

[Open in app](#)[Get started](#)

You have **2** free member-only stories left this month.

[Sign up for Medium and get an extra one](#)



K O R K R I D (KYLE)

Follow

Jan 14, 2020 · 6 min read · · Listen

Save



Mount an ADLS Gen 2 to Databricks File System Using a Service Principal and OAuth 2.0 (Ep. 5)

Demonstrate how to mount an Azure Data Lake Storage Gen2 (ADLS Gen 2) account to Databricks File System (DBFS), authenticating using a service principal and OAuth 2.0.

Disclaimer: The views and opinions expressed in this article are those of the author's and do not necessarily reflect the official policy or position of current or previous employer, organization, committee, other group or individual. Analysis performed within this article is based on limited dated open source information. Assumptions made within the analysis are not reflective of the position of any previous or current employer.





Open in app

Get started

TECHNOLOGY | AZURE | DATABRICKS

Mount an ADLS Gen 2 to Databricks File System Using a Service Principal and OAuth 2.0

Presented by Korkrid Akepanidtaworn

Prerequisites

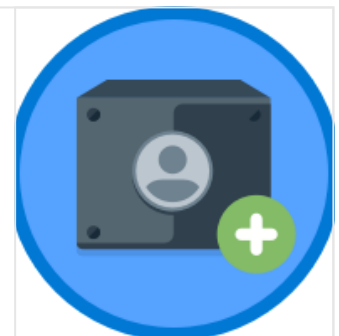
Before moving on, you should understand:

- How to provision Azure Data Lake Storage Gen 2 (ADLS Gen 2) and Azure Databricks.
- How to create container (file system) in Data Lake Gen2 File systems (Use latest **Azure Storage Explorer**.)

Create an Azure Storage account - Learn

Create an Azure Storage account with the correct options for your business needs.

docs.microsoft.com



Azure Databricks documentation

Learn Azure Databricks, an Apache Spark-based analytics platform with one-click setup, streamlined workflows, and an...

docs.microsoft.com





Open in app

Get started

manage Azure blobs, files, queues, and tables, as well as...

azure.microsoft.com

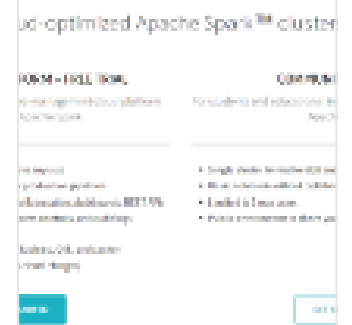


Previous Episodes

Databricks Community Edition: Apache Sparks for All (Ep. 1)

Learn to use cloud-based Spark platform on Databricks community edition, upload data to FileStore, and run your first...

medium.com



Why Databricks on Azure? A Customer's Perspective (Ep. 2)

Azure Databricks x Learn how Shell, Nielsen, and Magneti Marelli leverage Databricks Unified Analytics Platform

medium.com



How to Create and Deploy a Databricks Workspace Using the Azure Portal (Ep. 3)

The first step to using Azure Databricks is to create and deploy a Databricks workspace. You can do this in the Azure...

medium.com



Bye Pandas, Meet Koalas: Pandas APIs on Apache Spark (Ep. 4)

Keynote summary from the official announcement of Koalas by Reynold Xin at Spark + AI Summit


medium.com



[Open in app](#)[Get started](#)

connection is required. You need to also ensure that you have the rights to register apps. If the app registrations is configured to “yes” in your Azure Active Directory, then non-admin users may register custom-developed applications for use within this directory.

[Home](#) > [Microsoft](#) > [Users - User settings](#)

 **Users - User settings**
Microsoft - Azure Active Directory

All users

Deleted users

Password reset

User settings

Diagnose and solve problems

Activity

Sign-ins

Audit logs

Bulk operation results (Preview)

Troubleshooting + Support

New support request

« Save Discard

Enterprise applications

[Manage how end users launch an application](#)

App registrations

Users can register applications ⓘ

Yes

No

Administration portal

Restrict access to Azure AD administration portal ⓘ

Yes

No

LinkedIn account connections

If this option is set to yes, then non-admin users may register custom-developed applications for use within this directory.
If this option is set to no, then only users with an administrator role may register these types of applications.





Open in app

Get started



Microsoft App registrations
Azure Active Directory



Search (Ctrl+/)



Overview



Getting started



Diagnose and solve problems

Manage



Users



Groups



Organizational relationships



Roles and administrators



Enterprise applications



Devices



App registrations



Identity Governance



Application proxy



Licenses



Azure AD Connect



Custom domain names

Home > Microsoft - App registrations



Microsoft - App registrations
Azure Active Directory



Search (Ctrl+/)

Overview
Getting started
Diagnose and solve problems

Manage

Users

+ New registration | Endpoints | Troubleshooting | App registrations (Legacy) | Got feedback?

Well, New registration and improved App registrations (now Generally Available). See what's new and learn more on how it's changed. →

If you are building an application for external users that will be distributed by Microsoft, you must register as a first party application to meet all security, privacy, and compliance policies. [Read our decision guide](#)

All applications | Owned applications

Start typing a name or Application ID to filter these results



[Open in app](#)[Get started](#)

registrations” to proceed.). Click “New Registration.”

2. Fill in the required information for the application, e.g., application name, application type (web app/API by default.)

3. Copy “Client ID”, “Object ID”, and “Directory ID”

Home > Microsoft - App registrations > Register an application

Register an application

Warning If you are building an application for external users that will be distributed by Microsoft, you must register as a first party application to meet all security, privacy, and compliance policies. [Read our decision guide](#)

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 (Microsoft 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)

[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

Understanding different account types

Accounts in this organizational directory only (Microsoft only - Single tenant)
All user and guest accounts in your directory can use your application or API.
Use this option if your target audience is internal to your organization.

Accounts in any organizational directory (Any Azure AD directory - Multitenant)
All users with a work or school account from Microsoft can use your application or API. This includes schools and businesses that use Office 365.
Use this option if your target audience is business or educational customers and to enable multitenancy.

Accounts in any organizational directory (Any Azure AD directory - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)
All users with a work or school, or personal Microsoft account can use your application or API. It includes schools and businesses that use Office 365 as well as personal accounts that are used to sign in to services like Xbox and Skype.
Use this option to target the widest set of Microsoft identities and to enable multitenancy.

Due to temporary differences in functionality, you may see errors if you attempt to switch between supported audiences after registering the application.

Home > Microsoft - App registrations > appformvpdatabricks

appformvpdatabricks

Search (Ctrl+J)

Overview
Quickstart
Manage
Branding
Authentication
Certificates & secrets
Token configuration (preview)
API permissions
Expose an API
Owners
Roles and administrators (Preview)
Manifest
Support + Troubleshooting
Troubleshooting
New support request

Delete Endpoints

Got a second? We would love your feedback on Microsoft identity platform (previously Azure AD for developer). →

Display name	: appformvpdatabricks	Supported account types	: My organization only
Application (client) ID	: Copy this value (Client ID)	Redirect URIs	: 1 web, 0 public client
Directory (tenant) ID	: Copy this value (Directory ID)	Application ID URI	: Add an Application ID URI
Object ID	: Copy this value (Object ID)	Managed application in ...	: appformvpdatabricks

Welcome to the new and improved App registrations. Looking to learn how it's changed from App registrations (Legacy)? [Learn more](#)

Call APIs

Build more powerful apps with rich user and business data from Microsoft services and your own company's data sources.

[View API permissions](#)

Sign in users in 5 minutes

Documentation

- Microsoft identity platform
- Authentication scenarios
- Authentication libraries
- Code samples
- Microsoft Graph
- Glossary
- Help and Support

You also need to generate app secrets (authentication keys) to access the app you just created.



Open in app

Get started

Home > Microsoft - App registrations > appformvdpatabricks - Certificates & secrets

appformvdpatabricks - Certificates & secrets

Search (Ctrl+/)

- Overview
- Quickstart
- Manage
 - Branding
 - Authentication
 - Certificates & secrets
 - Token configuration (preview)
 - API permissions
 - Expose an API
 - Owners
 - Roles and administrators (Previ...
 - Manifest
- Support + Troubleshooting
 - Troubleshooting

Add a client secret

Description
For my test connection with ADLS Gen 2 (MVP Databricks)

Expires
☒ In 1 year
☐ In 2 years
☐ Never

Add Cancel

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

+ New client secret

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

Copy the new client secret value. You will not be able to retrieve it after you perform another operation or leave this blade.

. . .

Let's next store the key retrieved from the registered application in the Azure service called **Azure Key Vault**, "a cloud service that provides a secure store for secrets. You can securely store keys, passwords, certificates, and other secrets. Azure key vaults may be created and managed through the Azure portal." Please follow the steps under "**Add a secret to Key Vault**" section.

Azure Quickstart - Set and retrieve a secret from Key Vault using Azure portal

Azure Key Vault is a cloud service that provides a secure store for secrets. You can securely store keys, passwords...

docs.microsoft.com





Open in app

Get started

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Events (preview)
- Settings
 - Keys
 - Secrets
 - Certificates
 - Access policies
 - Firewalls and virtual networks
 - Properties
 - Locks
 - Export template
- Monitoring

Name	Type	Status	Expiration
kvdatabricks001		✓ Enabled	

Upload options

Manual

Name * ⓘ

kvdatabricks001

Value * ⓘ

Application Secret (a.k.a. Client Secret)

Content type (optional)

Set activation date? ⓘ

☐

Set expiration date? ⓘ

☐

Enabled?

Yes

No





Open in app

Get started

ew	Name	nextgenkv
y log	SKU (Pricing tier)	Standard
: control (IAM)	Location	southeastasia
use and solve problems	DNS Name	Copy this Value
(preview)	Resource ID	Copy this Value
	Subscription ID	Copy this Value
	Subscription Name	Not Revealed
s	Directory ID	Not Revealed
ates	Directory Name	Not Revealed
: policies		
ils and virtual networks		
ties		

• • •

Managing secrets begins with creating a secret scope. A secret scope is collection of secrets identified by a name.

Secret Scopes - Azure Databricks

Managing secrets begins with creating a secret scope. A secret scope is collection of secrets identified by a name...

docs.microsoft.com



You will need to create an Azure Key Vault-backed secret scope.





Open in app

Get started

Create Secret Scope |

A store for secrets that is identified by a name and backed by a specific store type. [Learn more](#)

Scope Name

Manage Principal

Azure Key Vault

DNS Name

[DNS Name in Azure Key Vault]

Resource ID

[Resource ID in Azure Key Vault]

Microsoft Azure

PORTAL koakepan@microsoft.com

HomePage / Create Secret Scope

Create Secret Scope |

A store for secrets that is identified by a name and backed by a specific store type. [Learn more](#)

Scope Name

Manage Principal

Azure Key Vault

DNS Name

[DNS Name in Azure Key Vault]

Resource ID

[Resource ID in Azure Key Vault]

The secret scope named dbscopekeyvault001 has been added.

Manage secrets in this scope in Azure KeyVault with manage principal = creator

• • •

By now, you should have the following information:

- Client ID (a.k.a. Application ID)
- Client Secret (a.k.a. Application Secret)



[Open in app](#)[Get started](#)

- Key Name for Service Credentials (from Azure Key vault)
- File System Name
- Storage Account Name
- Mount Name

I have hosted my sample script to mount ADLS Gen 2 to DBFS on GitHub.

```
1 #####
2 # Set the configurations. Here's what you need:
3 ## 1.) Client ID (a.k.a Application ID)
4 ## 2.) Client Secret (a.k.a. Application Secret)
5 ## 3.) Directory ID
6 ## 4.) File System Name
7 ## 5.) Storage Account Name
8 ## 6.) Mount Name
9 #####
10 configs = {"fs.azure.account.auth.type": "OAuth",
11            "fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.Client
12            "fs.azure.account.oauth2.client.id": "<client-id>",
13            "fs.azure.account.oauth2.client.secret": dbutils.secrets.get(scope = "<scope-name>",
14            "fs.azure.account.oauth2.client.endpoint": "https://login.microsoftonline.com/<direc
15
16 #####
17 # Optionally, you can add <directory-name> to the source URI of your mount point.
18 #####
19 dbutils.fs.mount(
20     source = "abfss://<file-system-name>@<storage-account-name>.dfs.core.windows.net/",
21     mount_point = "/mnt/<mount-name>",
22     extra_configs = configs)
```

where:

- `<mount-name>` is a DBFS path that represents where the Data Lake Store or a folder inside it (specified in `source`) will be mounted in DBFS.
- `dbutils.secrets.get(scope = "<scope-name>", key = "<key-name-for-service-"`



[Open in app](#)[Get started](#)

- If you do not want to use the `dbutils.secrets` API, you could simply paste the client secret in `fs.azure.account.oauth2.client.secret` parameter.

```
1 dbutils.fs.ls('/mnt/adltestfs/')

```

```
Out[13]: [FileInfo(path='dbfs:/mnt/adltestfs/01_Keras_R_Basics.R', name='01_Keras_R_Basics.R', size=6094),
FileInfo(path='dbfs:/mnt/adltestfs/02_Keras_PreTrained_ResNet50_Predict_Elephant.R', name='02_Keras_PreTrained_ResNet50_Predict_Elephant.R', size=44313),
FileInfo(path='dbfs:/mnt/adltestfs/03_Keras_PreTrained_VCG16_FeatureExtraction.R', name='03_Keras_PreTrained_VCG16_FeatureExtraction.R', size=1474),
FileInfo(path='dbfs:/mnt/adltestfs/sample-elephant-image.jpg', name='sample-elephant-image.jpg', size=169034)]

```

• • •

Important Glossary

- **Mounting** — To mount blob storage to DBFS is treated as if they were on the local file system. Importantly, all users have **read and write access** to the objects in blob storage mounted to DBFS. For information on how to mount and unmount AWS S3 buckets, see [Mount S3 Buckets with DBFS](#). For information on encrypting data when writing to S3 through DBFS, see [Encrypt data in S3 buckets](#). For information on how to mount and unmount Azure Blob Storage containers and Azure Data Lake Storage accounts, see [Mount Azure Blob Storage containers to DBFS](#), [Mount Azure Data Lake Storage Gen1 resource using a service principal and OAuth 2.0](#), and [Mount an Azure Data Lake Storage Gen2 account using a service principal and OAuth 2.0](#).

```
mount(source: String, mountPoint: String, encryptionType: String =
"", owner: String = null, extraConfigs: Map = Map.empty[String,
String]): boolean -> Mounts the given source directory into DBFS at
the given mount point
```



18



1

```
mounts: Seq -> Displays information about what is mounted within
DBFS
```

```
refreshMounts: boolean -> Forces all machines in this cluster to
refresh their mount cache, ensuring they receive the most recent
information
```

```
unmount(mountPoint: String): boolean -> Deletes a DBFS mount point
```

• • •



[Open in app](#)[Get started](#)

workspace and available on...

docs.databricks.com

Azure Data Lake Storage Gen2

Azure Data Lake Storage Gen2 (also known as ADLS Gen2) is a next-generation data lake solution for big data analytics...

docs.databricks.com

exercise01-blob - Databricks

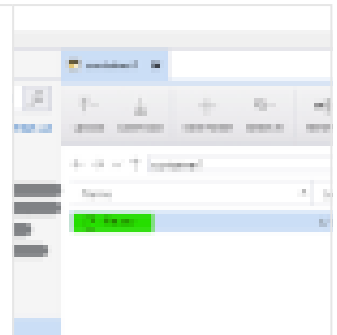
Edit description

tsmatz.github.io

Access to Azure Data Lake Storage Gen 2 from Databricks Part 1: Quick & Dirty

Support to existing documentation

medium.com



[About](#) [Help](#) [Terms](#) [Privacy](#)

Get the Medium app





[Open in app](#)

Get started

