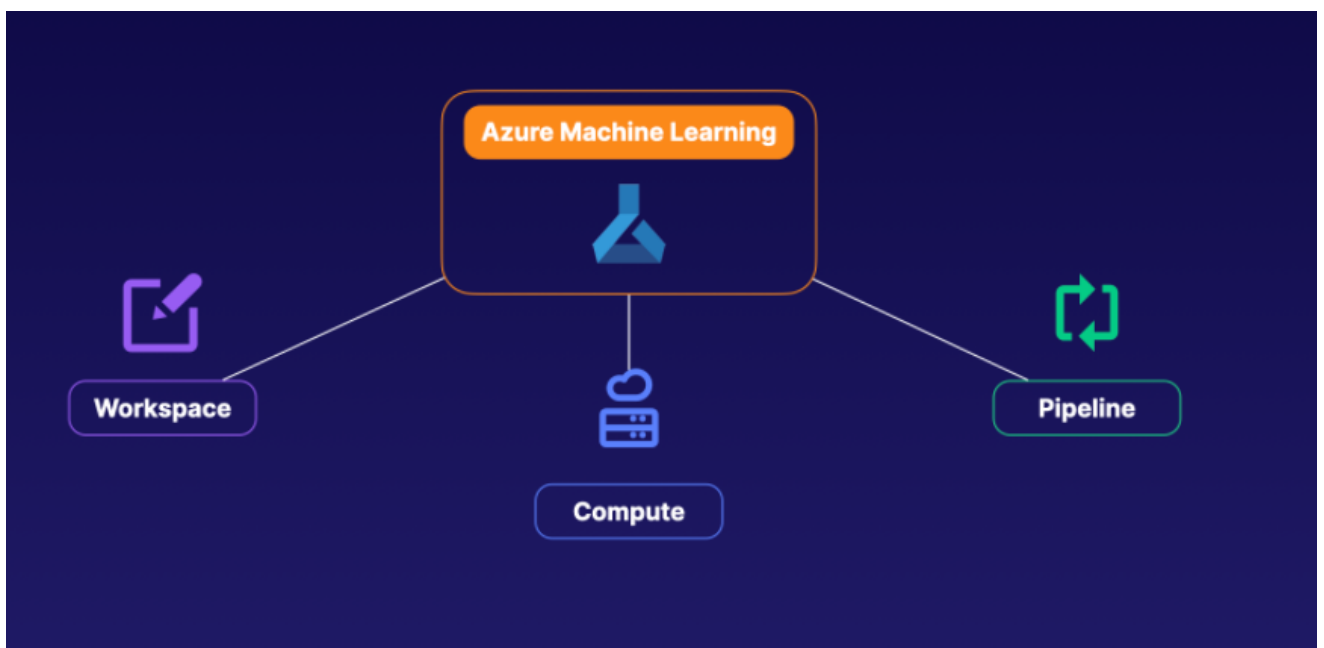


Create a Machine Learning Pipeline Using Azure SDK

Introduction

In this hands-on lab, you will become familiar with creating a machine learning workspace, creating compute resources, and programmatically developing a multi-step pipeline in Azure Machine Learning.



Create a Workspace

1. In the Azure portal, in the Search bar, enter "machine learning".
2. Under Services, click Machine learning.
3. Click the Create machine learning workspace button.
4. Set the following options for the new workspace:
 - Leave the Subscription at the default
 - Select the Resource group provided with the lab
 - For the Workspace name, enter "acml"
 - Use the same location as your lab provided resource group.
 - Leave all the other options at the default
5. Click Review + Create.

6. Once it has passed validation, click Create.

Note: It may take some time for the resource to be deployed.

Create Compute Resources

1. On the Overview page for the resource, click the Go to resource button.
2. Inside the acml resource, on the Overview page, click the Launch studio button.
3. The Azure Machine Learning Studio opens in a new browser tab. If necessary, dismiss the pop-up notification that appears.
4. In the menu on the left, under Manage, click Compute.
5. Click the + New button.
6. Leave the Virtual machine type at CPU.
7. Use the same location as your lab provided resource group.
8. From the Virtual machine size options, select Standard_DS3_v2.
9. Click Next.
10. In the Compute name field, enter "ac-aml".

Note: If Azure prompts you that this name is not unique enough, add a few characters to the end of the name to make it more unique.

11. Click Create.

Note: It may take some time for the resource to be deployed.

Create and Run a Pipeline

Note: The following walks you through using an existing Jupyter notebook with all of the necessary steps to create and run the pipeline already preconfigured, but you could create your own notebook and work through the steps on your own to complete this objective.

Clone the Provided GitHub Repository With the Preconfigured Jupyter Notebook

1. In the menu on the left, click the Notebooks icon.

Note: It is the fourth icon from the top, and looks like a spiral-bound notebook.

2. Dismiss any pop-up notifications that may appear.
3. Under the Files menu at the left, click the Open terminal icon.
4. In the terminal, enter the following to clone the GitHub repository, which includes the preconfigured Jupyter notebook:

git clone <https://github.com/hosniah/course-dp100-azure-data-scientist-associate>

5. Once that code has run, click the Refresh icon on the left to refresh the Files list and display the cloned notebook.
6. Click the cloned repository to expand it, and then click Hands-on Labs to expand that folder.
7. Click the Create a Machine Learning Pipeline using Azure SDK notebook.
8. Dismiss any pop-up notifications that may appear.
9. Click the Authenticate button at the top of the screen in the notification bar.

Follow the Steps in the Preconfigured Notebook to Build and Run the Pipeline

1. Under Connect to your workspace, review the information included in the Python script for this step.
2. Hover over the code cell and click on the Run cell button that appears to execute the code and load the workspace from the saved config file.
3. Under Prepare data, review the information included in the Python script for this step.
4. Hover over the code cell and click on the Run cell button to execute the code and create and register the diabetes dataset included in the cloned repo.
5. Under Create a folder for our python scripts, review the information included in the Python script for this step.
6. Hover over the code cell and click the Run cell button to execute the code and create the diabetes_pipeline experiment folder.
7. Under Create a data preparation script, review the information included in the Python script for this step.
8. Hover over the code cell and click the Run cell button to execute the code and create the prep_diabetes.py script that includes all of the necessary actions for preparing the data.
9. Under Create a script to train the model, review the information included in the Python script for this step.

10. Hover over the code cell and click the Run cell button to execute the code and create the `train_diabetes.py` script that includes all of the necessary actions for training the ML model.
11. Under Assign the compute target, review the information included in the Python script for this step.
12. Hover over the code cell and click the Run cell button to execute the code and assign the existing ac-aml compute resource that will be used as the compute target for the pipeline.
13. Under Ensure the necessary packages are installed on your compute, review the information included in the Python script for this step.
14. Hover over the code cell and click the Run cell button to execute the code and run the prescribed configuration on the compute resource.
15. Under Define the pipeline, review the information included in the Python script for this step.
16. Hover over the code cell and click the Run cell button to execute the code and define the steps of the pipeline, which includes running the `prep_diabetes.py` and `train_diabetes.py` scripts created previously.
17. Under Build the pipeline from the defined steps and run, review the information included in the Python script for this step.
18. Hover over the code cell and click the Run cell button to execute the code and build and run the pipeline as it was defined.
19. In the output that is returned in the code cell, click the link provided to the Azure Machine Learning portal.
20. The Azure Machine Learning portal opens in a new browser tab, where you can view the pipeline's progress. As necessary, click the Refresh button to refresh its progress.
Note: It may take some time for the pipeline to run.
21. Once it has completed, you can view the information presented about the pipeline in the Pipeline run overview section.