Lab 6 – Explore Synapse pipelines

Synapse pipelines provide a data integration experience closely modelled on that of Azure Data Factory. The two environments are not fully interchangeable – for example, Synapse pipelines support a small number of Synapse-specific activities not available in ADF – but are in most cases very similar. This lab explores the relationship between Synapse and ADF pipelines.

Lab 6.1 – Provision a Synapse Analytics workspace

In common with other Azure resources, a Synapse Analytics workspace can be provisioned online using the Azure portal.

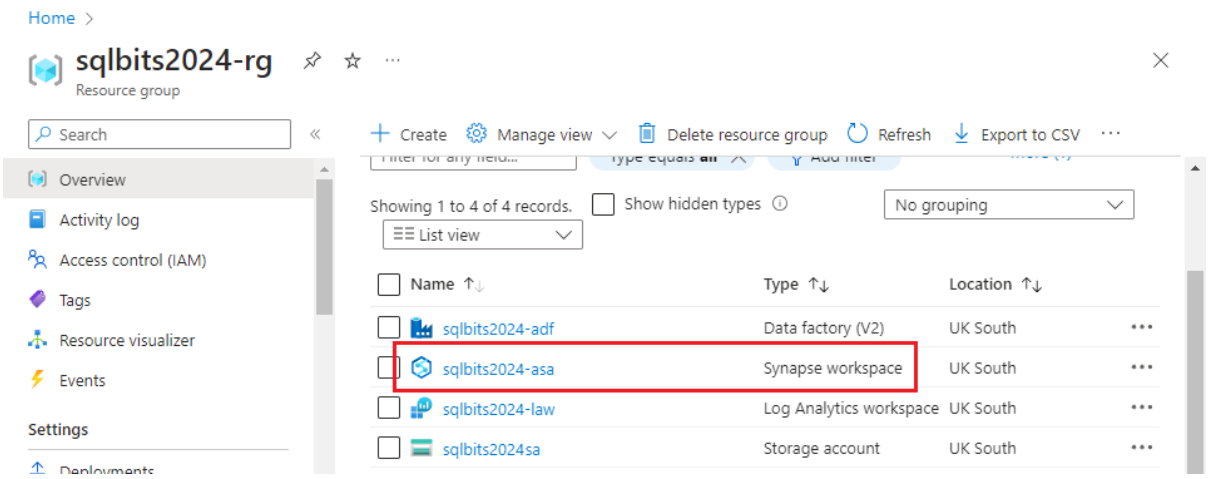
1. In the portal, click “Create a resource” and search for “Azure Synapse Analytics”. Click “Create” on the overview screen.

2. Complete the **Basics** tab like this:

   - Choose the subscription and resource group you have been using throughout the lab series.
   - Provide a name for the workspace’s **Managed resource group** – this resource group is a container for resources that are managed for you by Azure Synapse Analytics.
   - Provide a **Workspace name**.
   - Choose the same region as you have been using throughout the lab series.
   - Every Synapse workspace has an attached data lake – for convenience, we will continue to use the data lake you created in Lab 1.2. Choose your data lake storage account from the **Account name** dropdown, then select the “lakeroot” container from the **File system name** dropdown.

3. Click “Next: Security >” to open the “Security” tab (or open the tab by clicking on it directly). Ensure that the **Authentication method** “Use only Microsoft Entra authentication” is selected.

4. Click “Review + create” to accept defaults on the remaining tabs and skip to the “Review + create” tab. Click “Create”. When workspace is complete, click on “Go to resource group” – you will see the new Synapse workspace in the list of resources, alongside the data lake, data factory and Log Analytics workspace you created in earlier labs.

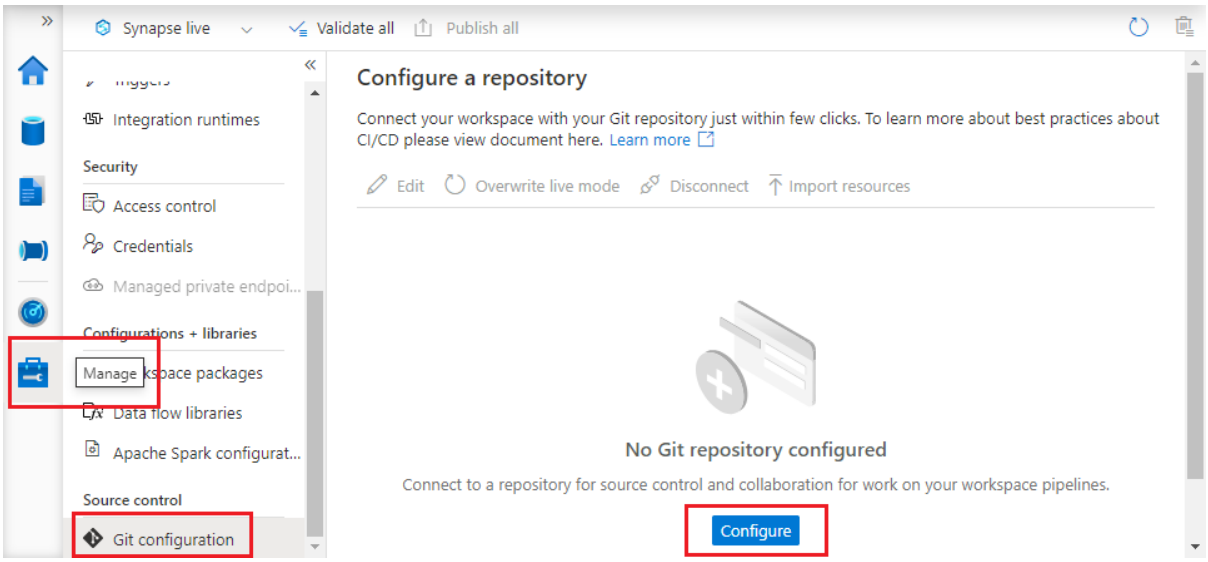


5. Click on the Synapse workspace resource, then on the "Open Synapse Studio" tile to launch Synapse Studio.

Lab 6.2 – Connect workspace to a GitHub repository

Like Azure Data Factory, a Synapse workspace can be connected to a GitHub (or Azure DevOps) repository so that workspace artifacts can be stored under version control. Connecting a Synapse workspace to a GitHub repository is a similar experience to connecting an ADF instance.

1. Navigate to Synapse Studio's Manage hub (toolbox icon) on the leftmost sidebar, then click the "Configure" button on the "Git configuration" page.



2. Choose "GitHub" from the "Repository type" dropdown list, then enter your GitHub account name into the "GitHub repository owner" field and click "Continue".

3. GitHub prompts you to sign in and to authorise the "Azure Synapse" application. Follow the instructions to sign into your account and confirm authorisation.

4. Once authorised, choose the same repository name, collaboration branch, **and root folder** as you selected in Lab 1.3, then click "Apply".

   In practice you would be very unlikely to take this approach! When hosting multiple Synapse workspaces and/or Azure Data Factory instances in the same repository, each workspace or factory should have a different root folder. You might go further and choose to host each workspace or factory in a repository of its own.

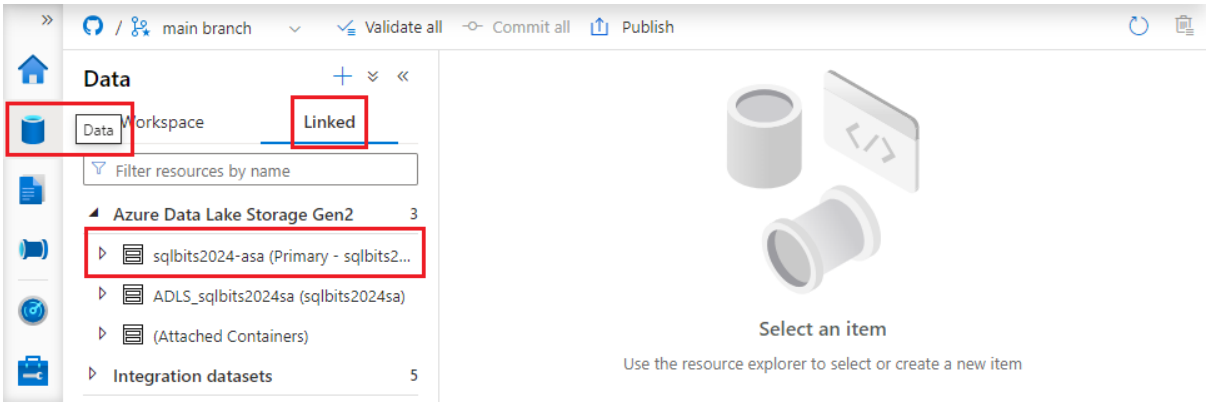
   The reason we have connected your Synapse workspace to your existing ADF repository is to demonstrate the close similarity between pipeline artifacts.

5. Finally, click "Refresh" in the top right corner to load artifacts from the linked repository into Synapse Studio.

Lab 6.3 – Explore Synapse Studio

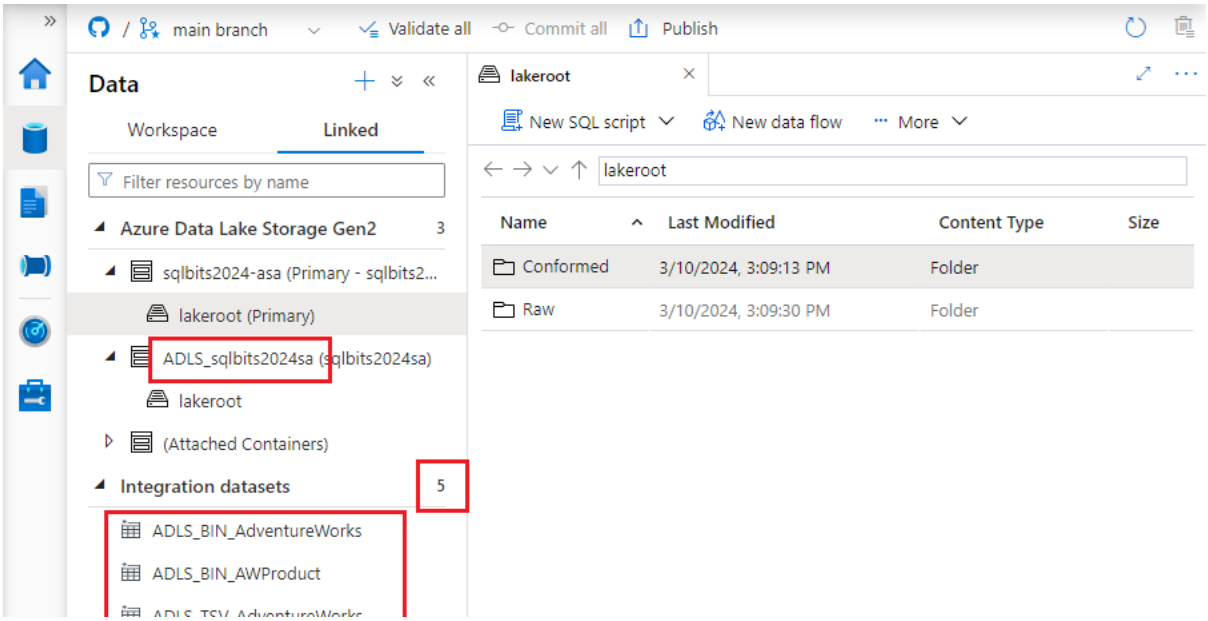
Now that you have connected your Synapse workspace to your ADF repository, all the artifacts you created in Labs 1-4 are available in the workspace. Synapse Studio organises them differently – the purpose of this lab is to help you locate them.

1. Open the Data hub (database icon) from the leftmost toolbar, and select the “Linked” tab. Expand the “Azure Data Lake Storage Gen2” node. The node includes a linked service with the same name as the Synapse workspace (indicated in the screenshot below, taken from workspace “sqlbits2024-asa”). This is created automatically when the workspace is provisioned, and provides access to the data lake and file system specified during creation (on the “Basics” tab in Lab 6.1).



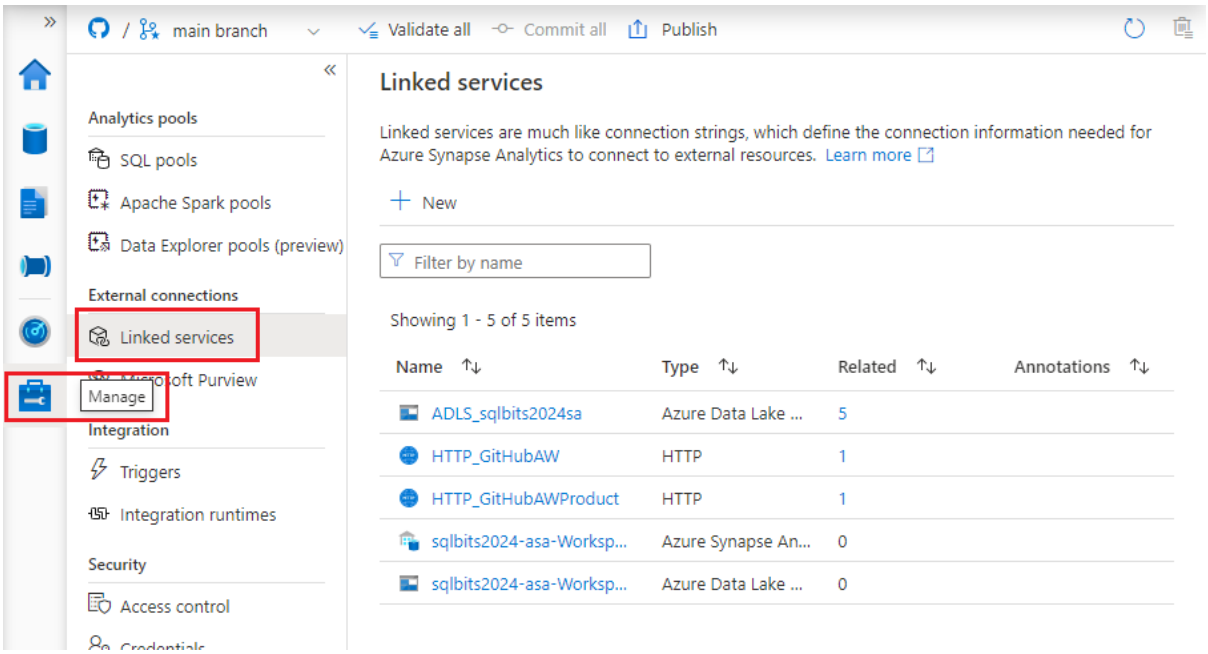
2. Expand the linked service and click on the “lakeroot” file system – the container’s contents are displayed in a new tab to the right. The workspace’s managed identity is automatically granted access to its integrated data lake during provisioning, so (unlike in the case of ADF) no further authorisation action is required.

3. Notice that “Azure Data Lake Storage Gen2” also includes the data lake linked service you created in Lab 1.4, indicated in the screenshot below. (This particular linked service duplicates the data lake linked service created automatically by Synapse). Synapse defines linked services in the same way as ADF, so ADF linked services are valid here. Similarly, “Integration datasets” lists the five datasets you created in earlier labs – expand the node to inspect the dataset definitions.



4. All linked services – including the data lake linked services above, the HTTP linked service you created in ADF, and a serverless SQL pool connection created by Synapse – can be

inspected in the Manage hub. Choose the “Linked services” item in the hub sidebar’s “External connections” group.



5. Navigate to the Develop hub (notebook icon) – the data flow you created in Lab 4 can be found here.

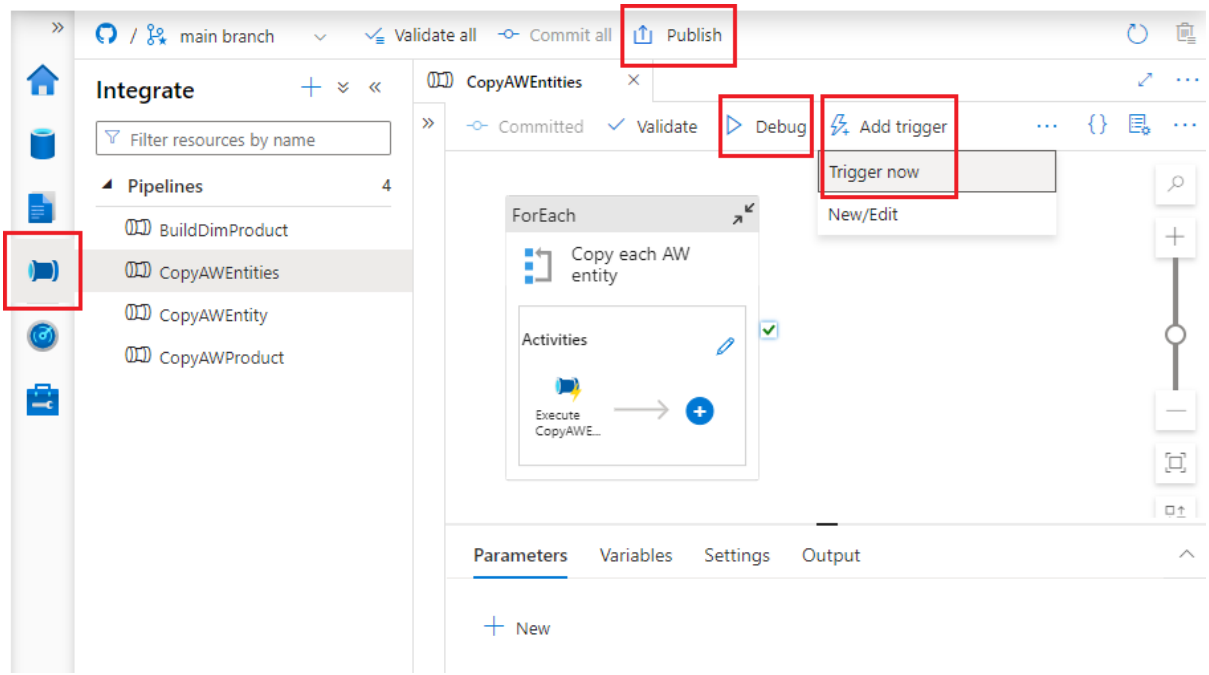
6. Open the Integrate hub (pipeline icon) – here you will find the four pipelines you defined in labs 2, 3 & 4.

Lab 6.4 – Run Synapse pipelines

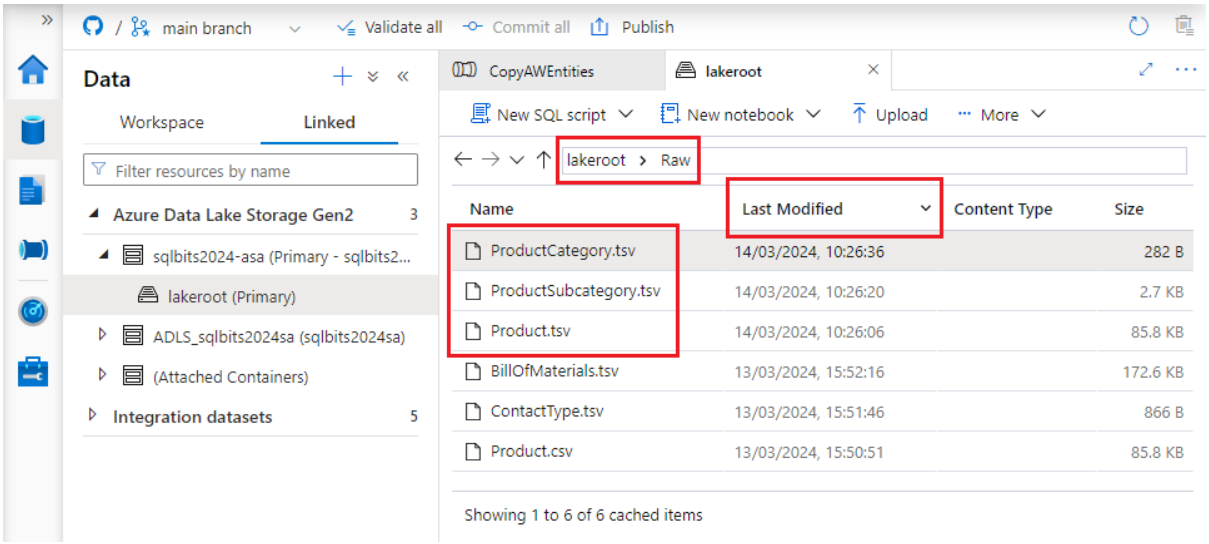
In this lab you will run – in your Synapse Analytics workspace – one of the pipelines you developed in ADF, and demonstrate that the pipelines are functionally equivalent.

1. Remain in Synapse Studio’s Integrate hub and open the “CopyAWEntities” pipeline you created in Lab 3.4

2. Run the pipeline. Either:
   - Use the “Debug” button to run the pipeline in debug mode
   - Publish Synapse workspace artifacts using the “Publish” button, then use the “Trigger now” button on the “Add trigger” dropdown



3. Return to the Data hub, and open the workspace data lake to browse the contents of the "Raw" folder – notice that the most recently updated files are the three files freshly copied by the "CopyAWEntities" pipeline.



Recap

In Lab 6 you:

- provisioned a new Synapse Analytics workspace
- connected the workspace to the GitHub repository you used in Labs 1 to 5
- observed that artifacts developed in ADF can be ported to a Synapse workspace and managed in Synapse Studio
- demonstrated that some pipelines developed in ADF can be run directly in Synapse Analytics with equivalent results.