

Azure Synapse Workspace

Azure Synapse Workspace is a key component of **Azure Synapse Analytics**, which is Microsoft's cloud-based analytics platform. It provides a unified environment for data integration, data exploration, big data analytics, and data visualization. Here's an overview of its features and capabilities:

Key Features:

1. Data Integration:

- Enables seamless integration of data from various sources using **Data Pipelines**.
- Combines both structured and unstructured data from cloud and on-premises systems.
- Supports data transformation using **Data Flow** and integration with **Azure Data Factory**.

2. Analytics and Big Data:

- **SQL Pools (formerly SQL Data Warehouse)**: Optimized for data warehousing and large-scale analytics with T-SQL.
- **Spark Pools**: Supports big data processing using Apache Spark for machine learning and advanced analytics.

3. Unified Environment:

- Combines big data and data warehousing within a single workspace.
- Supports querying data stored in Azure Data Lake or Azure Blob Storage using **serverless SQL**.
- Allows for exploration and visualization of data in an integrated development environment.

4. Collaboration and Development:

- Provides a **notebook experience** for code-first data exploration using languages like Python, Scala, and SQL.
- Enables collaborative development with integration into CI/CD pipelines using Azure DevOps or GitHub.

5. Security and Monitoring:

- Offers robust security with Azure Active Directory, data encryption, and firewall rules.
- Features monitoring capabilities through Azure Monitor and built-in auditing tools.

6. Integration with Power BI:

- Native integration with Power BI for data visualization and reporting.

Use Cases:

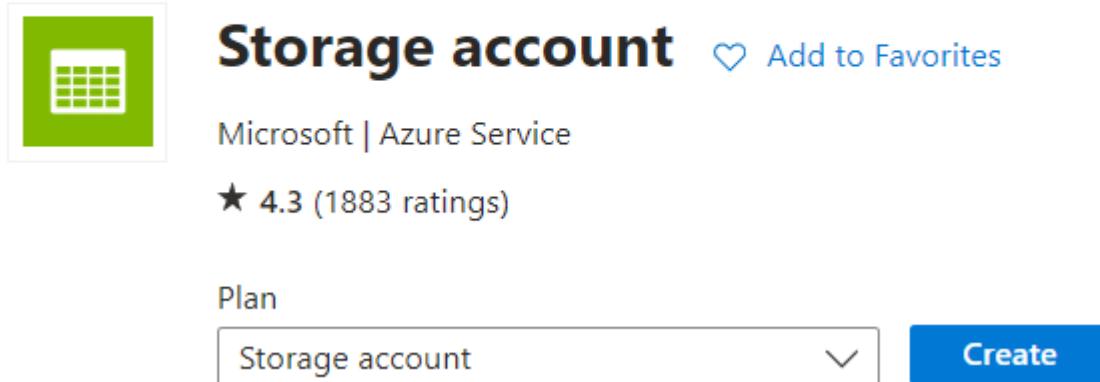
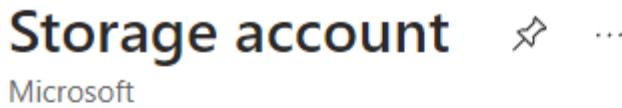
- **Data Warehousing:** Consolidate and query massive datasets with high performance.
- **Big Data Processing:** Process and analyze large datasets using Spark or other distributed frameworks.
- **Data Lake Exploration:** Query and transform data stored in Azure Data Lake.
- **Business Intelligence:** Build reports and dashboards directly connected to your Synapse environment.

Benefits:

- **Scalability:** Scale resources up or down based on your workload needs.
- **Cost Efficiency:** Serverless options allow you to pay only for what you use.
- **Unified Analytics:** Bridges the gap between traditional data warehousing and big data systems.

To begin with the Lab:

1. In this lab, we will provision an Azure Storage account and Azure Synapse Workspace. In your Azure Portal from the marketplace search for Storage account and click on Create.



2. First, you need to choose your resource group if you don't have any then you can create a new one.

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *	MSDN Platforms Subscription
Resource group *	<input type="text" value="new-rg"/> Create new

3. Then you have to give a unique name to your storage account, choose your region, in redundancy choose LRS, and move to the next step.

Instance details

Storage account name * ⓘ	storageaccount120
Region * ⓘ	(Europe) North Europe Deploy to an Azure Extended Zone
Primary service ⓘ	Select a primary service
Performance * ⓘ	<input checked="" type="radio"/> Standard: Recommended for most scenarios (general-purpose v2 account) <input type="radio"/> Premium: Recommended for scenarios that require low latency.
Redundancy * ⓘ	Locally-redundant storage (LRS)

4. Here in step 2, i.e. Advanced, you have to scroll down and enable the Hierarchical Namespace feature. Enabling this feature will help your storage account to be a Data Lake Gen2 storage account.
5. Now directly jump to the review page and create your storage account.

Hierarchical Namespace

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs) [Learn more](#)

Enable hierarchical namespace ⓘ	<input checked="" type="checkbox"/>
---------------------------------	-------------------------------------

6. On a new page, from the marketplace search for Synapse Analytics and click on Create.

Azure Synapse Analytics

Microsoft



...



Azure Synapse Analytics

Add to Favorites

Microsoft | Azure Service

★ 4.2 (72 ratings)

Plan

Azure Synapse Analytics



Create

7. Now you have to choose your subscription and create a new resource group. If you already have a resource group then you can choose that as well.

*Basics *Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription * ⓘ

MSDN Platforms Subscription

The Synapse and SQL resource providers are now registered with this subscription.

Resource group * ⓘ

(New) new-RG

[Create new](#)

Managed resource group ⓘ

Enter managed resource group name



8. Then you need to give a unique name to your workspace, choose your region, and choose to create a new Data Lake Storage Gen2. Give a unique name to it and click on Create, the same thing for the file system name.
9. We have to create this storage because Azure Synapse will use it for temporary storage.

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

 ✓

Region *

 ▼

Select Data Lake Storage Gen2 * ⓘ

From subscription Manually via URL

Account name * ⓘ

▼
[Create new](#)

File system name *

▼
[Create new](#)

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

10. On the next page, i.e. Security, keep the SQL Server admin login name to default, just give a SQL password for the authentication.
11. Directly jump to the review page and click on Create.

* Basics * **Security** Networking Tags Review + create

Configure security options for your workspace.

Authentication

Choose the authentication method for access to workspace resources such as SQL pools. The authentication method can be changed later on. [Learn more ↗](#)

Authentication method ⓘ

Use both local and Microsoft Entra ID authentication
 Use only Microsoft Entra ID authentication

SQL Server admin login * ⓘ

SQL Password ⓘ

 ✓

Confirm password

 ✓

12. Once your workspace is created now, we will navigate to the storage account that we created at the start of this lab.
13. Go to containers and click create container option. Create a new container with name data as you can see below in the snapshot.

+ Container ✓ Change access level ✓ Restore containers ✓ Refresh ✓ Delete ✓ Give feedback

Search containers by prefix ✓ Show deleted containers ✓

Name	Last modified	Anonymous access level	Lease state	...
<input type="checkbox"/> \$logs	12/13/2024, 10:38:31 AM	Private	Available	...
<input type="checkbox"/> data	12/13/2024, 12:23:51 PM	Private	Available	...

14. Then from GitHub you have to download the folder whose name is Synapse Files. In this folder, you will find a CSV file with the name ActivityLog01. You have to upload it inside your data container.

A screenshot of the Azure Storage Blob Container interface. At the top, there are navigation buttons: Upload, Add Directory, Refresh, Rename, Delete, Change tier, Acquire lease, Break lease, and Give feedback. Below these are sections for Authentication method (Access key) and Location (data). A search bar says "Search blobs by prefix (case-sensitive)". There is a toggle switch for "Show deleted objects". A table lists the blob "ActivityLog01.csv" with columns: Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. The blob details are: Name - ActivityLog01.csv, Modified - 12/13/2024, 12:26:21..., Access tier - Hot (Inferred), Archive status - Not yet archived, Blob type - Block blob, Size - 1.91 MiB, Lease state - Available. There is also a "... more options" button.

15. Now you need to come back to Synapse Analytics and from its overview page, you have to Open Synapse Studio using the link given there.

Getting started

A screenshot of the Synapse Analytics Getting Started page. It features two main sections: "Open Synapse Studio" (with a blue hexagonal icon) and "Read documentation" (with a blue document icon). The "Open Synapse Studio" section includes the text: "Start building your fully-integrated analytics solution and unlock new insights." and a "Open" button. The "Read documentation" section includes the text: "Learn how to be productive quickly. Explore concepts, tutorials, and samples." and a "Learn more" button.

16. In a new tab, Synapse Analytics Workspace will be opened. From the left pane, you have to choose the Data section or tab.

A screenshot of the Microsoft Azure Synapse Analytics workspace. The left sidebar has tabs: Home, Data (which is highlighted with a red box), Develop, Integrate, Monitor, and Manage. The main area shows the workspace title "synapseworkspace120". It has three main cards: "Ingest" (Perform a one-time or scheduled data load.), "Explore and analyze" (Learn how to get insights from your data.), and "Visualize" (Build interactive reports with Power BI capabilities.). Below these are sections for "Discover more" (Knowledge center, Browse partners) and "Recent resources".

17. Here you need to click on the Plus icon and choose to connect to external data.

Data

Workspace Link **Workspace**

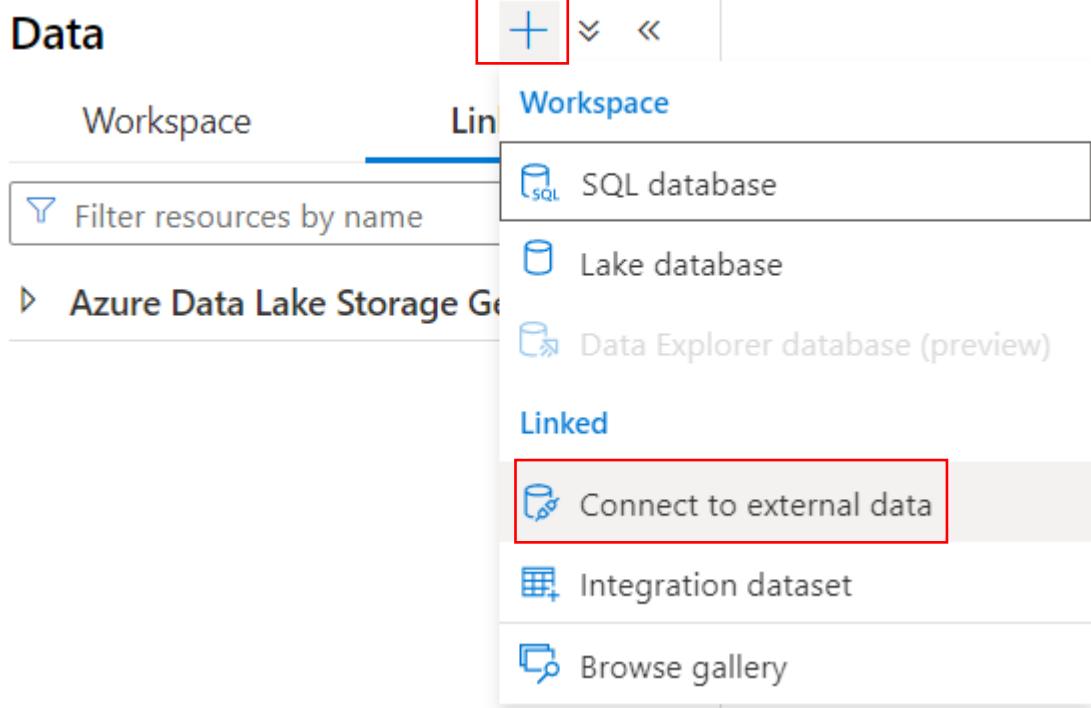
Filter resources by name

▷ Azure Data Lake Storage Gen2

SQL database
Lake database
Data Explorer database (preview)

Linked

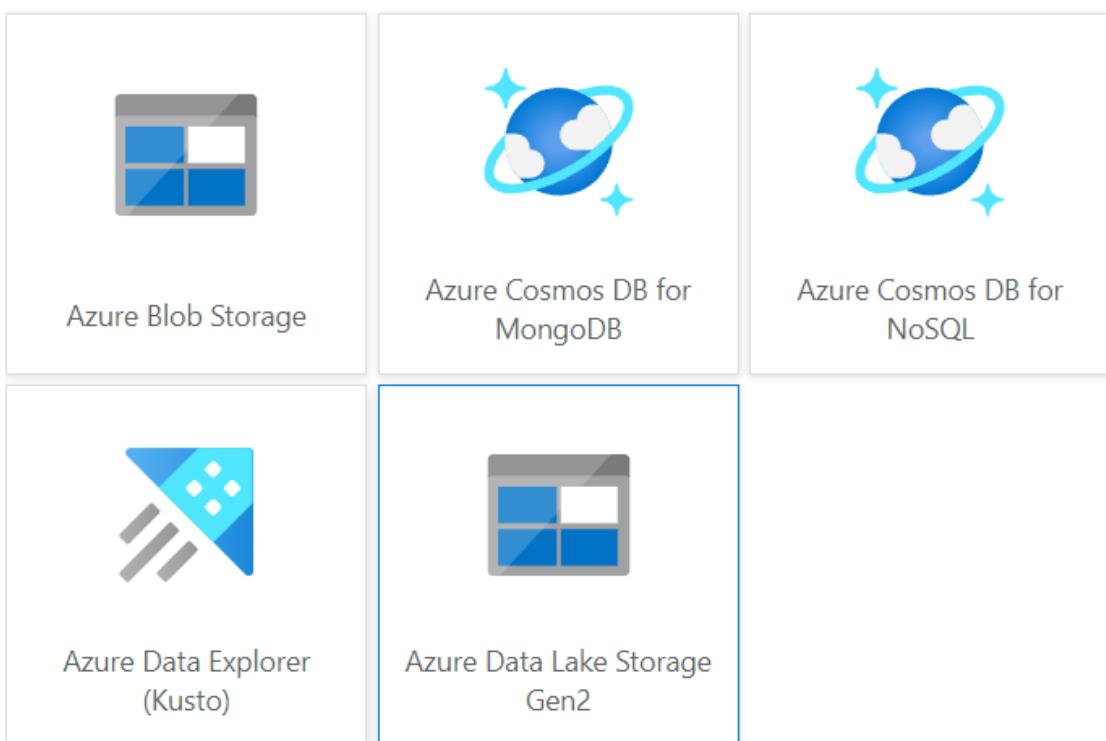
Connect to external data
Integration dataset
Browse gallery



18. For the external data you have to choose Azure Data Lake Storage Gen2.

Connect to external data

Once a connection is created, the underlying data of that connection will be available for analysis in the Data hub or for pipeline activities in the Integrate hub.



19. Then you have to create a new linked service for the Data Lake. First, you have to give the name to it, for that choose the same name as your storage account then scroll down.

New linked service

 Azure Data Lake Storage Gen2 [Learn more](#) 

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

Name *

thestorageaccount1201

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime



Authentication type

Account key



20. Now choose your subscription and the storage account, then click on test connection to verify proper availability. In the end click on Create.

Account selection method ⓘ

From Azure subscription Enter manually

Azure subscription ⓘ

MSDN Platforms Subscription (d6549a66-c45c-4979-840c-3b356da446b0) ▼

Storage account name *

thestorageaccount1201 ▼ ⟳

Test connection ⓘ

To linked service To file path

Annotations

+ New

> Parameters

> Advanced ⓘ

✓ Connection successful

🔗 Test connection Cancel

Create Back

21. After that if you go to the Linked section in data tab and refresh it you will see that the storage account has been linked to your Synapse workspace.

Data + ⌂ <

Linked Workspace

Filter resources by name

◀ **Azure Data Lake Storage Gen2** ...

▶ █ synapseworkspace120 (Primary - w...
▶ █ (Attached Containers)
▶ █ thestorageaccount1201 (thestorag...

22. Also, click on the storage account and then click on the data container you will be able to see the files inside it.

A screenshot of the Azure Data Lake Storage Gen2 Data workspace. On the left, there's a sidebar with 'Data' selected, showing 'Workspace' and 'Linked' tabs. Under 'Linked', it shows 'Azure Data Lake Storage Gen2' with three items: 'synapseworkspace120 (Primary - w...', '(Attached Containers)', and 'thestorageaccount1201 (thestorag...'. Below that is a 'data' folder. On the right, there's a preview pane titled 'data' with a list of files. One file, 'ActivityLog01.csv', is highlighted. The details show 'Name: ActivityLog01.csv', 'Last Modified: 12/13/2024, 12:26:21 PM', 'Content Type: CSV', and 'Size: 1.9 MB'.

23. Right-click on the ActivityLog01 file and from the New SQL Script choose to select the TOP 100 rows.

A screenshot of a context menu for 'ActivityLog01.csv'. The top bar shows the file name and last modified date: 'ActivityLog01.csv' and '12/13/2024, 12:26:21 PM'. The menu has several options: 'Preview', 'New SQL script' (which is highlighted with a red box), 'New notebook', 'New data flow', 'New integration dataset', 'Manage access...', 'Rename...', 'Download', 'Delete', and 'Properties...'. A submenu for 'New SQL script' is open, showing 'Select TOP 100 rows' (also highlighted with a red box) and 'Create external table' and 'Bulk load'.

24. It will take you to the canvas where you can write SQL scripts. Here you will see that a script has already been written to display the first 100 values from your CSV file. Click on Run.

A screenshot of the Synapse live canvas. At the top, there are buttons for 'Synapse live', 'Validate all', and 'Publish all'. Below that is a toolbar with 'Run' (highlighted with a red box), 'Undo', 'Publish', 'Query plan', 'Connect to' (set to 'Built-in'), 'Use database' (set to 'master'), and a refresh icon. The main area contains a SQL script editor with the following code:

```
1 -- This is auto-generated code
2
3 SELECT
4     TOP 100 *
5 FROM
6     OPENROWSET(
7         BULK 'https://thestorageaccount1201.dfs.core.windows.net/data/ActivityLog01.csv',
8         FORMAT = 'CSV',
9         PARSER_VERSION = '2.0'
10    ) AS [result]
```

25. You will see that you are getting the error because the Storage account doesn't have the required permission to display the data inside your Synapse workspace.

26. Because we need to authorize ourselves to be able to access the data within our storage account. There are a lot of security mechanisms in place when it comes to services on the Azure platform.

Messages ^

12:44:36 PM Started executing query at Line 1

File 'https://thestorageaccount1201.dfs.core.windows.net/data/ActivityLog01.csv' cannot be opened because it does not exist or it is used by another process. Visit this article to learn more about this error

Total execution time: 00:00:02.758

27. So, go back to your storage account and open Access Control (IAM). Here you need to click on Add to add a role assignment.

The screenshot shows the 'Check access' tab selected in the top navigation bar. Below it, there's a section titled 'My access' with a 'View my access' button. Further down, there's a 'Check access' section with a 'Check access' button. The entire interface is framed by a red border.

28. Then you need to search for the Storage Blob Data Reader role, select it, and move to the next step to add the member.

The screenshot shows the 'Role' tab selected in the top navigation bar. A search bar contains 'storage blob data reader'. The results table shows one result: 'Storage Blob Data Reader' with a description 'Allows for read access to Azure Storage blob containers and data'. The 'Details' column shows 'BuiltInRole', 'Storage', and 'View'. The entire interface is framed by a red border.

29. Here you need to click on Select Members to add yourself to have sufficient permission to be able to access the data inside your storage account.

Role **Members** • Conditions Review + assign

Selected role Storage Blob Data Reader

Assign access to User, group, or service principal
 Managed identity

Members [+ Select members](#)

30. Once you have added yourself then move to the review page and assign this role.

Role **Members** Conditions Review + assign

Selected role Storage Blob Data Reader

Assign access to User, group, or service principal
 Managed identity

Members [+ Select members](#)

Name	Object ID	Type
Ritesh Behal(Guest)	f203f45f-8be8-4716-8697-b840ce0a35e3	User

Description

31. After that come back to the synapse workspace and run the script again you will see your data here.

```
» data SQL script 1
Run Undo Publish Query plan Connect to Built-in Use database master ⚡
1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://thestorageaccount1201.dfs.core.windows.net/data/ActivityLog01.csv',
7         FORMAT = 'CSV',
8         PARSE_VERSION = '2.0'
9     ) AS [result]
--
```

Results Messages

View Table Chart Export results ▾

Search

C1	C2	C3	C4	C5	C6	C7	C8
Correlationid	Operationname	Status	Eventcategory	Level	Time	Subscription	Eventinitial
c5cf5ce4-d26f...	Delete Storage ...	Succeeded	Administrative	Informational	2024-05-30T15:...	387407e5-94af...	cloudlearn
c5cf5ce4-d26f...	Delete website	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af...	cloudlearn
c5cf5ce4-d26f...	Delete Namesp...	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af...	cloudlearn
c5cf5ce4-d26f...	Delete website	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af...	cloudlearn
c5cf5ce4-d26f...	Delete SQL dat...	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af...	cloudlearn
c5cf5ce4-d26f...	Delete Storage ...	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af...	cloudlearn

00:00:12 Query executed successfully.