Microsoft

Microsoft Azure AI Fundamentals

Exam Ref AI-900

Julian Sharp

Microsoft

Exam Ref AI-900 Microsoft Azure AI Fundamentals

Julian Sharp

- Azure Data Lake Gen 1
- Azure Data Lake Gen 2
- Azure SQL database
- Azure PostgreSQL
- Azure MySQL database

If you have existing data in Azure SQL database, you supply the details of the Azure SQL database, as shown in Figure 2-26.

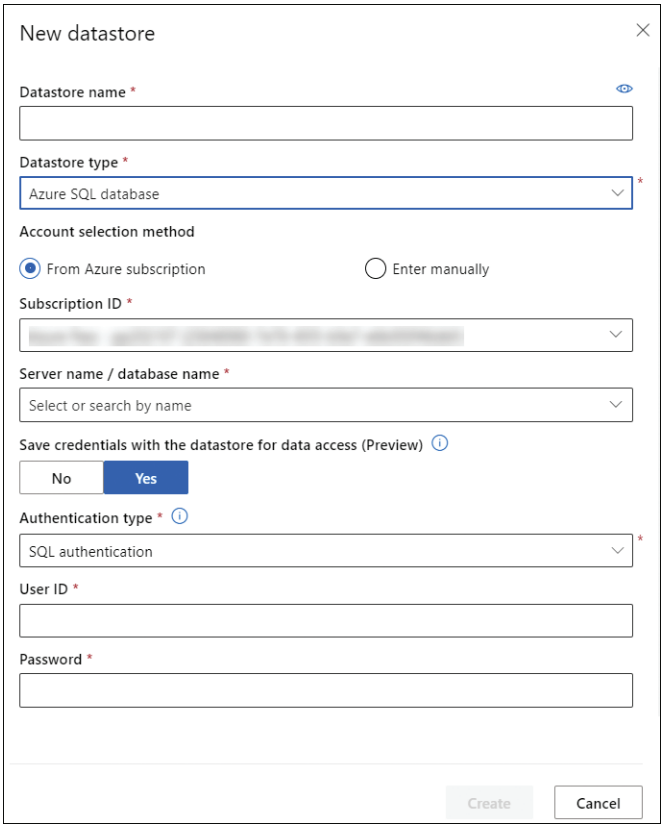

FIGURE 2-26 Add Azure SQL database

If you want to import your data into the Azure Machine Learning workspace, you register datasets using Machine Learning studio. You can create a dataset from:

- Files uploaded from your local computer
- A datastore associated with the workspace
- Files accessed via an HTTP URL
- Open datasets

When you import data, you must define the dataset type as either tabular or file:

- **Tabular** Files containing data in a tabular format. You can create a tabular dataset from CSV, TSV, Parquet, JSON files, and from the output of an SQL query.
- **File** A collection of file references in datastores or public URLs. A file dataset references single or multiple files.

NOTE **CREATING DATASETS**

You can create a dataset from the output of a SQL query but not from a SQL query itself. You can create a dataset from files in an Azure Blob container but not from an Azure Blob container itself. You can create a dataset from files in a folder but not from a folder itself.

The student data we referenced earlier in this chapter is held in a CSV file on a local computer. After clicking on + Create dataset, the Create dataset from the local files wizard is displayed, as shown in Figure 2-27.

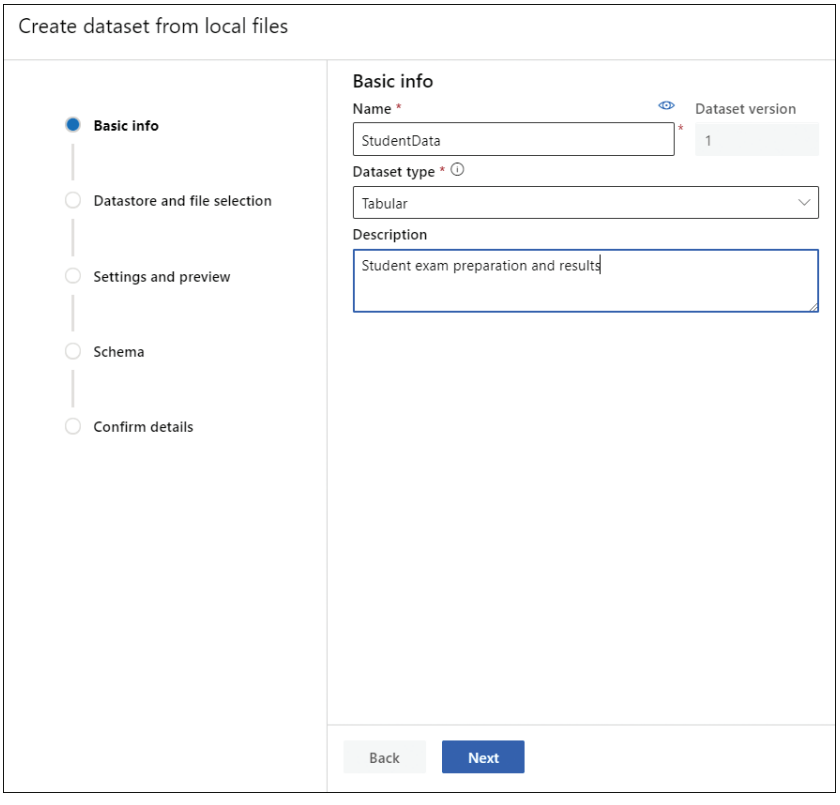

FIGURE 2-27 Create a dataset from the local files wizard, step 1

You must enter the name of the dataset, select the dataset type as either Tabular or File. Clicking on Next displays the next step in the wizard, as shown in Figure 2-28.

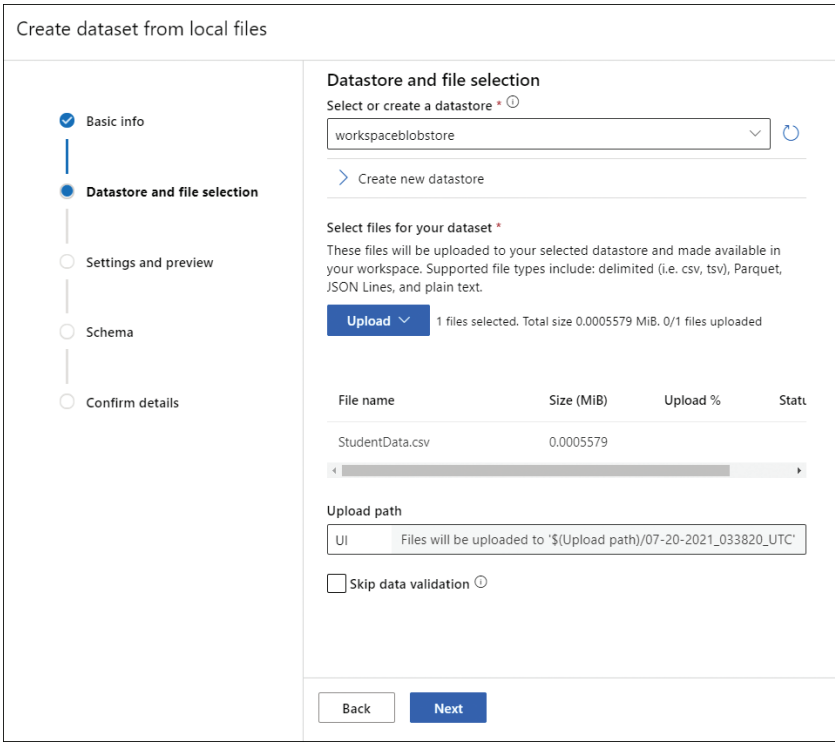

FIGURE 2-28 Create a dataset from the local files wizard, step 2

You select the datastore where the data will be imported into. This will normally be the built-in blob datastore, but you can add your own datastore if required. You can then upload a single file or select an entire folder of files to upload. Clicking on Next will parse the selected file(s) and display the next step in the wizard, as shown in Figure 2-29.

You can review the tabular data to ensure it has been parsed correctly. Clicking on Next displays the next step in the wizard, as shown in Figure 2-30.

You can exclude columns from being imported and correct any data type for each column. Clicking on Next displays the final step in the wizard, as shown in Figure 2-31.

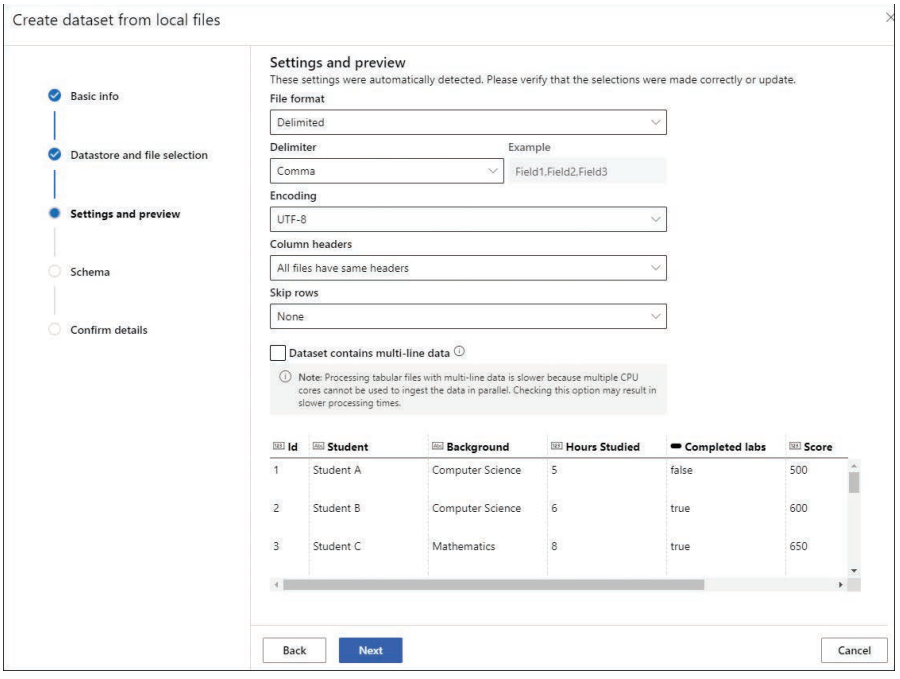

FIGURE 2-29 Create a dataset from the local files wizard, step 3

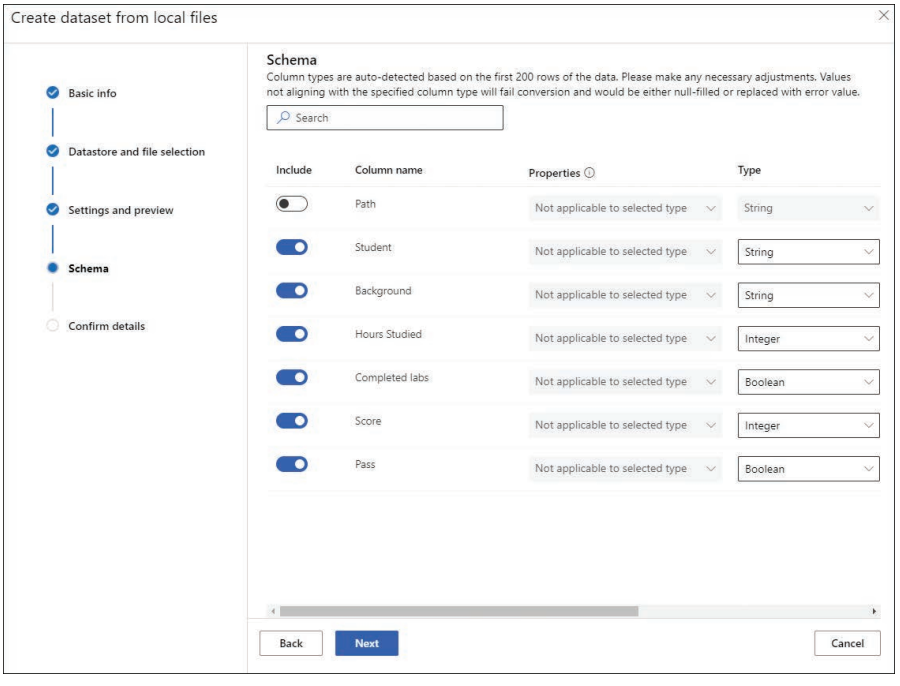

FIGURE 2-30 Create a dataset from the local files wizard, step 4

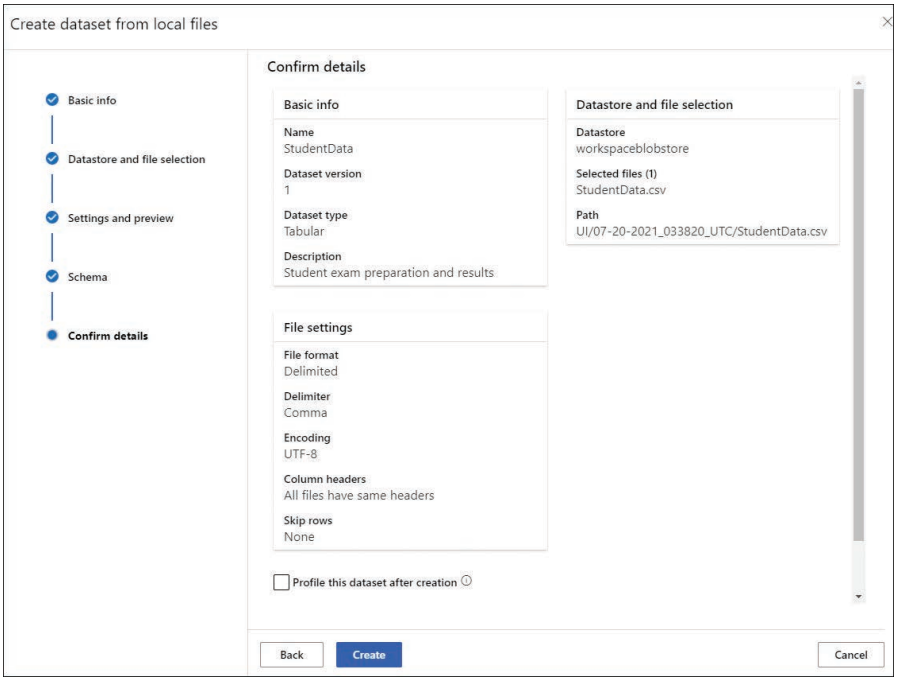

FIGURE 2-31 Create a dataset from the local files wizard, step 5

Clicking on Create will import the data and register the dataset, as shown in Figure 2-32.

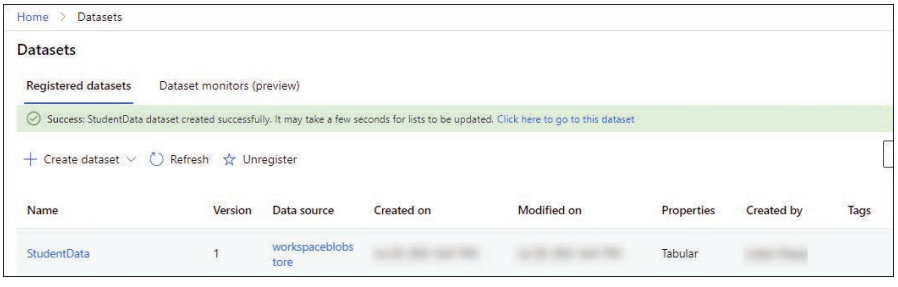

FIGURE 2-32 Registered datasets

You can use publicly available data from URLs such as daily bike rental data that can be found at http://aka.ms/bike-rentals. You can import this dataset, as shown in Figure 2-33.

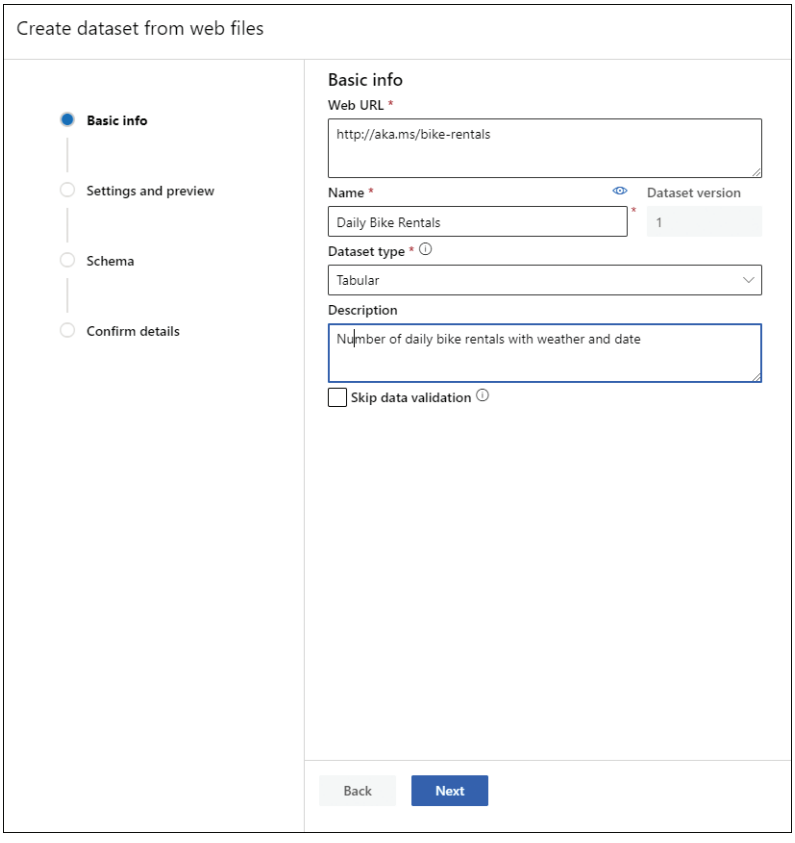

FIGURE 2-33 Create a dataset from the local files wizard, step 1

Azure Open datasets are curated datasets publicly available on Azure that you can import into your machine learning model, as shown in Figure 2-34.

You can simply add these datasets to your workspace. To find out more about these and other publicly available datasets, see https://docs.microsoft.com/azure/open-datasets/dataset-catalog.

> ***NOTE*** **SAMPLE DATASETS**
>
> **There are a further 16 sample datasets available in the Azure Machine Learning designer tool (https://docs.microsoft.com/azure/machine-learning/samples-designer#datasets).**