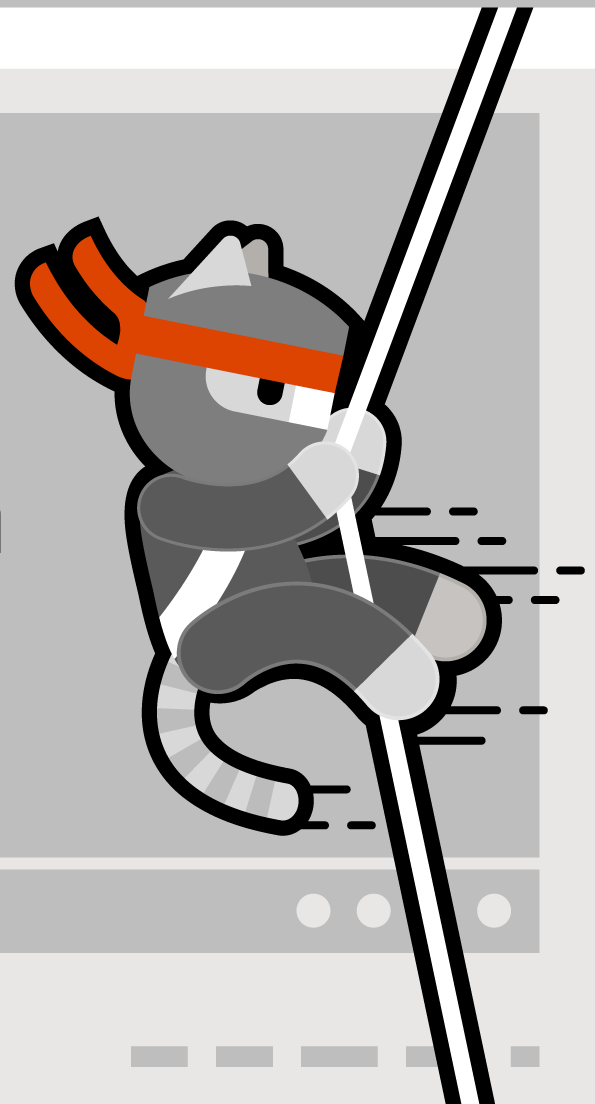
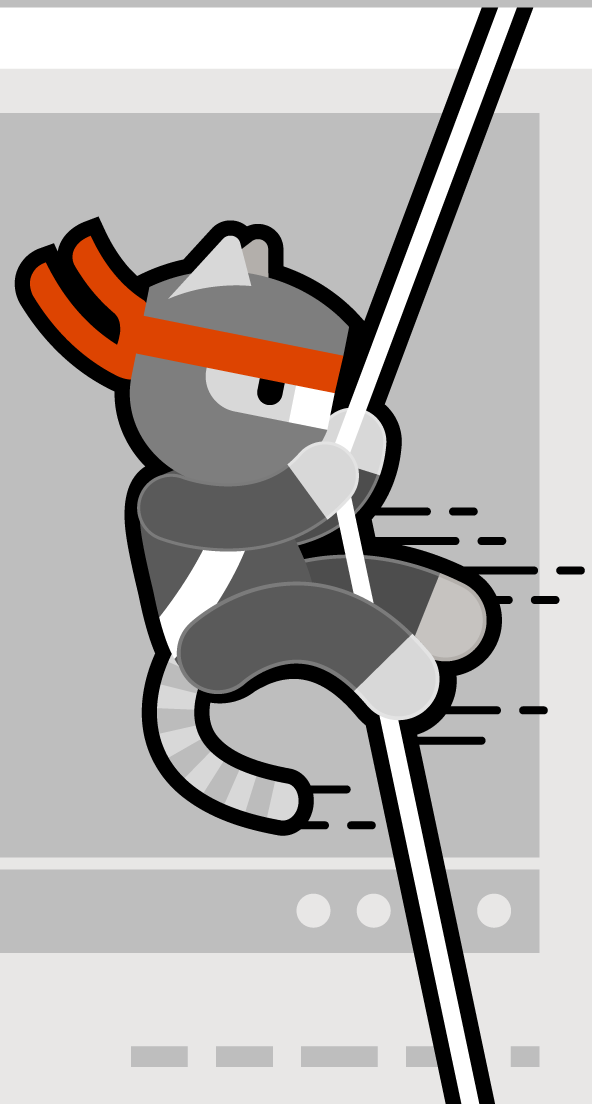
Working with Data



Eng Teong Cheah
Microsoft MVP



Working with Datastores

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 the Azure Machine Learning.

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)

Working with Datasets

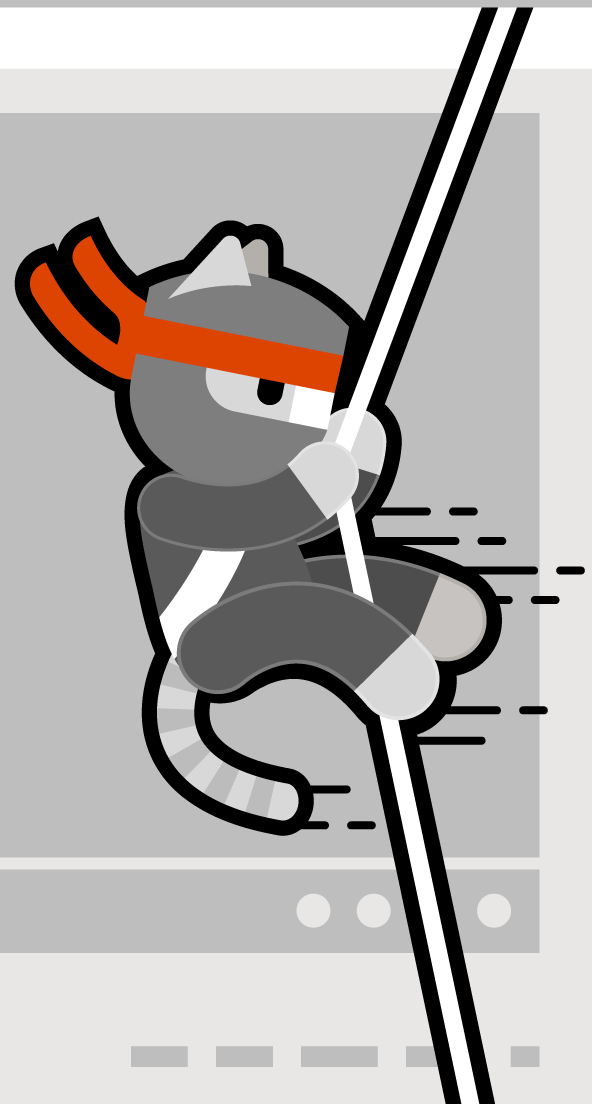
Datasets are versioned packaged data objects that can be easily consumed in experiments and pipelines. Datasets are the recommended what to work with data, and re the primary mechanism for advanced Azure Machine Learning capabilities like data labeling and data drift monitoring.

Types of Datasets

Datasets are typically on files in a datastores, though they can also be based on URLs and other sources. You can create the following types of datasets:

- Tabular
  The data is read from the dataset as a table. You should use this type of dataset when your data is consistently structured and you want to work with it in common tabular data structures, such as Pandas dataframes.
- File
  The dataset presents a list of file paths that can be read as though from the file system. Use this type of dataset when your data is unstructured, or when you need to process the data at the file level(for example, to train a convolutional neural network from a set of image files).

Demo
Work with Data



References

Microsoft Docs