Create compute resources

100 XP

3 minutes

After you have created an Azure Machine Learning workspace, you can use it to manage the various assets and resources you need to create machine learning solutions. At its core, Azure Machine Learning is a platform for training and managing machine learning models, for which you need compute on which to run the training process.

Create compute targets

Compute targets are cloud-based resources on which you can run model training and data exploration processes.

In Azure Machine Learning studio, view the **Compute** page (under **Manage**). This is where you manage the compute targets for your data science activities. There are four kinds of compute resource you can create:

- **Compute Instances**: Development workstations that data scientists can use to work with data and models.
- **Compute Clusters**: Scalable clusters of virtual machines for on-demand processing of experiment code.
- **Inference Clusters**: Deployment targets for predictive services that use your trained models.
- **Attached Compute**: Links to existing Azure compute resources, such as Virtual Machines or Azure Databricks clusters.

 Note

Compute instances and clusters are based on standard Azure virtual machine images. For this module, the *Standard_DS11_v2* image is recommended to achieve the optimal balance of cost and performance. If your subscription has a quota that does not include this image, choose an alternative image; but bear in mind that a larger image may incur higher cost and a smaller image may not be sufficient to complete the tasks. Alternatively, ask your Azure administrator to extend your quota.

1. On the **Compute Instances** tab, add a new compute instance with the following settings. You'll use this as a workstation from which to test your model:

   - **Compute name**: *enter a unique name*
   - **Virtual Machine type**: CPU
   - **Virtual Machine size**:
     - Choose **Select from all options**
     - Search for and select **Standard_DS11_v2**

2. While the compute instance is being created, switch to the **Compute Clusters** tab, and add a new compute cluster with the following settings. You'll use this to train a machine learning model:

- **Location**: *Select the same as your workspace. If that location is not listed, choose the one closest to you*
- **Virtual Machine tier**: Dedicated
- **Virtual Machine type**: CPU
- **Virtual Machine size**:
  - Choose **Select from all options**
  - Search for and select **Standard_DS11_v2**
- **Compute name**: *enter a unique name*
- **Minimum number of nodes**: 0
- **Maximum number of nodes**: 2
- **Idle seconds before scale down**: 120
- **Enable SSH access**: Unselected

 Tip

After completing each module, be sure to follow the **Clean Up** instructions at the end of the module to stop your compute resources. Stopping your compute ensures your subscription won't be charged for compute resources.

The compute targets will take some time to be created. You can move onto the next unit while you wait.