

Configuring Azure ADLS Gen2 with Qlik Data Integration

TABLE OF CONTENTS

1. Create or Configure Azure Data Lake Storage (ADLS) Gen2	2
2. Create an Active Directory and Perform “App Registration”	6
3 Finalize Configuration of the ADLS-2 Storage Account	11
4. Create a Target Directory in the File System	14

SUMMARY

This document was created to supplement Qlik Replicate and Qlik Compose Documentation. This guide is for customers intending to use Qlik Data Integration and Azure ADLS Gen2 and Azure Databricks. The official documentation can be found at

- <https://help.qlik.com/en-US/compose/Content/Compose/Home.htm>
- <https://help.qlik.com/en-US/replicate/Content/Replicate/Home.htm>

Azure ADLS Gen2 is foundation piece of Azure Technology and is the Object Store underpinning Azure Data Lake service.

Please see official documentation about ADLS Gen2 here:

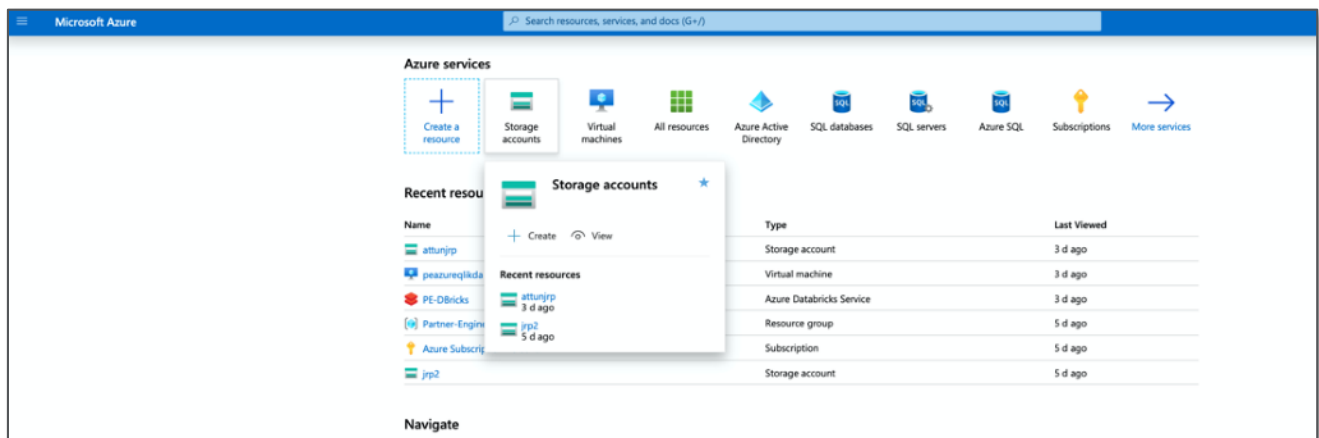
- <https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction>

Configuring Azure Components

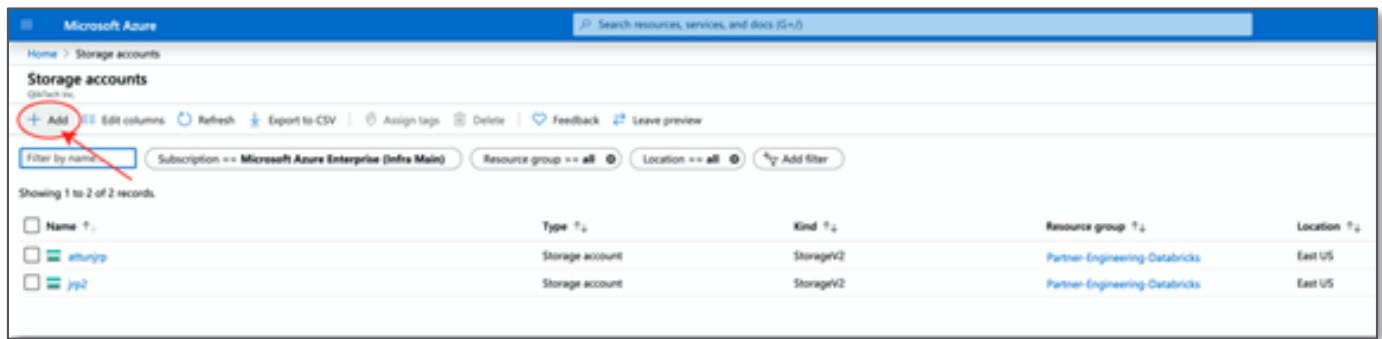
1. Create or Configure Azure Data Lake Storage (ADLS) Gen2

First, we need to set up the Azure storage account that Qlik Replicate will use to map data into Azure Databricks. We will setup Azure Data Lake Storage (ADLS Gen2) to manage the external tables.

If you do not already have an ADLS Gen2 storage account, we need to create one now. From your Azure portal home page, click on Storage Accounts.



Inside Storage Accounts click “+ Add”.



Select your Azure subscription and an existing resource group or select “Add” to create a new storage account. Next choose a name for your storage account and a location. For “Account kind” be sure that it says StorageV2 (general purpose v2). Click “Next: Networking” at the bottom of the page. Make note of the name of the “Storage account name” on this page.

*Refer to your IT Policies for Subscriptions and Resource Groups

Home > Storage accounts > Create storage account

Create storage account

Basics Networking Advanced Tags Review + create

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

Subscription * Microsoft Azure Enterprise (Infra Main)

Resource group * Partner-Engineering-Databricks

[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 * pedatabricks

Location * (US) East US

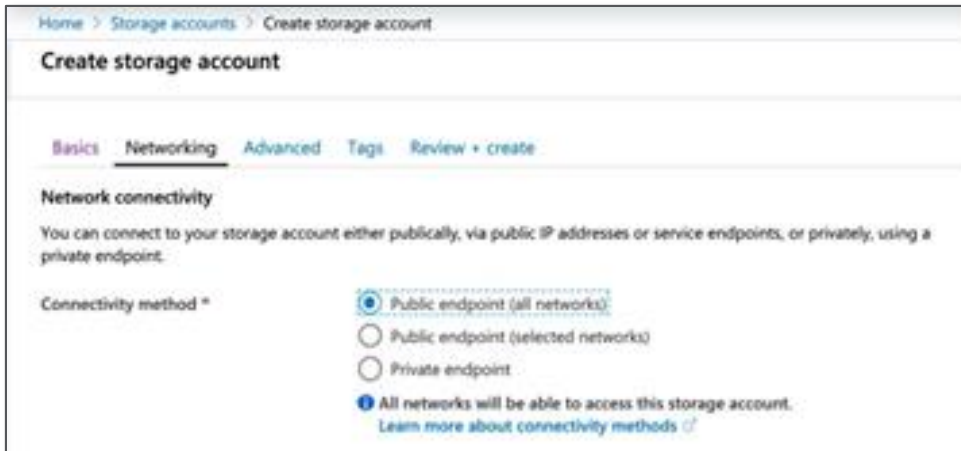
Performance Standard Premium

Account kind StorageV2 (general purpose v2)

Replication Read-access geo-redundant storage (RA-GRS)

Access tier (default) Cool Hot

For this setup we want to make Storage open to public, but you can change the setup to fit your organization needs. If Storage Account Network is configured Private endpoints with Private virtual networks, make sure to add 2 subnets that Azure Databricks creates so Storage Account is accessible to Azure Databricks. Upon launching Azure Databricks, the service will create “Databricks-Private” and “Databricks-Public” subnets.



Home > Storage accounts > Create storage account

Create storage account

Basics **Networking** Advanced Tags Review + create

Network connectivity

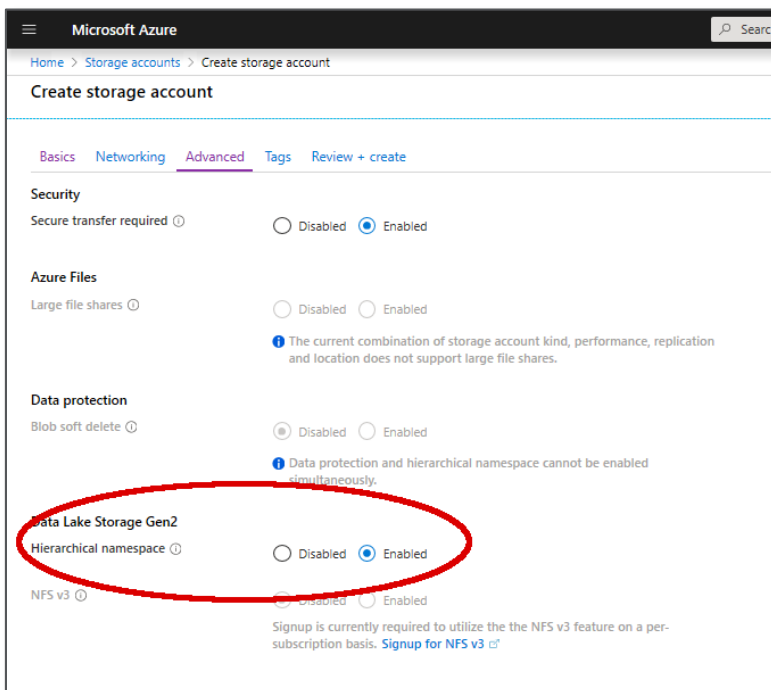
You can connect to your storage account either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Connectivity method *

- ☒ Public endpoint (all networks)
- ☐ Public endpoint (selected networks)
- ☐ Private endpoint

All networks will be able to access this storage account. [Learn more about connectivity methods](#)

Click “Next: Advanced” at the bottom of the page. On the Advanced tab, we want to be sure that Data Lake Storage Gen2 “Hierarchical namespace” is enabled.



Microsoft Azure

Home > Storage accounts > Create storage account

Create storage account

Basics Networking **Advanced** Tags Review + create

Security

Secure transfer required ☐ Disabled ☒ Enabled

Azure Files

Large file shares ☐ Disabled ☐ Enabled

The current combination of storage account kind, performance, replication and location does not support large file shares.

Data protection

Blob soft delete ☒ Disabled ☐ Enabled

Data protection and hierarchical namespace cannot be enabled simultaneously.

Data Lake Storage Gen2

Hierarchical namespace ☐ Disabled ☒ Enabled

NFS v3 ☐ Disabled ☐ Enabled

Signup is currently required to utilize the the NFS v3 feature on a per-subscription basis. [Signup for NFS v3](#)

Configuration is complete. Press Review + create at the bottom of the page. Azure will validate what you have configured.

Microsoft Azure

Home > Storage accounts > Create storage account

Create storage account

✓ Validation passed

Basics Networking Advanced Tags **Review + create**

Basics

Subscription	Microsoft Azure Enterprise (Infra Main)
Resource group	Partner-Engineering-Databricks
Location	(US) East US
Storage account name	pedatabricks
Deployment model	Resource manager
Account kind	StorageV2 (general purpose v2)
Replication	Read-access geo-redundant storage (RA-GRS)
Performance	Standard
Access tier (default)	Hot

Networking

Connectivity method	Public endpoint (all networks)
---------------------	--------------------------------

Advanced

Secure transfer required	Enabled
Blob soft delete	Disabled
Blob change feed	Disabled
Hierarchical namespace	Enabled
NFS v3	Disabled

Tags

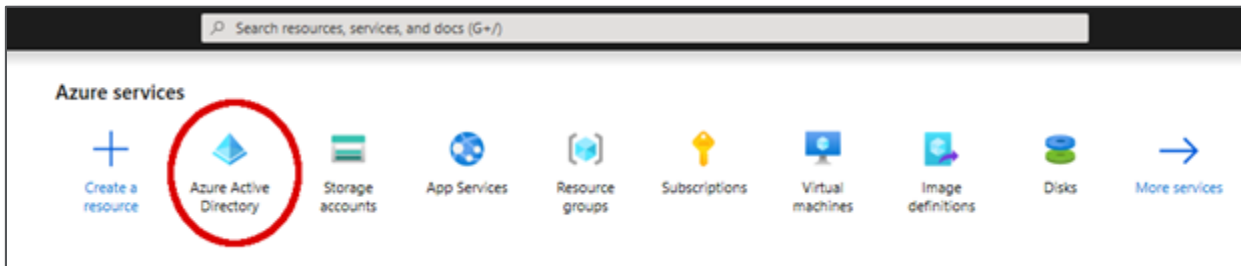
owner	John Park (Storage account)
-------	-----------------------------

[Create](#) [< Previous](#) [Next >](#) [Download a template for automation](#)

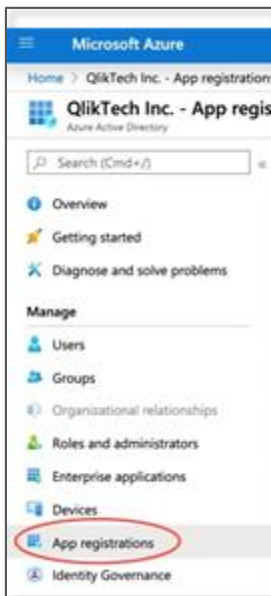
Assuming your configuration passed the validation step, press “Create” button at the bottom of the page. It will take Azure a few minutes to create your storage account.

2. Create an Active Directory and perform “App Registration”

Now we need to create an Azure “App Registration” with appropriate permissions that Qlik Replicate will use when writing data to ADLS Gen2 storage. From the Azure portal home page, click on “Azure Active Directory”.



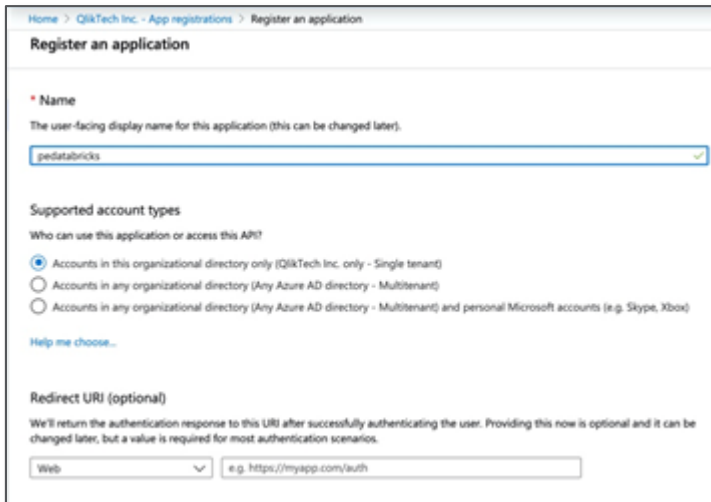
Click on “App registrations” on the left side of the screen.



Select “App registrations” from side menu and click “+ New registration”.

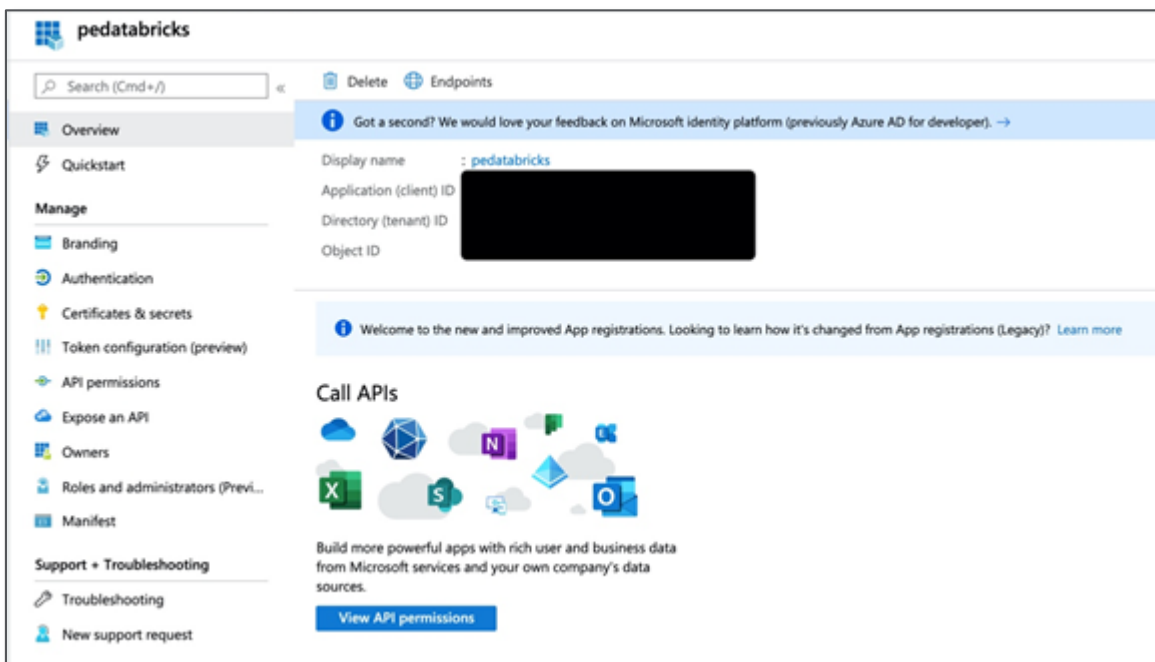


Choose a name for your App and press “Register” at the bottom of the page. You do not need to enter anything for the “Redirect URI”. It is optional and not required in our case.



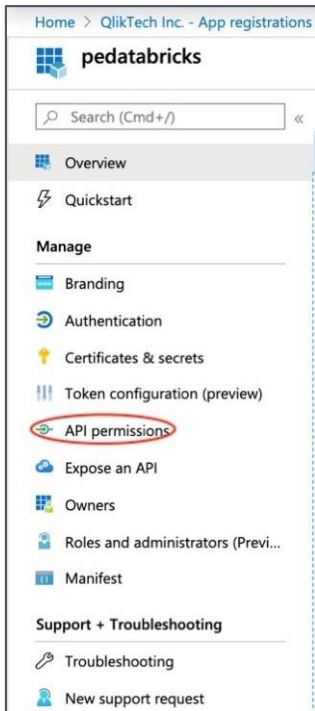
The screenshot shows the 'Register an application' page in the Azure portal. The breadcrumb navigation at the top reads 'Home > QlikTech Inc. > App registrations > Register an application'. The page title is 'Register an application'. Under the 'Name' section, there is a text input field containing 'pedatabricks' with a checkmark icon to its right. Below this, the 'Supported account types' section asks 'Who can use this application or access this API?' and has three radio button options: 'Accounts in this organizational directory only (QlikTech Inc. only - Single tenant)' (selected), 'Accounts in any organizational directory (Any Azure AD directory - Multitenant)', and 'Accounts in any organizational directory (Any Azure AD directory - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)'. A 'Help me choose...' link is below. The 'Redirect URI (optional)' section has a dropdown menu set to 'Web' and a text input field containing 'e.g. https://myapp.com/auth'.

Registration is almost immediate. Make note of the location of the “Application (client) ID” and “Directory (tenant) ID” fields at the top of the page. You will need this information later when configuring Qlik Replicate™ to land data in Azure Databricks.

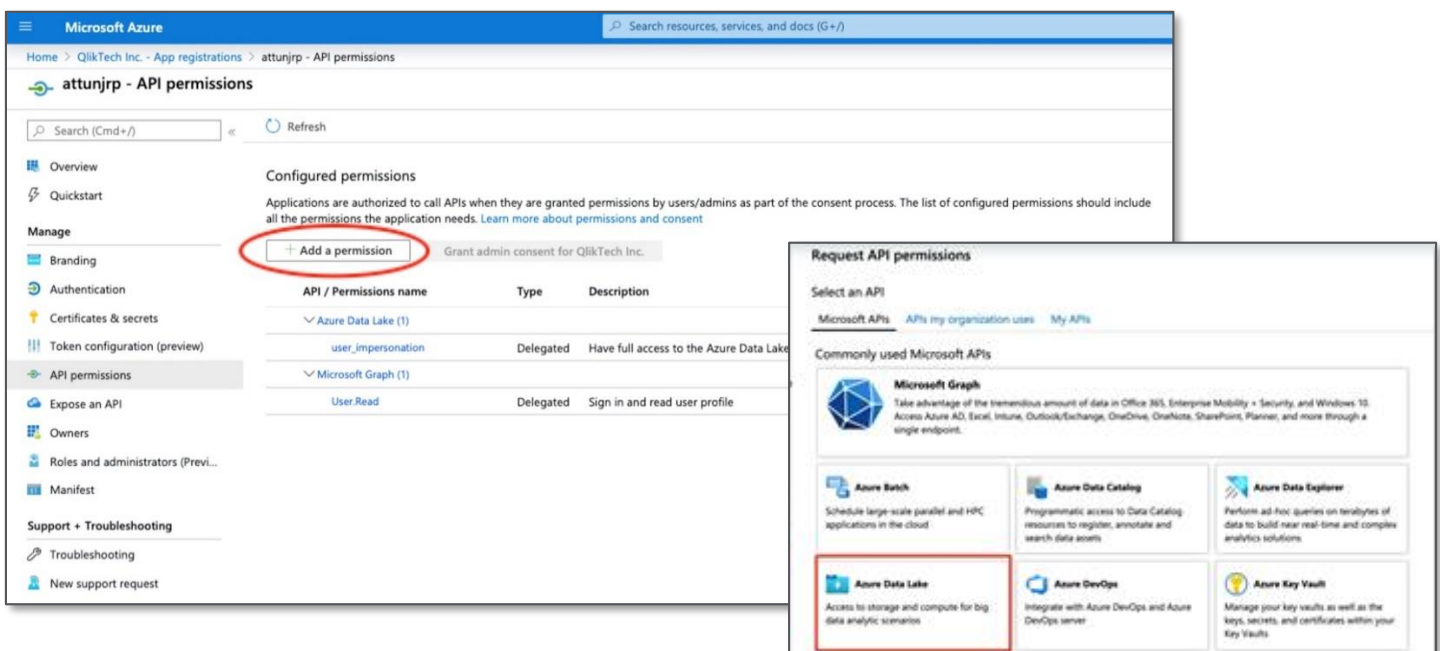


The screenshot shows the 'pedatabricks' application overview page in the Azure portal. The left sidebar contains a navigation menu with sections: 'Overview' (selected), 'Quickstart', 'Manage' (with sub-items: Branding, Authentication, Certificates & secrets, Token configuration (preview), API permissions, Expose an API, Owners, Roles and administrators (Previ...), Manifest), and 'Support + Troubleshooting' (with sub-items: Troubleshooting, New support request). The main content area shows the application details: 'Display name' is 'pedatabricks', 'Application (client) ID' is a blacked-out field, 'Directory (tenant) ID' is a blacked-out field, and 'Object ID' is a blacked-out field. Below this is a message: 'Welcome to the new and improved App registrations. Looking to learn how it's changed from App registrations (Legacy)? Learn more'. At the bottom, there is a 'Call APIs' section with icons for various Microsoft services (Azure, Databricks, OneDrive, OneNote, Outlook, Excel, SharePoint, Teams, Word) and a text block: 'Build more powerful apps with rich user and business data from Microsoft services and your own company's data sources.' with a 'View API permissions' button.

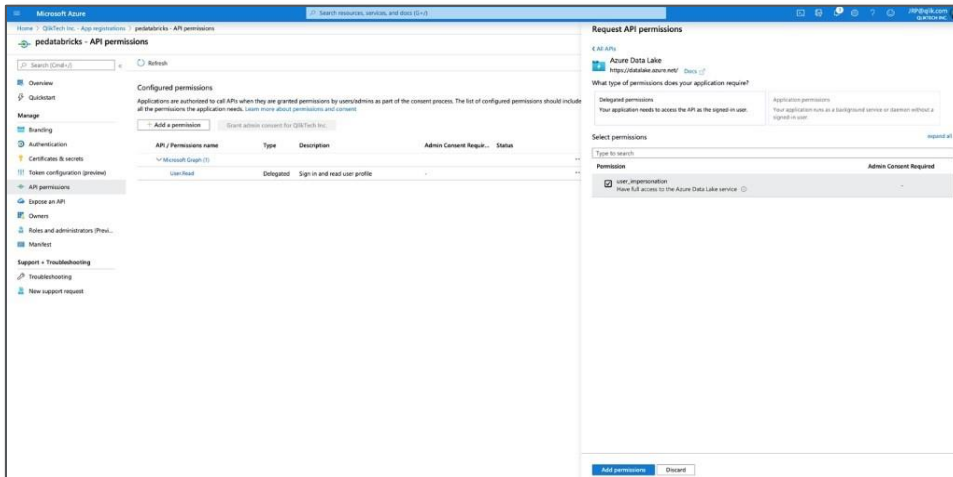
Next, we need to grant this application some basic permissions. Click on API permissions on the left side of the page.



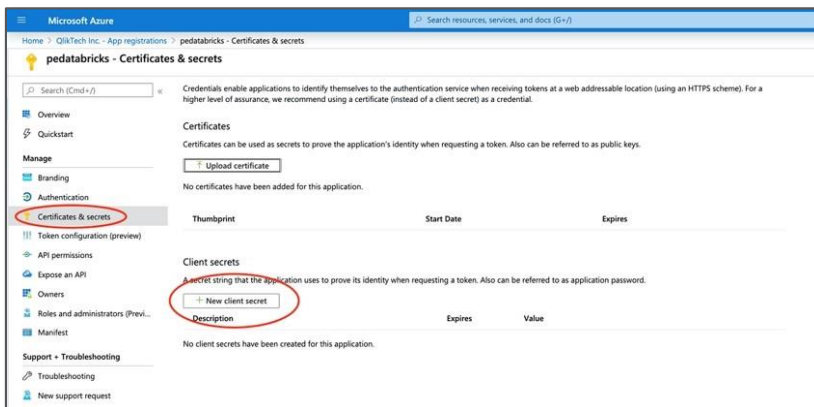
Click "+ Add permission" followed by Azure Data Lake under "Microsoft APIs" on the right side of the screen.



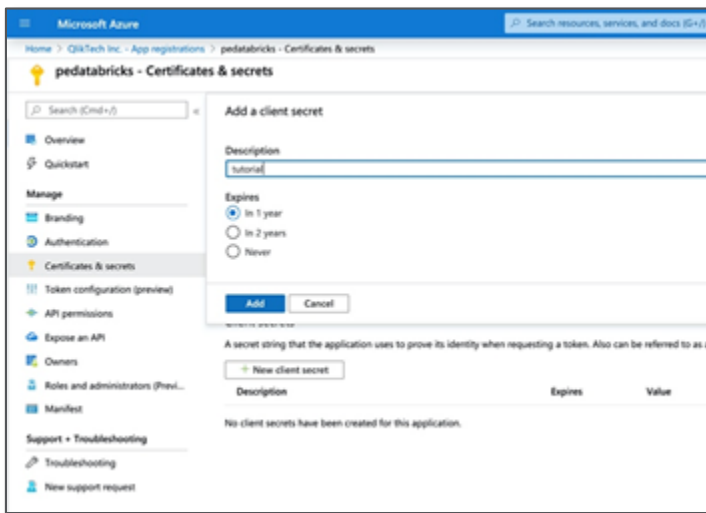
From there, check “user_impersonation” under “Delegated permissions” and then press “Add permissions” button at the bottom of the page



Now we need to create a “secret” (essentially a password) for this API. Click on “Certificates & secrets” on the left side of the screen and click on “+ New client secret”.



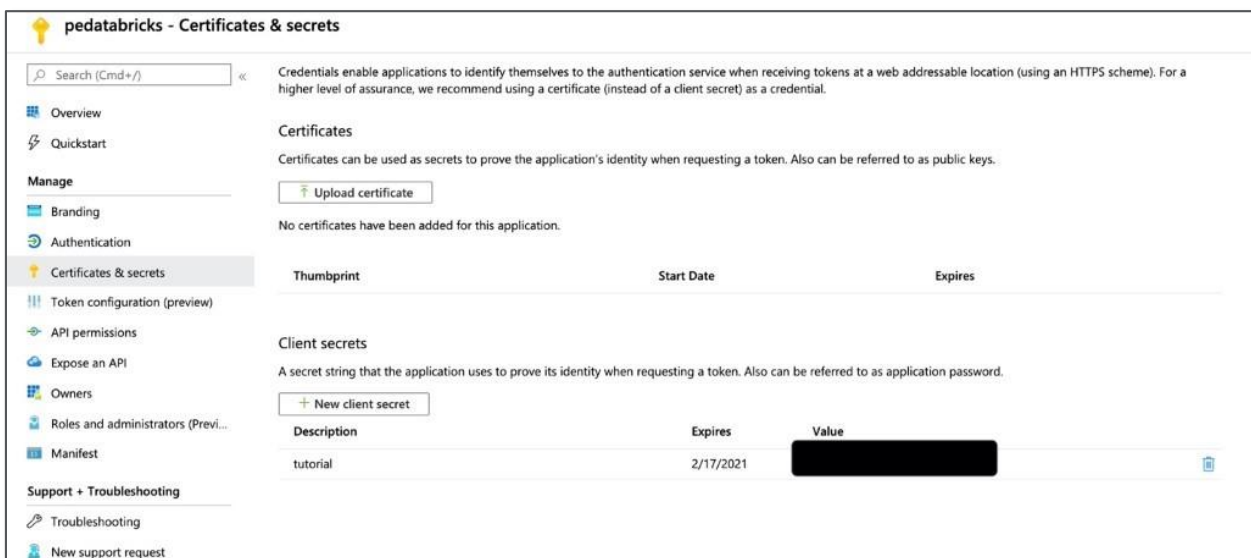
Enter a description, choose an expiration, and press “Add.”



The screenshot shows the 'Add a client secret' dialog in the Azure portal. The dialog has a 'Description' field with the value 'tutorial', an 'Expires' section with radio buttons for 'In 1 year' (selected), 'In 2 years', and 'Never', and 'Add' and 'Cancel' buttons. Below the dialog, a table shows no client secrets have been created for this application.

Description	Expires	Value
No client secrets have been created for this application.		

IMPORTANT NOTE: Make Note of value of “Secret” you generated. You must save the value of the secret you created before you leave this page. You will not be able to retrieve it later.



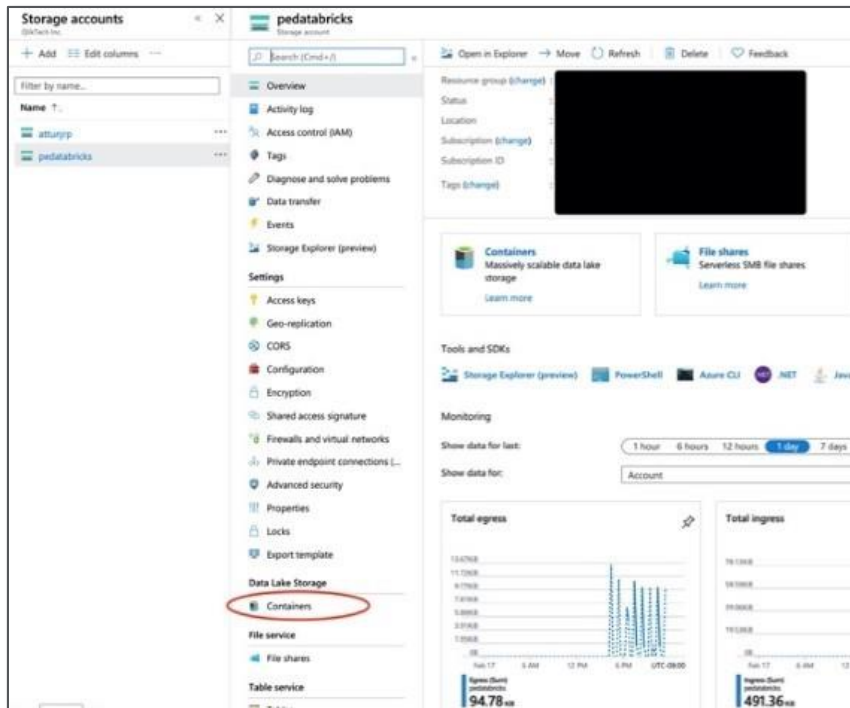
The screenshot shows the 'Certificates & secrets' page in the Azure portal. It includes sections for 'Certificates' and 'Client secrets'. The 'Client secrets' section shows a table with one secret.

Description	Expires	Value
tutorial	2/17/2021	[REDACTED]

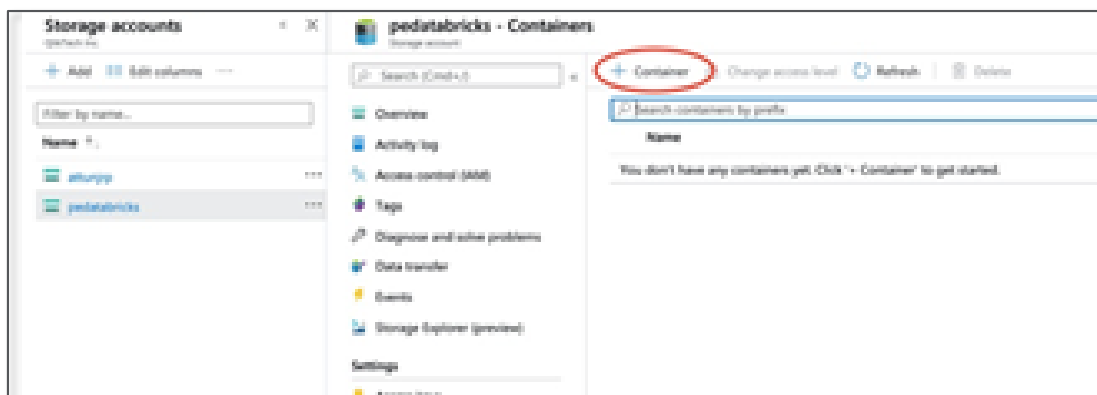
3. Finalizing Configuration of the ADLS-2 Storage Account

Now that we have created the Azure Active Directory App Registration, we need to return to the ADLS-2 storage account and finalize the configuration, so it is ready for use by Qlik Replicate. Return to the Azure Portal home page and then drill in to get back to the storage account we created.

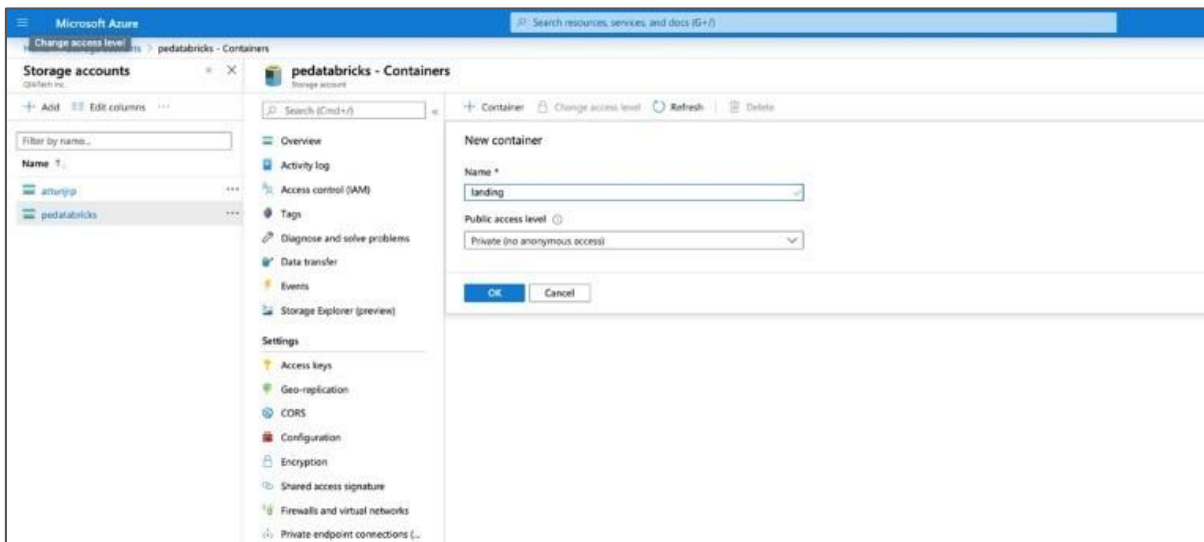
First, we need to create a “Container” for us to use for the Qlik Replicate. Select “Containers” on the left side of the storage account screen.



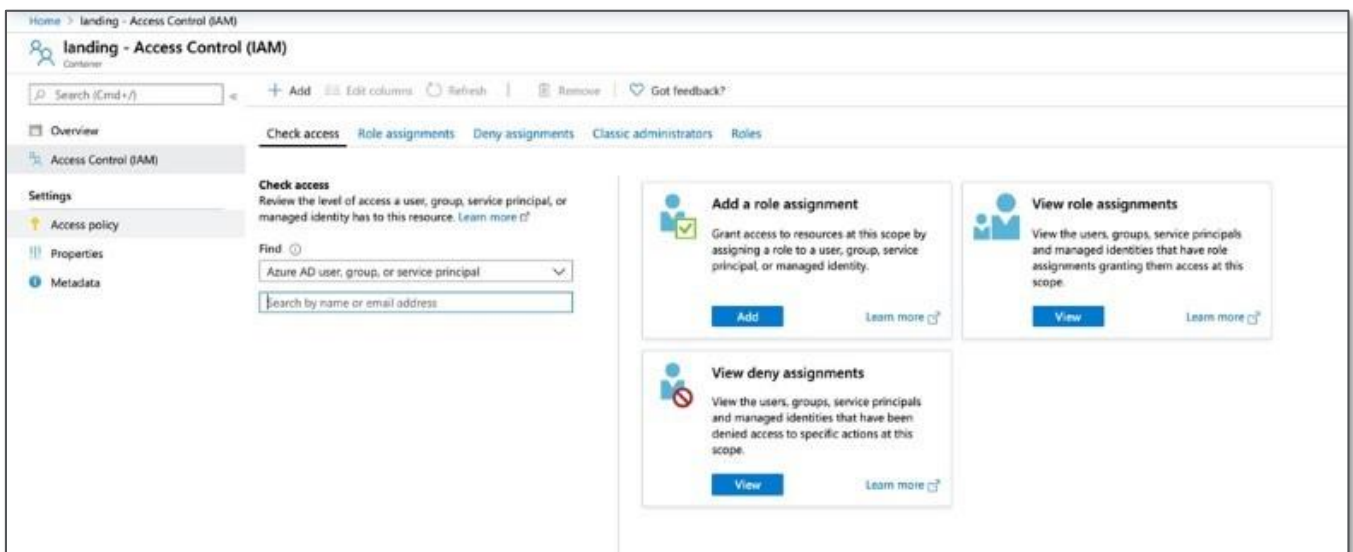
Inside your Storage account “+ Container” at the top of the page.



Enter a name for the Container and press “OK” to save it.



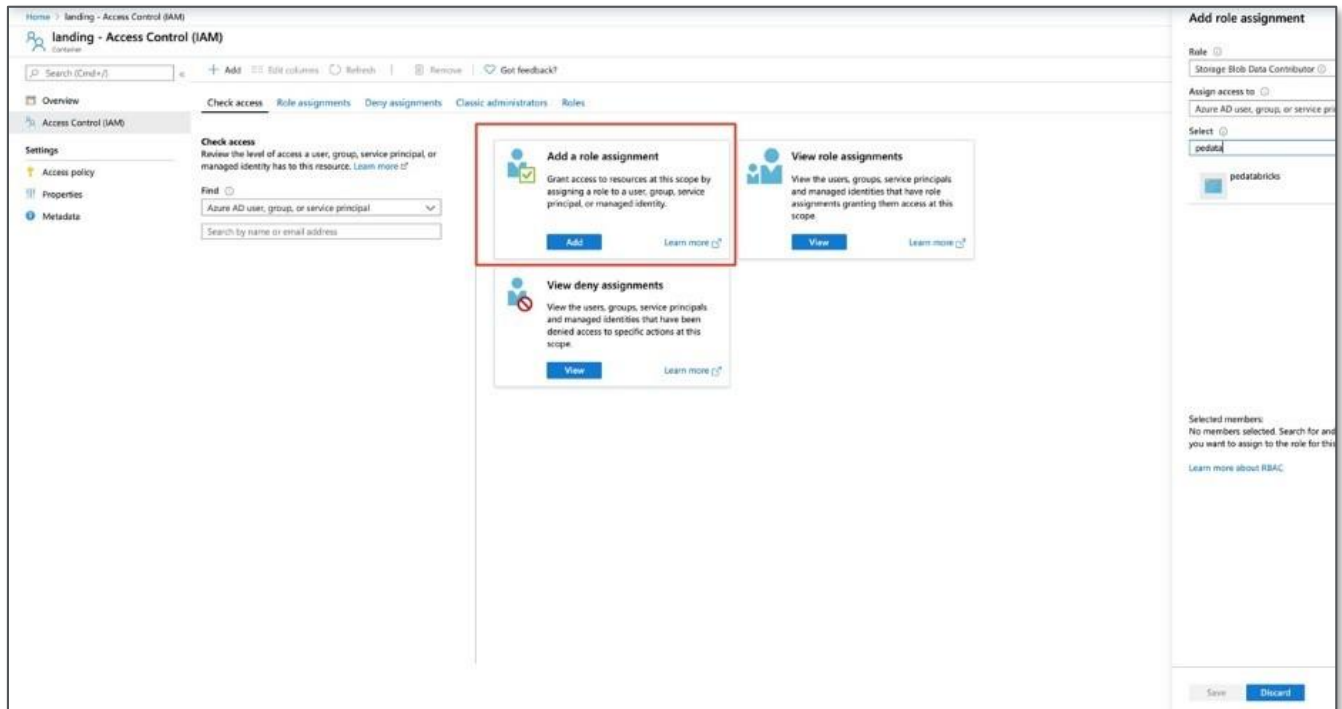
As a final step, we need to grant access to the storage account to the “App” that we created / registered previously. Click on “Access control (IAM)” on the left side of the page. Then click on “Add a role assignment” on right.



Once in the Screen on the right side of the screen select:

- Role: Storage Blob Data Contributor
- Assign access to: ` Azure AD user, group, or service principal
- Select: enter the name of the App you registered previously. (Figure A.2.4)

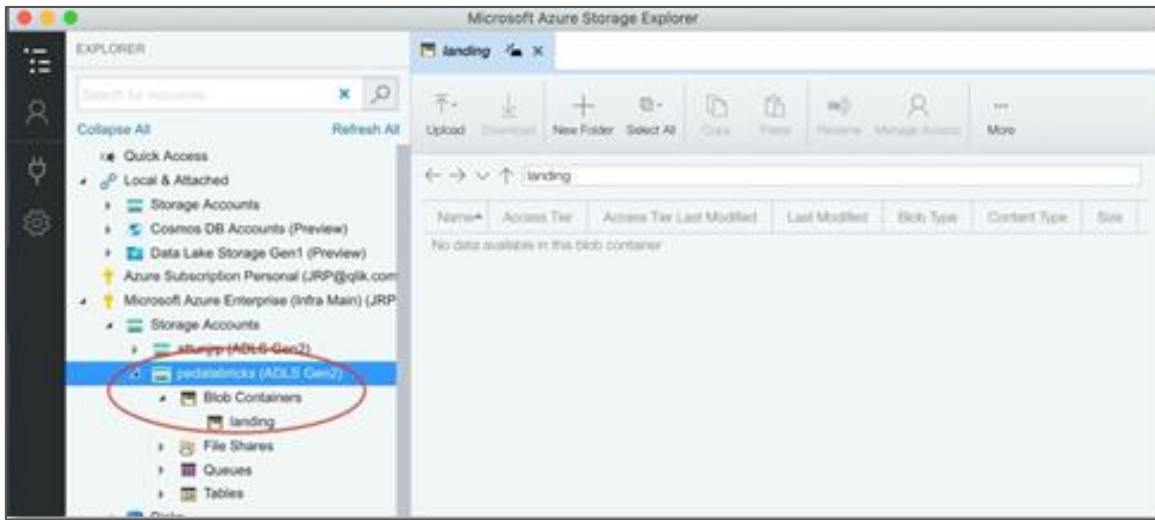
and press “Save” button at the bottom of the page.



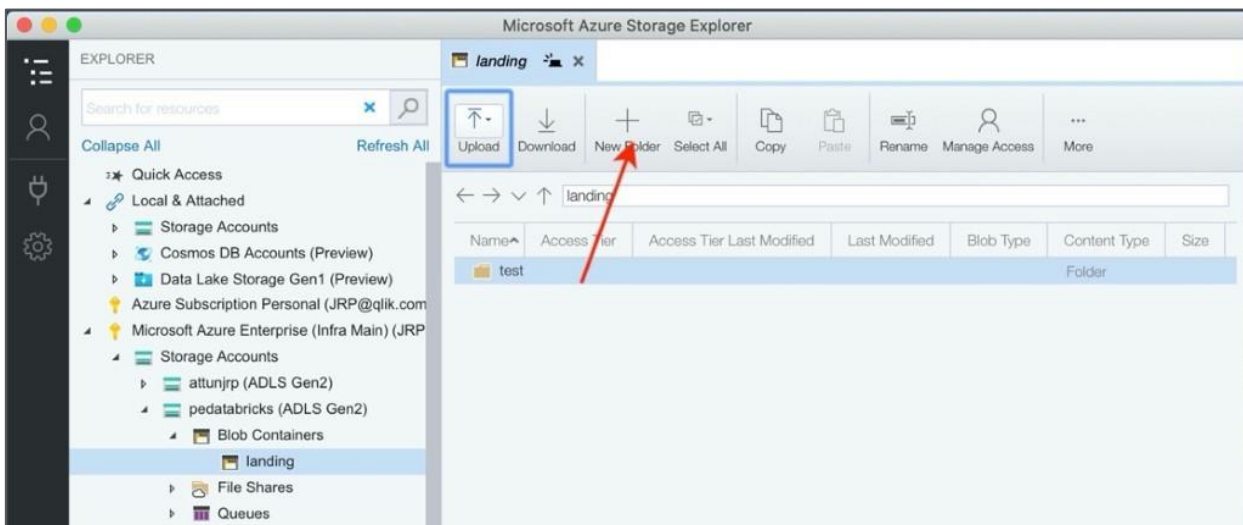
4. Create a Target Directory in the File System

To complete this part of the guide, you will need to create a target directory where Qlik Replicate will deliver data for the Databricks external tables.

To get started, go to your ADLS-2 storage account and click on Storage Explorer (preview). This will download Storage Explorer application on your computer. Select Storage Accounts and select Blob Container created



From there click on Blob Container Created and Add New Folders for Qlik Replicate (In this example we used “test” as folder name)



Read “Configuring Qlik Replicate with Azure Databricks” to configure Azure Databricks and use this storage to ingest data delivered by Qlik Replicate.



About Qlik

Qlik's vision is a data-literate world, one where everyone can use data to improve decision-making and solve their most challenging problems. Only Qlik offers end-to-end, real-time data integration and analytics solutions that help organizations access and transform all their data into value. Qlik helps companies lead with data to see more deeply into customer behavior, reinvent business processes, discover new revenue streams, and balance risk and reward. Qlik does business in more than 100 countries and serves over 50,000 customers around the world.

[qlik.com](https://www.qlik.com)

© 2020 QlikTech International AB. All rights reserved. All company and/or product names may be trade names, trademarks and/or registered trademarks of the respective owners with which they are associated.