



## Azure Synapse Workspace

Azure Synapse Analytics is Microsoft's cloud-based analytics service. It's designed to handle big data and data warehousing tasks. Formerly known as Azure SQL Data Warehouse, it's been expanded to include big data analytics and data integration capabilities.

Azure Synapse Workspace is the environment where you can manage and work with your Synapse resources. It provides a unified experience for developing, monitoring, managing, and analyzing your data solutions. Within the workspace, you can create and manage data pipelines, explore and analyze data, build machine learning models, and collaborate with your team members.

The key components of Azure Synapse Workspace include:

1. **Synapse Studio:** This is the main interface for working with Azure Synapse Analytics. It provides a unified experience for data engineers, data scientists, and business analysts to collaborate on data projects.
2. **Data Integration:** Azure Synapse Workspace includes tools for ingesting, preparing, and transforming data from various sources. This includes capabilities for building data pipelines, integrating with other Azure services like Azure Data Factory, and working with data lakes.
3. **Big Data Analytics:** With Azure Synapse Analytics, you can analyze large volumes of data using familiar tools like SQL and Apache Spark. It provides both on-demand and provisioned resources for running analytics workloads at scale.
4. **Data Warehousing:** Azure Synapse Analytics includes a fully managed data warehouse service that allows you to store and query petabytes of data using standard SQL queries. It's optimized for analytics workloads and provides features like automatic scaling and workload management.
5. **Machine Learning:** Azure Synapse Workspace integrates with Azure Machine Learning to enable data scientists to build, train, and deploy machine learning models directly within the environment.



### Use cases of Azure Synapse:

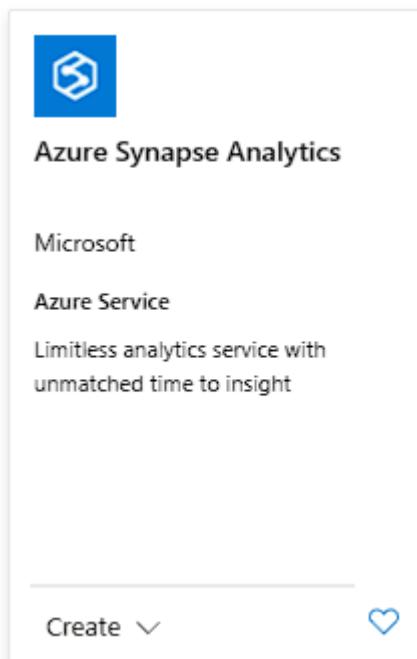
Azure Synapse Workspace can be leveraged across various industries and scenarios for a wide range of use cases. Here are some examples:

1. **Retail Analytics:** Retailers can use Azure Synapse Workspace to analyze sales data, customer demographics, and inventory levels. They can build machine learning models to predict customer behavior, optimize pricing strategies, and manage supply chain operations more effectively.
2. **Financial Services:** Financial institutions can use Azure Synapse Workspace to analyze transaction data, detect fraud patterns, and perform risk modeling. They can also build predictive models to assess creditworthiness, optimize investment portfolios, and comply with regulatory requirements.

3. **Healthcare Analytics:** Healthcare organizations can use Azure Synapse Workspace to analyze patient data, medical records, and clinical trials. They can build predictive models to identify at-risk patients, personalize treatment plans, and improve operational efficiency in hospitals and clinics.
4. **Manufacturing Optimization:** Manufacturers can use Azure Synapse Workspace to analyze production data, equipment sensor readings, and supply chain logistics. They can build predictive maintenance models to reduce downtime, optimize production schedules, and improve product quality.
5. **Marketing Campaign Optimization:** Marketing teams can use Azure Synapse Workspace to analyze customer engagement data, social media trends, and advertising campaigns. They can build recommendation engines to personalize marketing messages, optimize ad spend, and measure the effectiveness of marketing initiatives.
6. **Energy and Utilities:** Energy companies can use Azure Synapse Workspace to analyze sensor data from power plants, smart grids, and renewable energy sources. They can build predictive models to optimize energy production, detect equipment failures, and improve energy efficiency.
7. **Government and Public Sector:** Government agencies can use Azure Synapse Workspace to analyze data related to public safety, transportation, and urban planning. They can build predictive models to forecast traffic congestion, prevent crime, and allocate resources more efficiently.

### To begin with the Lab:

1. In your Azure Portal click on Create Resources and then search for Synapse. Choose this service accordingly.



2. First you have to choose your resource group.

\*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 *	Azure Pass - Sponsorship (9e3f0cae-8274-4931-b16b-95242092e301)
Resource group *	demo-resource-group
	<a href="#">Create new</a>
Managed resource group	Enter managed resource group name

3. Then you are going to give a unique name to your workspace.
4. After that you are going to create a new Data Lake Storage Gen2 and a file system. For that click on Create New.

#### 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 *	datasynapse1234
Region *	North Europe
Select Data Lake Storage Gen2 *	<input checked="" type="radio"/> From subscription <input type="radio"/> Manually via URL
Account name *	<a href="#">Create new</a>
File system name *	<a href="#">Create new</a>
Select Data Lake Storage Gen2 *	<input checked="" type="radio"/> From subscription <input type="radio"/> Manually via URL
Account name *	(New) datalakedemo12345
File system name *	(New) demodata
<input checked="" type="checkbox"/> Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.	

5. After that you need to go to security and keep the authentication method to default and just give it a password.

\* 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 ⓘ	<input checked="" type="radio"/> Use both local and Microsoft Entra ID authentication <input type="radio"/> Use only Microsoft Entra ID authentication
SQL Server admin login *	sqladminuser
SQL Password ⓘ	***** <span style="float: right;">✓</span>
Confirm password	***** <span style="float: right;">✓</span>

6. After that move to review page and create your resources.

7. Once your deployment is completed then click on go to resources.

### ✓ Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20240330153844	Start time : 3/30/2024, 3:47:16 PM
Subscription : Azure Pass - Sponsorship (9e3f0cae-8274-4931-b16b-95242092e301)	Correlation ID : ce99bd35-96c3-418b-a805-8c7462a2316b
Resource group : demo-resource-group	

> Deployment details  
▼ Next steps

[Go to resource group](#)

[Give feedback](#)

[Tell us about your experience with deployment](#)

8. And in the essentials, you can pretty much everything.

[+ New dedicated SQL pool](#) [+ New Apache Spark pool](#) [+ New Data Explorer pool \(preview\)](#) [⟳ Refresh](#) [✍ Reset SQL admin password](#) | [✖ Delete](#)

^ Essentials

Resource group ( <a href="#">move</a> ) : <a href="#">demo-resource-group</a>	Networking : <a href="#">Show firewall settings</a>
Status : Succeeded	Primary ADLS Gen2 account : <a href="https://datalakedemo112211.dfs.core.windows.net">https://datalakedemo112211.dfs.core.windows.net</a>
Location : North Europe	Primary ADLS Gen2 file system : demodatalake
Subscription ( <a href="#">move</a> ) : <a href="#">Azure Pass - Sponsorship</a>	SQL admin username : sqladminuser
Subscription ID : 9e3f0cae-8274-4931-b16b-95242092e301	SQL Microsoft Entra admin : <a href="mailto:live.com#behal.ritesh@gmail.com">live.com#behal.ritesh@gmail.com</a>
Managed virtual network : No	Dedicated SQL endpoint : <a href="#">datasynapse1234.sql.azuresynapse.net</a>
Managed identity object ID : 3f2192e0-e53b-43ed-bad7-09c65465fd0d	Serverless SQL endpoint : <a href="#">datasynapse1234-ondemand.sql.azuresynapse.net</a>
Workspace web URL : <a href="https://web.azure-synapse.net/workspace=%2bsubscriptions%2f9e3f0cae-8274-4931-b16b-9524209...">https://web.azure-synapse.net/workspace=%2bsubscriptions%2f9e3f0cae-8274-4931-b16b-9524209...</a>	Development endpoint : <a href="#">https://datasynapse1234.dev.azuresynapse.net</a>
Tags ( <a href="#">edit</a> ) : <a href="#">Add tags</a>	

9. Let's go ahead and view some data that we have in our Azure Data Lake Gen2 Storage account.
10. Now in the previous Lab we have created a data lake gen2 storage account and uploaded 2 different files in 2 different containers.
11. Now we are going to connect that storage account with our Azure Synapse Workspace.
12. Now on your Synapse Workspace Overview page you will see an option for opening Synapse Studio.
13. Click on open and you will be directed towards a new page or tab.

+ New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) ⏪ Refresh ⚙️ Reset SQL admin password | 🗑 Delete

^ Essentials

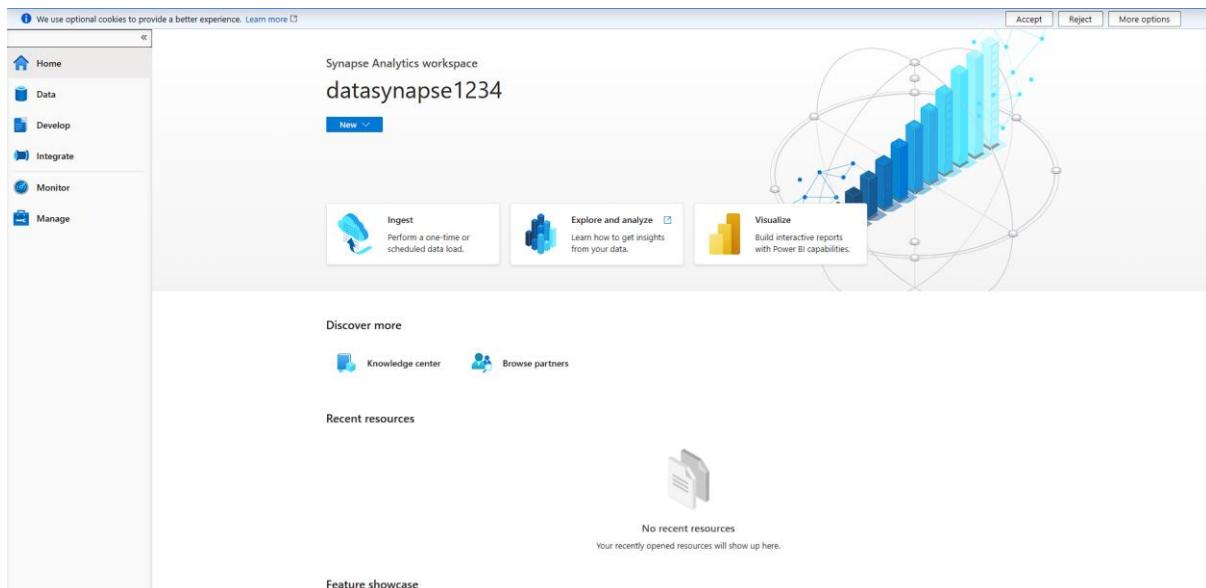
Resource group ( <a href="#">move</a> )	: <a href="#">demo-resource-group</a>	Networking	: <a href="#">Show firewall settings</a>
Status	: Succeeded	Primary ADLS Gen2 acco...	: <a href="https://datalakedemo112211.dfs.core.windows.net">https://datalakedemo112211.dfs.core.windows.net</a>
Location	: North Europe	Primary ADLS Gen2 file s...	: demodatalake
Subscription ( <a href="#">move</a> )	: <a href="#">Azure Pass - Sponsorship</a>	SQL admin username	: sqladminuser
Subscription ID	: 9e3f0cae-8274-4931-b16b-95242092e301	SQL Microsoft Entra admin	: <a href="mailto:live.com#behal.ritesh@gmail.com">live.com#behal.ritesh@gmail.com</a>
Managed virtual network	: No	Dedicated SQL endpoint	: <a href="http://datasynapse1234.sql.azuresynapse.net">datasynapse1234.sql.azuresynapse.net</a>
Managed Identity object ...	: f3f2192e0-e53b-43ed-bad7-09c65465fd0d	Serverless SQL endpoint	: <a href="http://datasynapse1234-ondemand.sql.azuresynapse.net">datasynapse1234-ondemand.sql.azuresynapse.net</a>
Workspace web URL	: <a href="https://web.azuresynapse.net?workspace=%2fsubscriptions%2f9e3f0cae-8274-4931-b16b-9524209...">https://web.azuresynapse.net?workspace=%2fsubscriptions%2f9e3f0cae-8274-4931-b16b-9524209...</a>	Development endpoint	: <a href="https://datasynapse1234.dev.azuresynapse.net">https://datasynapse1234.dev.azuresynapse.net</a>
Tags ( <a href="#">edit</a> )	: <a href="#">Add tags</a>		

Getting started

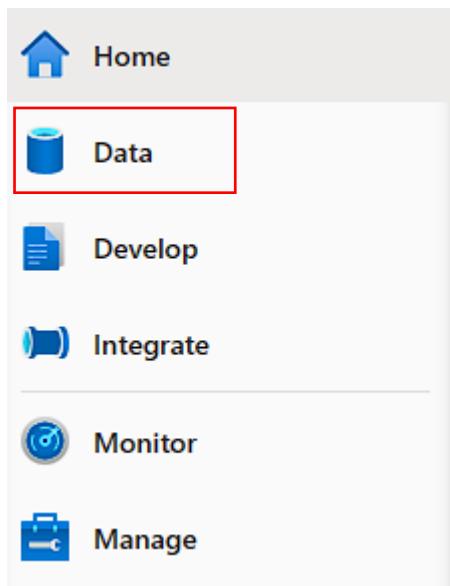
 Open Synapse Studio  
Start building your fully-integrated analytics solution and unlock new insights.  
[Open](#)

 Read documentation  
Learn how to be productive quickly. Explore concepts, tutorials, and samples.  
[Learn more](#)

14. Once you have clicked on Open then you will be on this page.
15. This gives you a browser-based experience when it comes to working with your data when it comes to viewing the tables that you could have in your SQL data warehouse. So, all of the tools that you need are available in Azure Synapse itself in this particular studio.



16. Now on the left pane you can see some options to play with. For now, click on data.



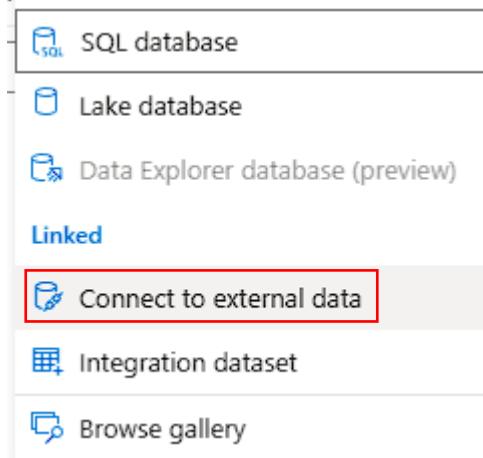
17. Now in the data in the linked section you can see your gen2 storage and currently you don't have any attached containers.

This screenshot shows the 'Data' section in Azure Data Studio. The 'Linked' tab is selected. It lists 'Azure Data Lake Storage Gen2' with two items: 'datasynapse1234 (Primary - data...)' and 'demodatalake (Primary)'. Below this is a section for '(Attached Containers)'.

18. Now if you go onto the Develop section, here we can create SQL scripts, etcetera. In the Integrate section, we can build our pipelines.
19. Now we want to bring our other data lake gen2 storage account closer to our workspace.
20. For that go to your data section then in your workspace click on the plus icon which is highlighted.

This screenshot shows the 'Data' section in Azure Data Studio. The 'Workspace' tab is selected. A red box highlights the blue plus icon located at the top right of the workspace area.

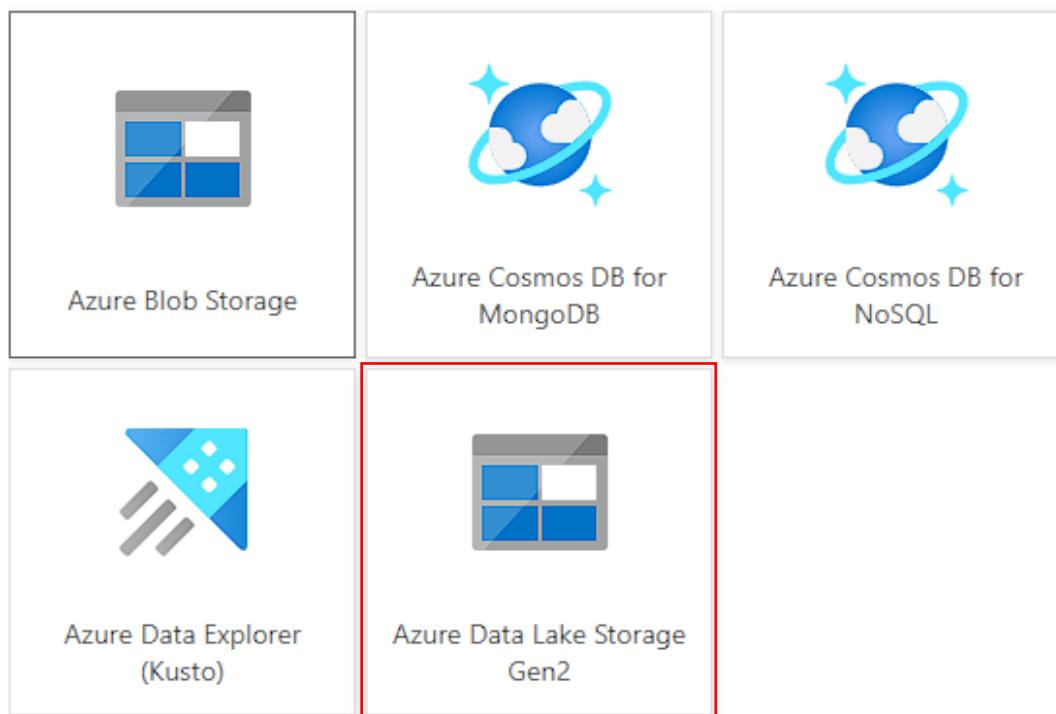
21. Then you will get some options and you have to click on connect to external data.



22. After that you will see some more options, from here 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.



23. First you have to give it the same, you can give the same name as your Azure storage.  
24. Then keep choose your subscription then your storage account name and click on create.

Name \*

Description

Connect via integration runtime \* ⓘ

AutoResolveIntegrationRuntime ✎

Authentication type

Account key ▼

Account selection method ⓘ

From Azure subscription  Enter manually

Azure subscription ⓘ

Azure Pass - Sponsorship (9e3f0cae-8274-4931-b16b-95242092e301) ▼

Storage account name \*

sqlstorage1010 ↻

Test connection ⓘ

To linked service  To file path

25. Once you are done now you have to go to the linked section and refresh it.
26. Then you will be able to see your storage account in place.

Data + ▾ <

Workspace Linked

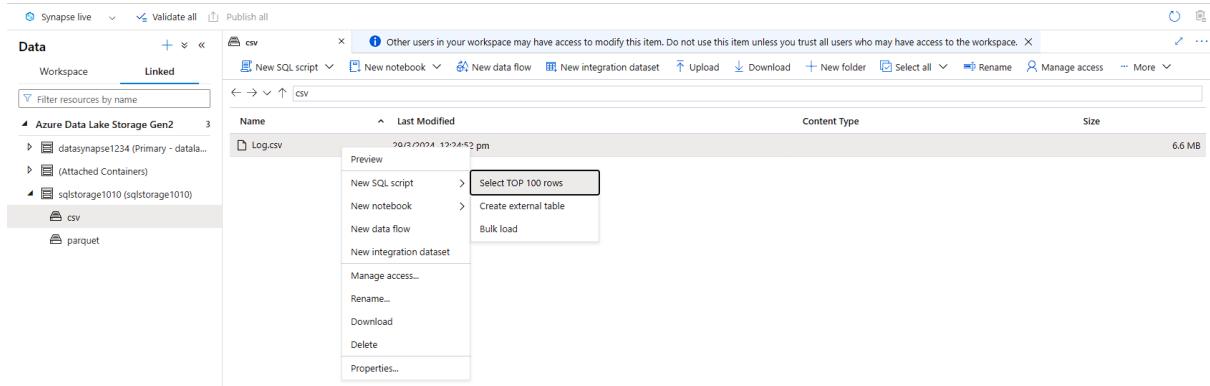
▲ Azure Data Lake Storage Gen2 3

▷  datasynapse1234 (Primary - data...  
▷  (Attached Containers) ...

▲  sqlstorage1010 (sqlstorage1010)

 CSV  
 parquet

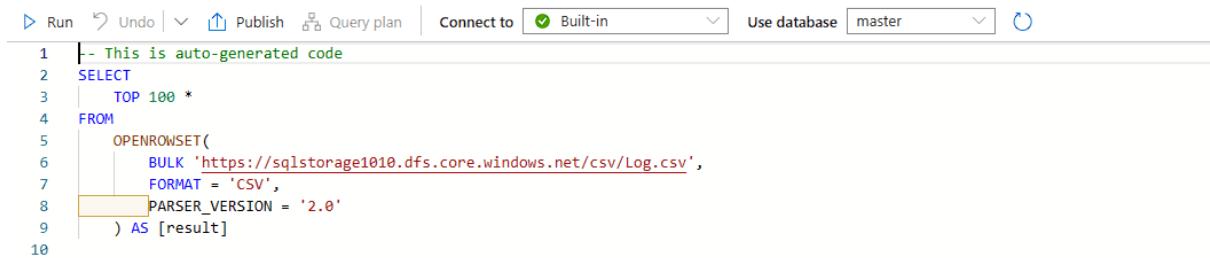
27. Now you can click on CSV then right-click on the Log.csv file and from the New SQL Script click on select Top 100 rows.



A screenshot of the Azure Synapse Studio interface. On the left, there's a sidebar with 'Data' selected, showing 'Workspace' and 'Linked' sections. Under 'Linked', there's a tree view for 'Azure Data Lake Storage Gen2' with 'sqlstorage1010 (sqlstorage1010)' expanded, showing 'csv' and 'parquet' under it. In the main area, there's a table view with one item: 'Log.csv' (Last Modified: 20/3/2024 - 12:58:52 pm). A context menu is open over this item, with the 'Select TOP 100 rows' option highlighted.

28. Now here we are trying to view the data within a CSV file in our data lake. If we had to execute a SQL query against a CSV file, can we do this by default? **Well, the answer is no, right?** So, you can bring your CSV file down, open it in Microsoft Excel, and perform analysis. But can you use the structured query language as it is against a CSV file? **The direct answer is no.** You need to have structured data in a relational database system or some sort of database engine that can accept SQL queries and give you the result accordingly.

29. But Azure Synapse is giving you now the capability of being able to execute typical SQL query language statements against a CSV file located in your Azure Data Lake Gen2 Storage account.



A screenshot of the 'Query plan' editor in Azure Synapse Studio. The top bar shows 'Run', 'Undo', 'Publish', 'Query plan', 'Connect to' (set to 'Built-in'), 'Use database' (set to 'master'), and a refresh button. The code pane contains the following T-SQL:

```
1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://sqlstorage1010.dfs.core.windows.net/csv/Log.csv',
7         FORMAT = 'CSV',
8         PARSER_VERSION = '2.0'
9     ) AS [result]
10
```

30. Now you have to click on Run and try to execute your query. But you will see that you are facing an error which says that it does not exist.

31. 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.

4:37:37 pm [Started executing query at Line 1](#)  
File 'https://sqlstorage1010.dfs.core.windows.net/csv/Log.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: 00:00:04.240

32. So, now you have to navigate your Gen2 storage account and open IAM. Then you have to click on Add then choose add a role assignment.

33. So, I want to give permissions, actually to myself, to be able to access data. Now, you might be thinking that I am the Azure admin, why do I need to give myself access? Well, this is the way that it works. See, this is data within some sort of service on the Azure platform. It should be granted that anyone who has admin access can also access the data. So, now there are specific permissions for the data within the storage account.

The screenshot shows the 'Access Control (IAM)' blade for a storage account named 'sqlstorage1010'. At the top, there's a search bar and a red box highlights the '+ Add' button. Below the header, there are tabs: 'Check access' (which is selected), 'Role assignments', 'Roles', 'Deny assignments', and 'Classic administrators'. Under 'Check access', there are two main sections: 'My access' (with a 'View my access' button) and 'Check access' (with a 'Check access' button). Further down, there are three cards: 'Grant access to this resource' (with a 'Add role assignment' button), 'View access to this resource' (with a 'View' button), and 'View deny assignments' (with a 'View' button).

34. Now you have to filter for role which one you need. In our case search for storage blob data reader.

The screenshot shows the 'Review + assign' page for a role definition. The 'Members' tab is selected. At the top, there are tabs: 'Role', 'Members' (which is selected), 'Conditions', and 'Review + assign'. Below the tabs, there's a note about role definitions being collections of permissions. Under 'Job function roles', there's a section for 'Privileged administrator roles'. A table lists a single result: 'Storage Blob Data Reader' (Type: BuiltInRole, Category: Storage, Details: View). The table has columns: Name, Description, Type, Category, and Details.

35. Then on the next page you have to add the user in our case you are the member.

36. Once this is done then move to the review page and assign this role for yourself.

37. Now it takes some minutes for the role to be assigned. So, wait for sometime.

The screenshot shows the 'Assign access to' page for the 'Storage Blob Data Reader' role. The 'Members' tab is selected. At the top, there are tabs: 'Role', 'Members' (selected), 'Conditions', and 'Review + assign'. Below the tabs, there's a section for 'Selected role' (Storage Blob Data Reader). Under 'Assign access to', there are two radio buttons: 'User, group, or service principal' (selected) and 'Managed identity'. Below that, there's a 'Members' section with a '+ Select members' button. A table shows a single member: 'Ritesh Behal' (Object ID: f203f45f-8be8-4716-8697-b840ce0a35e3, Type: User). At the bottom, there's a 'Description' field with the placeholder 'Optional'.

38. Now here you can see that data accordingly.

Data      + <>      Workspace      Linked

Filter resources by name

▲ Azure Data Lake Storage Gen2      3

  ▷ dataspase1234 (Primary - data...  
  ▷ (Attached Containers)  
  ◀ sqlstorage1010 (sqlstorage1010)

  CSV  
  parquet

csv      SQL script 2      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://sqlstorage1010.dfs.core.windows.net/csv/Log.csv',
7     FORMAT = 'CSV',
8     PARSE_VERSION = '2.0'
9   ) AS [result]
10
```

Results      Messages

View      Table      Chart      Export results ▾

Search

C1	C2	C3	C4	C5	C6	C7
Correlation id	Operation name	Status	Event category	Level	Time	Subscription
99fe9c3a-e36e-...	Update SQL da...	Succeeded	Administrative	Informational	2023-04-25T03...	6912d7a0-bc2
99fe9c3a-e36e-...	Create Deploy...	Started	Administrative	Informational	2023-04-25T03...	6912d7a0-bc2
99fe9c3a-e36e-...	Create Deploy...	Accepted	Administrative	Informational	2023-04-25T03...	6912d7a0-bc2
99fe9c3a-e36e-...	Registers the M...	Started	Administrative	Informational	2023-04-25T03...	6912d7a0-bc2
99fe9c3a-e36e-...	Registers the M...	Succeeded	Administrative	Informational	2023-04-25T03...	6912d7a0-bc2

00:00:11 Query executed successfully.

Properties

General      Related (0)

Name \*      SQL script 2

Description

Type      sql script

Size      220 bytes

Results settings per query

First 5000 rows (default)  
 All rows