



# Microsoft Azure AI Fundamentals

Exam Ref AI-900

Julian Sharp

# **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.

The screenshot shows a wizard titled "Create dataset from local files". On the left is a vertical sidebar with five steps: "Basic info" (selected with a blue dot), "Datastore and file selection", "Settings and preview", "Schema", and "Confirm details". The main area is titled "Basic info" and contains the following fields:

- Name \***: A text input field containing "StudentData". To its right is a "Dataset version" field with the value "1".
- Dataset type \***: A dropdown menu with "Tabular" selected.
- Description**: A text area containing "Student exam preparation and results".

At the bottom of the wizard are two buttons: "Back" (disabled) and "Next" (active).

**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.

The screenshot shows a web-based wizard titled "Create dataset from local files". On the left is a vertical navigation pane with five steps: "Basic info" (checked), "Datastore and file selection" (active), "Settings and preview", "Schema", and "Confirm details". The main content area is titled "Datastore and file selection" and contains the following elements:

- A section "Select or create a datastore" with a dropdown menu showing "workspaceblobstore" and a refresh icon.
- A link "> Create new datastore".
- A section "Select files for your dataset" with a note: "These files will be uploaded to your selected datastore and made available in your workspace. Supported file types include: delimited (i.e. csv, tsv), Parquet, JSON Lines, and plain text."
- An "Upload" button with a dropdown arrow, followed by the text "1 files selected. Total size 0.0005579 MiB. 0/1 files uploaded".
- A table with columns: "File name", "Size (MiB)", "Upload %", and "Status". It contains one row for "StudentData.csv" with a size of "0.0005579" and a progress bar.
- An "Upload path" section with a text input field containing "UI" and a note: "Files will be uploaded to '\${Upload path}/07-20-2021\_033820\_UTC'".
- A checkbox labeled "Skip data validation" with an information icon.
- At the bottom are "Back" and "Next" buttons.

**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.

Basic info

Datastore and file selection

Settings and preview

Schema

Confirm details

### Settings and preview

These settings were automatically detected. Please verify that the selections were made correctly or update.

**File format**

Delimited

**Delimiter** Comma **Example** Field1.Field2.Field3

**Encoding** UTF-8

**Column headers** All files have same headers

**Skip rows** None

☐ Dataset contains multi-line data

Note: Processing tabular files with multi-line data is slower because multiple CPU cores cannot be used to ingest the data in parallel. Checking this option may result in slower processing times.

Id	Student	Background	Hours Studied	Completed labs	Score
1	Student A	Computer Science	5	false	500
2	Student B	Computer Science	6	true	600
3	Student C	Mathematics	8	true	650

Back Next Cancel

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

Basic info

Datastore and file selection

Settings and preview

Schema

Confirm details

### Schema

Column types are auto-detected based on the first 200 rows of the data. Please make any necessary adjustments. Values not aligning with the specified column type will fail conversion and would be either null-filled or replaced with error value.

Search

Include	Column name	Properties	Type
<input type="checkbox"/>	Path	Not applicable to selected type	String
<input checked="" type="checkbox"/>	Student	Not applicable to selected type	String
<input checked="" type="checkbox"/>	Background	Not applicable to selected type	String
<input checked="" type="checkbox"/>	Hours Studied	Not applicable to selected type	Integer
<input checked="" type="checkbox"/>	Completed labs	Not applicable to selected type	Boolean
<input checked="" type="checkbox"/>	Score	Not applicable to selected type	Integer
<input checked="" type="checkbox"/>	Pass	Not applicable to selected type	Boolean

Back Next Cancel

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

Create dataset from local files

Confirm details

Basic info

Name: StudentData

Dataset version: 1

Dataset type: Tabular

Description: Student exam preparation and results

Datastore and file selection

Datastore: workspaceblobstore

Selected files (1): StudentData.csv

Path: UJ/07-20-2021\_033820\_UTC/StudentData.csv

File settings

File format: Delimited

Delimiter: Comma

Encoding: UTF-8

Column headers: All files have same headers

Skip rows: None

☐ Profile this dataset after creation ⓘ

Back Create Cancel

**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.

Home > Datasets

Datasets

Registered datasets Dataset monitors (preview)

Success: StudentData dataset created successfully. It may take a few seconds for lists to be updated. [Click here to go to this dataset](#)

+ Create dataset Refresh Unregister

Name	Version	Data source	Created on	Modified on	Properties	Created by	Tags
StudentData	1	workspaceblobstore	2021-07-20 03:38:20 UTC	2021-07-20 03:38:20 UTC	Tabular	user@company.com	

**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.

Create dataset from web files

**Basic info**

Web URL \*

http://aka.ms/bike-rentals

Name \*

Daily Bike Rentals

Dataset version

1

Dataset type \*

Tabular

Description

Number of daily bike rentals with weather and date

☐ Skip data validation

Back Next

**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>).