Lab 5 – Integrate on-premises data

In previous labs you have been working with data already available on the internet, but integrating data from on-premises or other private networks is a common requirement. In this lab you will use a Self-Hosted Integration Runtime to allow ADF to copy a file from your computer and into the data lake.

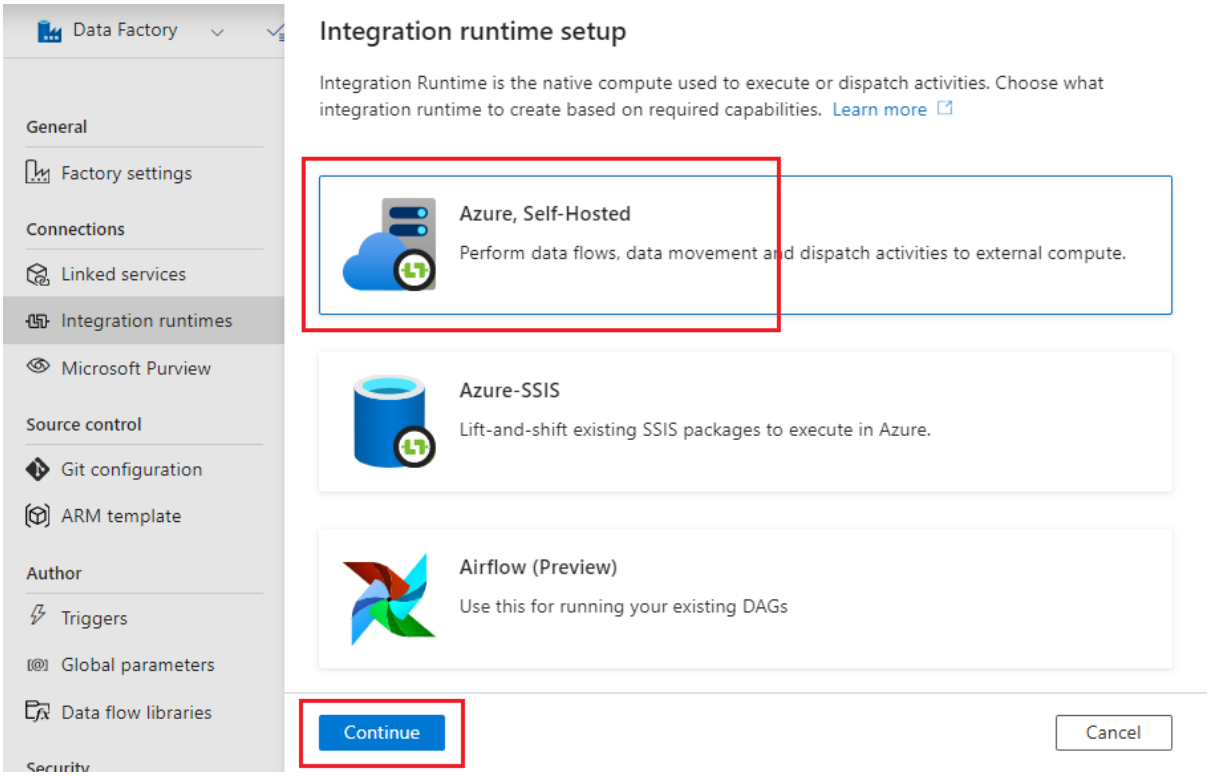
Note: If you already have a self-hosted IR installed on your computer, you cannot install another one – to complete this lab you will need first to remove the existing installation.

Lab 5.1 – Create a Self-Hosted Integration Runtime

Integration runtimes provide the compute power required for data movement and other activities. In this lab you will create a Self-Hosted Integration Runtime (SHIR) which can support data movement to and from your own computer.

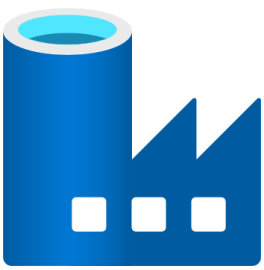
1. Navigate to ADF Studio's management hub, then select the "Integration runtimes" item from the "Connections" section of the hub sidebar. In the main Integration runtimes pane click "+ New".

2. On the "Integration runtime setup" flyout, select the "Azure, Self-Hosted" tile and click "Continue".

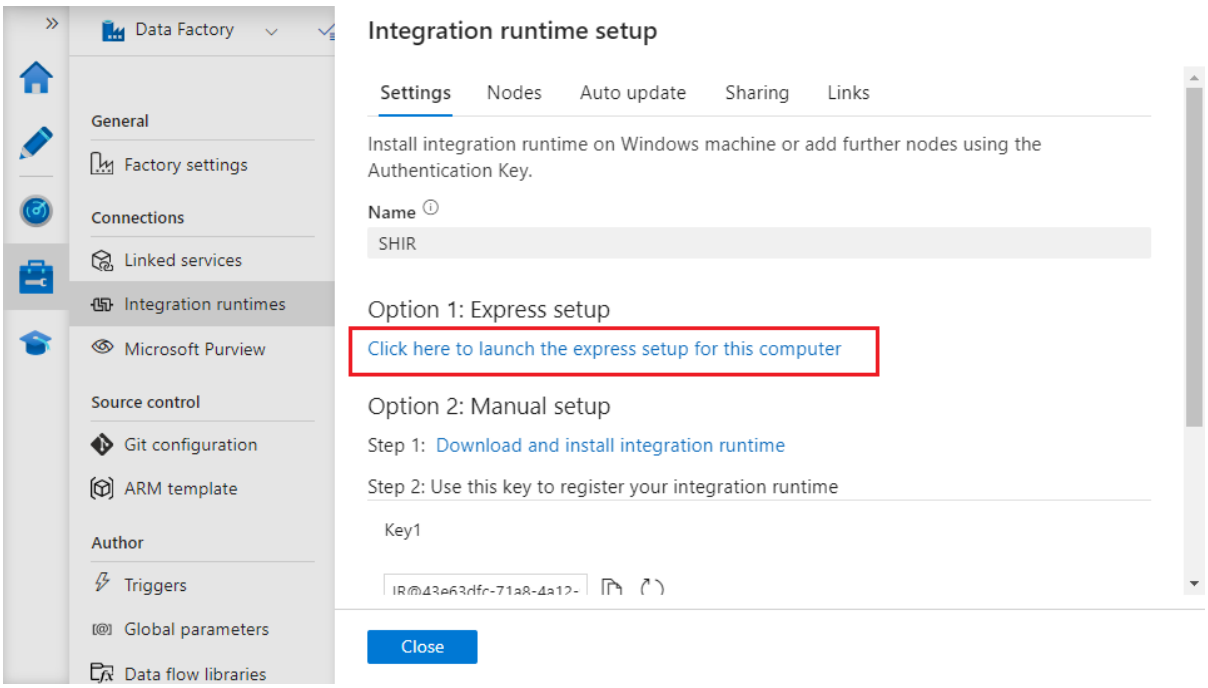


3. On the following page, choose the "Self-Hosted" tile (under the "Network environment" heading – do **not** choose "Linked Self-Hosted"). Click continue.

4. Provide a name for your SHIR and click "Create".



5. Steps 1-4 have created the ADF side of the connection to your computer. The next step starts the process of installing the other end of the connection. Click the link under "Option 1: Express setup" to download a customised installer.



6. Run the downloaded .exe file to install and register the SHIR on your computer. This requires administrator permissions and may take a few minutes to complete.

7. When installation has finished, return to ADF Studio. Close the Integration runtime setup flyout and verify that the new SHIR appears in the list of integration runtimes and has status "Running".

Lab 5.2 – Disable local folder path validation

The SHIR you created in Lab 5.1 is designed to provide a gateway to on-premises systems reachable from your computer, but to demonstrate the connection we're going to copy a file directly from your machine. By default, an SHIR has no access to its host's file system – to copy a file, you must first permit this using the SHIR diagnostic tool.

1. Open a Windows command or PowerShell prompt **as an administrator**.

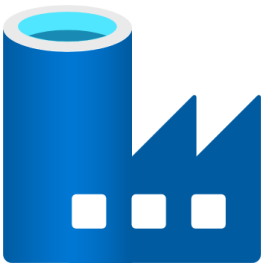
2. **cd** to the folder containing the diagnostic tool:

```
cd "C:\Program Files\Microsoft Integration Runtime\5.0\Shared"
```

3. Run the diagnostic tool with the **DisableLocalFolderPathValidation** switch:

```
.\dmgcmd.exe -DisableLocalFolderPathValidation
```

Disabling local folder path validation disables security validation, enabling the SHIR to access your local machine's file system. We will undo this measure at the end of Lab 5.



Lab 5.3 – Create a file system linked service

Although the SHIR provides a gateway to systems reachable from your computer, it does not represent any specific target system. Just as you created linked services to connect ADF to GitHub or your data lake, linked services are also required when connecting to systems via the SHIR.

Rather than use your new SHIR to connect to other on-premises systems, in this lab you will create a linked service to enable ADF to connect to your computer's file system.

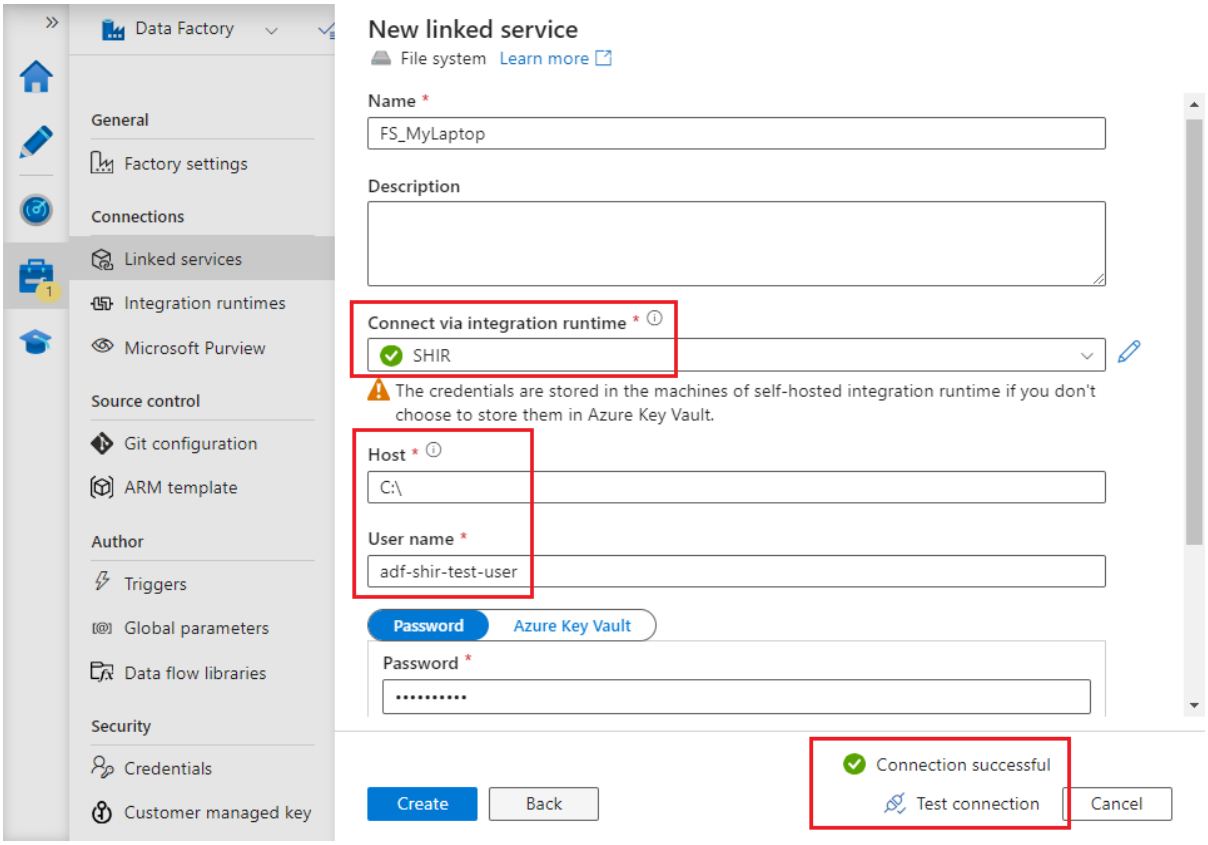
1. Using the "Linked services" item in ADF Studio's management hub, click "+ New" to open the "New linked service" flyout.

2. Search for, then select, the "File system" data store type. Click "Continue".

3. Configure your new file system as follows:

   - Give it a **name**.
   - Use the **Connect via integration runtime** dropdown to select your new self-hosted IR.
   - Set **Host** to a file path on your computer, e.g. C:\
   - Provide a **user name** and **password** to access the computer.

   Click "Test connection" to verify that ADF can connect to your computer's file system. If you receive the error message "Access to '*<your-file-location>*' is not allowed", check that you have correctly enabled file system access in Lab 5.2.



4. When the connection test succeeds, click "Create" to close the flyout.

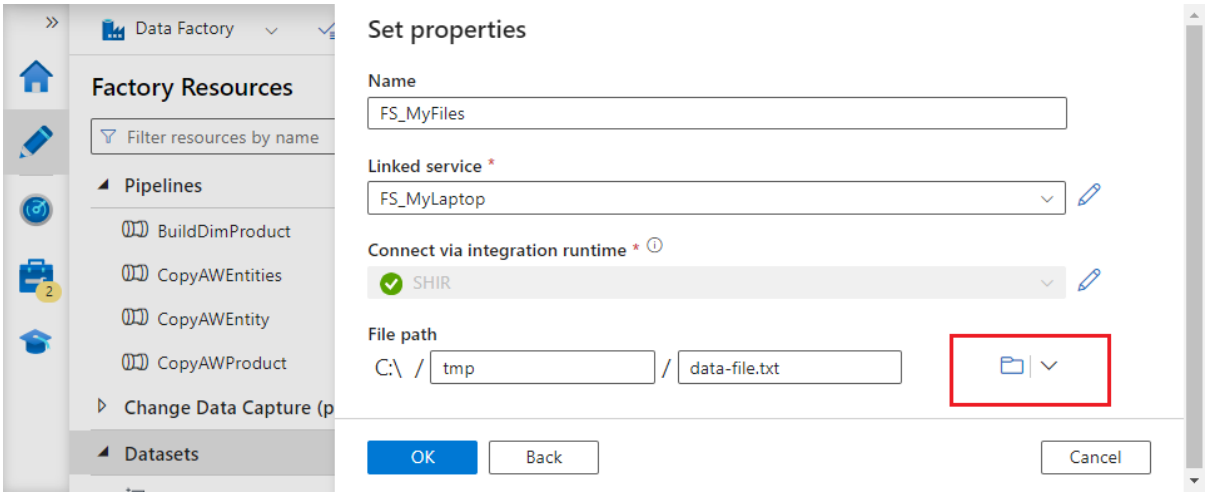
Lab 5.4 – Create a file system dataset

The linked service you created in Lab 5.3 represents a connection to your computer's file system. To copy files from the computer, you will need a file system dataset to represent them.

1. In the ADF Authoring experience, create a new dataset using the "File system" data store and "Binary" format".

2. Name it "FS_MyFiles" and select the file system linked service you created in Lab 5.3.

3. Use the browse icon to select a file from your computer, then click OK.



4. Publish your changes to the ADF service.

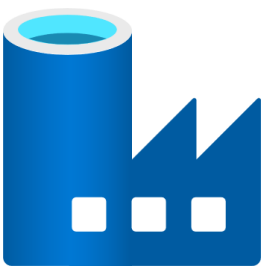
Lab 5.5 – Copy a file from your computer to ADF

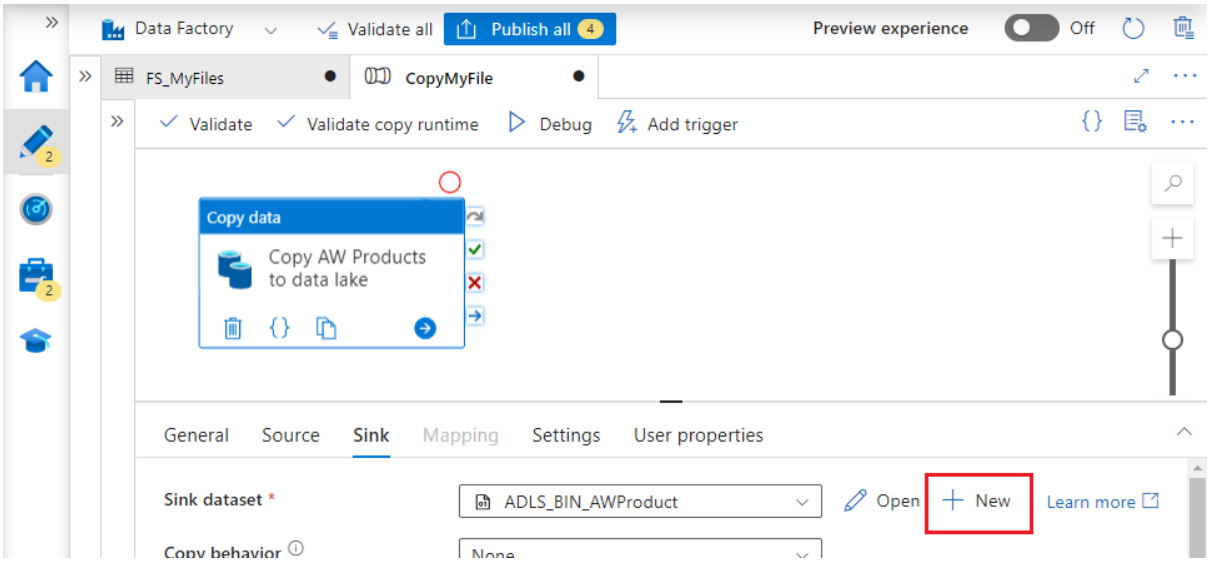
In this lab you will copy the file from your computer into the data lake, using datasets you have created and a new pipeline.

1. The new pipeline has similar requirements to the pipeline you created in Lab 1. Create the new pipeline by cloning the Lab 1 pipeline, then rename it appropriately.

2. Replace the Copy activity's HTTP source dataset with the file system dataset you created in Lab 5.4.

3. Replace the data lake sink dataset with a new dataset. You can create new datasets directly from the Copy activity's sink tab, using the + New button:





4. The new sink should use your data lake (data store type ADLS Gen 2) and the binary file format. Copy the file into the lakeroot/Raw folder with a name of your choice.

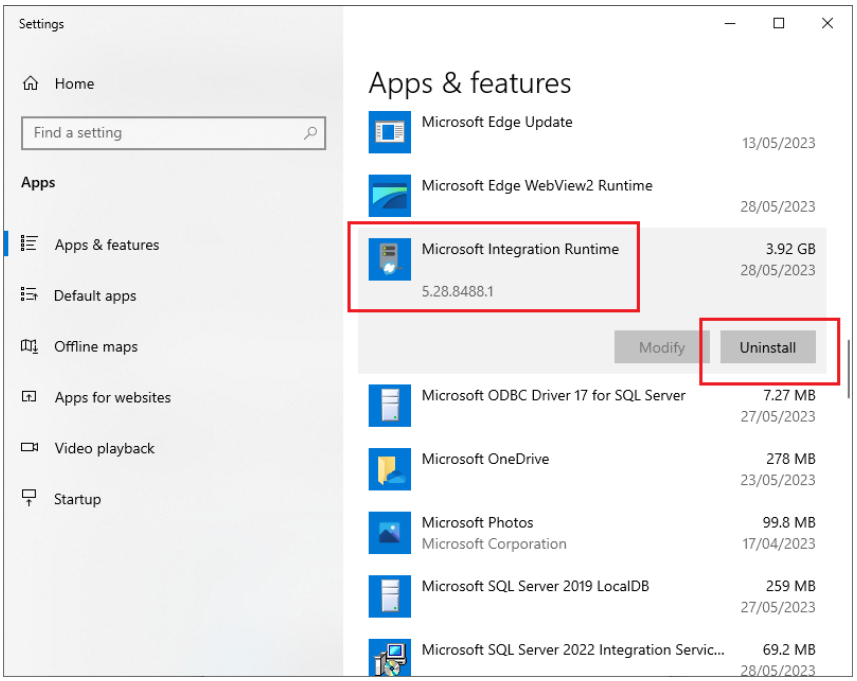
5. Run the pipeline in debug mode and verify that a copy of the file stored on your computer is now present in your data lake.

Lab 5.6 – Clean up

Disable the connection to your computer from ADF by uninstalling the SHIR:

1. Open “Apps & Features” on your computer.

2. Locate and select “Microsoft Integration Runtime”.



3. Click “Uninstall”.

When uninstallation is complete, ADF will no longer be able to connect to your computer.

Recap

In Lab 5 you have:

- Created a self-hosted integration runtime on your own computer.
- Allowed the runtime to access the computer's file system. This is not the default behaviour – SHIR's are normally used as gateways to gain access to other on-premises systems.
- Copied a file from your computer into the data lake.
- Removed the SHIR from your computer, ensuring that ADF can no longer connect to it.