How-To Library

Informatica

Prerequisites to Create a Microsoft Azure Data Lake Storage Gen2 Connection

Abstract

You can use PowerExchange® for Microsoft Azure Data Lake Storage Gen2 to connect to Microsoft Azure Data Lake Storage Gen2 from Informatica. This article explains the prerequisite tasks that you must complete before you create a Microsoft Azure Data Lake Storage Gen2 Connection.

Supported Versions

- Informatica® PowerExchange® for Microsoft Azure Data Lake Storage Gen2

Table of Contents

Overview

You can use PowerExchange® for Microsoft Azure Data Lake Storage Gen2 to connect to Microsoft Azure Data Lake Storage Gen 2 using Azure Active Directory (AAD) principal-based authentication.

Before you create a Microsoft Azure Data Lake Storage Gen2 connection, complete the following prerequisite tasks:

1. Create a storage account to use with Microsoft Azure Data Lake Storage Gen2, enable **Hierarchical namespace**, and provide **Contributor** or **Reader** role to users.
   The contributor role grants you full access to manage all resources in the storage account, but does not allow you to assign roles.

   The reader role allows you to view all resources in the storage account, but does not allow you to make any changes.

   Note: To add or remove role assignments, you must have write and delete permissions, such as an Owner role.
2. Create a Blob container in the storage account.
3. Register a new application in Azure Active Directory to authenticate users to access the Microsoft Azure Data Lake Storage Gen2 account. Provide **Storage Blob Data Contributor** or **Storage Blob Data Reader** role to the application.
   The Storage Blob Data Contributor role lets you read, write, and delete Azure Storage containers and blobs in the storage account.

   The Storage Blob Data Reader role lets you only read and list Azure Storage containers and blobs in the storage account.

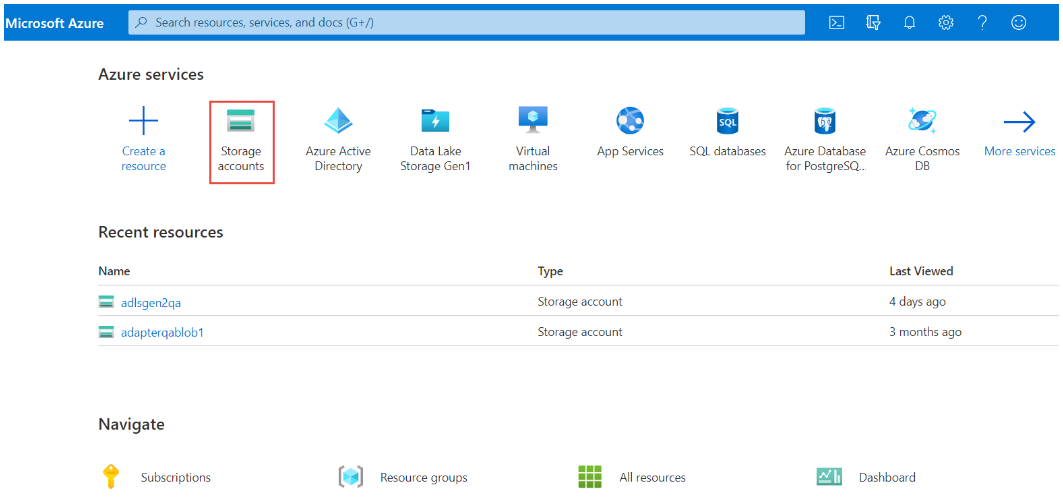
   Note: To write to or delete Azure Storage containers and blobs, you must have the Contributor role either at the storage account level or the container level.
4. Set the Access Control List to provide the read, write, and execute permissions to Microsoft Azure Data Lake Storage Gen2.

For more information about Microsoft Azure Data Lake Storage Gen2 Connector, see the *Informatica Cloud® Data Integration Microsoft Azure Data Lake Storage Gen2 Connector User Guide*.

Creating a Storage Account to use with Microsoft Azure Data Lake Storage Gen2

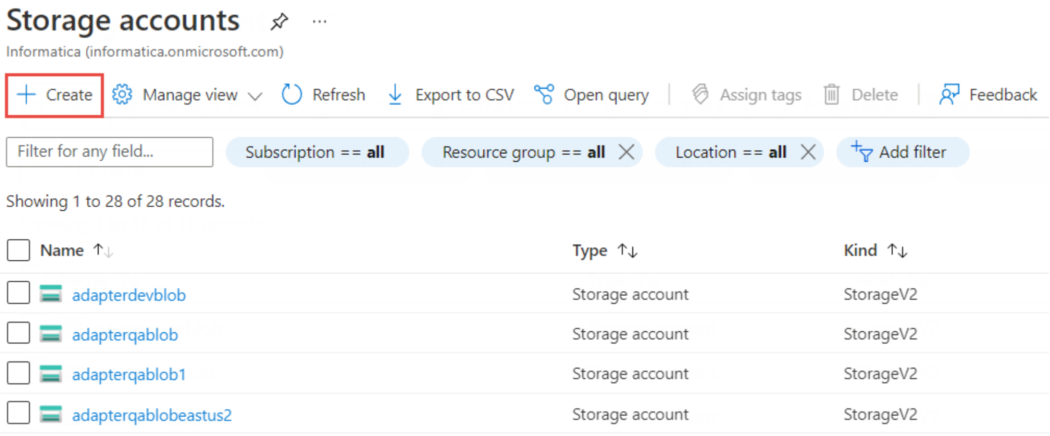
Perform the following steps to create a storage account:

1. Log in to the following Azure portal: https://portal.azure.com/
2. Under Azure Services, click **Storage accounts**.



Name	Type	Last Viewed
adlsgen2qa	Storage account	4 days ago
adapterqablob1	Storage account	3 months ago

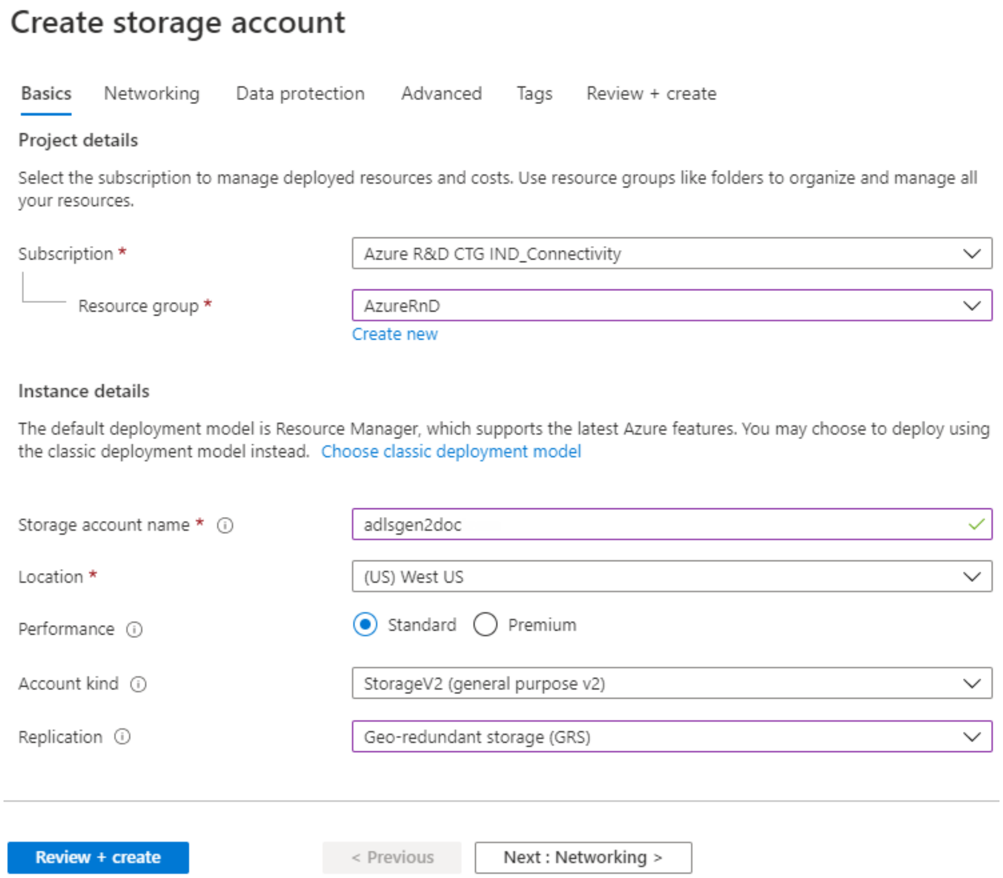
3. On the **Storage accounts** page, click **Create** to create a new storage account.



Showing 1 to 28 of 28 records.

Name	Type	Kind
adapterdevblob	Storage account	StorageV2
adapterqablob	Storage account	StorageV2
adapterqablob1	Storage account	StorageV2
adapterqablobeastus2	Storage account	StorageV2

4. On the **Basics** tab, enter the project and instance details.



Create storage account

Project details

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

Subscription * : Azure R&D CTG IND_Connectivity

Resource group * : AzureRnD

Create new

Instance details

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. Choose classic deployment model

Storage account name * : adlsgen2doc ✓

Location * : (US) West US

Performance : ● Standard ○ Premium

Account kind : StorageV2 (general purpose v2)

Replication : Geo-redundant storage (GRS)

a. In the **Subscription** field, select the subscription in which you want to create the storage account.

b. In the **Resource group** field, select the resource group in which the Azure resources are deployed and managed.

c. In the **Storage account name** field, enter a name for your storage account.

   Note: The name must be unique across Azure, between 3 and 24 characters in length, and must include only numbers and lowercase letters.

d. In the **Location** field, select a location for your storage account, or use the default location.

e. In the **Performance** field, select **Standard**.

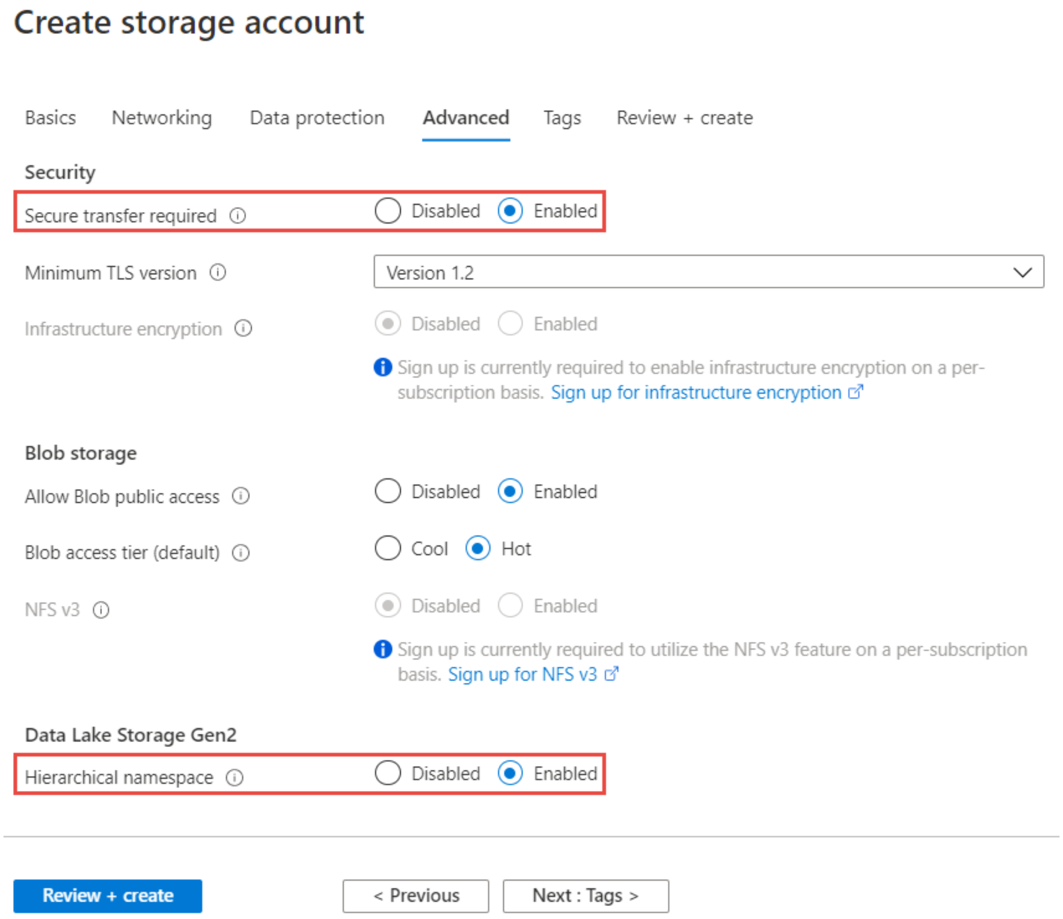
f. In the **Account kind** field, select **StorageV2 (general purpose v2)**.

   A general-purpose v2 storage account provides access to all the Azure Storage services, such as blobs, files, queues, tables, and disks.

g. In the **Replication** field, select **Geo-redundant storage (GRS)**.

   The replication type specifies how the storage account will be replicated.

5. On the **Advanced** tab, set the **Secure transfer required** and **Hierarchical namespace** fields to **Enabled**.



Create storage account

Security

Secure transfer required ⓘ ○ Disabled ● Enabled

Minimum TLS version ⓘ Version 1.2

Infrastructure encryption ⓘ ● Disabled ○ Enabled

ⓘ Sign up is currently required to enable infrastructure encryption on a per-subscription basis. Sign up for infrastructure encryption

Blob storage

Allow Blob public access ⓘ ○ Disabled ● Enabled

Blob access tier (default) ⓘ ○ Cool ● Hot

NFS v3 ⓘ ● Disabled ○ Enabled

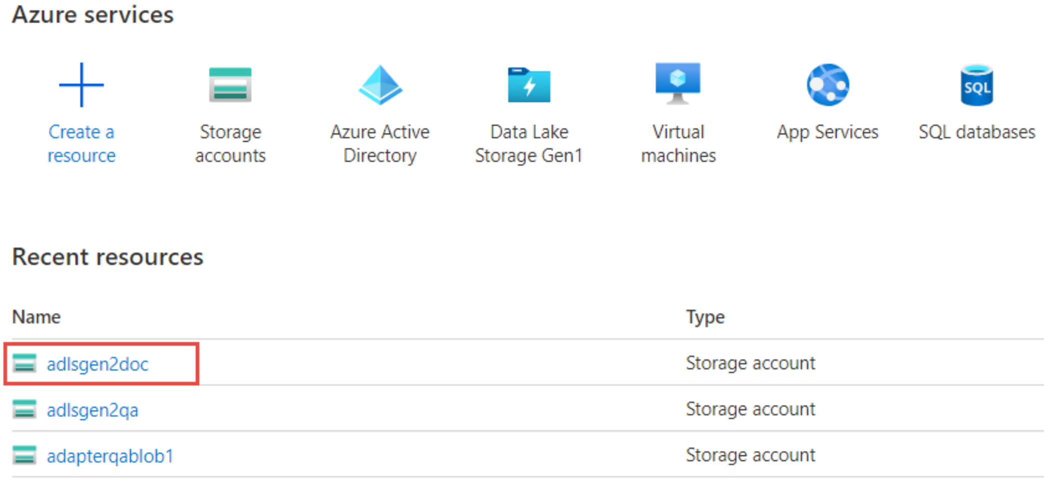
ⓘ Sign up is currently required to utilize the NFS v3 feature on a per-subscription basis. Sign up for NFS v3

Data Lake Storage Gen2

Hierarchical namespace ⓘ ○ Disabled ● Enabled

Note: When you run a mapping on the Databricks Spark engine, the request to the Azure Blob Filesystem (abfs) is always made over secure connections whether you enable the **Secure transfer required** option or not.

6. Click **Review + Create** > **Create**.
7. Click on the newly created storage account name.

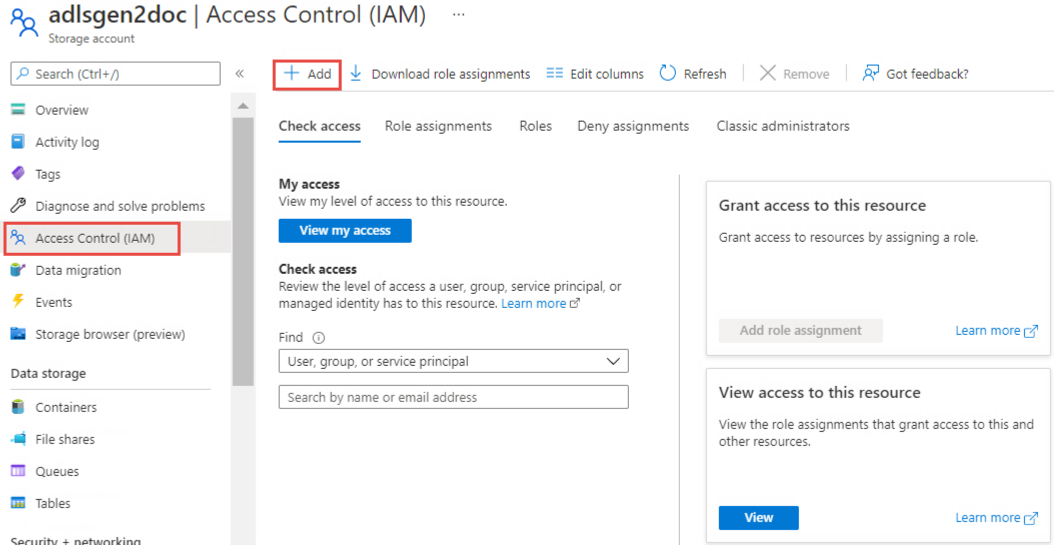


Azure services

Recent resources

Name	Type
adlsgen2doc	Storage account
adlsgen2qa	Storage account
adapterqablob1	Storage account

8. Click **Access control (IAM)** > **Add**.



View my level of access to this resource.

Review the level of access a user, group, service principal, or managed identity has to this resource. Learn more

Grant access to resources by assigning a role.

View the role assignments that grant access to this and other resources.

9. On the **Add role assignment** page, assign a role to the users.

   a. In the **Role** field, select **Reader** or **Contributor**.

      Note: To add or remove role assignments, you must have write and delete permissions, such as an Owner role.

   b. In the **Assign access to** field, select **Azure AD user, group or service principal**.

   c. In the **Select** field, select the user that requires access to the storage account.

   d. Click **Save**.

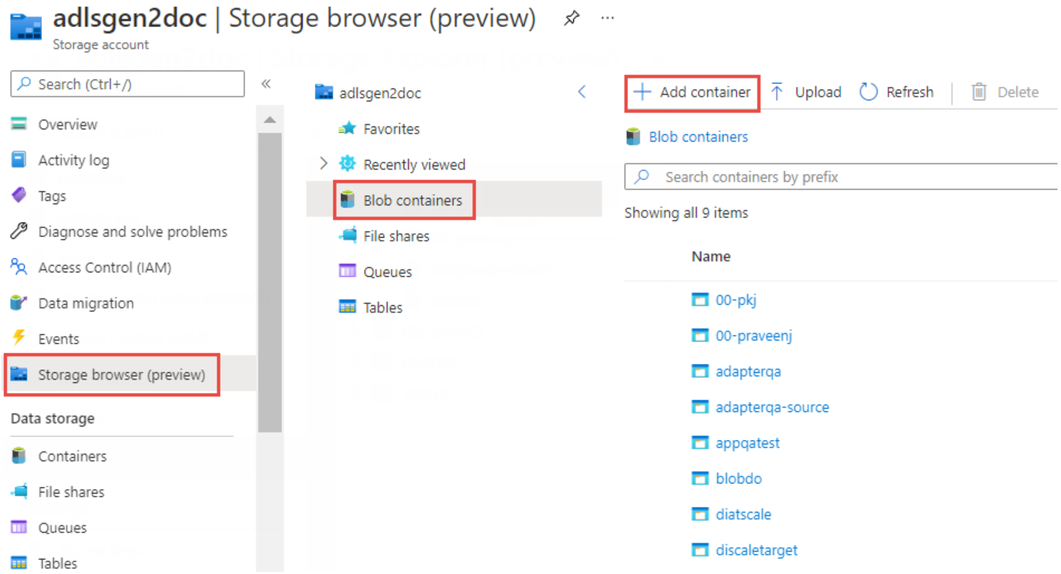
      Note: If you want to add multiple users to access the storage account, you must perform the same steps for each user.

Creating a Blob Container in the Storage Account

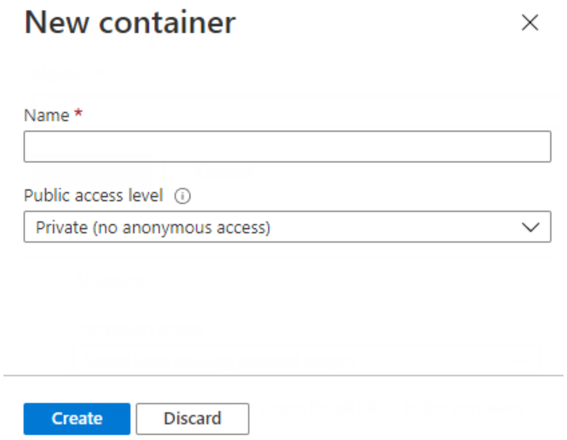
Perform the following steps to create a Blob container in Microsoft Azure Data Lake Storage Gen2:

1. Log in to the Azure portal.
2. Open the storage account that you created.
3. Click **Storage browsers** > **Blob containers**.

4. Click **Add container**.



5. Enter a name for the new container.



New container

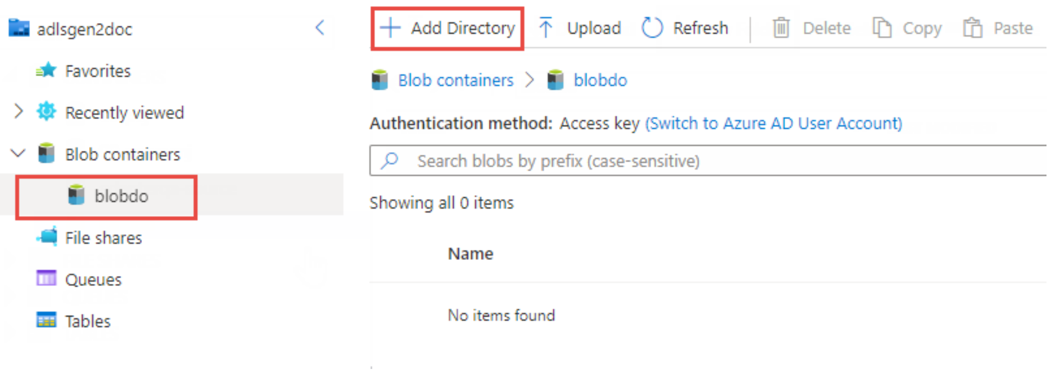
Name *

Public access level: Private (no anonymous access)

Note: You can only use lowercase letters, numbers, and hyphens when you enter the names of the folder and the file system. The names must begin and end with a letter or number. Do not add consecutive hyphens when you enter the names of the folder and the file system.

6. Select the **Public access level** as **Private**.
7. Click **Create**.
8. Click the container that you created.

9. Click **Add Directory** to create a new Microsoft Azure Data Lake Storage Gen2 folder within the container that you created.

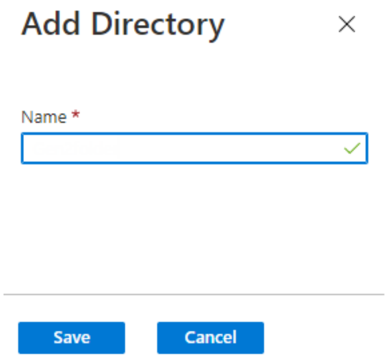


Authentication method: Access key (Switch to Azure AD User Account)

Showing all 0 items

No items found

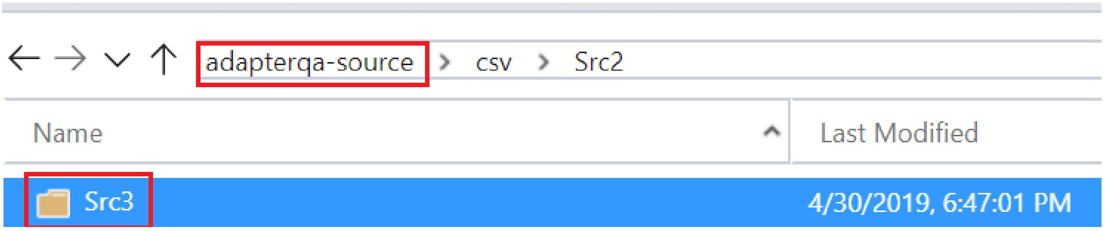
10. Enter a name for the Microsoft Azure Data Lake Storage Gen2 folder and then click **Save**.



To get the Microsoft Azure Data Lake Storage Gen2 folder path, select the Microsoft Azure Data Lake Storage Gen2 folder. Then, right-click on the folder and select **Properties** option to copy the folder path.

For example, create a `Src3` Microsoft Azure Data Lake Storage Gen2 folder within the `adapterqa-source` file system name. When you select the folder, the folder path is displayed as `/csv/src2/src3` . Right-click on the folder and select **Properties** option to copy the folder path.

The following image shows the path of the `Src3` Microsoft Azure Data Lake Storage Gen2 folder within the file system:



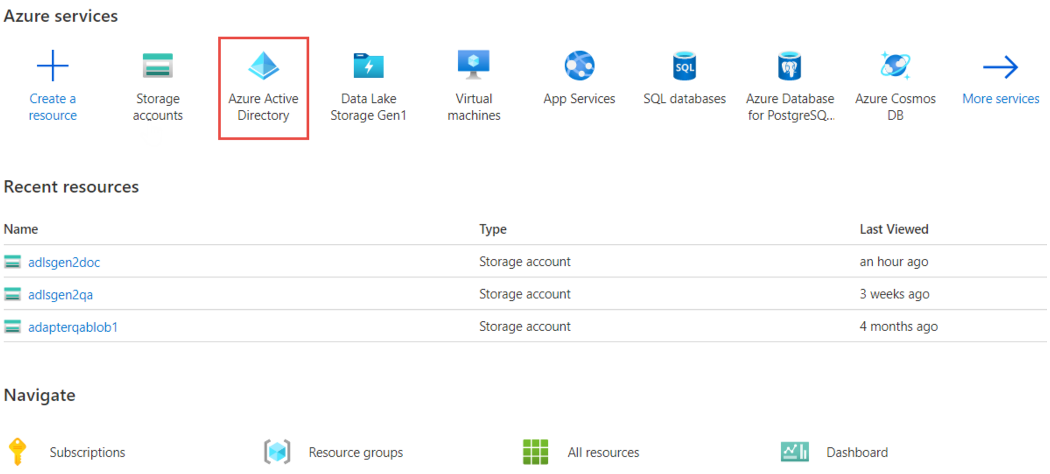
Name	Last Modified
Src3	4/30/2019, 6:47:01 PM

Registering an Application in Azure Active Directory

Register a new application in Azure Active Directory to authenticate access to the storage account.

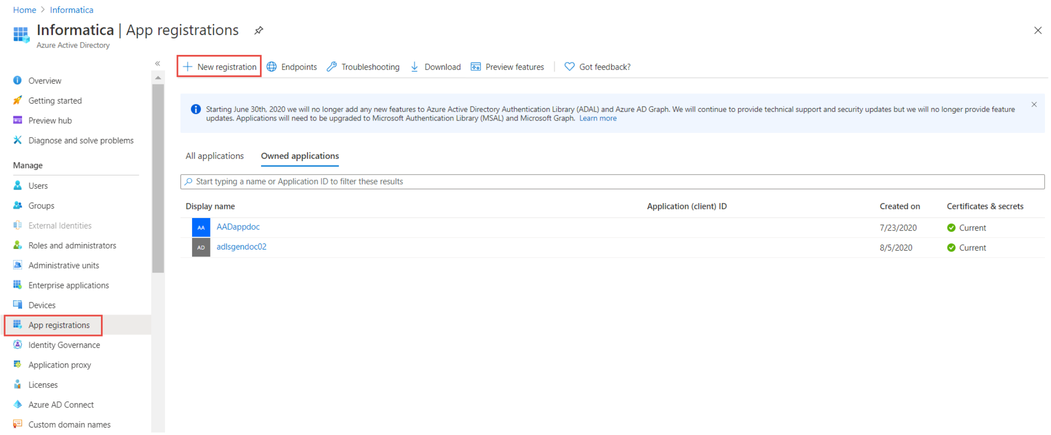
1. Log in to the Azure portal.

2. Click **Azure Active Directory**.



Name	Type	Last Viewed
adlsgen2doc	Storage account	an hour ago
adlsgen2qa	Storage account	3 weeks ago
adapterqablob1	Storage account	4 months ago

3. In the **Manage** section, click **App registrations**.



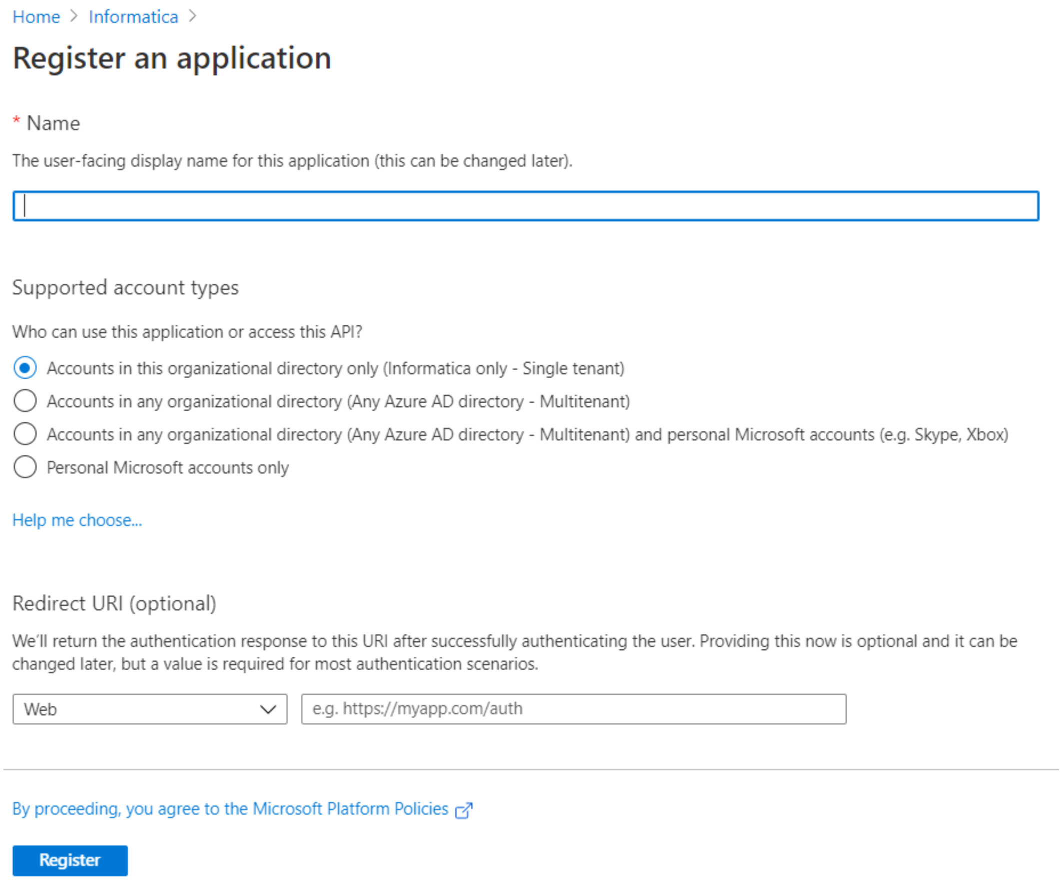
Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library (ADAL) and Azure AD Graph. We will continue to provide technical support and security updates but we will no longer provide feature updates. Applications will need to be upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. Learn more

Start typing a name or Application ID to filter these results

Display name	Application (client) ID	Created on	Certificates & secrets
AA AADappdoc		7/23/2020	✓ Current
AD adlsgendoc02		8/5/2020	✓ Current

4. Click **New registration** to create a new Azure Active Directory application.

5. On the **Register an application** page, enter the details for the new application.



Home > Informatica >

Register an application

* Name

The user-facing display name for this application (this can be changed later).

Supported account types

Who can use this application or access this API?

● Accounts in this organizational directory only (Informatica only - Single tenant)

○ Accounts in any organizational directory (Any Azure AD directory - Multitenant)

○ Accounts in any organizational directory (Any Azure AD directory - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)

○ Personal Microsoft accounts only

Help me choose...

Redirect URI (optional)

We'll return the authentication response to this URI after successfully authenticating the user. Providing this now is optional and it can be changed later, but a value is required for most authentication scenarios.

Web | e.g. https://myapp.com/auth

By proceeding, you agree to the Microsoft Platform Policies

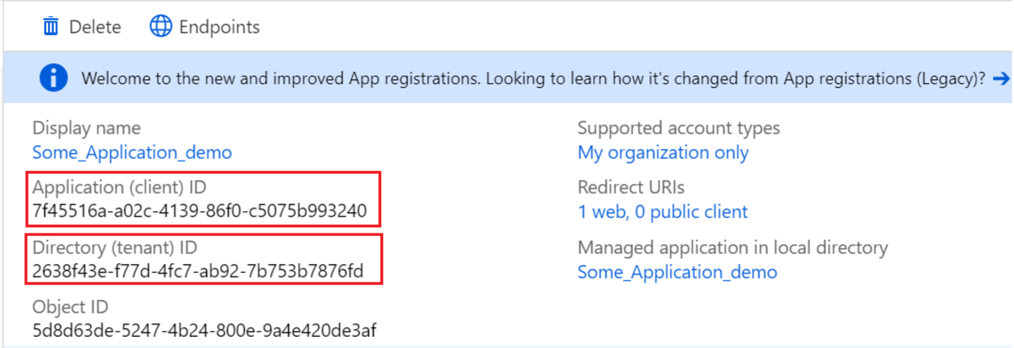
Register

   a. In the **Name** field, enter the application name.

   b. In the **Redirect URI** section, select **Web** as the type of the application and enter the URL of the application.

   c. Click **Register**.

      The details of the newly created Azure Active Directory application page are displayed.



Delete   Endpoints

Welcome to the new and improved App registrations. Looking to learn how it's changed from App registrations (Legacy)? →

Display name
Some_Application_demo

Application (client) ID
7f45516a-a02c-4139-86f0-c5075b993240

Directory (tenant) ID
2638f43e-f77d-4fc7-ab92-7b753b7876fd

Object ID
5d8d63de-5247-4b24-800e-9a4e420de3af

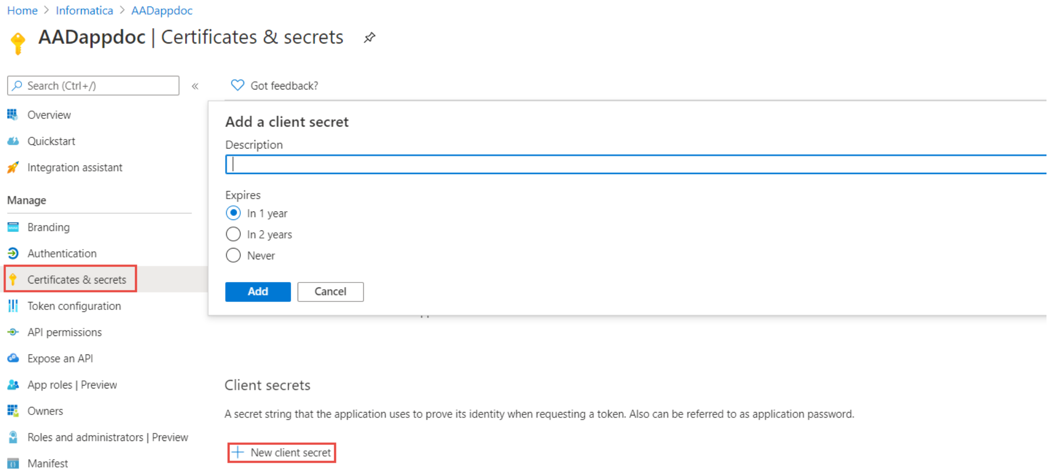
Supported account types
My organization only

Redirect URIs
1 web, 0 public client

Managed application in local directory
Some_Application_demo

6. In the Manage section, click **Certificates & secrets** section.

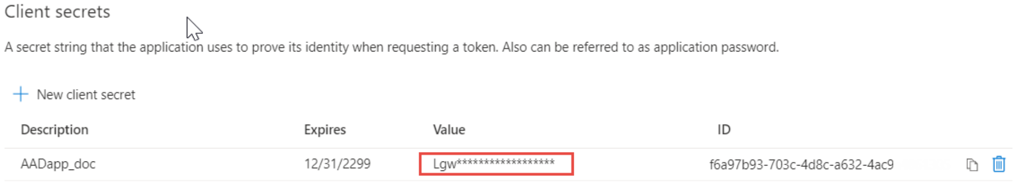
7. Click **New client secret**.



Expires: ● In 1 year ○ In 2 years ○ Never

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

8. In the **Add a client secret** page, perform the following steps:
   a. Enter a name for the client secret in the **Description** field.
   b. In the **Expires** field, you can select the duration of the key as **Never**(Recommended).
   c. Click **Add**.
   d. The value of the key is generated and displayed in the **Value** field.



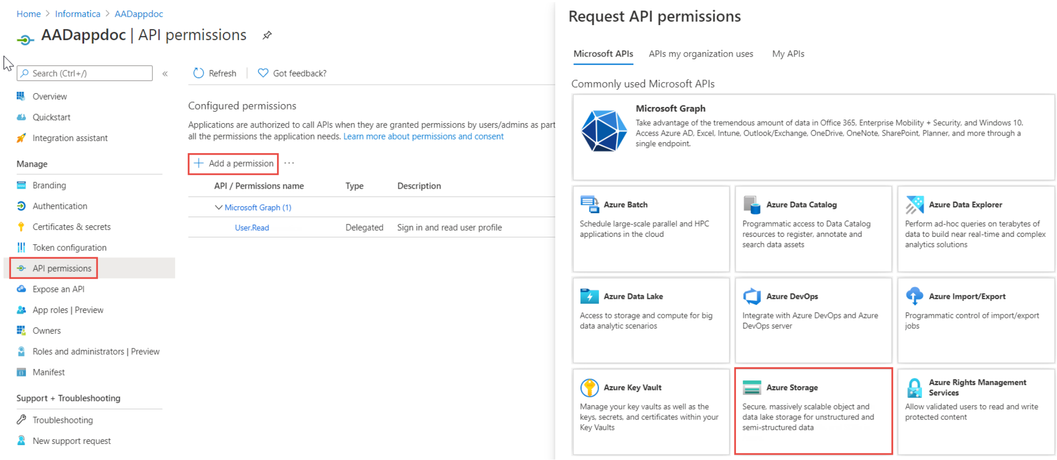
A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

Description	Expires	Value	ID
AADapp_doc	12/31/2299	Lgw******************	f6a97b93-703c-4d8c-a632-4ac9

Note: You must copy the key value as you cannot retrieve the value once you leave the page. Ensure that the client secret does not contain special characters.

9. In the Manage section, click **Owners**.
10. Click **Add owner**.
11. In the **Search** field, search for the owner name or email address that you used to login to Azure portal.
12. Select the owner name or email address and click **Select**.
13. In the Manage section, click **API permissions**.

    The configured permissions are displayed.



Configured permissions

Applications are authorized to call APIs when they are granted permissions by users/admins as part all the permissions the application needs. Learn more about permissions and consent

API / Permissions name	Type	Description
Microsoft Graph (1)		
User.Read	Delegated	Sign in and read user profile

Commonly used Microsoft APIs

Microsoft Graph
Take advantage of the tremendous amount of data in Office 365, Enterprise Mobility + Security, and Windows 10. Access Azure AD, Excel, Intune, Outlook/Exchange, OneDrive, OneNote, SharePoint, Planner, and more through a single endpoint.

Azure Batch
Schedule large-scale parallel and HPC applications in the cloud

Azure Data Catalog
Programmatic access to Data Catalog resources to register, annotate and search data assets

Azure Data Explorer
Perform ad-hoc queries on terabytes of data to build near real-time and complex analytics solutions

Azure Data Lake
Access to storage and compute for big data analytic scenarios

Azure DevOps
Integrate with Azure DevOps and Azure DevOps server

Azure Import/Export
Programmatic control of import/export jobs

Azure Key Vault
Manage your key vaults as well as the keys, secrets, and certificates within your Key Vaults

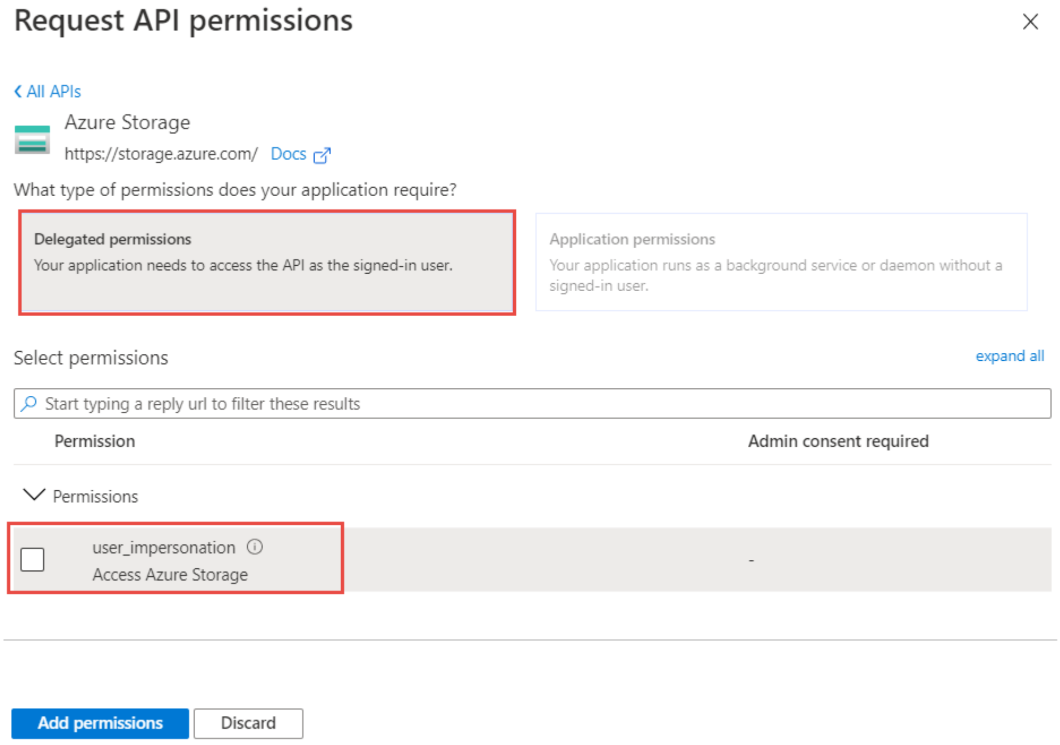
Azure Storage
Secure, massively scalable object and data lake storage for unstructured and semi-structured data

Azure Rights Management Services
Allow validated users to read and write protected content

14. Click **Add a permission**.

    The **Request API permissions** page appears.

15. In the Microsoft APIs section, click **Azure Storage**.
16. Select **Delegated permissions** as the type of permissions.
17. Select **Access Azure Storage** from the listed permissions.



What type of permissions does your application require?

Delegated permissions
Your application needs to access the API as the signed-in user.

Application permissions
Your application runs as a background service or daemon without a signed-in user.

Select permissions

Start typing a reply url to filter these results

Permission	Admin consent required
Permissions	
□ user_impersonation ⓘ Access Azure Storage	-

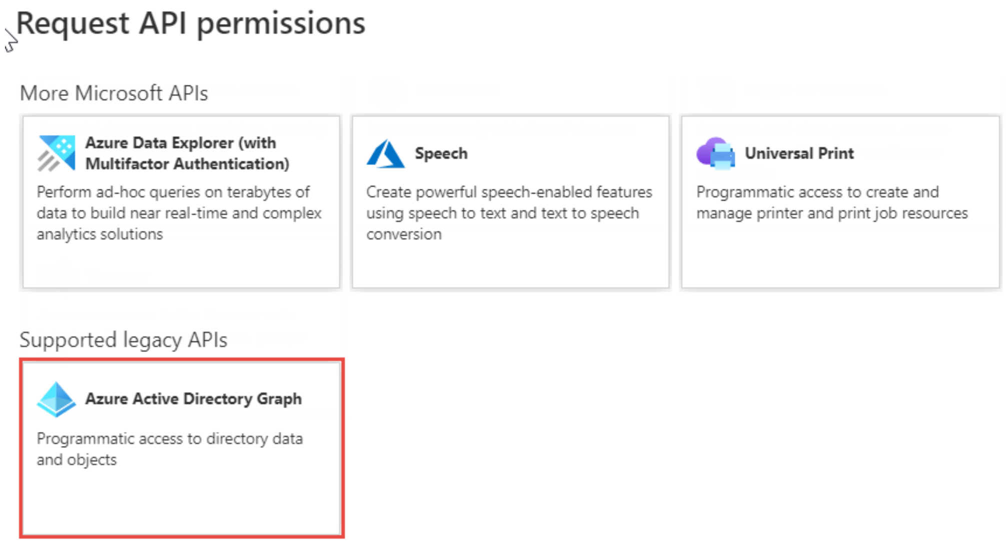
18. Click **Add permissions**.
19. In the **Configured permissions**, select **Azure Active directory** and ensure that the **Sign in and read user profile** option is enabled in the **Delegated permissions** section.

If **Azure Active directory** is not listed under the **Configured permissions**, perform the following steps:

a. Click **Add a permission**.

   The **Request API permissions** page appears.

b. In the Microsoft APIs section, click **Azure Active Directory Graph**.



Request API permissions

More Microsoft APIs

Azure Data Explorer (with Multifactor Authentication)
Perform ad-hoc queries on terabytes of data to build near real-time and complex analytics solutions

Speech
Create powerful speech-enabled features using speech to text and text to speech conversion

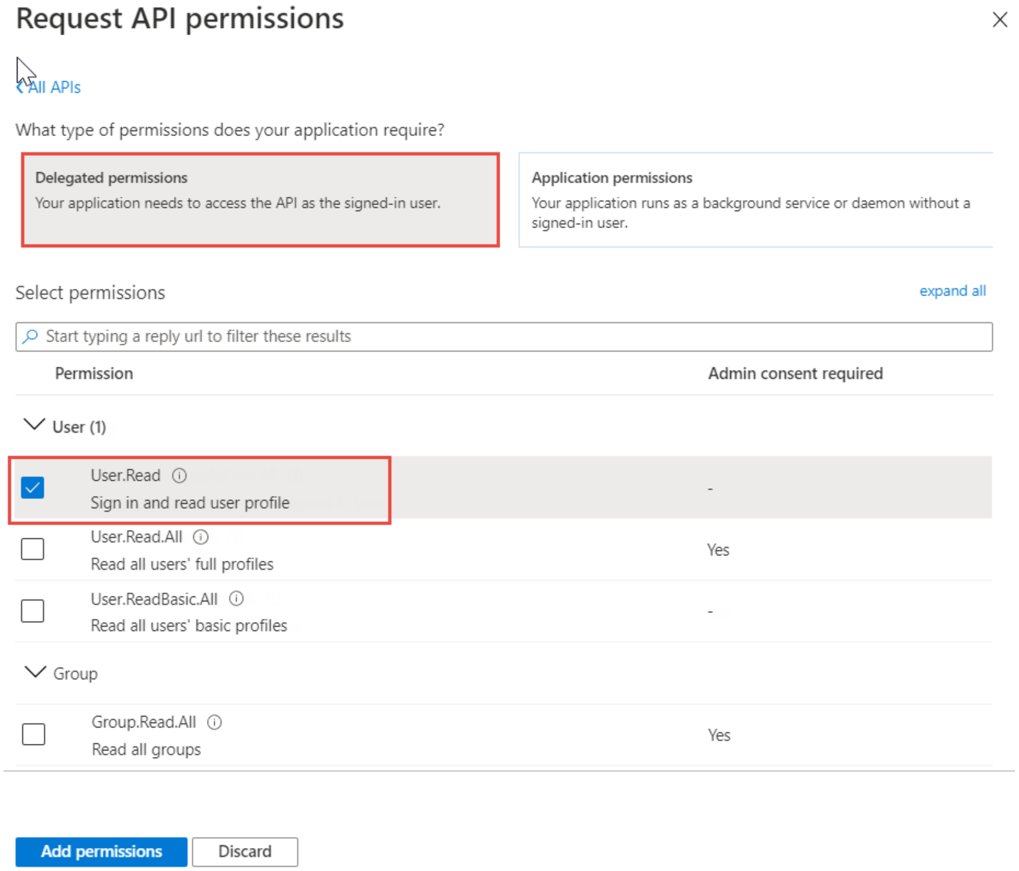
Universal Print
Programmatic access to create and manage printer and print job resources

Supported legacy APIs

Azure Active Directory Graph
Programmatic access to directory data and objects

c. Select **Delegated permissions** as the type of permissions.

d. Select **Sign in and read user profile** from the listed permissions.



Request API permissions

< All APIs

What type of permissions does your application require?

Delegated permissions
Your application needs to access the API as the signed-in user.

Application permissions
Your application runs as a background service or daemon without a signed-in user.

Select permissions

	Permission	Admin consent required
	User (1)	
☑	User.Read Sign in and read user profile	-
□	User.Read.All Read all users' full profiles	Yes
□	User.ReadBasic.All Read all users' basic profiles	-
	Group	
□	Group.Read.All Read all groups	Yes

20. Go to the home page and in the Storage Account section, select the Microsoft Azure Data Lake Storage Gen2 account that you created.

21. Click **Access control (IAM)** > **Add**.

22. In the **Add role assignment** page, provide the **Storage Blob Data Contributor** or the **Storage Blob Data Reader** role to the application.

    Note: To write to or delete Azure Storage containers and blobs, you must have the Contributor role either at the storage account level or the container level.

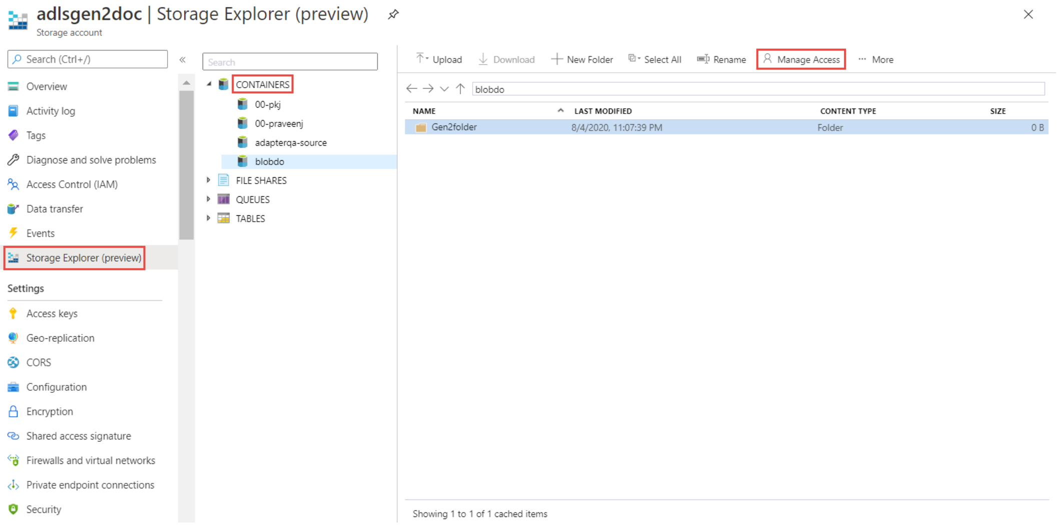
Setting Permissions for Microsoft Azure Data Lake Store Gen2 (Access Control List)

Set the Access Control List to provide permissions to Microsoft Azure Data Lake Store Gen2. To access objects from an HDI 4.0 Kerberised cluster, configure the impersonation user details into your Microsoft Azure Data Lake Storage Gen2 account. Provide Contributor role and full access, for the container used in the internal storage account of the HDInsight Data Lake Storage Gen2 cluster, to the impersonation user.

1. Log in to the Azure portal.

2. In the **Storage Accounts** section, select the Microsoft Azure Data Lake Storage Gen2 account that you created.

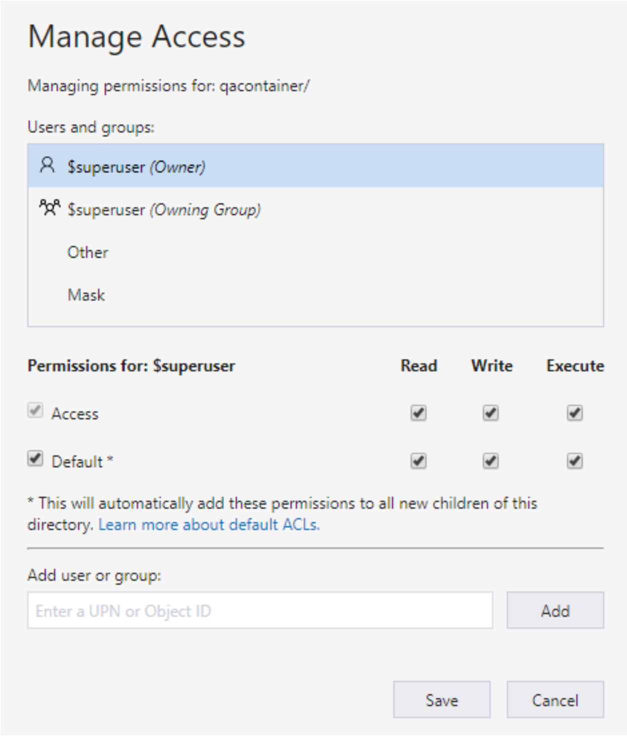
3. Click **Storage Explorer** > **CONTAINERS**.



NAME	LAST MODIFIED	CONTENT TYPE	SIZE
Gen2folder	8/4/2020, 11:07:39 PM	Folder	0 B

4. Select the file system that you created and then click **Manage Access**.

The **Manage Access** dialog box appears.



Manage Access

Managing permissions for: qacontainer/

Users and groups:

- $superuser *(Owner)*
- $superuser *(Owning Group)*
- Other
- Mask

Permissions for: $superuser	Read	Write	Execute
☑ Access	☑	☑	☑
☑ Default *	☑	☑	☑

* This will automatically add these permissions to all new children of this directory. Learn more about default ACLs.

Add user or group:

Enter a UPN or Object ID

5. In the **Permission for:** section, select both the **Access** and **Default** check boxes. Enable the **Read**, **Write**, and **Execute** permissions.
6. In the **Add user or group** field, enter the Azure Active Directory application **Object ID** and click **Add**.

   For information about getting the **Object ID**, see "Registering an Application in Azure Active Directory" on page 8.

   Note: If you enter an Azure Active Directory group name, all the users within the group will have the same permissions.
7. Click **Save**.

When you set the permissions of a file system, all the folders within that file system have the same permissions. However, if you create a folder within a file system before setting the permissions of the file system, you must perform the same steps to set the Access Control List for that folder.

Setting the Connection Properties to Create a Microsoft Azure Data Lake Storage Gen2 Connection

When you complete all the prerequisite tasks, perform the following steps to create a Microsoft Azure Data Lake Store Gen2 Connection:

1. In the Developer tool, click **Windows** > **Preferences**.
2. Select **Informatica** > **Connections**.
3. Expand the domain in the Available Connections.
4. Select the connection type **File System** > **Microsoft Azure Data Lake Storage Gen2**, and click Add.
5. Enter a connection name and an optional description.

6. Select Microsoft Azure Data Lake Storage Gen2 as the connection type.
7. Click **Next**.
8. Enter the **Storage account name** in the **Account Name** connection property.
9. Enter the **Application ID** in the **Client ID** connection property.
10. Enter the **Value** of the key in the **Client Secret** connection property.
11. Enter the **Directory ID** in the **Tenant ID** connection property.
12. Enter the **File System Name**.
13. Enter the path of the folder in the **Directory Path** connection property.
14. Click **Test Connection** to validate the connection and then click **Finish**.

Author

Adrija Pandya

Acknowledgements

The author would like to acknowledge A.S. Megha, for her technical assistance with this article.