setup

Complete the following steps within your provided lab environment:

1. Sign in to the Azure portal (<https://portal.azure.com>) with the lab credentials.
2. Open the lab resource group, select the Azure Databricks workspace, then select **Launch Workspace**. The Azure Databricks workspace will open in a new browser tab and automatically authenticate using your Azure AD account.

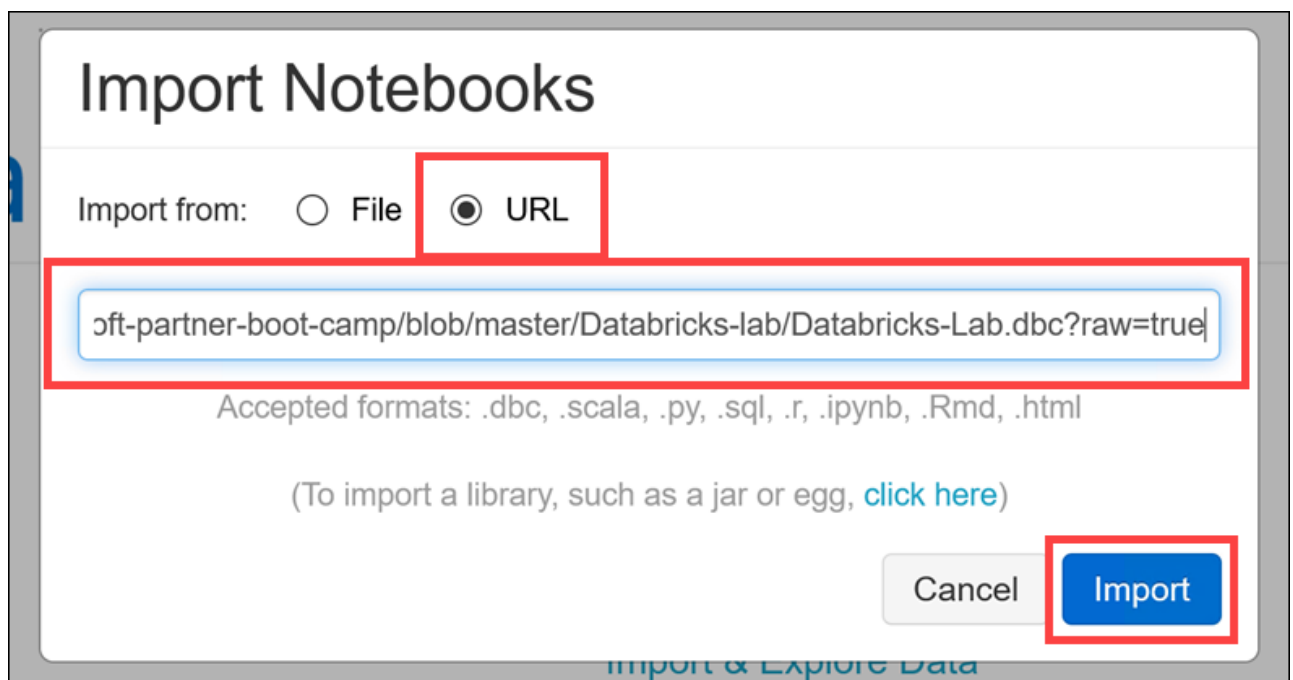


3. Select **Workspace** in the left-hand menu, then select your user. Select the down arrow next to your username, then select **Import** in the context menu.

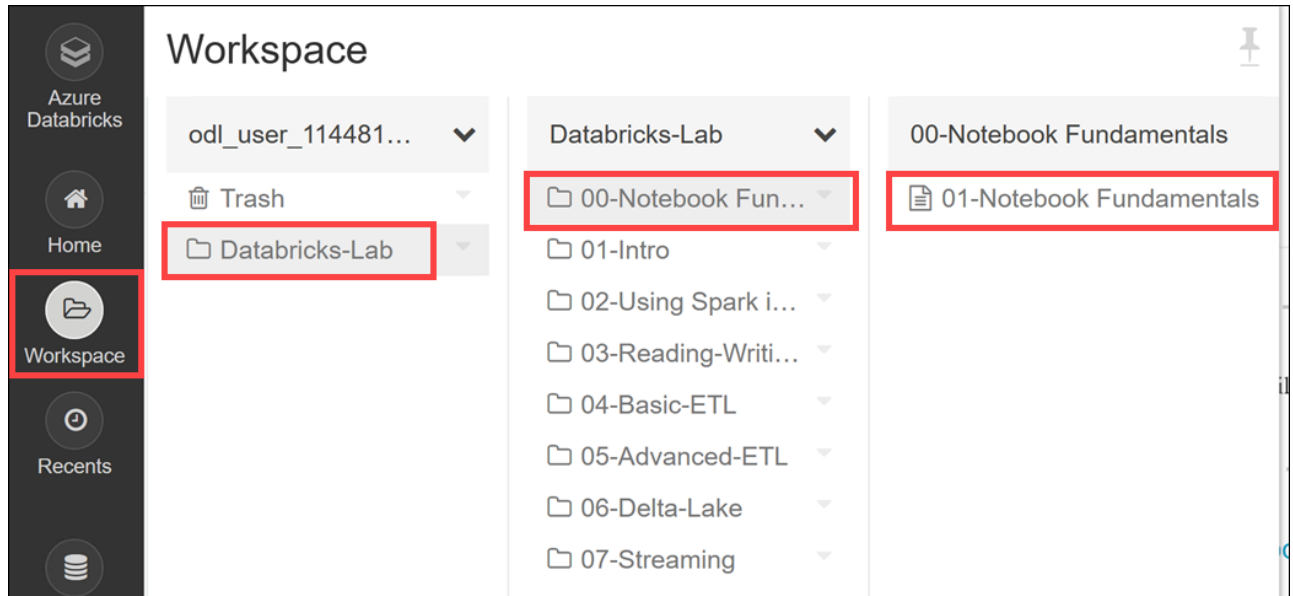


4. Within the *Import Notebooks* dialog, select **URL** and paste the following into the textbox:

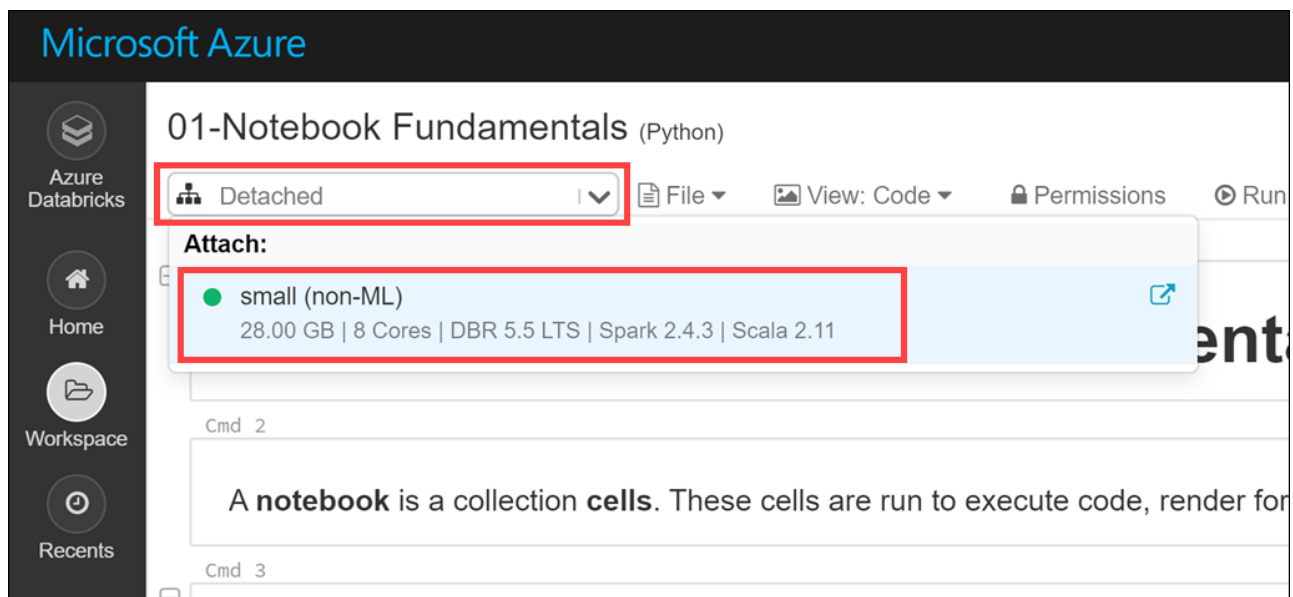
<https://github.com/solliancenet/microsoft-partner-boot-camp/blob/master/Databricks-lab/Databricks-Lab.dbc?raw=true>. Select **Import**.



5. After importing the notebooks, select the **Databricks-Lab** folder, the **00-Notebook Fundamentals** sub-folder, then select the **01-Notebook Fundamentals** notebook.



6. Before you can run the notebook, you first need to attach your cluster. To do this, select **Detached** within the notebook's task bar, then select your cluster. If the cluster is stopped, you will see an option to start it after you attach it to the notebook.



7. Carefully read the instructions throughout each notebook to learn the concepts and what you are executing. There are several places where you must wait for something to complete before continuing, and there are some places where you need to complete challenges. To execute a cell, press **Ctrl+Enter**, or to execute a cell and move to the next one automatically, press **Shift+Enter**.
8. At the bottom of each notebook, you will find a section titled **Next Steps**. This section includes a link to the next notebook, end-to-end. This way, you will not need to navigate through the notebooks in your workspace.

## 01-Notebook Fundamentals (Python)

Detached

Clear Results

Clear State

Clear State & Results

Clear State & Run All

clear the underlying state.

- Clear -> Clear Results (removes the displayed output for all cells)
- Clear -> Clear State (removes all cell states)

You typically do this when you want to cleanly re-run a notebook you have been working on and eliminate any accidental changes to the state that may have occurred while you were authoring the notebook.

Read more about execution context [here](#).

Cmd 52

### Next Steps

Start the next lesson, [Why Databricks and Apache Spark?](#).

Shift+Enter to run [shortcuts](#)