Highlight Note

What is a Pipeline?

In Azure Machine Learning, a *pipeline* is a workflow of machine learning tasks in which each task is implemented as a *step*.

Note: The term *pipeline* is used extensively in machine learning, often with different meanings. For example, in Scikit-Learn, you can define pipelines that combine data preprocessing transformations with a training algorithm; and in Azure DevOps, you can define a build or release pipeline to perform the build and configuration tasks required to deliver software. The focus of this module is on Azure Machine Learning pipelines, which encapsulate steps that can be run as an experiment. However, bear in mind that it's perfectly feasible to have an Azure DevOps pipeline with a task that that initiates an Azure Machine Learning pipeline, which in turn includes a step that trains a model based on a Scikit-Learn pipeline!

Steps can be arranged sequentially or in parallel, enabling you to build sophisticated flow logic to orchestrate machine learning operations. Each step can be run on a specific compute target, making it possible to combine different types of processing as required to achieve an overall goal.

Pipelines as Executable Processes

A pipeline can be executed as a process by running the pipeline as an experiment. Each step in the pipeline runs on its allocated compute target as part of the overall experiment run.

You can publish a pipeline as a REST endpoint, enabling client applications to initiate a pipeline run. You can also define a schedule for a pipeline, and have it run automatically at periodic intervals.