

# Work with Data

---

Although it's fairly common to work with data on their local file system, in an enterprise environment it can be more effective to store the data in a central location where multiple data scientists and machine learning engineers can access it.

In this exercise, you'll explore *datastores* and *datasets*, which are the primary objects used to abstract data access in Azure Machine Learning.

## Before You start

If you have not already done so, complete the [Create an Azure Machine Learning Workspace](#) exercise to create an Azure Machine Learning workspace and compute instance, and clone the notebooks required for this exercise.

## Open Jupyter

While you can use the **Notebooks** page in Azure Machine Learning studio to run notebooks, it's often more productive to use a more fully-featured notebook development environment like *Jupyter*.

1. In [Azure Machine Learning studio](#), view the **Compute** page for your workspace; and on the **Compute Instances** tab, start your compute instance if it is not already running.
2. When the compute instance is running, click the **Jupyter** link to open the Jupyter home page in a new browser tab.

## Work with datastores and datasets

In this exercise, the code to work with data is provided in a notebook.

1. In the Jupyter home page, browse to the **Users/mslearn-dp100** folder where you cloned the notebook repository, and open the **Work with Data** notebook.
2. Then read the notes in the notebook, running each code cell in turn.
3. When you have finished running the code in the notebook, on the **File** menu, click **Close and Halt** to close it and shut down its Python kernel. Then close all Jupyter browser tabs.

## Clean-up

If you're finished working with Azure Machine Learning for now, in Azure Machine Learning studio, on the **Compute** page, on the **Compute Instances** tab, select your compute instance and click **Stop** to shut it down. Otherwise, leave it running for the next lab.