How-To Library

Informatica

Prerequisites to Create a Microsoft Azure Data Lake Storage Gen2 Connection

© Copyright Informatica LLC 2019, 2022. Informatica, the Informatica logo, and Informatica PowerExchange are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at https://www.informatica.com/trademarks.html.

Abstract

You can use Microsoft Azure Data Lake Storage Gen2 Connector to connect to Microsoft Azure Data Lake Storage Gen2 from Cloud Data Integration. This article explains the prerequisite tasks that you must complete before you create a Microsoft Azure Data Lake Storage Gen2 Connection.

Supported Versions

- Informatica Cloud® Data Integration Microsoft Azure Data Lake Storage Gen2 Connector

Table of Contents

Overview

You can use Microsoft Azure Data Lake Storage Gen2 Connector 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 and enable **Hierarchical namespace** in the Azure portal.
   You can use role-based access control or access control lists to authorize the users to access the resources in the storage account.
   - **Role-based access control**
     If you use role-based access control, assign the Contributor role or Reader role to the 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.
   - **Access control lists**
     If you use access control lists, you can provide read, write, and execute permissions to each directory and file for users.
2. Create a Blob container in the storage account.
3. Register an application in Azure Active Directory to authenticate users to access the Microsoft Azure Data Lake Storage Gen2 account.

You can use role-based access control or access control lists to authorize the application.

- **Role-based access control**
  If you use role-based access control, assign the 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.

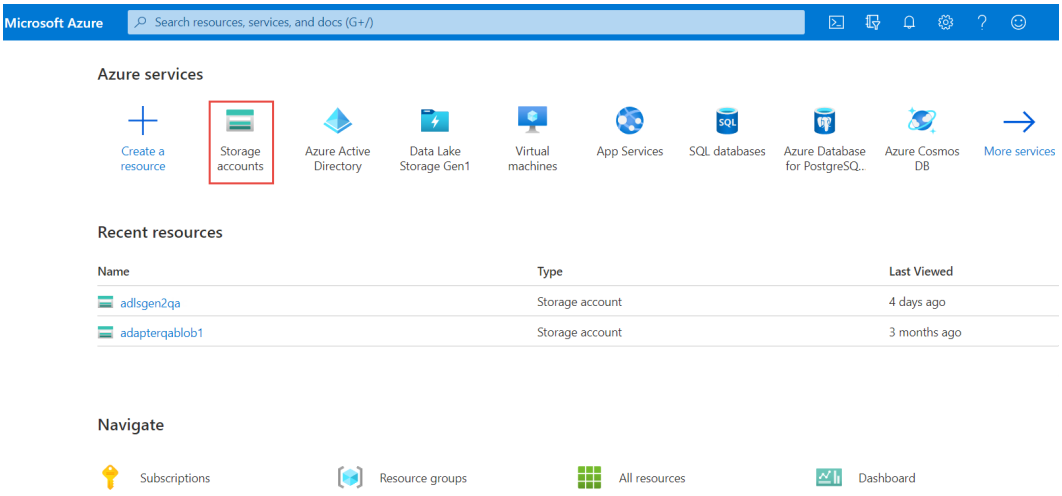
- **Access control lists**
  If you use access control lists, you can provide read, write, and execute permissions to each directory and file in the container.

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*.

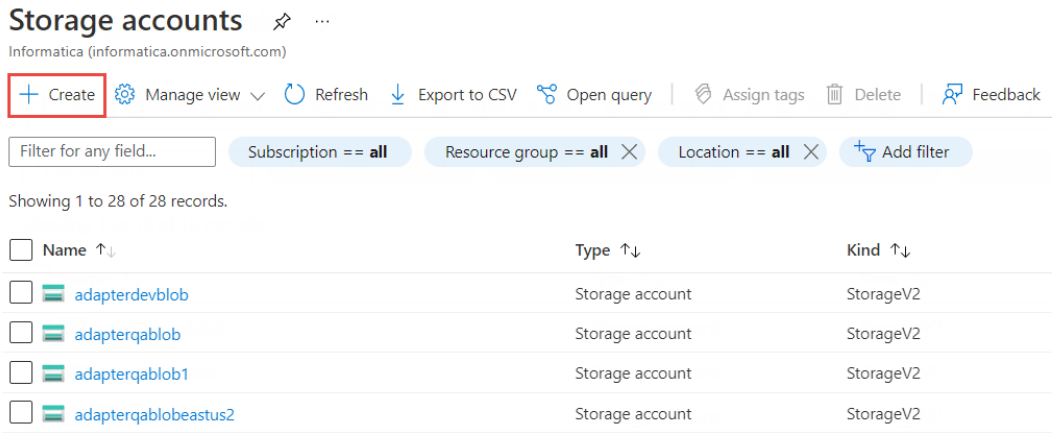
Creating a Storage Account to use with Microsoft Azure Data Lake Storage Gen2

Perform the following steps to create a storage account:

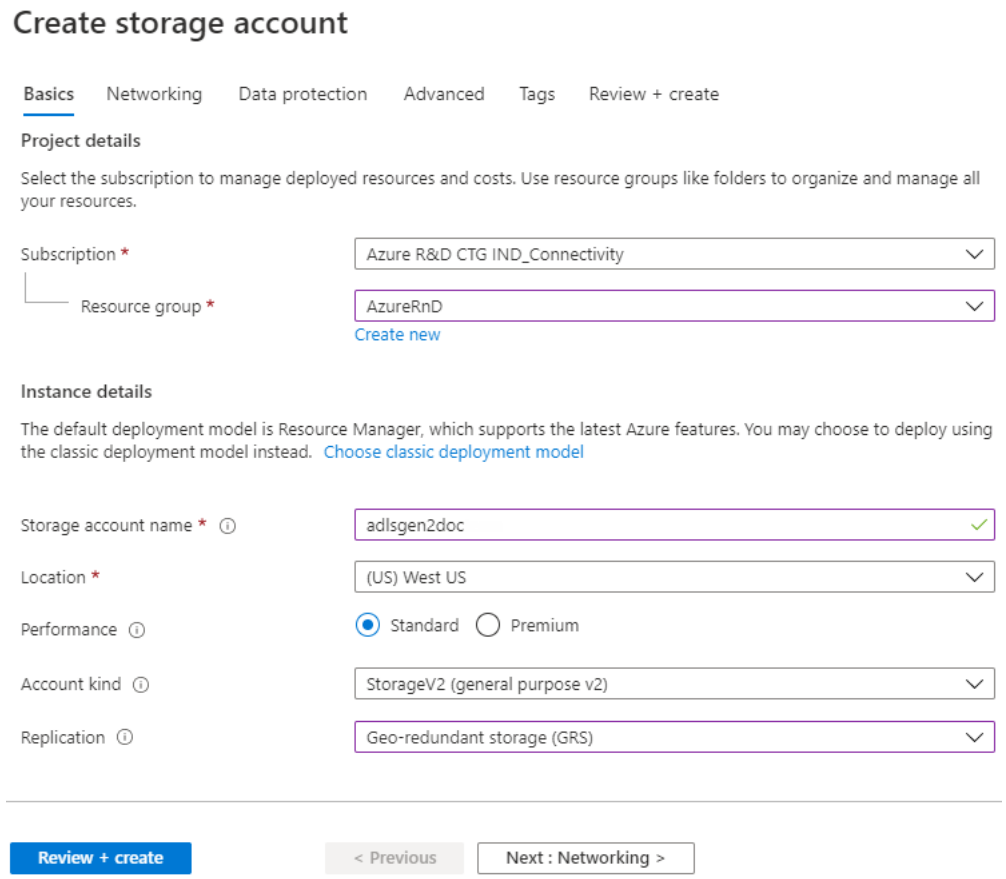
1. Log in to the following Azure portal: https://portal.azure.com/
2. Under Azure Services, click **Storage accounts**.



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



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



   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**.

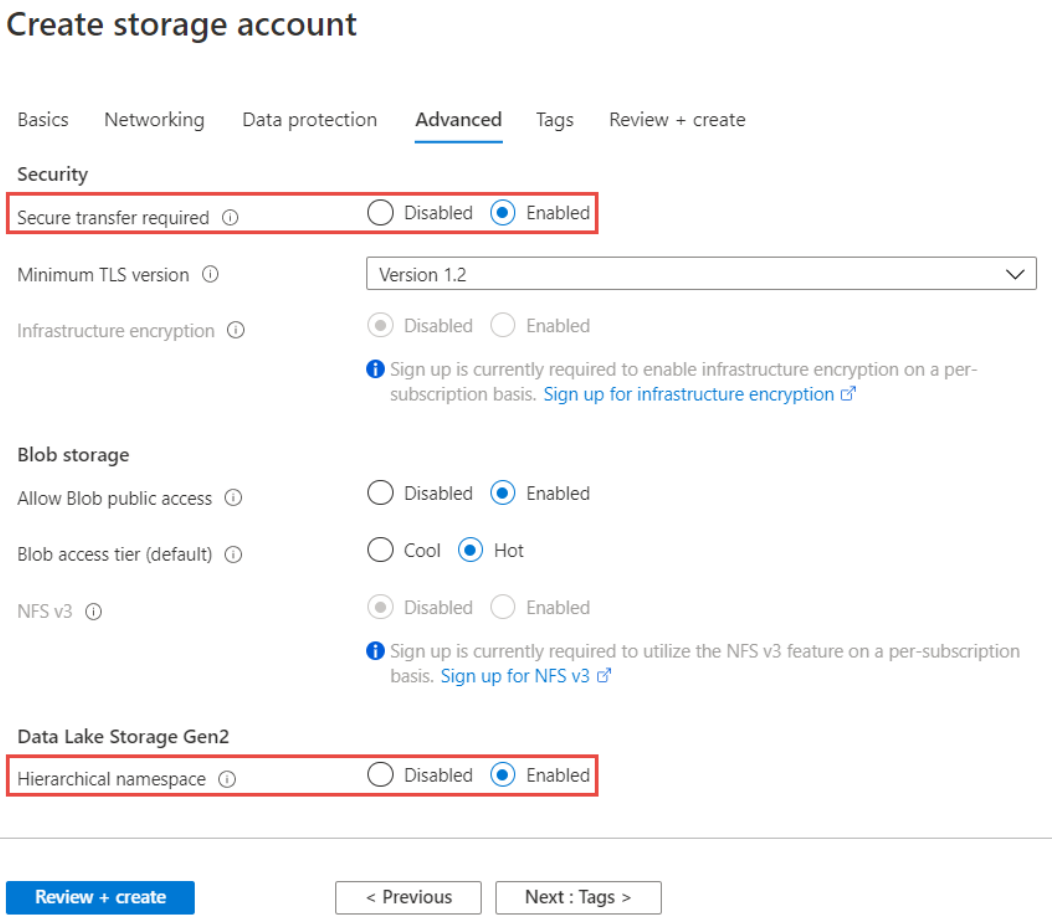
f. In the **Account kind** field, select **StorageV2 (general purpose v2)**.

   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)**.

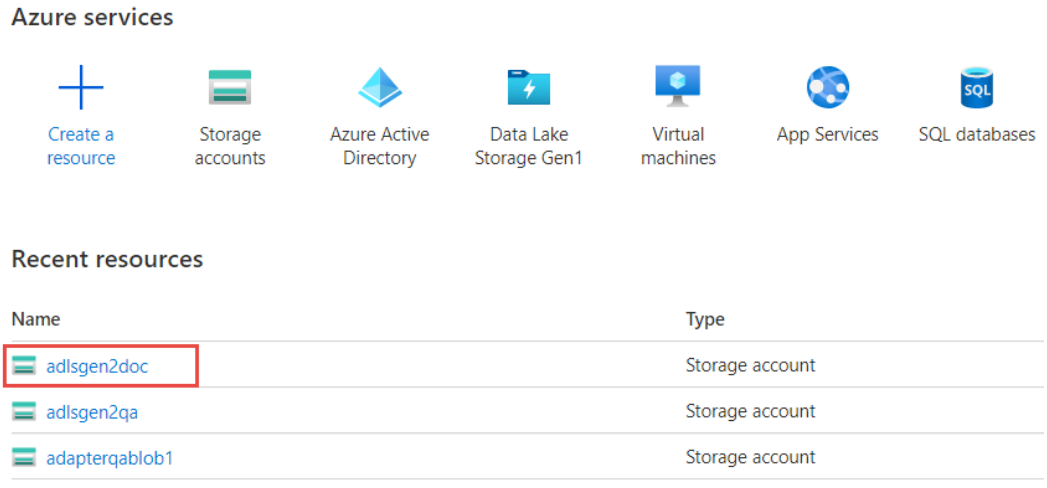
   The replication type specifies how the storage account will be replicated.

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

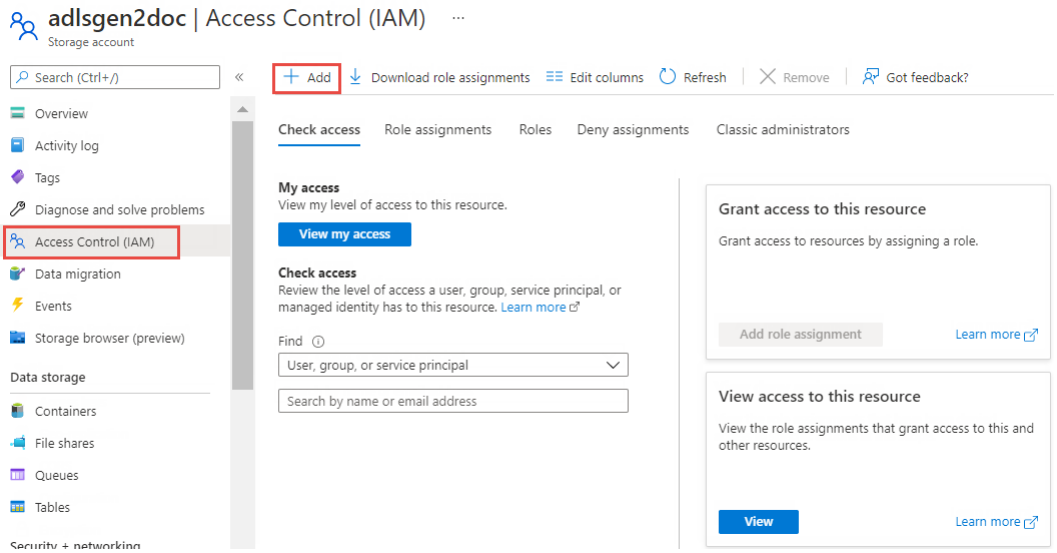


6. Click **Review + Create** > **Create**.

7. Click on the newly created storage account name.



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



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.

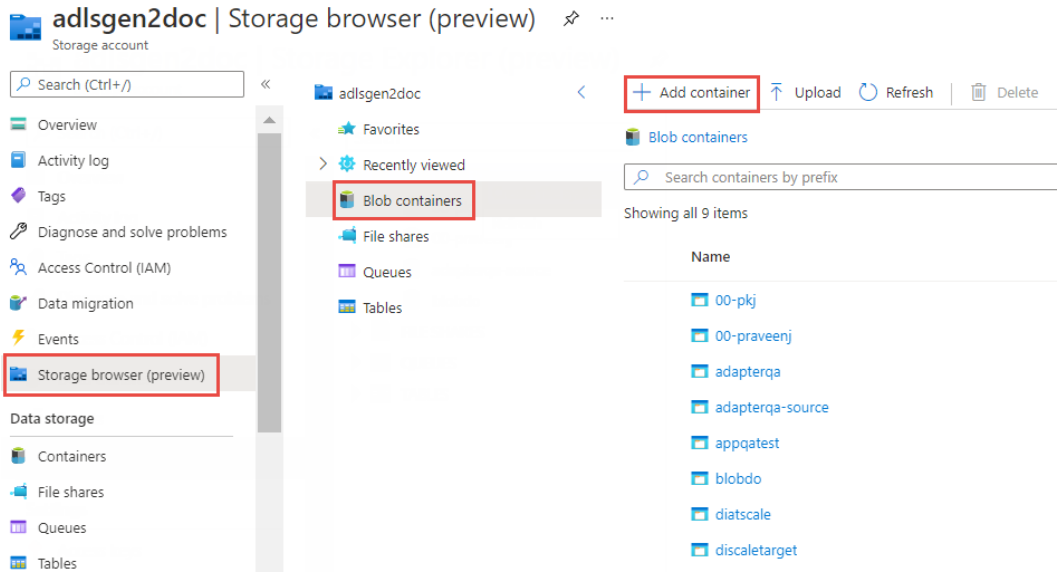
   b. In the **Assign access to** field, select **Azure AD user, group or service principal**.
   c. In the **Select** field, select the user that requires access to the storage account.
   d. Click **Save**.

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

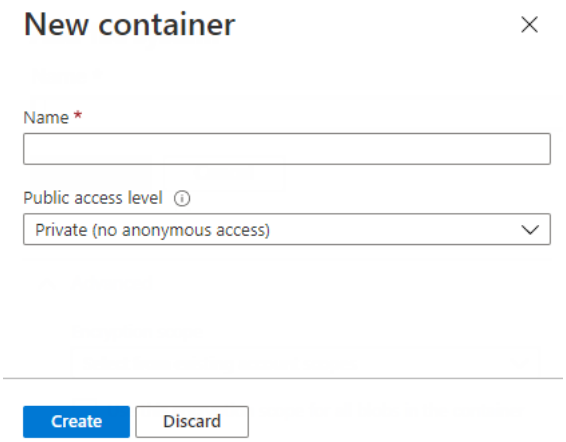
Creating a Blob Container in the Storage Account

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**.



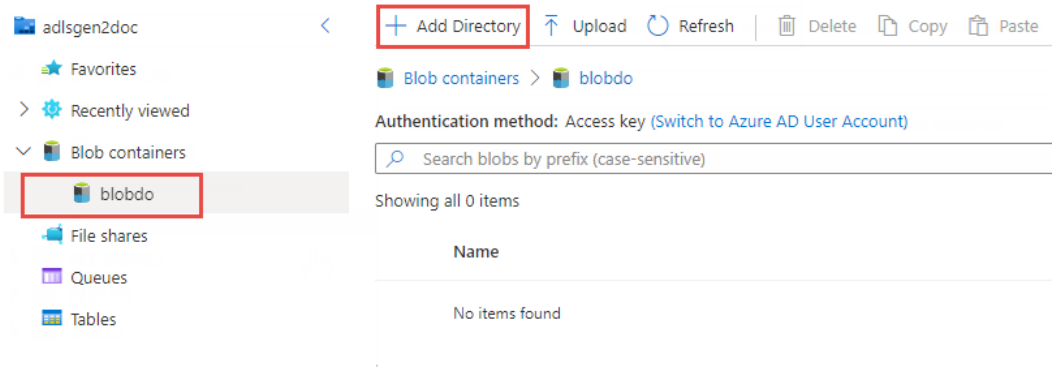
5. Enter a name for the new container.



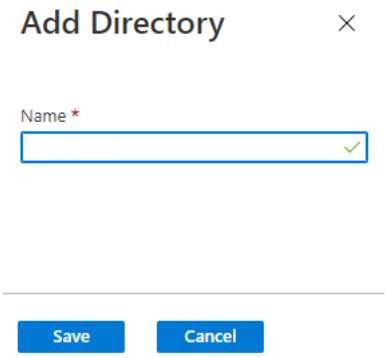
Note: You can only use lowercase letters, numbers, and hyphens when you enter the names of the folder and the file system. The names must begin and end with a letter or number. Do not add consecutive hyphens when you enter the names of the folder and the file system.

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.



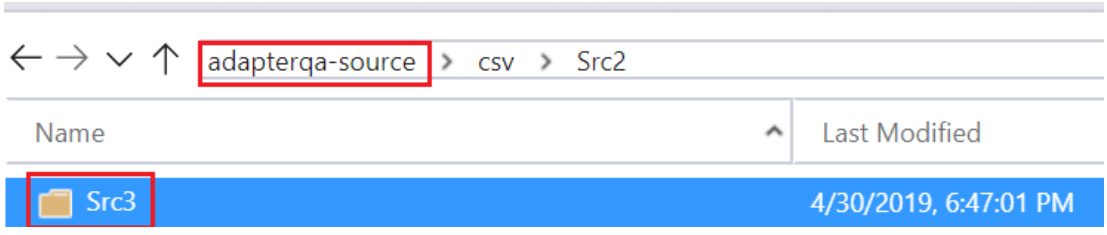
10. Enter a name for the Microsoft Azure Data Lake Storage Gen2 folder and then click **Save**.



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:

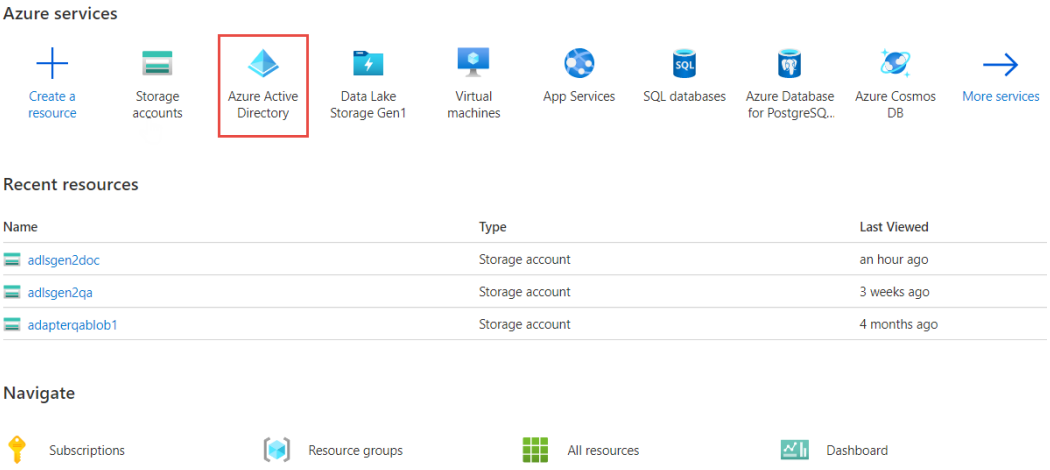


Registering an Application in Azure Active Directory

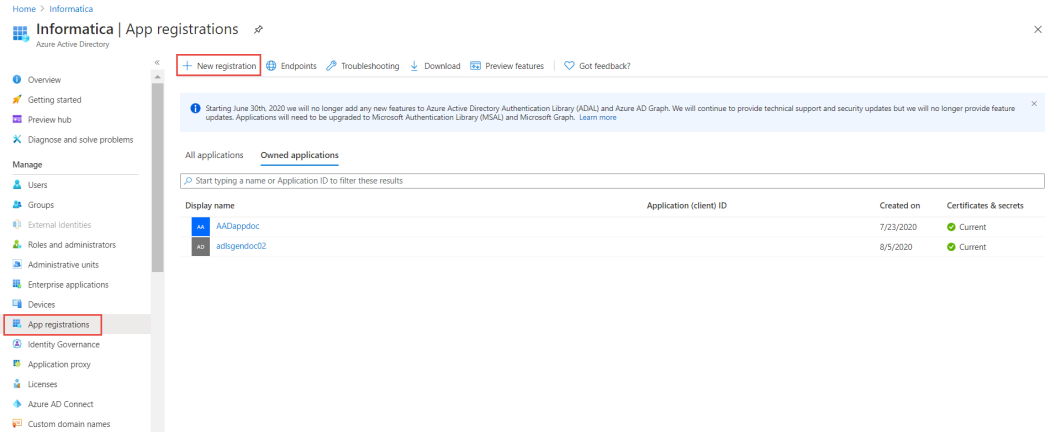
Register a new application in Azure Active Directory to authenticate access to the storage account.

1. Log in to the Azure portal.

2. Click **Azure Active Directory**.

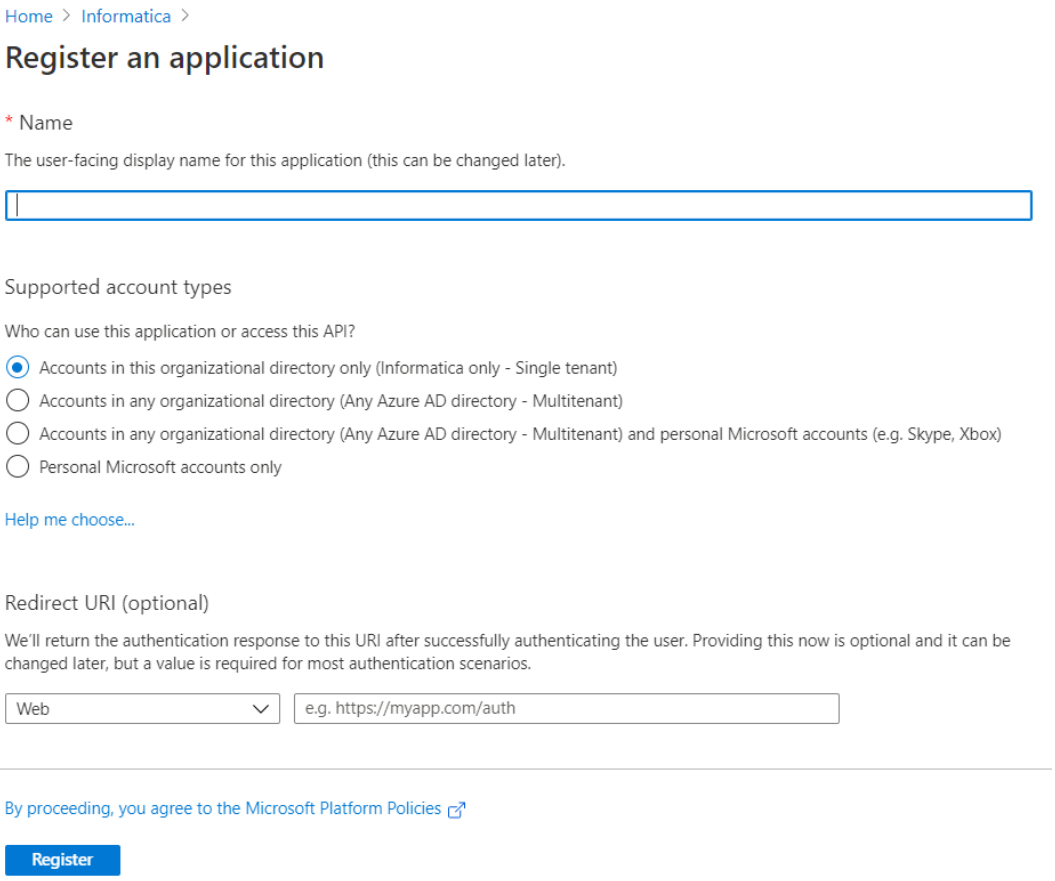


3. In the **Manage** section, click **App registrations**.



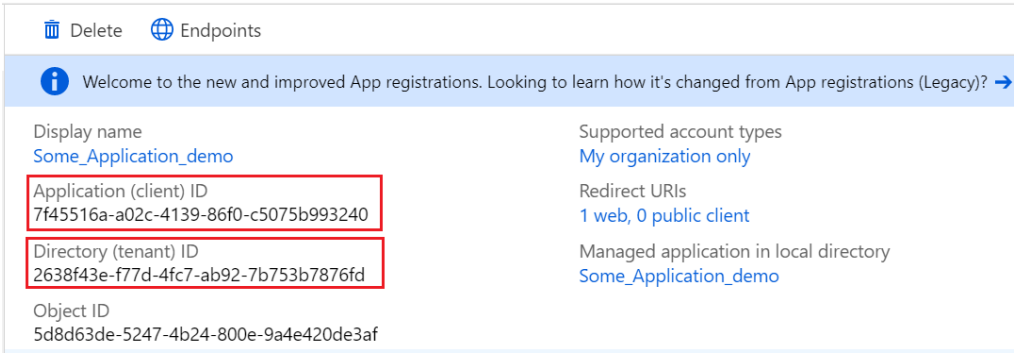
4. Click **New registration** to create a new Azure Active Directory application.

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



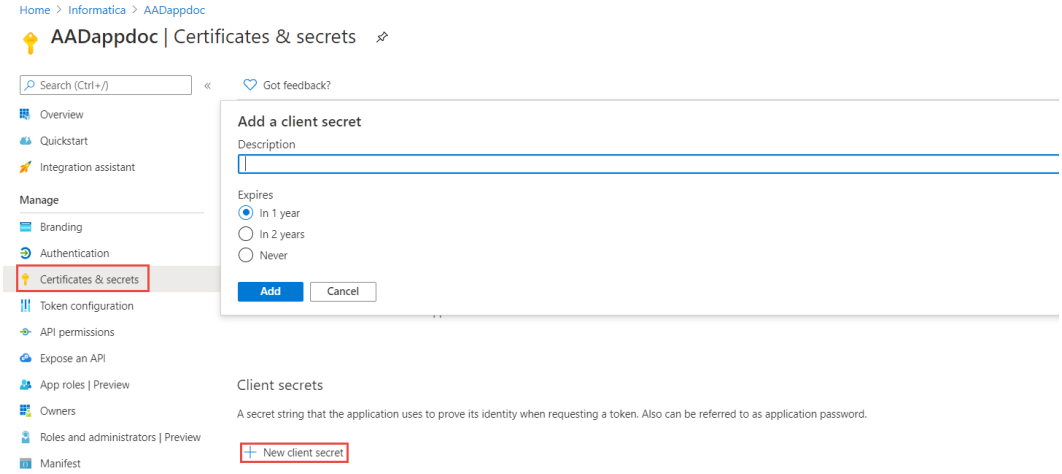
   a. In the **Name** field, enter the application name.
   b. In the **Redirect URI** section, select **Web** as the type of the application and enter the URL of the application.
   c. Click **Register**.

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

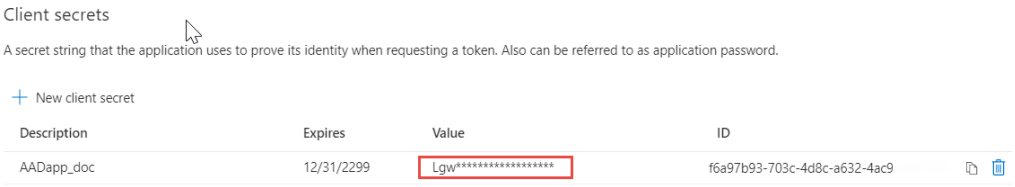


6. In the Manage section, click **Certificates & secrets** section.

7. Click **New client secret**.

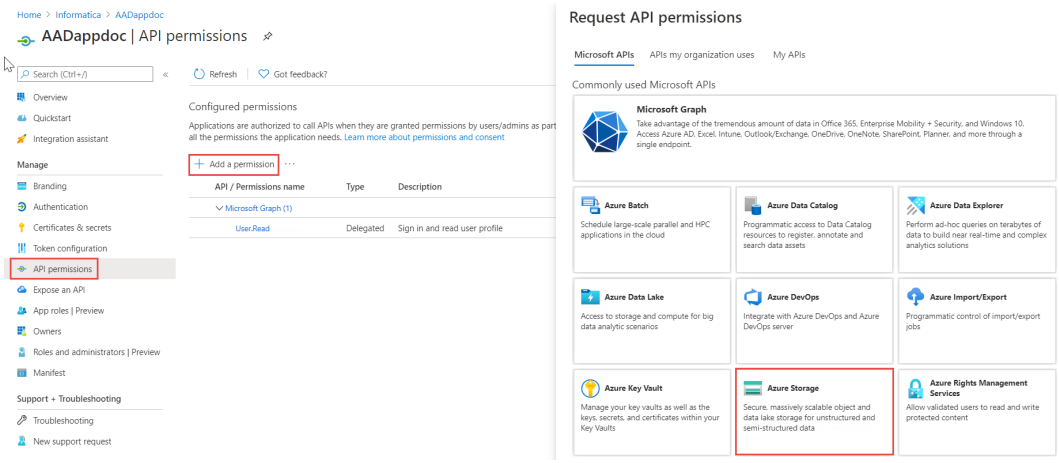


8. In the **Add a client secret** page, perform the following steps:
   a. Enter a name for the client secret in the **Description** field.
   b. In the **Expires** field, you can select the duration of the key as **Never**(Recommended).
   c. Click **Add**.
   d. The value of the key is generated and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



   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.



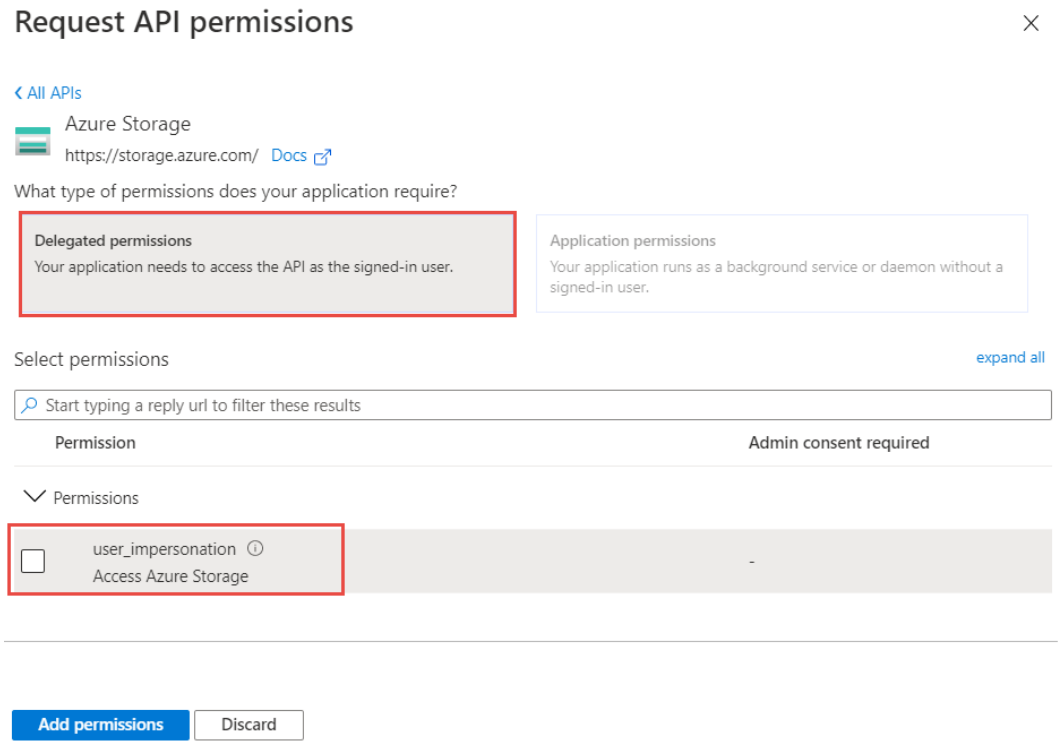
14. Click **Add a permission**.

    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.



18. Click **Add permissions**.

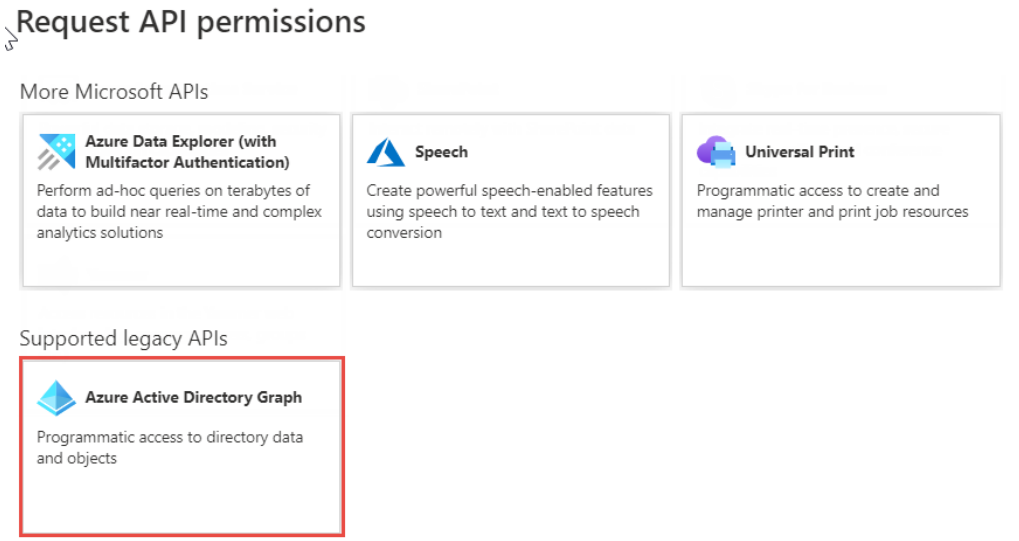
19. In the **Configured permissions**, select **Azure Active directory** and ensure that the **Sign in and read user profile** option is enabled in the **Delegated permissions** section.

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

a. Click **Add a permission**.

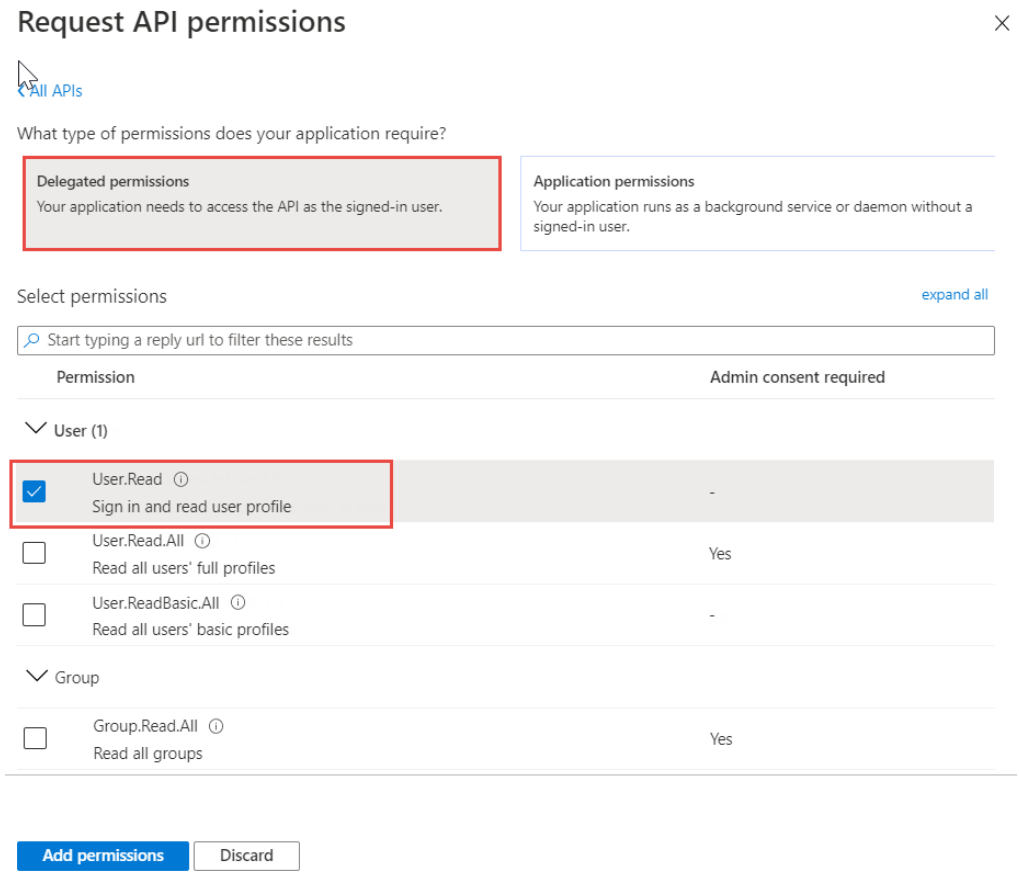
The **Request API permissions** page appears.

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



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

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



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 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. Log in to Informatica Intelligent Cloud Services.
2. Click **Administrator**.
3. Click **Connections** > **New Connection**.
4. Configure the following connection properties:

Property	Description
Connection Name	Name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + -, Maximum length is 255 characters.
Description	Description of the connection. Maximum length is 4000 characters.
Type	The Microsoft Azure Data Lake Storage Gen2 connection type.
Runtime Environment	The name of the runtime environment where you want to run the tasks. Select a Secure Agent, Hosted Agent, or serverless runtime environment.
Account Name	Microsoft Azure Data Lake Storage Gen2 account name or the service name.
Authentication Type	Authentication type to access the Microsoft Azure Data Lake Storage Gen2 account. Select one of the following options: - Service Principal Authentication. Uses the client ID, client secret, and tenant ID to connect to Microsoft Azure Data Lake Storage Gen2. - Shared Key Authentication. Uses the account key to connect to Microsoft Azure Data Lake Storage Gen2. - Managed Identity Authentication. Select to authenticate using identities that are assigned to applications in Azure to access Azure resources in Microsoft Azure Data Lake Storage Gen2.

Property	Description
Client ID	Applies to Service Principal Authentication and Managed Identity Authentication. The client ID of your application. To use service principal authentication, specify the application ID or client ID for your application registered in the Azure Active Directory. To use managed identity authentication, specify the client ID for the user-assigned managed identity. If the managed identity is system-assigned, leave the field empty.
Client Secret	Applies to Service Principal Authentication. The client secret key to complete the OAuth authentication in the Azure Active Directory.
Tenant ID	Applies to Service Principal Authentication. The directory ID of the Azure Active Directory.
Account Key	Applies to Shared Key Authentication. The account key for the Microsoft Azure Data Lake Storage Gen2 account.
File System Name	The name of the file system in the Microsoft Azure Data Lake Storage Gen2 account.
Directory Path	The path of an existing directory without the file system name. You can select one of the following syntax: - / for root directory - /dir1 - dir1/dir2 There is no default directory.
Adls Gen2 End-point	The type of Microsoft Azure endpoints. Select one of the following endpoints: - core.windows.net. Connects to Azure endpoints. - core.usgovcloudapi.net. Connects to US government Microsoft Azure Data Lake storage Gen2 endpoints. - core.chinacloudapi.cn. Connects to Microsoft Azure Data Lake storage Gen2 endpoints in the China region. Default is core.windows.net. **Note:** You cannot configure the Azure Government endpoints in elastic mappings.

5. Click **Test Connection** to validate the connection.
6. Click **Save**.

Author

Adrija Pandya

Acknowledgements

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