

# Lab 8 - End-to-end security with Azure Synapse Analytics

---

In this lab, you will learn how to secure a Synapse Analytics workspace and its supporting infrastructure. You will observe the SQL Active Directory Admin, manage IP firewall rules, manage secrets with Azure Key Vault and access those secrets through a Key Vault linked service and pipeline activities. You will understand how to implement column-level security, row-level security, and dynamic data masking when using dedicated SQL pools.

After completing this lab, you will be able to:

- Secure Azure Synapse Analytics supporting infrastructure
- Secure the Azure Synapse Analytics workspace and managed services
- Secure Azure Synapse Analytics workspace data

This lab will guide you through several security-related steps that cover an end-to-end security story for Azure Synapse Analytics. Some key take-aways from this lab are:

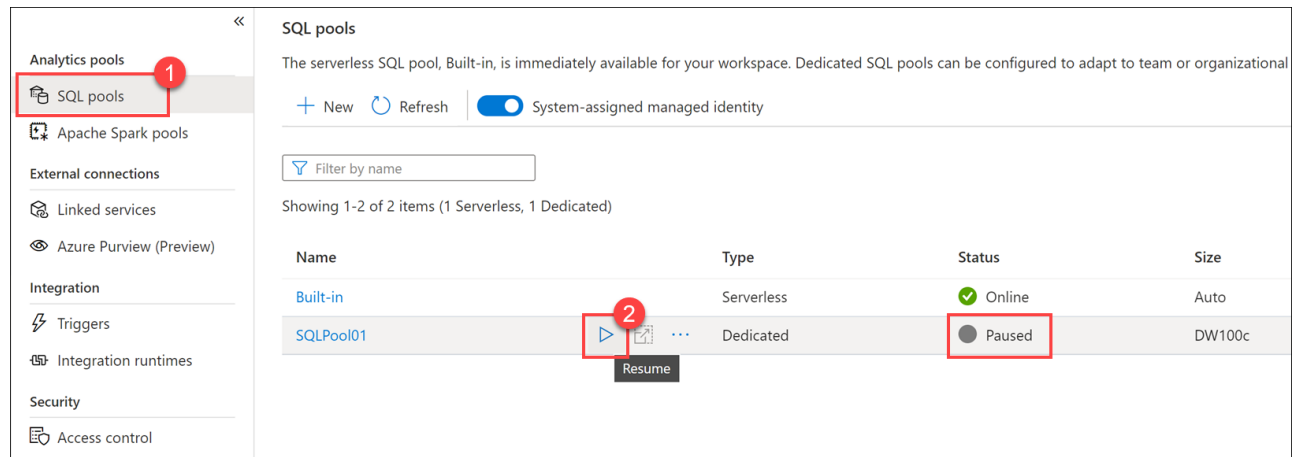
1. Leverage Azure Key Vault to store sensitive connection information, such as access keys and passwords for linked services as well as in pipelines.
2. Introspect the data that is contained within the SQL Pools in the context of potential sensitive/confidential data disclosure. Identify the columns representing sensitive data, then secure them by adding column-level security. Determine at the table level what data should be hidden from specific groups of users then define security predicates to apply row level security (filters) on the table. If desired, you also have the option of applying Dynamic Data Masking to mask sensitive data returned in queries on a column by column basis.

## Lab setup and pre-requisites

Before starting this lab, you must complete at least the setup steps in **Lab 4: Explore, transform, and load data into the Data Warehouse using Apache Spark**.

This lab uses the dedicated SQL pool you created in the previous lab. You should have paused the SQL pool at the end of the previous lab, so resume it by following these instructions:

1. Open Azure Synapse Studio (<https://web.azuresynapse.net/>).
2. Select the **Manage** hub.
3. Select **SQL pools** in the left-hand menu. If the **SQLPool01** dedicated SQL pool is paused, hover over its name and select ▷.



4. When prompted, select **Resume**. It will take a minute or two to resume the pool.

5. Continue to the next exercise while the dedicated SQL pool resumes.

**Important:** Once started, a dedicated SQL pool consumes credits in your Azure subscription until it is paused. If you take a break from this lab, or decide not to complete it; follow the instructions at the end of the lab to pause your SQL pool!

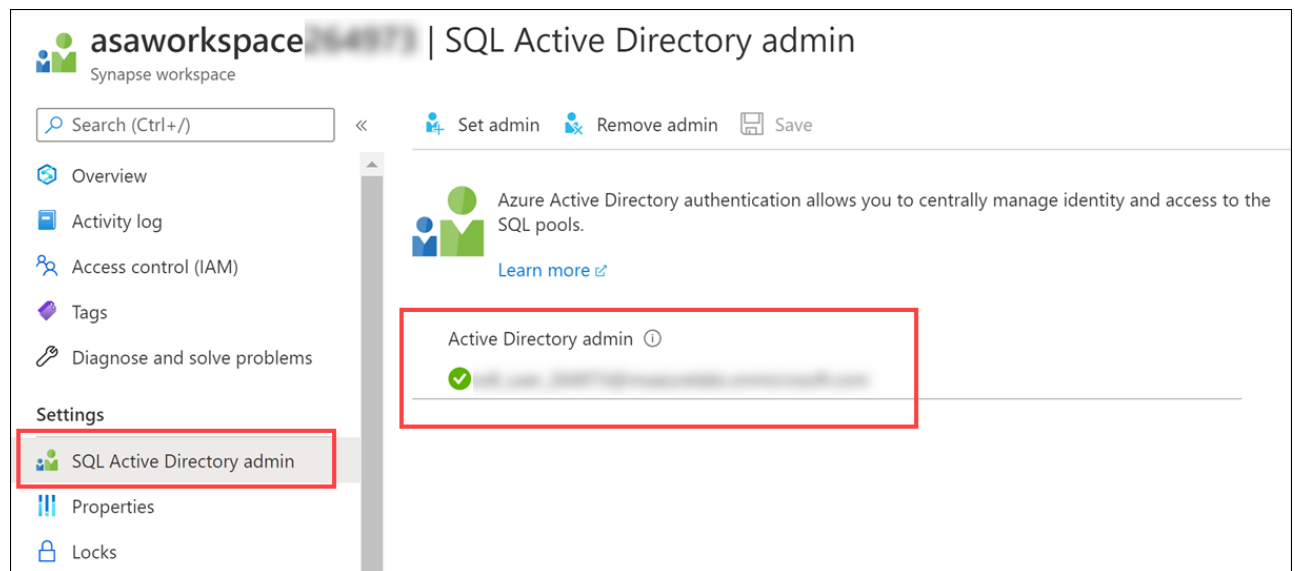
## Exercise 1 - Securing Azure Synapse Analytics supporting infrastructure

Azure Synapse Analytics (ASA) is a powerful solution that handles security for many of the resources that it creates and manages. In order to run ASA, however, some foundational security measures need to be put in place to ensure the infrastructure that it relies upon is secure. In this exercise, we will walk through securing the supporting infrastructure of ASA.

### Task 1 - Observing the SQL Active Directory admin

The SQL Active Directory Admin can be a user (the default) or group (best practice so that more than one user can be provided these permissions) security principal. The principal assigned to this will have administrative permissions to the SQL Pools contained in the workspace.

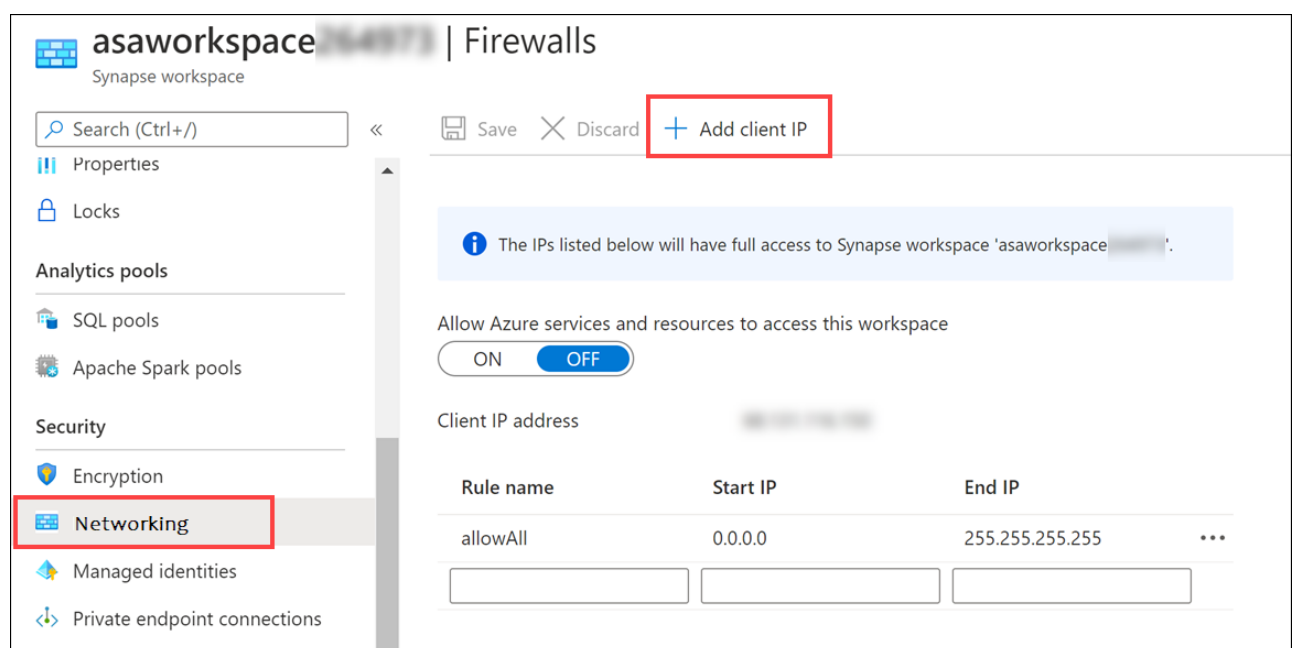
1. In the Azure Portal (<https://portal.azure.com>), browse to your lab resource group, and from the list of resources open your Synapse workspace (do not launch Synapse Studio).
2. On the left menu, select **SQL Active Directory admin** and observe who is listed as a SQL Active Directory Admin. Is it a user or group?



## Task 2 - Manage IP firewall rules

Having robust Internet security is a must for every technology system. One way to mitigate internet threat vectors is by reducing the number of public IP addresses that can access the Azure Synapse Analytics Workspace through the use of IP firewall rules. The Azure Synapse Analytics workspace will then delegate those same rules to all managed public endpoints of the workspace, including those for SQL pools and SQL Serverless endpoints.

1. In the Azure Portal, on the blade for your Synapse workspace, select **Networking**.
2. Notice that an IP Firewall rule of **Allow All** has already been created for you in the lab environment. If you wanted to add your specific IP address you would instead select **+ Add Client IP** from the taskbar menu (you do not need to do this now).



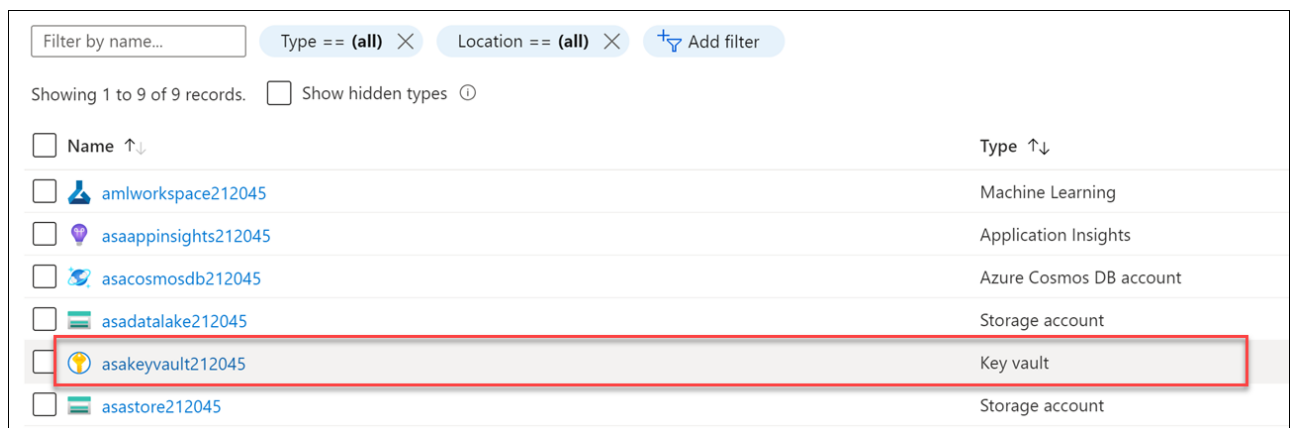
**Note:** When connecting to Synapse from your local network, certain ports need to be open. To support the functions of Synapse Studio, ensure outgoing TCP ports 80, 443, and 1143, and UDP port 53 are open.

## Exercise 2 - Securing the Azure Synapse Analytics workspace and managed services

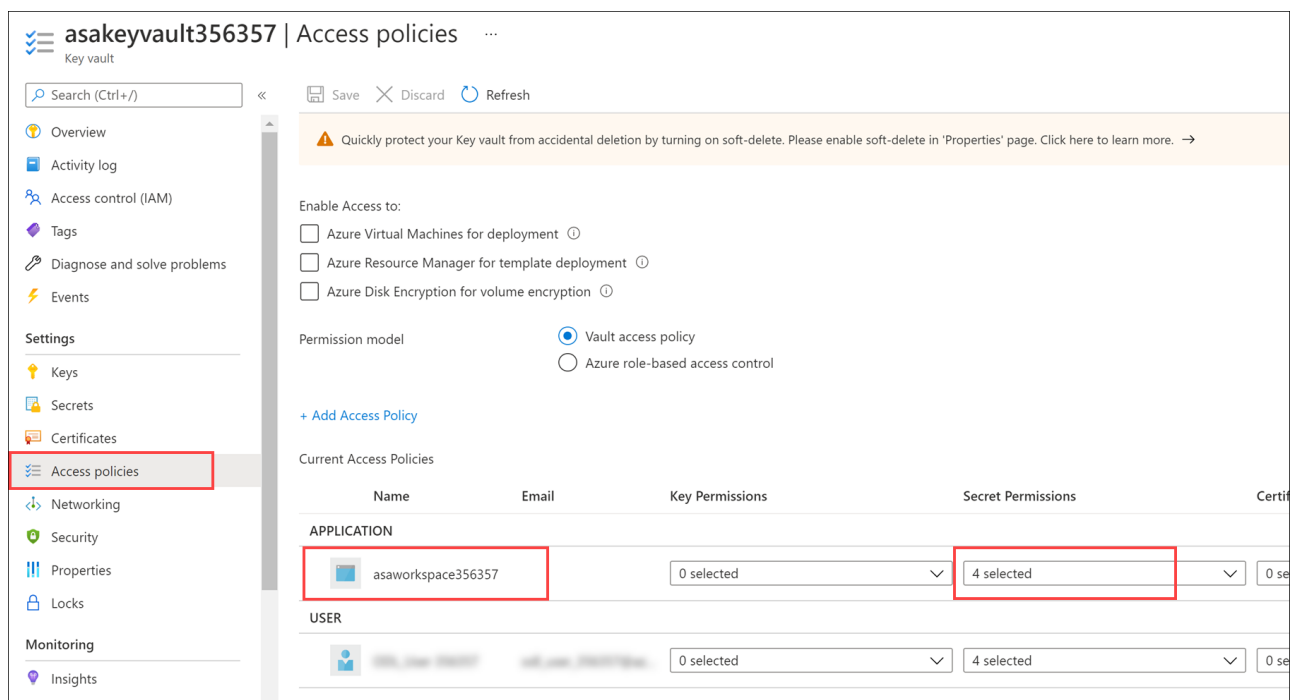
### Task 1 - Managing secrets with Azure Key Vault

When dealing with connectivity to external data sources and services, sensitive connection information such as passwords and access keys should be properly handled. It is recommended that this type of information be stored in an Azure Key Vault. Leveraging Azure Key Vault not only protects against secrets being compromised, it also serves as a central source of truth; meaning that if a secret value needs to be updated (such as when cycling access keys on a storage account), it can be changed in one place and all services consuming this key will start pulling the new value immediately. Azure Key Vault encrypts and decrypts information transparently using 256-bit AES encryption, which is FIPS 140-2 compliant.

1. In the Azure Portal, open the resource group for this lab, and from the list of resources, select the **Key vault** resource.



2. On the left menu, under Settings, select **Access Policies**.
3. Observe that Managed Service Identity (MSI) representing your Synapse workspace (it has a name similar to **asaworkspacexxxxxx**) has already been listed under Application and it has 4 selected Secret Management Operations.



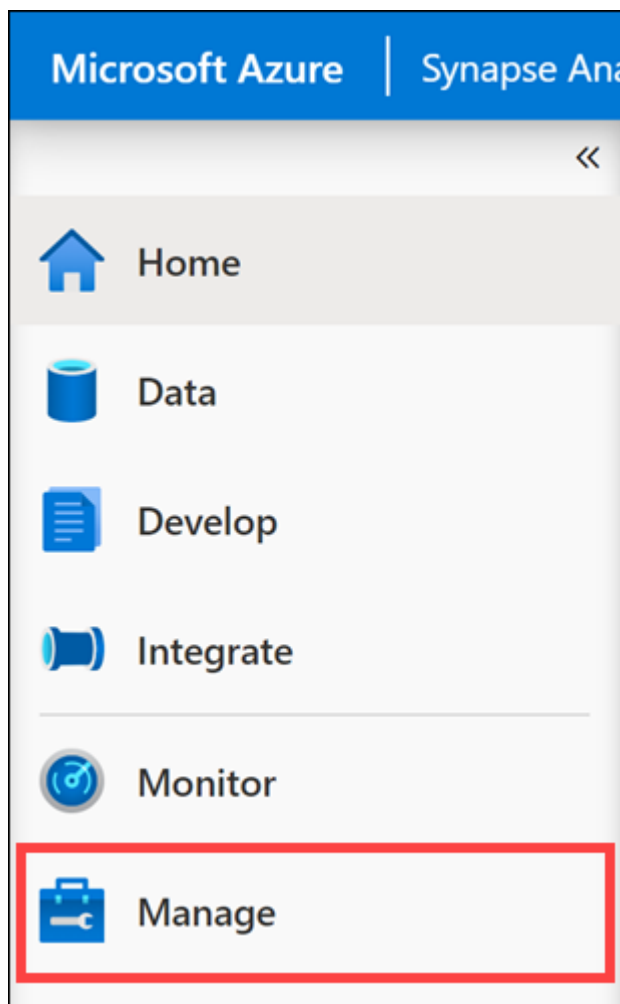
4. Select the drop-down that reads **4 selected** under **Secret Management Operations**, observe that **Get** (which allows your workspace to retrieve the values of secrets from Key Vault) and **List** (which allows your workspace to enumerate secrets) are set.

## Task 2 - Use Azure Key Vault for secrets when creating Linked Services

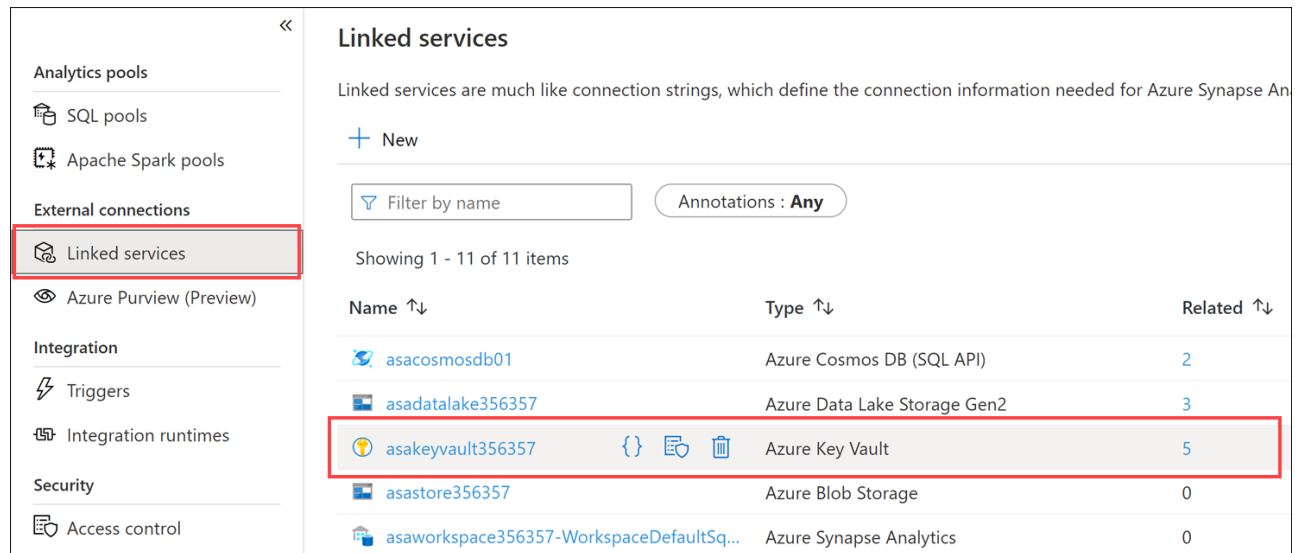
Linked Services are synonymous with connection strings in Azure Synapse Analytics. Azure Synapse Analytics linked services provides the ability to connect to nearly 100 different types of external services ranging from Azure Storage Accounts to Amazon S3 and more. When connecting to external services, having secrets related to connection information is almost guaranteed. The best place to store these secrets is the Azure Key Vault. Azure Synapse Analytics provides the ability to configure all linked service connections with values from Azure Key Vault.

In order to leverage Azure Key Vault in linked services, you must first add your key vault resource as a linked service in Azure Synapse Analytics.

1. In Azure Synapse Studio, select the **Manage** hub from the left menu.



2. Beneath **External Connections**, select **Linked Services**, observe that a Linked Service pointing to your Key Vault has been created in the environment.



**Linked services**

Linked services are much like connection strings, which define the connection information needed for Azure Synapse Analytics.

+ New

Filter by name Annotations : Any

Showing 1 - 11 of 11 items

Name ↑↓	Type ↑↓	Related ↑↓
asacosmosdb01	Azure Cosmos DB (SQL API)	2
asdatalake356357	Azure Data Lake Storage Gen2	3
asakeyvault356357	Azure Key Vault	5
asastore356357	Azure Blob Storage	0
asaworkspace356357-WorkspaceDefaultSq...	Azure Synapse Analytics	0

Since we have the Azure Key Vault set up as a linked service, we can leverage it when defining new linked services. Every New linked service provides the option to retrieve secrets from Azure Key Vault. The form requests the selection of the Azure Key Vault linked service, the secret name, and (optional) specific version of the secret.

## New linked service (Oracle)

**i** Choose a name for your linked service. This name cannot be updated later.

Name \*

Description

Connect via integration runtime \*

**i**

AutoResolveIntegrationRuntime

✓

Connection string

Azure Key Vault

AKV linked service \*

**i**

[Edit connection](#)

Secret name \*

**i**

Secret version

**i**

Use the latest version if left blank

Annotations

+ New

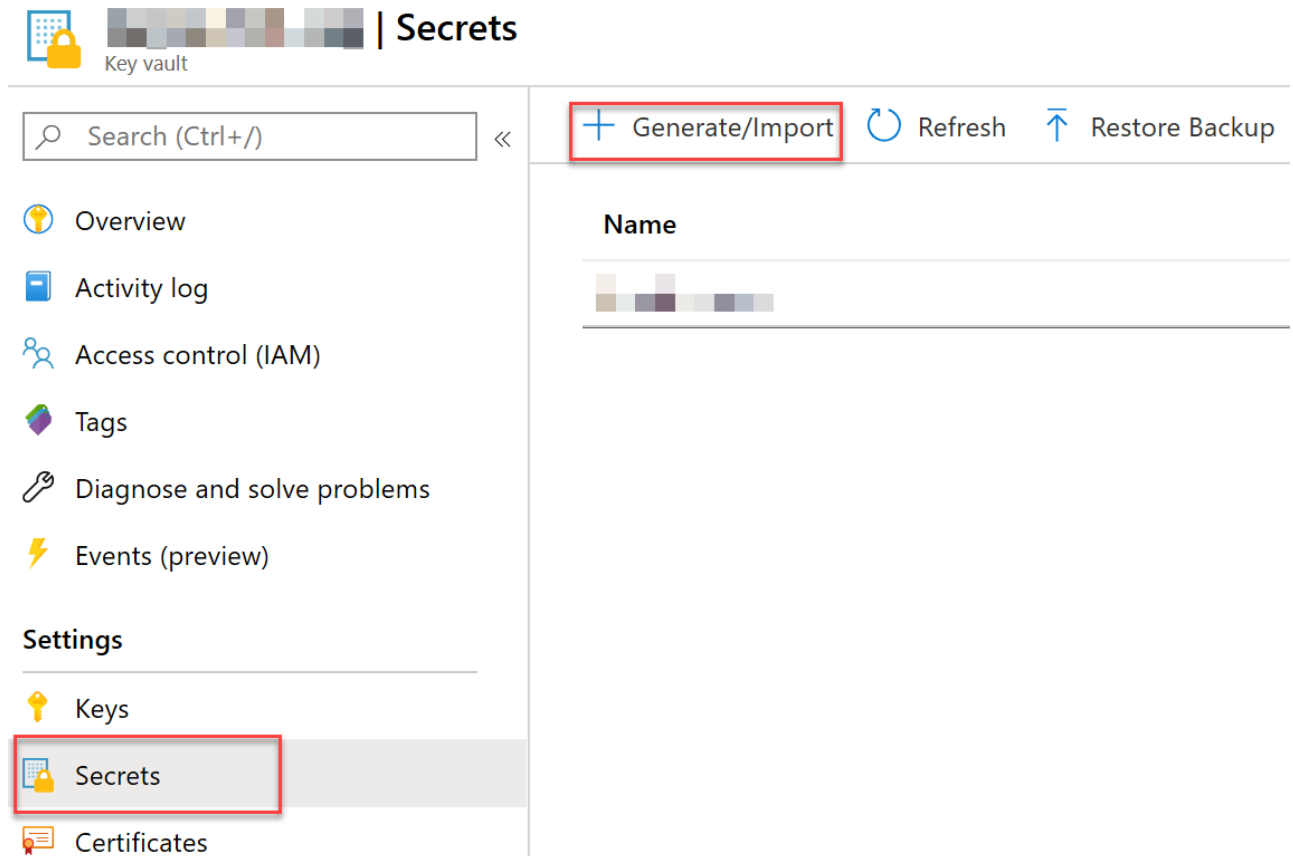
▸ Parameters

▸ Advanced **i**

### Task 3 - Secure workspace pipeline runs

It is recommended to store any secrets that are part of your pipeline in Azure Key Vault. In this task you will retrieve these values using a Web activity, just to show the mechanics. The second part of this task demonstrates using a Web activity in the pipeline to retrieve a secret from the Key Vault.

1. Return to the Azure portal.
2. In the blade for your **asakeyvaultxxxxxxx** Azure Key Vault resource, and select **Secrets** from the left menu. Then, in the top toolbar, select + **Generate/Import**.



Key vault | Secrets

Search (Ctrl+/) <<

+ Generate/Import Refresh Restore Backup

Name

Overview  
Activity log  
Access control (IAM)  
Tags  
Diagnose and solve problems  
Events (preview)

Settings

Keys  
Secrets  
Certificates

3. Create a secret, with the name `PipelineSecret` and assign it a value of `IsNotASecret`, and select the **Create** button.

## Create a secret

### Upload options

Manual

Name \* ⓘ

PipelineSecret

Value \* ⓘ

.....

Content type (optional)

Set activation date? ⓘ ☐

Set expiration date? ⓘ ☐

Enabled?

Yes

No



4. Open the secret that you just created, drill into the current version, and copy the value in the Secret Identifier field. Save this value in a text editor, or retain it in your clipboard for a future step.

Secret Version

Save Discard

Properties

Created 12/3/2020, 9:01:16 PM

Updated 12/3/2020, 9:01:16 PM

Secret Identifier

https://asakeyvault-...vault.azure.net/secrets/PipelineSecret/56e20...

Copy to clipboard

Settings

Set activation date? ⓘ

☐

Set expiration date? ⓘ

☐

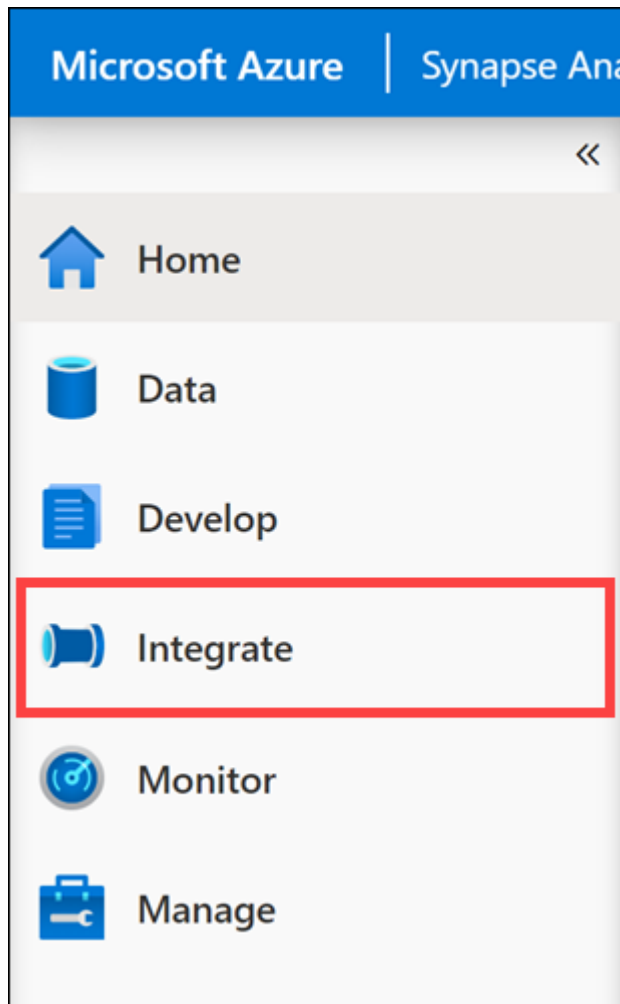
Enabled?

Yes No

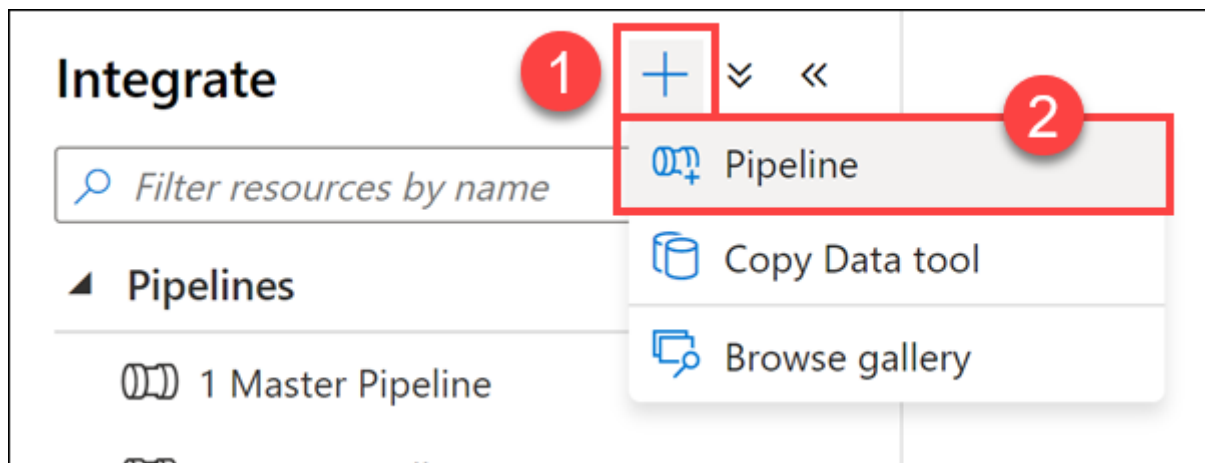
Tags

0 tags

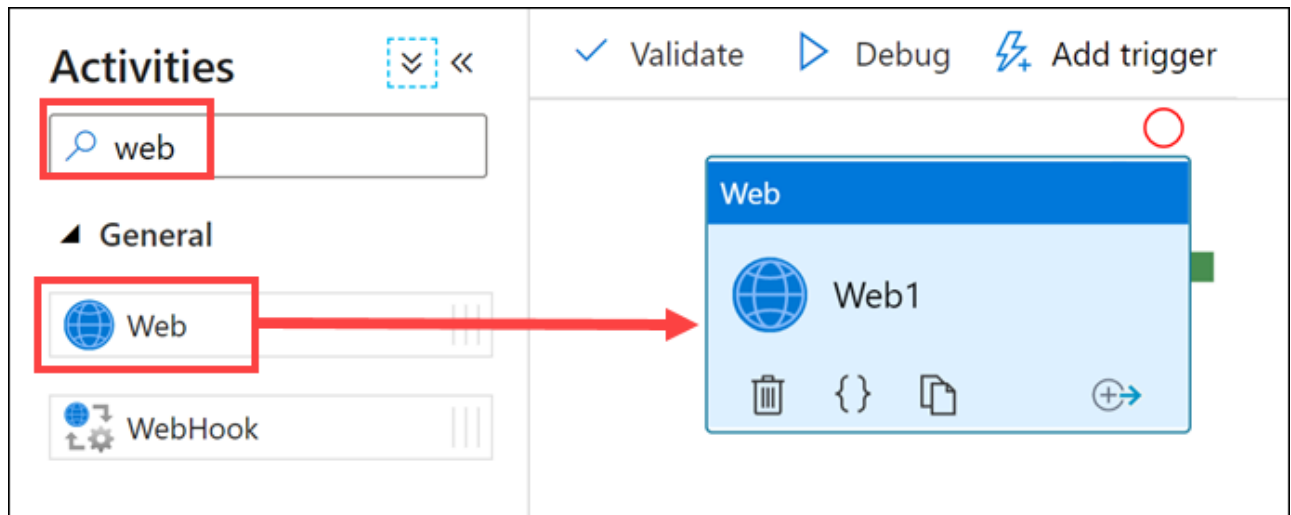
5. Switch back to Synapse Studio, then select the **Integrate** hub from the left menu.



6. On the **Integrate** pane, in the + menu, select **Pipeline**.



7. On the **Pipeline** tab, in the **Activities** pane search for **Web** and then drag an instance of a **Web** activity to the design area.



8. Select the **Web1** web activity, and select the **Settings** tab. Fill out the form as follows:

1. **URL:** Paste the Key Vault Secret Identifier value you copied in step 4 above, then **append** `?api-version=7.1` to the end of this value. For example, it should look something like:  
`https://asakeyvaultNNNNN.vault.azure.net/secrets/PipelineSecret/f808d4fa99d84861872010f6c8d25c68?api-version=7.1`.
2. **Method:** Select **Get**.
3. For **Authentication** select **Managed Identity**. We have already established an Access Policy for the Managed Service Identity of our Synapse workspace, this means that the pipeline activity has permissions to access the key vault via an HTTP call.
4. **Resource:** Enter <https://vault.azure.net>

Web

Web1

General **Settings** User properties

URL \*

Method \*

Headers [+ New](#)

Datasets [+ Add dataset reference](#)

Linked services [+ Add linked service reference](#)

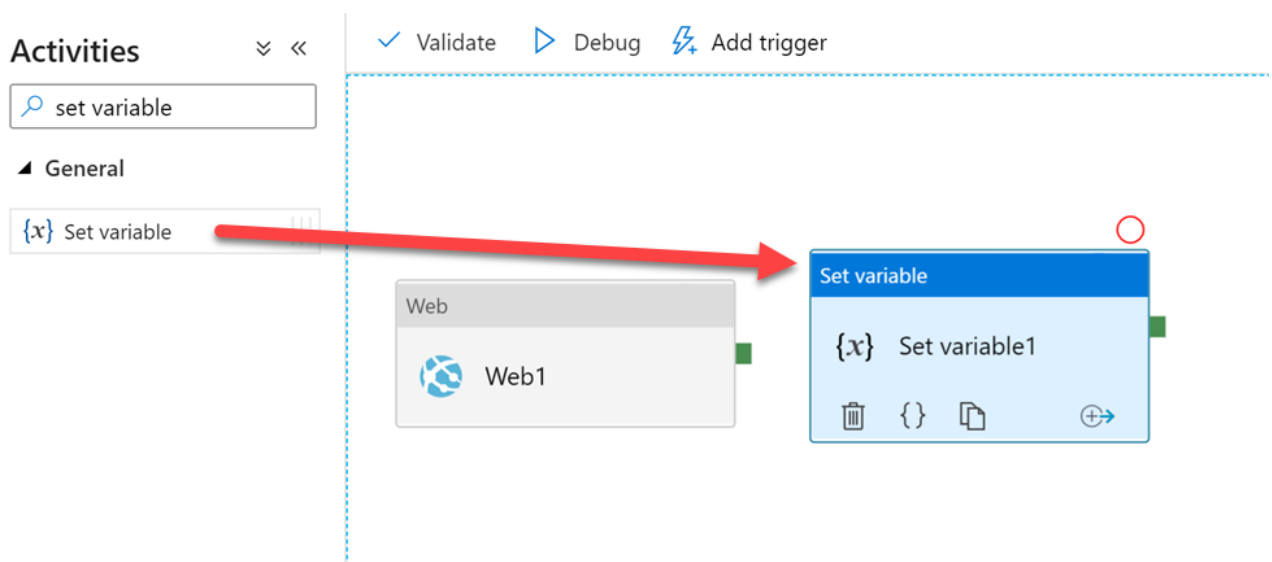
Integration runtime \*

**Advanced**

Authentication ☐ None ☐ Basic ☒ **Managed Identity** ☐ Client Certificate

Resource \*

9. From the Activities pane, add a **Set variable** activity to the design surface of the pipeline.



10. On the design surface of the pipeline, select the **Web1** activity and drag a **Success** activity pipeline connection (green box) to the **Set variable1** activity.

11. With the pipeline selected in the designer (e.g., neither of the activities are selected), select the **Variables** tab and add a new **String** parameter named **SecretValue**.

The screenshot shows a Synapse Studio pipeline with two activities: 'Web' (labeled Web1) and 'Set variable' (labeled Set variable1). The 'Variables' tab is selected, and a new variable named 'SecretValue' of type 'String' is being created with a default value of 'Value'.

12. Select the **Set variable1** activity and select the **Variables** tab. Fill out the form as follows:

1. **Name:** Select **SecretValue** (the variable that we just created).

2. **Value:** Enter `@activity('Web1').output.value`

✓ Validate   ▶ Debug   ⚙ Add trigger

The screenshot shows the 'Set variable' activity selected. The 'Variables' tab is selected, and the variable 'SecretValue' is configured with the value '@activity('Web1').output.value'.

13. Debug the pipeline by selecting **Debug** from the toolbar menu. When it runs observe the inputs and outputs of both activities from the **Output** tab of the pipeline.

✓ Validate **Debug** Add trigger

**Input**

```
{
  "variableName": "SecretValue",
  "value": "IsNotASecret"
}
```

variable1

☐

START

Input and output buttons are visible here on hover



Set variable1		SetVariable	2020-04-17T21:31
Web1		WebActivity	2020-04-17T21:31

**Note:** On the **Web1** activity, on the **General** tab there is a **Secure Output** checkbox that when checked will prevent the secret value from being logged in plain text, for instance in the pipeline run, you would see a masked value \*\*\*\*\* instead of the actual value retrieved from the Key vault. Any activity that consumes this value should also have their **Secure Input** checkbox checked.

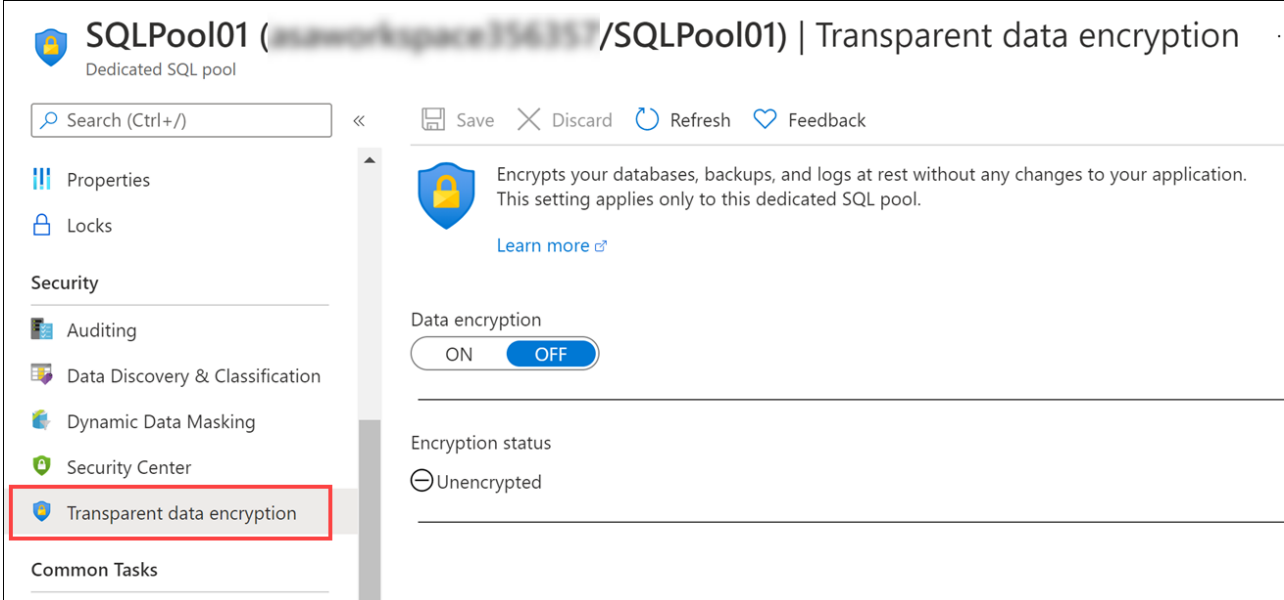
## Task 4 - Secure Azure Synapse Analytics dedicated SQL pools

Transparent Data Encryption (TDE) is a feature of SQL Server that provides encryption and decryption of data at rest, this includes: databases, log files, and back ups. When using this feature with Synapse Analytics dedicated SQL pools, it will use a built-in symmetric Database Encryption Key (DEK) that is provided by the pool itself. With TDE, all stored data is encrypted on disk, when the data is requested, TDE will decrypt this data at the page level as it's read into memory, and vice-versa encrypting in-memory data before it gets written back to disk. As with the name, this happens transparently without affecting any application code. When creating a dedicated SQL pool through Synapse Analytics, Transparent Data Encryption is not enabled. The first part of this task will show you how to enable this feature.

1. In the **Azure Portal**, open your resource group, then locate and open the **SqlPool01** dedicated SQL pool resource.

<input type="checkbox"/>	 asaworkspace264973	Synapse workspace
<input type="checkbox"/>	 SparkPool01 (asaworkspace264973/SparkPool01)	Apache Spark pool
<input type="checkbox"/>	 SQLPool01 (asaworkspace264973/SQLPool01)	Dedicated SQL pool

- On the **SQL pool** resource blade, select **Transparent data encryption** from the left-hand menu. **DO NOT** turn on data encryption.



**SQLPool01** (asaworkspace264973/SQLPool01) | Transparent data encryption

Dedicated SQL pool

Search (Ctrl+ /) Save Discard Refresh Feedback

Properties  
Locks  
Security  
Auditing  
Data Discovery & Classification  
Dynamic Data Masking  
Security Center  
**Transparent data encryption**  
Common Tasks

Encrypts your databases, backups, and logs at rest without any changes to your application. This setting applies only to this dedicated SQL pool.  
[Learn more](#)

Data encryption  
ON OFF

Encryption status  
Unencrypted

By default, this option is turned off. When you enable data encryption on this dedicated SQL pool, the pool is taken offline for a few minutes while TDE is applied.

## Exercise 3 - Securing Azure Synapse Analytics workspace data

### Task 1 - Column Level Security

It is important to identify data columns that hold sensitive information. Types of sensitive could be social security numbers, email addresses, credit card numbers, financial totals, and more. Azure Synapse Analytics allows you define permissions that prevent users or roles select privileges on specific columns.

- In **Azure Synapse Studio**, in the **Develop** hub, expand the **SQL scripts** section, and select **Column Level Security**.
- In the toolbar, connect to the **SQLPool01** database.
- In the query window, **run each step individually** by highlighting the statement(s) in the step in the query window, and selecting the **Run** button from the toolbar (or press **F5**).
- Close the script tab. If prompted select **Discard all changes**.

### Task 2 - Row level security

- In the **Develop** hub, in the **SQL scripts** section, select **Row Level Security**.
- In the toolbar, connect to the **SQLPool01** database.
- In the query window, **run each step individually** by highlighting the statement(s) for the step in the query window, and selecting the **Run** button from the toolbar (or press **F5**).
- Close the script tab. If prompted select **Discard all changes**.

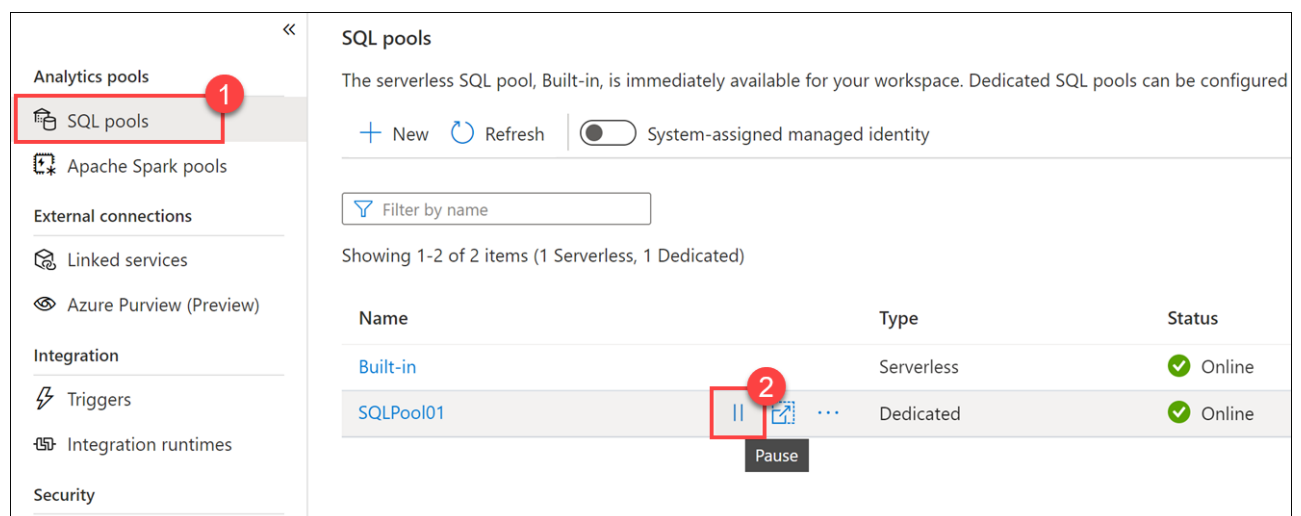
## Task 3 - Dynamic data masking

1. In the **Develop** hub, in the **SQL scripts** section, select **Dynamic Data Masking**.
2. In the toolbar, connect to the **SQLPool01** database.
3. In the query window, **run each step individually** by highlighting the statement(s) for the step in the query window, and selecting the **Run** button from the toolbar (or press **F5**).
4. Close the script tab. If prompted select **Discard all changes**.

## Important: Pause your SQL pool

Complete these steps to free up resources you no longer need.

1. In Synapse Studio, select the **Manage** hub.
2. Select **SQL pools** in the left-hand menu. Hover over the **SQLPool01** dedicated SQL pool and select **||**.



3. When prompted, select **Pause**.

## Reference

- [IP Firewalls](#)
- [Synapse Workspace Managed Identity](#)
- [Synapse Managed VNet](#)
- [Synapse Managed Private Endpoints](#)
- [Secure your Synapse Workspace](#)
- [Connect to your Synapse Workspace using private links](#)
- [Create a Managed private endpoint to your data source](#)
- [Granting Permissions to Workspace Managed Identity](#)

## Other Resources

- [Managing access to workspaces, data and pipelines](#)
- [Analyze with Apache Spark](#)
- [Visualize data with Power BI](#)
- [Control storage account access for SQL on-demand](#)