100 XP

Introduction to datastores

5 minutes

In Azure Machine Learning, *datastores* are abstractions for cloud data sources. They encapsulate the information required to connect to data sources. You can access datastores directly in code by using the Azure Machine Learning SDK, and use it to upload or download data.

Types of Datastore

Azure Machine Learning supports the creation of datastores for multiple kinds of Azure data source, including:

- Azure Storage (blob and file containers)
- Azure Data Lake stores
- Azure SQL Database
- Azure Databricks file system (DBFS)

> **Note**: For a full list of supported datastores, see the Azure Machine Learning documentation .

Built-in Datastores

Every workspace has two built-in datastores (an Azure Storage blob container, and an Azure Storage file container) that are used as system storage by Azure Machine Learning. There's also a third datastore that gets added to your workspace if you make use of the open datasets provided as samples (for example, by creating a designer pipeline based on a sample dataset)

In most machine learning projects, you will likely need to work with data sources of your own - either because you need to store larger volumes of data than the built-in datastores support, or because you need to integrate your machine learning solution with data from existing applications.

Next unit: Use datastores

Continue