

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

Wide World Importers is host to a plethora of data coming from many disparate sources. The idea of bringing all of their data together into Azure Synapse Analytics for them to query, gain insights, and consume in ways they have never done before is exhilarating! As much as it is an exciting game-changer for this business, it opens up a large amount of surface area for potential attack. Security must be established in the forefront at the time of design of this solution.

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.

### Resource naming throughout this lab

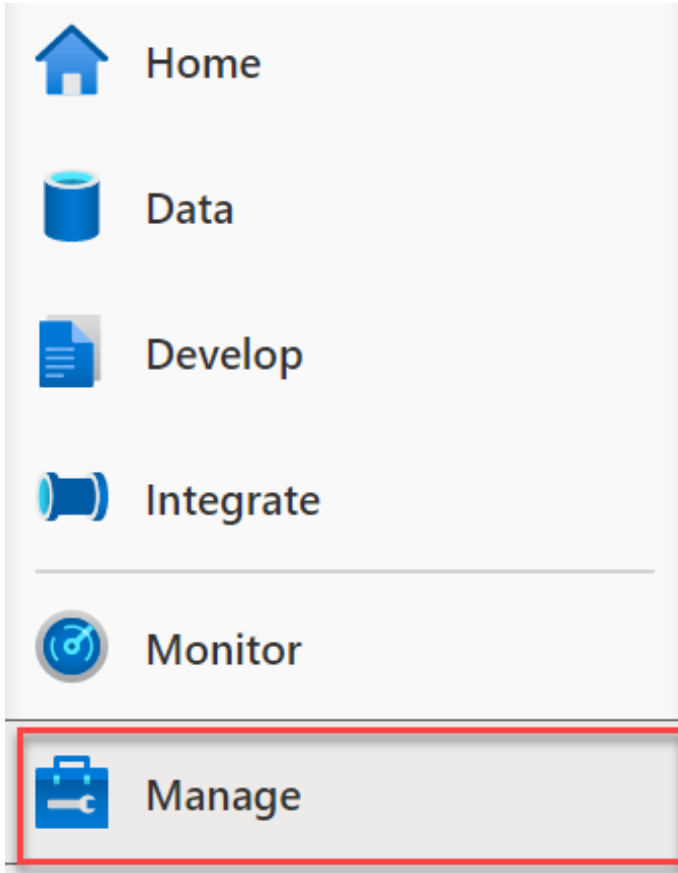
For the remainder of this guide, the following terms will be used for various ASA-related resources (make sure you replace them with actual names and values):

Azure Synapse Analytics Resource	To be referred to
Workspace resource group	WorkspaceResourceGroup
Workspace / workspace name	Workspace
Primary Storage Account	PrimaryStorage
Default file system container	DefaultFileSystem
SQL Pool	SqlPool01
SQL Serverless Endpoint	SqlServerless01
Active Directory Principal of New User	user@domain.com
SQL username of New User	newuser
Azure Key Vault	KeyVault01
Azure Key Vault Private Endpoint Name	KeyVaultPrivateEndpointName
Azure Subscription	WorkspaceSubscription
Azure Region	WorkspaceRegion

## Lab pre-requisite

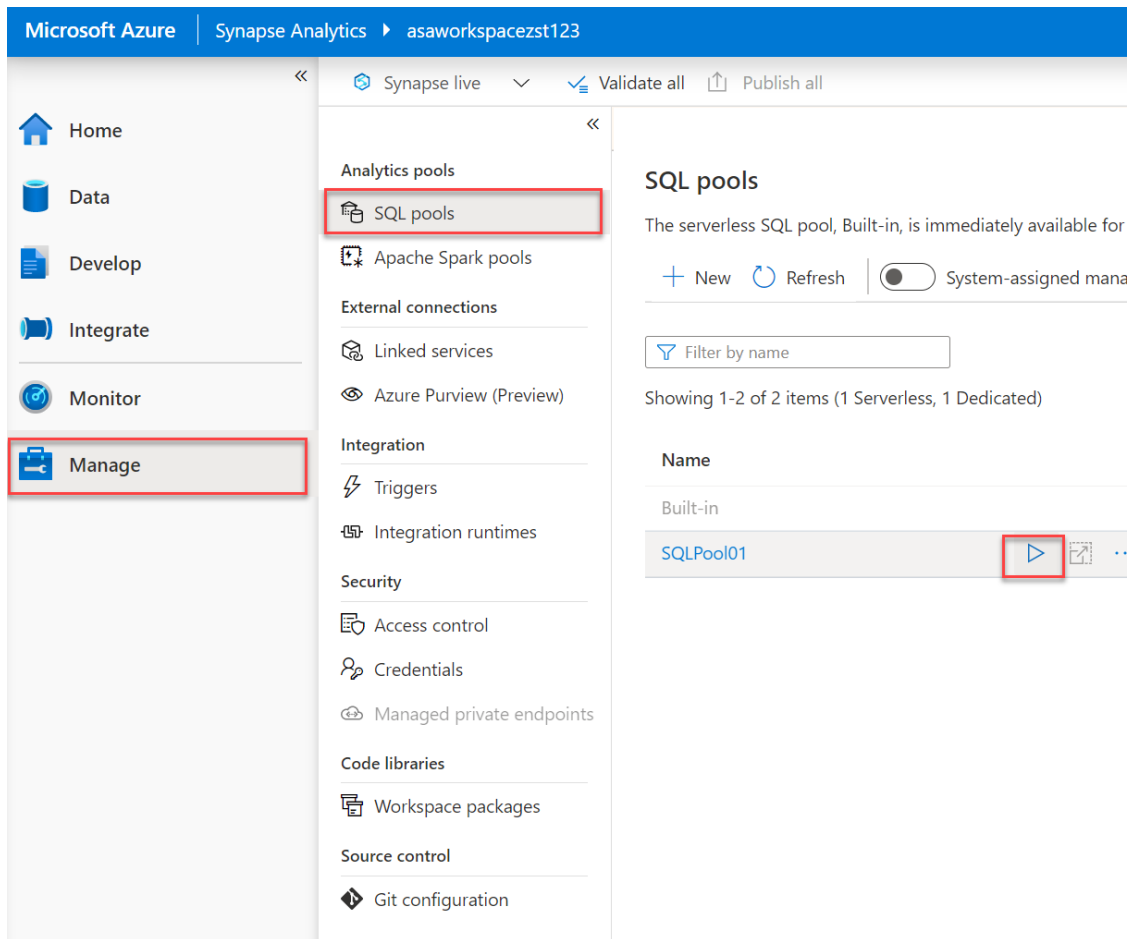
Start the SQL Pool in your lab environment.

1. Open the Synapse Studio workspace and navigate to the **Manage** hub.



*The Manage menu item is highlighted.*

2. From the center menu, select **SQL pools** from beneath the **Analytics pools** heading. Locate SQLPool01, and select the **Resume** button.



*The Manage menu item is selected, with SQL pools selected from the center menu. The resume button is selected next to the SQLPool01 item.*

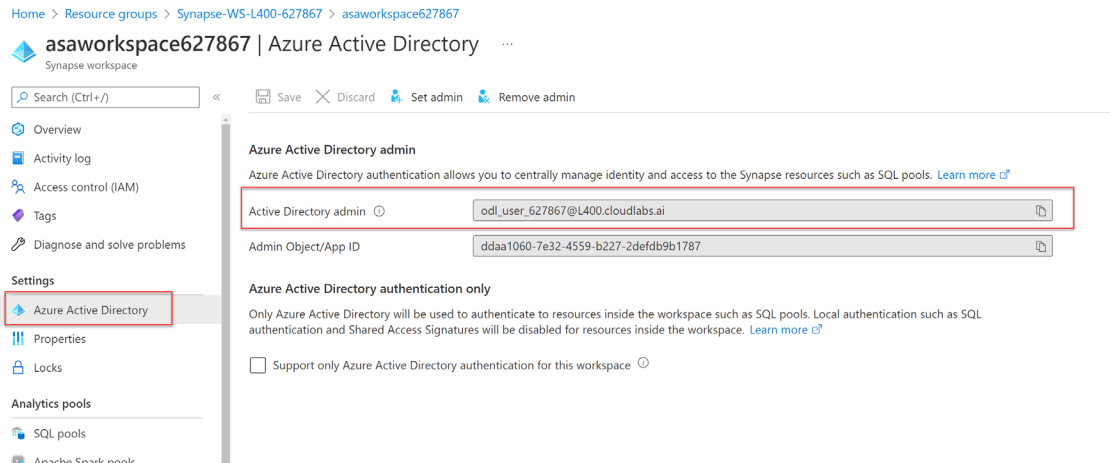
## Exercise 1 - Securing Azure Synapse Analytics supporting infrastructure

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**, browse to your L400 resource group and from the list of resources open your Synapse workspace (do not launch Synapse Studio).
2. From the left menu, select **Azure Active Directory** and observe who is listed as the Active Directory Admin. Is it a user or group?

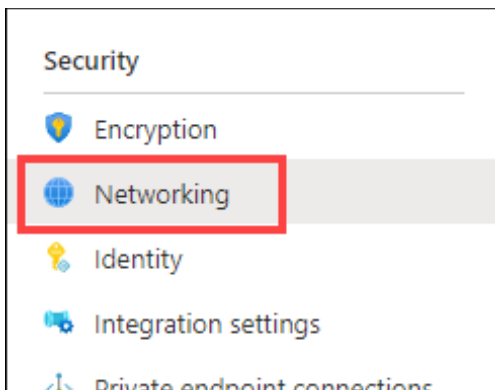


*On the SQL Active Directory Admin screen, SQL Active Directory admin is selected from the left menu, and the Active Directory Admin is highlighted.*

## Task 2 - Manage IP firewall rules

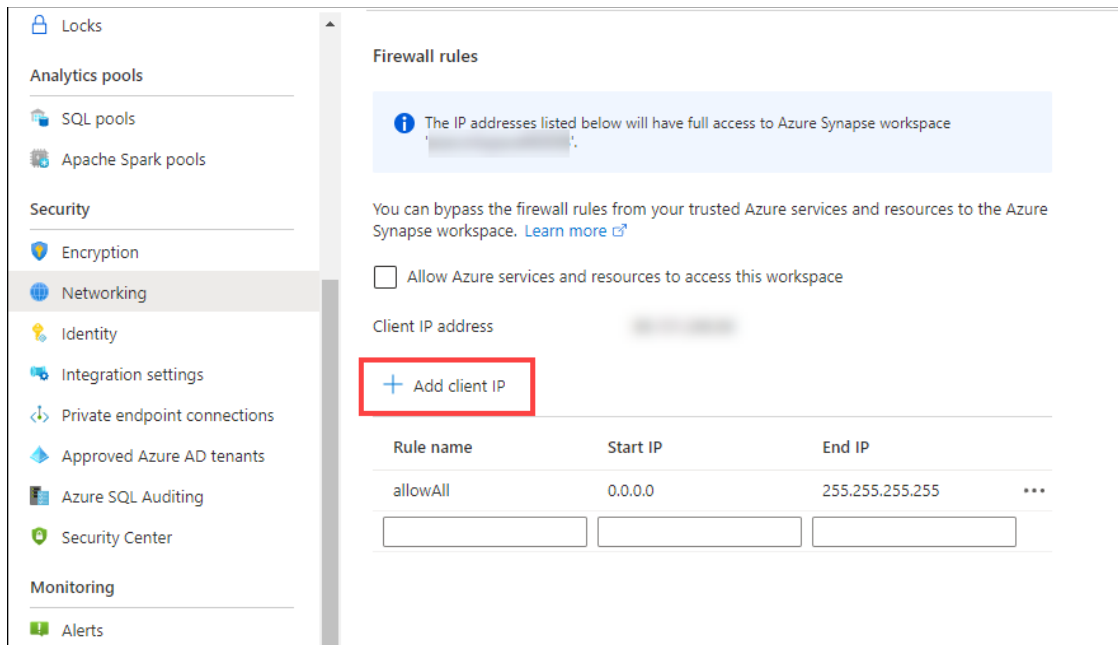
Having robust Internet security is a must for every technology system. One way to mitigate internet threat vectors is by reducing the number of public IP addresses that can access the Azure Synapse Analytics Workspace through the use of IP firewall rules. The Azure Synapse Analytics workspace will then 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**, open the Synapse workspace (do not launch Studio).
2. From the left menu of the **Azure Synapse Analytics** page, select **Networking**.



*The Synapse Workspace screen is displayed, with the Firewalls item selected from the left menu.*

3. 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 should not do this in this lab).



*On the Synapse Workspace screen, the + Add Client IP button is selected from the toolbar menu.*

**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 Synapse-WS-L400-SUFFIX resource group and from the list of resources and select the **Key vault** resource.

Filter by name...		Type == (all) X	Location == (all) X	+ Add filter
Showing 1 to 9 of 9 records. <input type="checkbox"/> Show hidden types ⓘ				
<input type="checkbox"/>	Name ↑↓			Type ↑↓
<input type="checkbox"/>	amlworkspace212045			Machine Learning
<input type="checkbox"/>	asaappinsights212045			Application Insights
<input type="checkbox"/>	asacosmosdb212045			Azure Cosmos DB account
<input type="checkbox"/>	asadatalake212045			Storage account
<input checked="" type="checkbox"/>	asakeyvault212045			Key vault
<input type="checkbox"/>	asastore212045			Storage account

*Key vault is selected in the resource group.*

- From the left menu, under Settings, select **Access Policies**.
- Observe that Managed Service Identity (MSI) representing your Synapse workspace (it has a name similar to asaworkspaceSUFFIX) has already been listed under Application and it has 4 selected Secret Permissions.
- Select the drop-down that reads 4 selected under Secret Permissions, 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 asakeyvaultSUFFIX as a linked service in Azure Synapse Analytics.

- In **Azure Synapse Studio** (<https://web.azuresynapse.net/>), select **Manage** from the left menu.
- Beneath **External Connections**, select **Linked Services**, observe that a Linked Service pointing to your Key Vault has been provided in the environment.

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

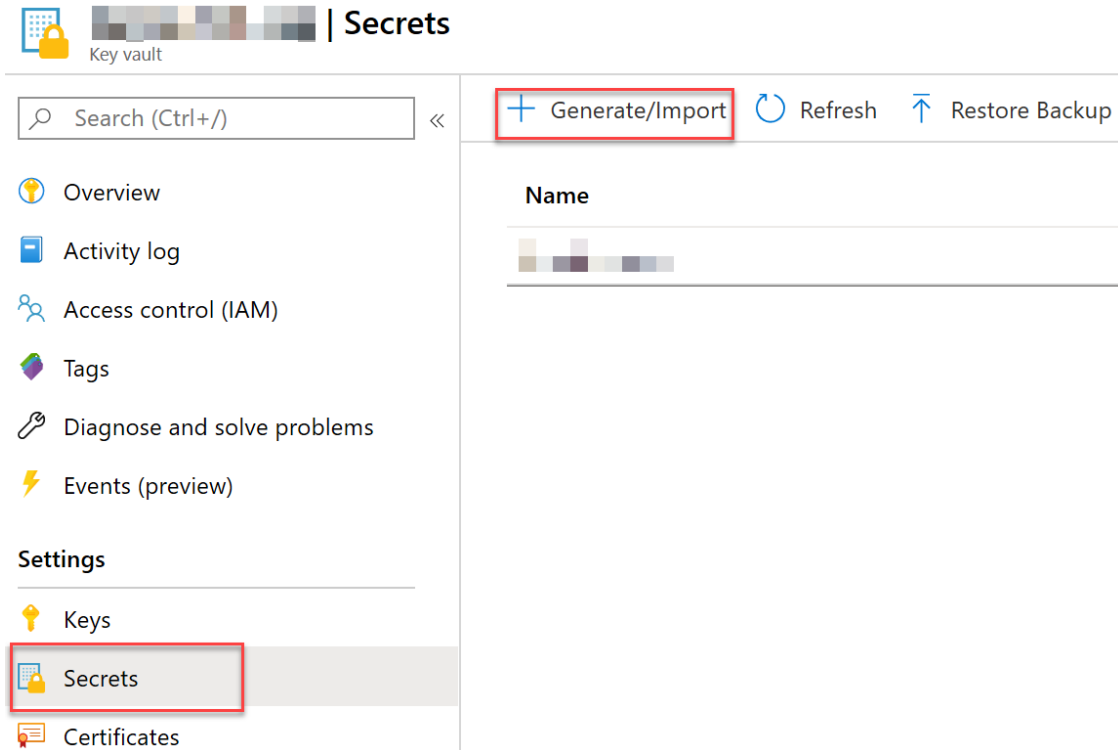
▸ Advanced **i**

*A New linked service form is displayed with the Azure Key Vault setting highlighted with the fields described in the preceding paragraph.*

### 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. Open the `asakeyvaultsUFFIX` Azure Key Vault resource, and select **Secrets** from the left menu. From the top toolbar, select **+ Generate/Import**.



*In Azure Key Vault, Secrets is selected from the left menu, and + Generate/Import is selected from the top toolbar.*

2. 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

*The Create a secret form is displayed populated with the specified values.*

3. 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 4/17/2020, 3:38:44 PM

Updated 4/17/2020, 3:38:44 PM

### Secret Identifier

https://[redacted].vault.azure.net/secrets/PipelineSecret/[redacted]



### Settings


Set activation date? ⓘ ☐

Set expiration date? ⓘ ☐

Enabled? Yes No

#### Tags

0 tags




### Secret

Content type (optional)

Show Secret Value

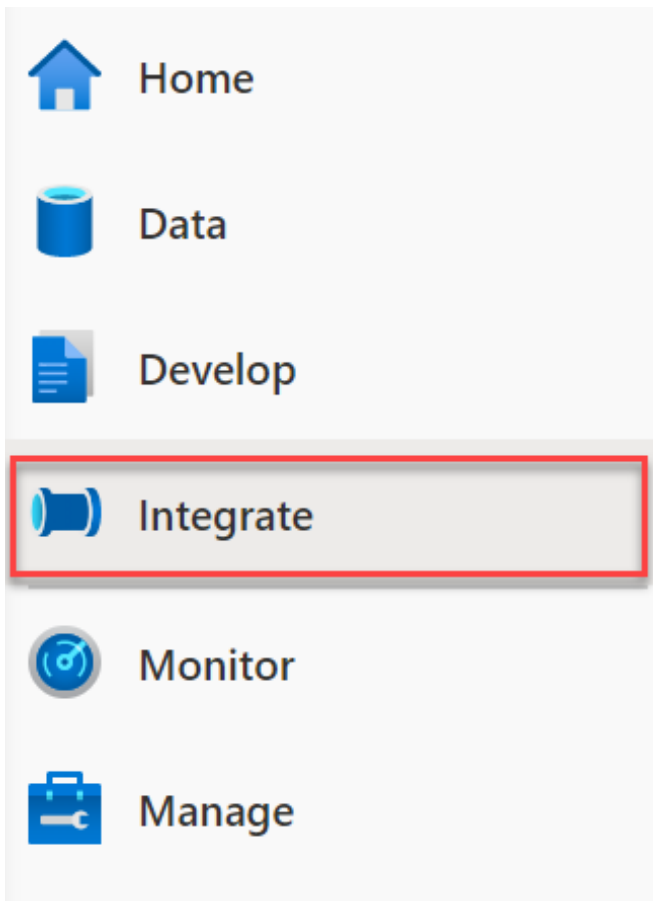
### Secret value

.....



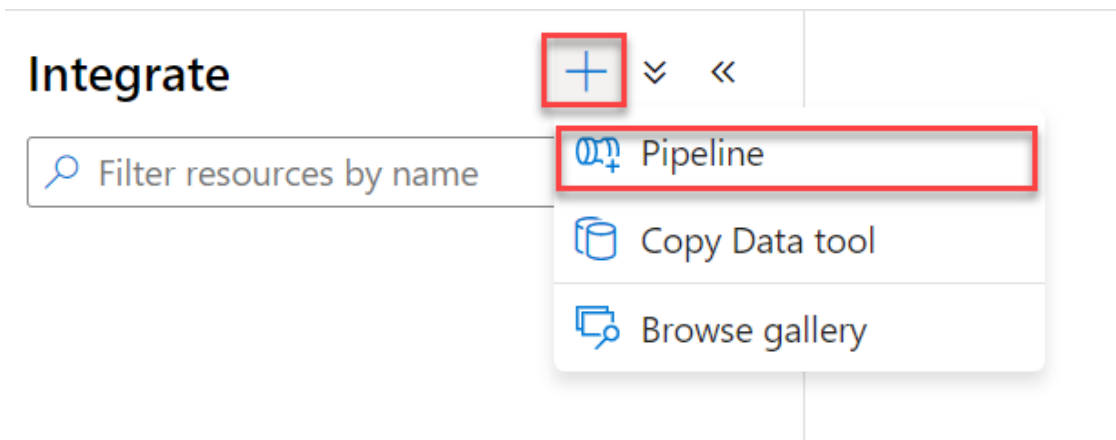
*On the Secret Version form, the Copy icon is selected next to the Secret Identifier text field.*

4. Open the Azure Synapse Analytics Studio, select **Integrate** from the left menu.



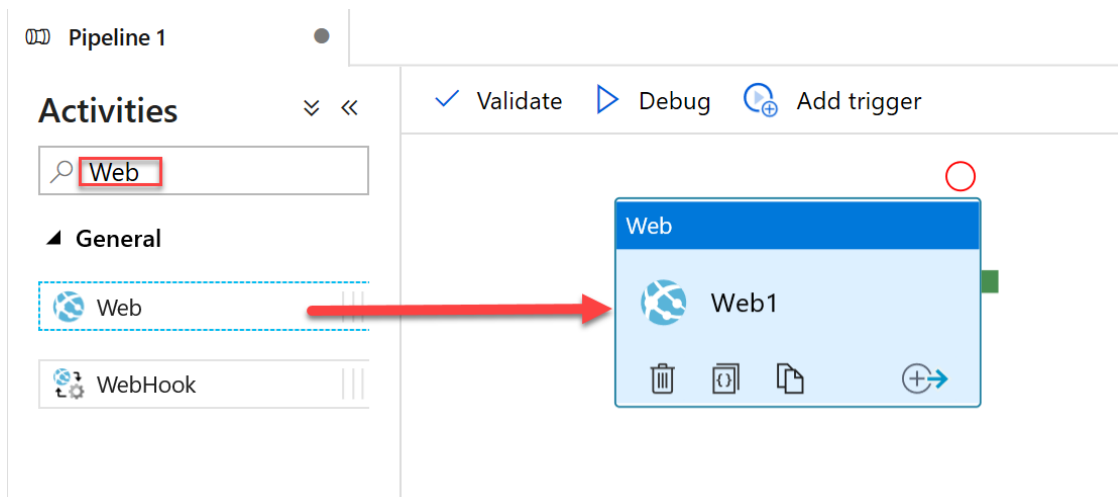
*The Azure Synapse Analytics Studio left menu is displayed with the Integrate item selected.*

5. From the **Integrate** blade, select the **+** button and add a new **Pipeline**.



*On the Integrate blade the + button is expanded with the Pipeline item selected beneath it.*

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



*In the Activities pane, Web is entered into the search field. Under General, the Web activity is displayed in the search results. An arrow indicates the drag and drop movement of the activity to the design surface of the pipeline. The Web activity is displayed on the design surface.*

7. Select the **Web1** web activity, and select the **Settings** tab. Fill out the form as follows:
  1. **URL:** Paste the Secret Identifier value for the secret **append** ?api-version=7.0 to this value.
  2. **Method:** Select **Get**.
  3. Expand the **Advanced** section, and 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>

The screenshot shows the 'Web' activity settings in Azure Data Studio. The 'Settings' tab is selected. The 'URL' is 'https://asakeyvault460008.vault.azure.net/s', 'Method' is 'GET', 'Integration runtime' is 'AutoResolveIntegrationRuntime', and 'Authentication' is 'Managed Identity'. The 'Resource' is 'https://vault.azure.net'. The 'Advanced' section is expanded.

Web

Web1

General **Settings** User properties

URL \*

Method \*

Headers [+ New](#)

Datasets [+ Add dataset reference](#)

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

Integration runtime \*  [Edit](#)

Disable certificate validation ☐

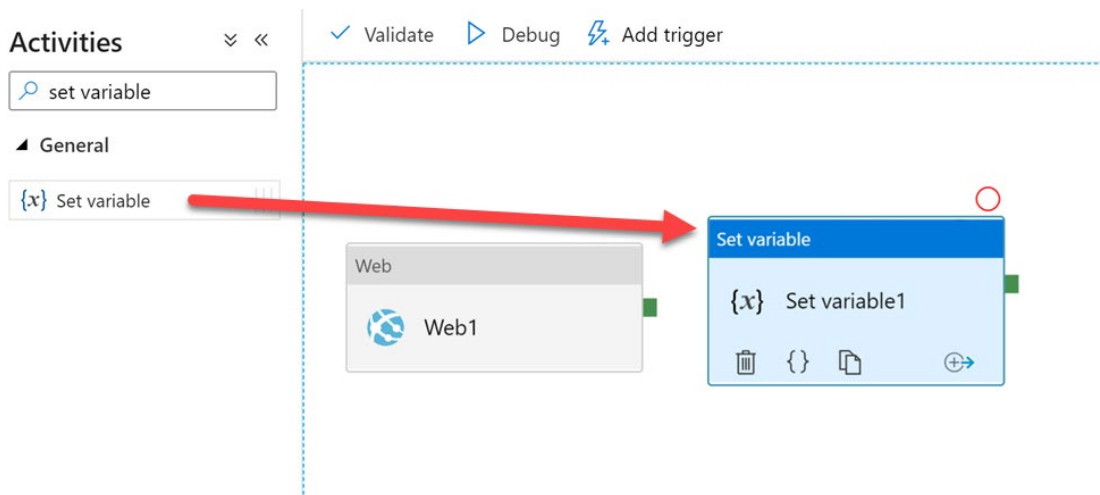
**Advanced**

Authentication ☐ None ☐ Basic ☒ Managed Identity ☐ Client Certificate

Resource \*

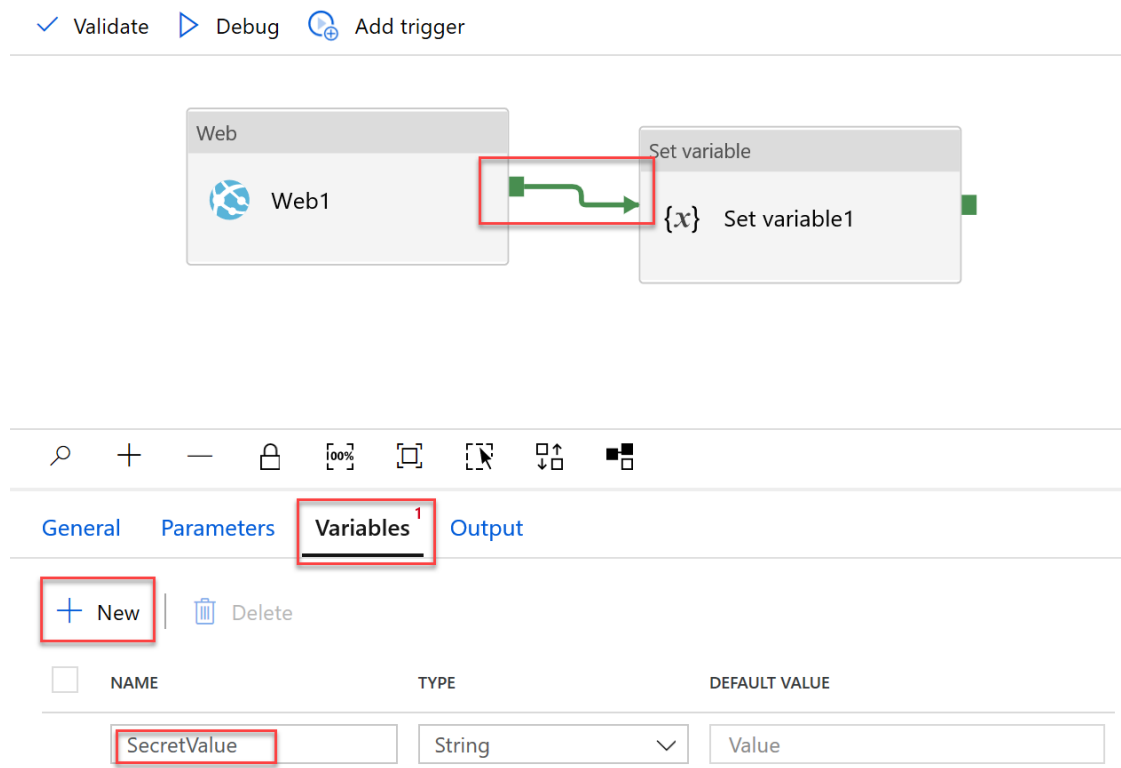
*The Web Activity Settings tab is selected and the form is populated with the values indicated above.*

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



*An arrow goes from the set variable item under Activities to the pipeline canvas.*

9. 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.
10. 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**.



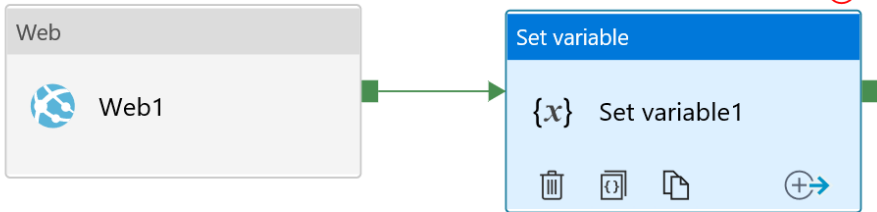
*The design surface of the pipeline is shown with a new pipeline arrow connecting the Web1 and Set variable1 activities. The pipeline is selected, and beneath the design*

surface, the **Variables** tab is selected with a variable with the name of **SecretValue** highlighted.

11. 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 on our pipeline).
2. **Value:** Enter `@activity('Web1').output.value`

✓ Validate ▶ Debug ⚙️ Add trigger



General **Variables** User properties

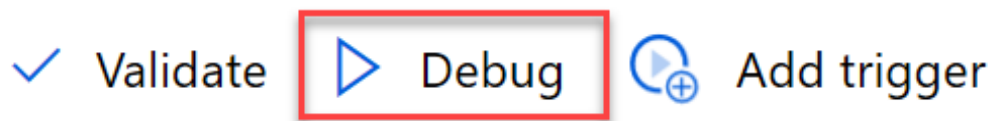
**Name \*** SecretValue ▼

**Value \*** @activity('Web1').output.value

[Add dynamic content \[Alt+P\]](#)

On the pipeline designer, the **Set Variable1** activity is selected. Below the designer, the **Variables** tab is selected with the form set the previously specified values.

12. 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.



The pipeline toolbar is displayed with the **Debug** item highlighted.

Input

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

variable1

START

Set variable1

Web1

SetVariable

WebActivity

2020-04-17T21:31

2020-04-17T21:31

Input and output buttons are visible here on hover

*In the output of the pipeline, the Set variable 1 activity is selected with its input displayed. The input shows the value of NotASecret that was pulled from the key vault being assigned to the SecretValue pipeline variable.*

**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 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 ASA 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 SQL Pool through ASA, 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"/> Name ↑↓	Type ↑↓
<input type="checkbox"/>  amiworkspace460008	Machine learning
<input type="checkbox"/>  asaappinsights460008	Application Insights
<input type="checkbox"/>  asacosmosdb460008	Azure Cosmos DB account
<input type="checkbox"/>  asadatalake460008	Storage account
<input type="checkbox"/>  asakeyvault460008	Key vault
<input type="checkbox"/>  asastore460008	Storage account
<input type="checkbox"/>  asaworkspace460008	Synapse workspace
<input type="checkbox"/>  SparkPool01 (asaworkspace460008/SparkPool01)	Apache Spark pool
<input type="checkbox"/>  SQLPool01 (asaworkspace460008/SQLPool01)	Dedicated SQL pool

*The SQLPool01 resource is highlighted in the Resource Group.*

2. On the **SQL pool** resource screen, select **Transparent data encryption** from the left



SQL pool



Search (Ctrl+ /)



Overview



Activity log



Access control (IAM)



Tags

## Settings



Maintenance schedule



Geo-backup policy



Connection strings



Properties



Locks



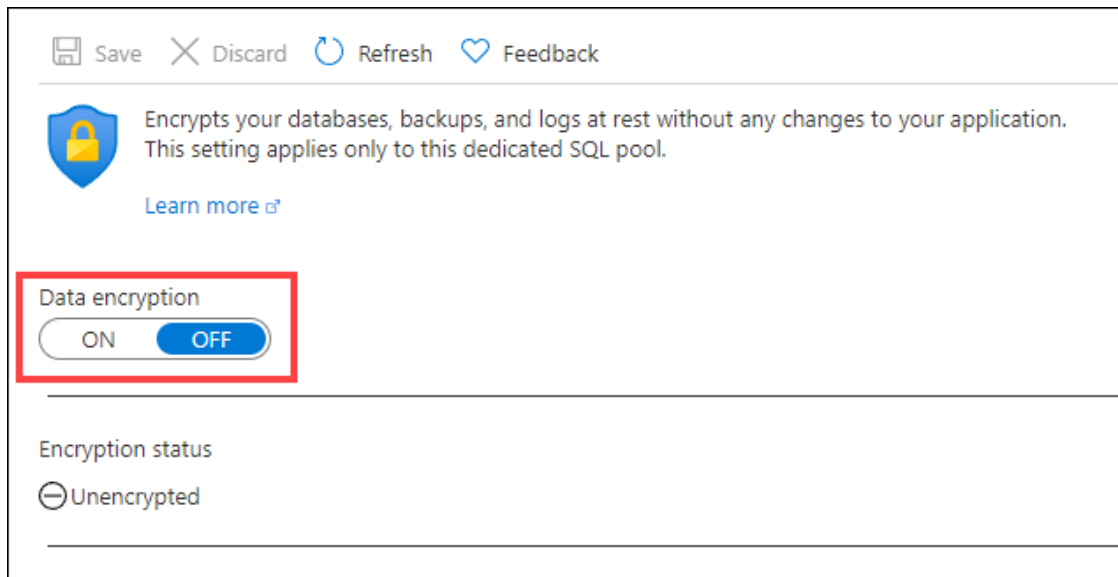
Export template

## Security



Transparent data encryption

3. This is where you see if your database is taking advantage of TDE. For the sake of time, we will **leave TDE off**.



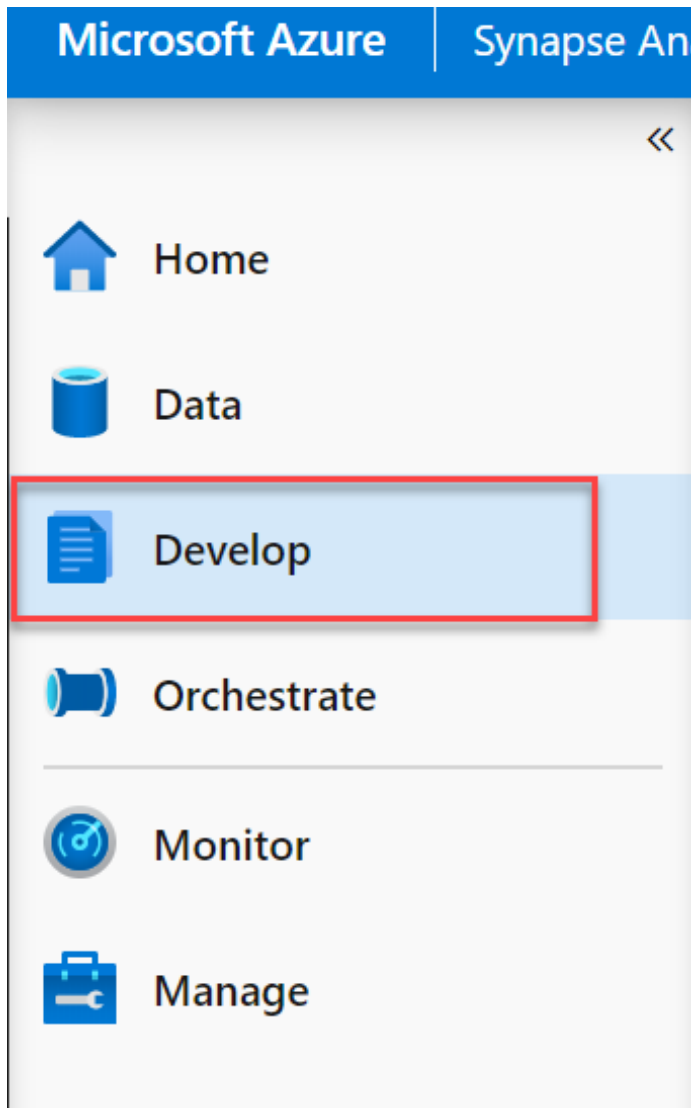
*On the SQL Pool Transparent Data Encryption screen, the Data Encryption toggle is set to the ON position and the Save button is highlighted in the toolbar.*

## Exercise 3 - Securing Azure Synapse Analytics workspace data

### Task 1 - Column Level Security

It is important to identify data columns of 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.

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



*In Azure Synapse Studio, the Develop item is selected from the left menu.*

2. From the **Develop** menu, expand the **SQL scripts** section, and select **Lab 05 - Exercise 3 - Column Level Security**.

## Develop




 Filter resources by name

### SQL scripts



 ASAL400 - Lab 05 - Exercise 3 - Column Level Security

 ASAL400 - Lab 05 - Exercise 3- Dynamic Data Masking

 ASAL400 - Lab 05 - Exercise 3 - Row Level Security



*In Synapse Studio the develop menu is displayed with SQL Scripts section expanded, Lab 05 - Exercise 3 - Column Level Security is selected from the context menu.*

3. In the toolbar menu, connect to the database on which you want to execute the query, SQLPool01.



*The Synapse Studio query toolbar is shown with the Connect to dropdown list field highlighted.*

4. 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.

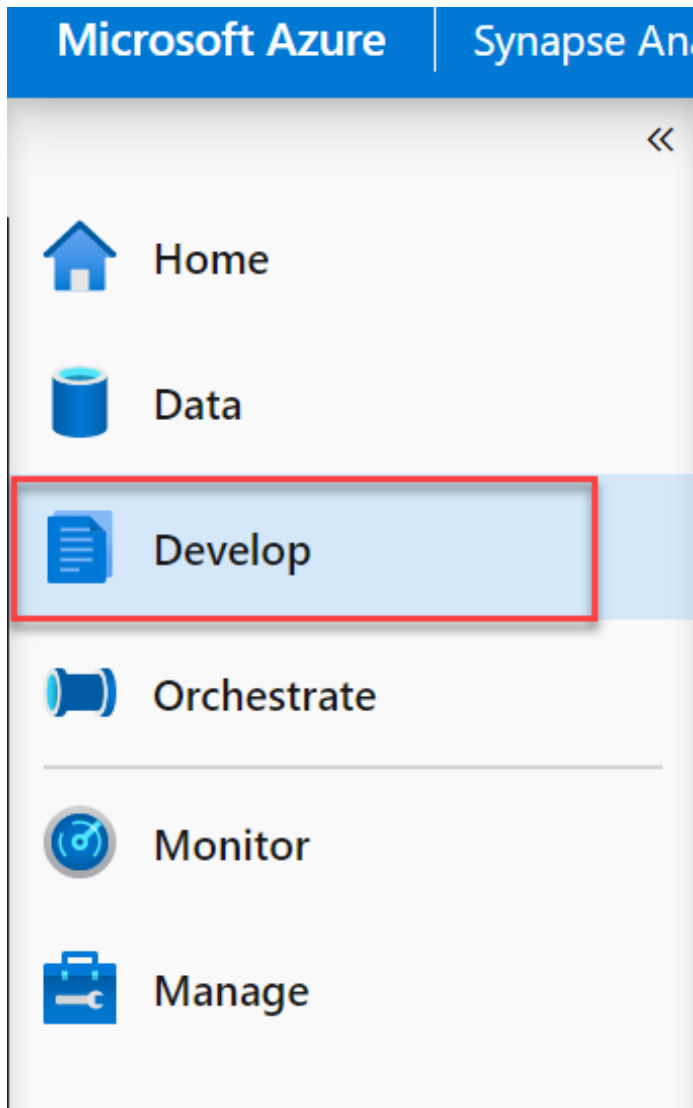


*The Synapse Studio toolbar is displayed with the Run button selected.*

5. You may now close the script tab, when prompted choose to **Discard all changes**.

## **Task 2 - Row level security**

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



*In Azure Synapse Studio, the Develop item is selected from the left menu.*

2. From the **Develop** menu, expand the **SQL scripts** section, and select **Lab05 - Exercise 3 - Row Level Security**.




## Develop





 Filter resources by name

### SQL scripts



 ASAL400 - Lab 05 - Exercise 3 - Column Level Security

 ASAL400 - Lab 05 - Exercise 3- Dynamic Data Masking

 ASAL400 - Lab 05 - Exercise 3 - Row Level Security



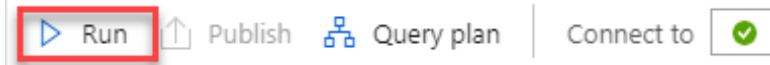
*In Synapse Studio the develop menu is displayed with SQL Scripts section expanded, Lab05 - Exercise 3 - Row Level Security is selected from the context menu.*

3. In the toolbar menu, connect to the database on which you want to execute the query, SQLPool01.



*The Synapse Studio query toolbar is shown with the Connect to dropdown list field highlighted.*

4. 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.

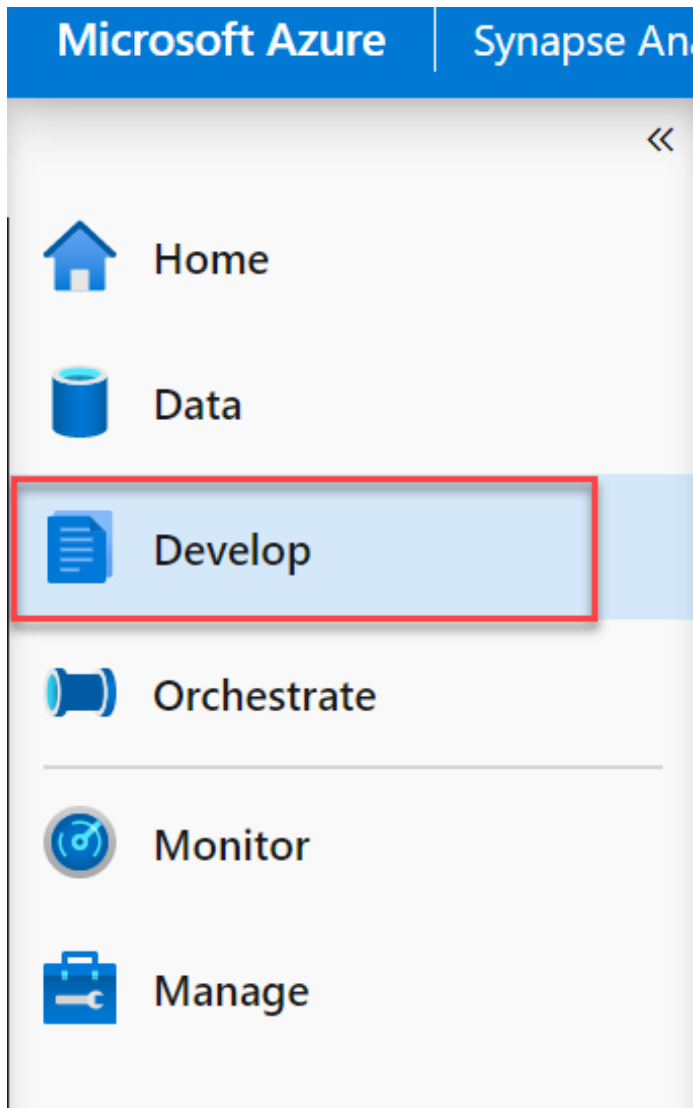


*The Synapse Studio toolbar is displayed with the Run button selected.*

5. You may now close the script tab, when prompted choose to **Discard all changes**.

### **Task 3 - Dynamic data masking**

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



*In Azure Synapse Studio, the Develop item is selected from the left menu.*

2. From the **Develop** menu, expand the **SQL scripts** section, and select **Lab05 - Exercise 3 - Dynamic Data Masking**.


## Develop




 Filter resources by name

### SQL scripts



 ASAL400 - Lab 05 - Exercise 3 - Column Level Security

 ASAL400 - Lab 05 - Exercise 3- Dynamic Data Masking

 ASAL400 - Lab 05 - Exercise 3 - Row Level Security



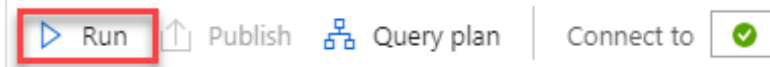
*In Synapse Studio the develop menu is displayed with SQL Scripts section expanded, Lab05 - Exercise 3 - Dynamic Data Masking is selected from the context menu.*

3. In the toolbar menu, connect to the database on which you want to execute the query, SQLPool01.



*The Synapse Studio query toolbar is shown with the Connect to dropdown list field highlighted.*

4. 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.

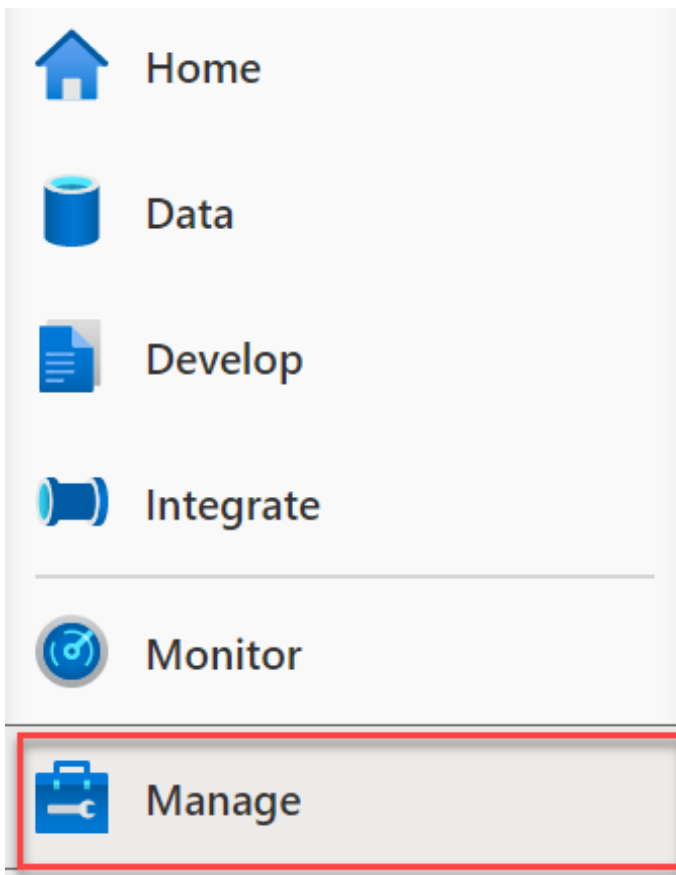


*The Synapse Studio toolbar is displayed with the Run button selected.*

5. You may now close the script tab, when prompted choose to **Discard all changes**.

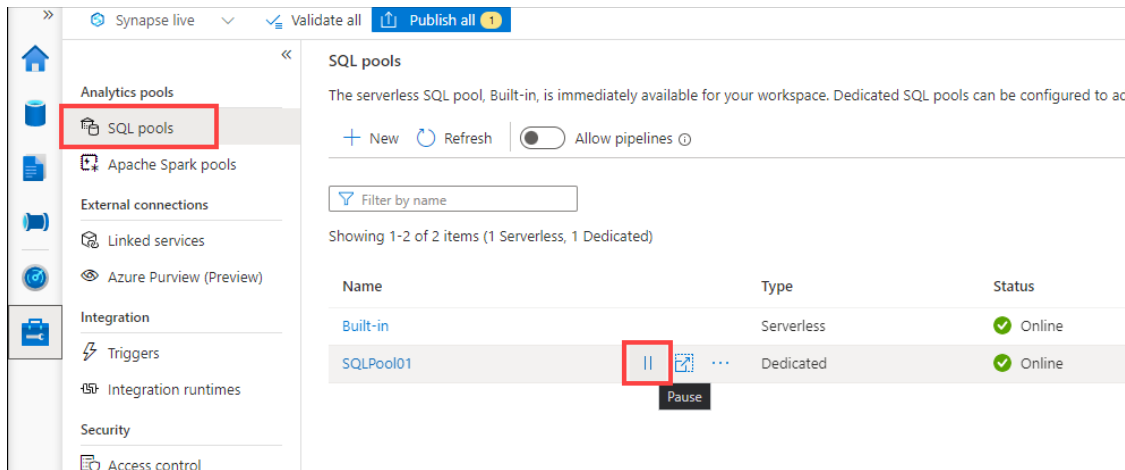
### Cleanup: Pause the dedicated SQL pool

1. Navigate to the **Manage** hub.



*The Manage menu item is highlighted.*

2. From the center menu, select **SQL pools** from beneath the **Analytics pools** heading. Locate SQLPool01, and select the **Pause** button.



*The Manage menu item is selected, with SQL pools selected from the center menu. The resume button is selected next to the SQLPool01 item.*

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 serverless SQL](#)