

Setup Guide

Onehouse Catalog Sync for Databricks Unity Catalog

Jan 2025

Contents

Introduction	2
Prerequisites	2
Create a Databricks workspace	3
Create a Unity Catalog	5
Create an S3 Bucket for the metastore	5
Create an IAM role to access the storage location	5
Create a Unity Catalog metastore	8
Configure the Unity Catalog for CORS	11
Connect the Workspace to the Onehouse data storage location	12
Create and attach IAM policy for Onehouse customer storage access	12
Create a Databricks Storage Credential	13
Create a Databricks External Location	15
Gather the external catalog information in Databricks	17
Create the Metadata Catalog Sync in Onehouse	19
Create the Databricks metadata catalog in Onehouse	19
Attach the Onehouse Databricks catalog to one or more Stream Captures.	20

Introduction

This guide provides a step-by-step recipe for setting up a Onehouse metadata catalog sync to <u>Databrick's Unity Catalog</u>.

This provides a unified governance solution for managing and securing data assets across cloud environments. With <u>Onehouse's OneTable</u> support, users can ingest, and store data in Delta Lake format, and these tables automatically sync to Unity Catalog. This ensures seamless management of metadata, permissions, and lineage in a centralized, secure catalog.

The setup for Databricks Unity Catalog sync setup in Onehouse is relatively easy – the majority of this guide describes how to set up the AWS IAM roles and policies, and Databricks external catalog location.

Prerequisites

You will need the following:

- A Onehouse account with Admin role, and a provisioned Onehouse project.
- At least one Onehouse table that has Delta Lake metadata. Databricks requires Delta Lake table metadata to work with Onehouse tables. To meet this requirement, be sure you have created a OneTable catalog with Delta Lake enabled, and have attached it to the relevant Stream Capture.
- Databricks 'Account admin' role on your account in order to be able to create a new Databricks Workspace and Catalog.
- Amazon AWS account administrator privileges to be able to create a VPC and IAM roles.

Create a Databricks workspace

Refer to the Databricks documentation for workspace creation here:

https://docs.databricks.com/en/admin/workspace/quick-start.html#create-a-workspace-usingthe-aws-quick-start-recommended

□ If you don't already have a Databricks workspace or wish to start with a new one, create a new Databricks workspace using the AWS quickstart method.

≡ 🤪 Account									0	C
器 Workspaces	Workspaces							1	Create worksp	pace
🗠 Usage	Filter workspaces		Q							
Se User management	Name	Status	Pricing tier	Region	Storage	Credential name	Created	Metastore		
Previews	Vishal3	Running	Create Workspac	•		×	09/06/2024	us-west-2	Open 🗗	:
Settings	Vishal2	Failed	Create Workspace				09/06/2024	-		:
	Vishal	Failed	A workspace is an environme	ent to access all o	of your Databricks data asset	s and computational	09/05/2024	-		:
	Andy	Running	resources. Learn more			06/05/2024	us-west-2	Open 🖪	:	
	PM Team	Running	How would you like to create	your workspace		07/16/2023	metastore-us-east-1	Open 🗗	:	
	Onehouse Benchm	Running					10/17/2022	-	Open 📑	:
	Ethan performance	Running	Quickstart (Recomm Use Quickstart to cre workspace in just a f	ended) ate a w clicks)	Manual Typically for adva AWS experience. create your own s credentials config	nced users with You will need to torage and urations Cancel Next	05/19/2022	•	Open t	

Note that if you intend to create a new Unity catalog, this will need to be done in an AWS region that has not already been used to host a Unity catalog.

🗏 🥪 Account						0	C
器 Workspaces	Workspaces					Create worksp	pace
💑 Catalog	Filter workspaces		Let's set up your workspace ×				
Se User management	Name	Status	We're going to send you to your AWS Console to configure your account	Created	Metastore		
Previews	Vishal3	Running	Once you size in well are populate a CloudEermation template that creater an IAM role and \$2	09/06/2024	us-west-2	Open 🗗	:
Settings	Vishal2	Failed	bucket for you, then deploys your workspace.	09/06/2024	-		:
	Vishal	Failed	If you encounter any errors during the process, reach out at onboarding-help.databricks.com for	09/05/2024	-		÷
	Andy	Running	troubleshooting guidance.	06/05/2024	us-west-2	Open 🗗	:
	PM Team	Running		07/16/2023	metastore-us-east-1	Open 🖸	:
	Onehouse Benchm	Running	workspace Name	10/17/2022		Open 🖪	:
	Ethan performance	Running	Acme	05/19/2022	-	Open 🗗	:
			Human readable name for your workspace AWS Region of the Databricks workspace us-east-2 AWS Region where the workspace will be created Cancel Start Quickstart				

In AWS, you'll need to acknowledge that roles will be created in your account. (The roles all start with 'databricks-workspace-stack-'.)

Public extensions	Capabilities
Activated extensions	
Publisher	() The following resource(s) require capabilities: [AWS::IAM::Role]
	This template contains Identity and Access Management (IAM) resources. Check that you want to create each of these resources and that they have the minimum required permissions. In addition, they have custom names. Check that the custom names are unique within your AWS account. Learn more [2]
Spotlight	I acknowledge that AWS CloudFormation might create IAM resources with custom names.
Feedback	h h
	Cancel Create change set Create stack

You'll see several CloudFormation events scroll by. When the stack shows 'COMPLETE', you can continue.

aws III Q Search	[Option-	-s] D	🗘 🕜 🕸 United States (Ohio	o) ▼ cameron@onehouse.ai @ 6546-5423-5321 ▼
<u>CloudFormation</u> > <u>Stacks</u> >	databricks-workspace-stack-1b7c9		N	0 0
CloudFormation <	🗆 Stacks (1) 🕜	databricks-workspace-stack-1b	97c9	© >
Stacks Stack details Drifts StackSets	Filter status Q Filter by stack name C View nested	< Stack info Events - up	Delete Update dated Resources Outputs	Stack actions Create stack Parameters Template
Exports	< 1 > Stacks databricks-workspace-stack-1b7c9	Resources (14)		⊘ < 1 > ⊗
laC generator	 2025-01-10 12:44:46 UTC-0800 CREATE_COMPLETE 	Logical ID cataloglamRole	AWS::IAM::Role	7 Status ▼ ② CREATE_COMPLETE
Hooks overview New		II catalogIamRolePolicy	AWS::IAM::Policy	CREATE_COMPLETE
		CopyZips	Custom::CopyZips	O CREATE_COMPLETE

Go back to Databricks, and refresh the browser page to see the new workspace in the list.

Create a Unity Catalog

You will also need to enable Unity Catalog metastore in the workspace.

Refer to the Databricks documentation 'Managing Unity Catalog' at: <u>https://docs.databricks.com/en/data-governance/unity-catalog/get-started.html</u>

Also, refer 'Create a Unity Catalog' at: <u>https://docs.databricks.com/en/data-governance/unity-catalog/create-metastore.html</u>

Create an S3 Bucket for the metastore

□ Name the S3 bucket something unique, but identifying it as a metastore.

Amazon 53 bucket Create bucket and amazon 53 bucket Create bucket and amazon 54 bucket So can countianers for data stored in 53. Cencer countiguous exast 2. bucket type Info Counting to for some cases and access patterns. General purpose buckets are the original 53 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 53 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 53 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 53 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 53 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 53 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 53 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 50 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 50 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 50 bucket type. Tripe Counting to for some cases and access patterns. General purpose buckets are the original 50 bucket transmite original 50 b	S Q Search	[Option+S]	D A	\$ Ø \$	United States (Ohio) 🔻	cameron@onehouse.ai @ 6546-5	423-53
Beach Bucket info Reserved and stored in 53. Beach configuration With Region US fast (bin) us-sat-2: Boach surport Oncomparing on Reserved and stores a starters. General purpore buckets are the original 53 bucket type. Thrip Image: Store S	Amazon S3 > Buckets > Create bucket					0	₽
General configuration WX Regin UX5 East (Ohio) us-east-2 UX5 East (Ohio) UX5 East (Ohio	reate bucket Info ckets are containers for data stored in S3.						
WS Region JS East (Chiol) us-east-2 Succet type info Info Consent purpose Necommended for low-latency use cases and access patterns. General purpose buckets are the original 53 bucket type. Thy allows ank of storage classes that redundantly store objects across multiple Availability Zone. Info Intercommended for low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses number of low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses number of low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses number of low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses number of low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses number of low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses number of low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses number of low-latency use cases. These buckets use only the 53 Express One Zone storage class, which arouses hubite the global namespace and follow the bucket naming rules. See rules for bucket naming in the Storage class that redundantly in the global namespace and follow the bucket naming rules. See rules for bucket naming in the Storage classes that redundant proves objects. Choose bucket Choose bucket Choose bucket Choose bucket for form-latency use cases to this bucket form other AWS accounts and the use of access control lists (AcLs). Object ownership determines who can specify access to objects. Choose bucket in this bucket are owned by this account. Access to this bucket and its objects is specified using original ACLs. Oblect in this bucket are owned by other AWS accounts. Ac	General configuration						
US East (Ohio) us-east-2 Backet type Info	AWS Region						
Backet type Info Original Synchronic Status asses and access patterns. General purpose buckets are the original S3 bucket type. They allow an ink of storage classes that redundantly store objects across multiple Availability Zones. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Status processing of data within a single Availability Zone. Original Synchronic Synchro	US East (Ohio) us-east-2						
General purpose Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones. Directory Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone. Bucket name info databricks-unity-catalog-acme sucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming [2] Copy settings from existing bucket - optional Daty the bucket settings in the following configuration are copied. Choose bucket format: s3://bucket/prefix Object Ownership Info Control ownership of objects written to this bucket from other AW5 accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects. O ACLs disabled (recommended) All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. O ACLs disabled (recommended) All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. O Action of the bucket are owned by this account. Access to this bucket and its objects is specified using only policies. O Action of the bucket are owned by this account. Access to this bucket and its objects is specified using only policies. O Action of the bucket are owned by this account. Access to this bucket and its objects is specified using only policies. O Action of the bucket are owned by other AWS accounts. Access to this bucket and its objects is appecified using only policies.	Bucket type Info						
Bucket name Info databricks-unity-catalog-acme Under the global namespace and follow the bucket naming rules. See rules for bucket naming C Copy settings from existing bucket - optional Only the bucket settings in the following configuration are copied. Choose bucket Permat: s3://bucket/perfix C C Copy Setting from C Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects. C ACLs disabled (recommended) Al objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only polices. C ACLs disabled (recommended) Al objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only polices. C ACLS disabled (recommended) Al objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only polices. C ACLS disabled (recommended) Al objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only polices. C ACLS disabled (recommended) Al objects in this bucket are owned by other AWS accounts. Access to this bucket and its objects can be specified Copy access to access to this bucket and its objects is appendixed on the access to this bucket and its objects can be specified Copy access to access to this bucket and its objects can be specified Copy access to access to this bucket and its objects can be specified Copy access Copy access to access to this bucket and its objects can be specified Copy access to copy access Copy access to copy access to this bucket and its objects can be specified Copy access Copy access to this bucket and its objects can be specified Copy access Copy ac	• General purpose Recommended for most use cases and access patterns. G allow a mix of storage classes that redundantly store obj	eneral purpose buckets are the original S3 bucket type. They ects across multiple Availability Zones.	O Directory Recommended for low-lu provides faster processin	atency use cases. Thes ng of data within a sing	e buckets use only the S3 Exp gle Availability Zone.	ress One Zone storage class, which	
databricks-unity-catalog-acme bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming [] Copy settings from existing bucket - optional Inty the bucket settings in the following configuration are copied. Choose bucket ormat: s3://bucket/prefix	Bucket name Info		\bigcirc				
Sector and must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming [] Copy settings from existing bucket - optional Only the bucket settings in the following configuration are copied. Choose bucket ormat: s3://bucket/prefix	databricks-unity-catalog-acme		Ŭ,				
Choose bucket	Bucket name must be unique within the global namespace and	follow the bucket naming rules. See rules for bucket naming	3				
Choose bucket settings in the following configuration are copied. Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket Choose bucket C	Copy settings from existing bucket - optional						
Choose bucket Tormat: s3://bucket/prefix	Only the bucket settings in the following configuration are copi	ed.					
Format: s3://bucket/prefix Object Ownership info Control ownership of objects written to this bucket from other AW5 accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects. Image: AcL disabled (recommended) All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. Image: Access to this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. Image: Access to this bucket are owned by other AW5 accounts. Access to this bucket and its objects can be specified using ACLs.	Choose bucket						
Object Ownership Info Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects. ACLs disabled (recommended) All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. Chain of the second of	Format: s3://bucket/prefix						
Object Ownership Info Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects. • ACLs disabled (recommended) All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. • Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using only using ACLs.							
Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects. All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. All objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using only using ACLs.	Object Ownership Info						
ACLs disabled (recommended) All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies. O ACLs enabled Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.	Control ownership of objects written to this bucket from	n other AWS accounts and the use of access control lis	sts (ACLs). Object ownership det	termines who can sp	pecify access to objects.		
8-1025 kararan Mikelanda - Baharan Tamar Kardin and	• ACLs disabled (recommended) All objects in this bucket are owned by this account. Acce policies.	ss to this bucket and its objects is specified using only	O ACLs enabled Objects in this bucket ca using ACLs.	in be owned by other A	WS accounts. Access to this b	ucket and its objects can be specified	
© 2026 Among Web Consists in a site officient States - Delana, Consist of the set							

Create an IAM role to access the storage location

□ In the Databricks account console, click on your username in the upper right corner, and copy the **Account ID** value.

- □ In AWS IAM, create a new role.
 - □ Type: Custom Trust policy
 - □ Paste in the following policy, substituting the Databricks Account #.

```
"Version": "2012-10-17",
       "Statement": [
         {
           "Effect": "Allow",
           "Principal": {
             "AWS": [
     arn:aws:iam::414351767826:role/unity-catalog-prod-UCMasterRole-14S5ZJ
    VKOTYTL"
           },
           "Action": "sts:AssumeRole",
           "Condition": {
             "StringEquals": {
                "sts:ExternalId": "<DATABRICKS_ACCT_ID>"
             }
         }
       1
Click 'Next', and 'Next' and name the role something similar to
   databricks-unity-catalog-access
Click 'Create Role'
Click into the new role, and copy the ARN for the role you just created:
```



Scroll down and click the 'Trust relationships' tab, then click 'Edit trust policy'

Add the ARN for the this role to the list of Principals. (This makes the role self-assumable.)



Click 'Update Policy'

Create a new IAM policy to allow access to the S3 metastore bucket created earlier. Be sure to replace the bucket name and also the assumable role. Name the policy something like: databricks-unity-catalog-s3-access

```
"Version": "2012-10-17",
   "Statement": [
          {
                "Action": [
                      "s3:GetObject",
                      "s3:PutObject",
                       "s3:DeleteObject",
                      "s3:ListBucket",
                      "s3:GetBucketLocation"
                ],
                "Resource": [
                       "arn:aws:s3:::databricks-unity-catalog-acme/*",
                       "arn:aws:s3:::databricks-unity-catalog-acme"
                ],
                "Effect": "Allow"
          },
          {
                "Action": [
                      "sts:AssumeRole"
                ],
                "Resource": [
arn:aws:iam::654654235321:role/databricks-unity-catalog-access"
```

```
"Effect": "Allow"
}
]
}
```

- Navigate back to the databricks-unity-catalog-access role you created and scroll down to the Permissions tab, and under 'Permissions policies, click 'Add permissions'.
- Attach the databricks-unity-catalog-s3-access policy you created to the role.

aws			Q Search	[Option+9	5]	I	2	¢ Ø	٢	Global 🔻	cameron@onehouse.ai @ 6546-5423-5321 ▼
≡	IAM	> Rol	es > databricks-unity-catalog-access > Add permissions								0 0
Att	ach p	oolic	y to databricks-unity-catalog-access								
	• Cu	rren	t permissions policies (0)								
(Othe	. per	missions policies (1/1073)								Ø
						Filter by Type					
	Q da	tabric	ks		×	All types		 1 match 			< 1 > 😂
	<	Polic	:y name	▲	Туре	(x)0				▼ Descript	ion
		÷	databricks-unity-catalog-s3-access		Custom	ner managed				-	
										(Cancel Add pomissions

Create a Unity Catalog metastore

□ In the Databricks Account console, click on 'Catalog' in the left navigation area, and then click 'Create metastore'

📃 🥪 Account					•
₩orkspaces Ar Catalog	Catalog				
Usage	Metastores				
🐣 User management	A metastore is the top-level container for catalog in Unit	/ Catalog. Within a metastore,	Unity Catalog provides a 3-level namespace for organizing data: ca	talogs, schemas (also called	d databases), and tables /
해 Cloud resources	views. Learn more 🖸				
Previews	Q Filter metastores				Create metastore
ASS Settings				=	
	Name	Region	Path	Created at =↓	Updated at
	us-west-2	us-west-2	s3://us-west-2-databricks-metastore-bucket/19d2f021-1	06/05/2024	01/02/2025
	metastore-us-east-1	us-east-1	s3://databricks-workspace-stack-bf4d5-metastore-bucket	07/16/2023	09/25/2023
	metastore-us-west-1	us-west-1	s3://databricks-workspace-stack-c2f31-metastore-bucket	07/16/2023	07/16/2023
		(х)—0		

Choose an AWS region that has not already been used to host a Unity catalog. Paste in the S3 metastore bucket URI. Paste in the IAM role ARN. Click 'Create'.

≡ 🤪 Account		? C
😹 Workspaces	Catalog > Create metastore >	
A Catalog	Create metastore	
🗠 Usage		
🐣 User management	Create metastore – (2) Assign to workspaces	
않 Cloud resources		
Previews	* Name	
Settings	acme-us-east-2	
	* Region	
	Ohio (us-east-2)	
	Select the region for your metastore. You will only be able to assign workspaces in this region to this metastore.	
	S3 bucket path (optional) ⑦	
	s3://databricks-unity-catalog-acme	
	Optional location for storing managed tables data across all catalogs in the metastore. Learn more 🖸	
	IAM role ARN (optional) ③	
	arn:aws:iam::654654235321:role/databricks-unity-catalog-access	
	Enter the IAM role that Databricks will use to access the S3 bucket. Learn more	
	X	
	Create Cancel	

□ When you are asked to 'Configure the IAM Role', note that you have already created the role, but double-check it to be sure that it meets the stated requirements.

😑 🥪 Account			3
a Workspaces	Catalog > Create metastore >		
合 Catalog	Create metastore		
🗠 Usage			
😤 User management	Create metastore —	2) Assign to workspaces	
양 Cloud resources			
Previews	* Name	Configure IAM Role ×	
Settings	acme-us-east-2	Please navigate to the AWS console and configure your IAM role with the below value as the externalID to establish a cross account trust relationship. Learn more	
	* Region	S3 bucket path	
		s3://databricks-unity-catalog-acme	
	Ohio (us-east-2) 🛆	IAM role ADN	
	This region already contains a metasto	amawsilam:654654235321:role/databricks-unity-catalog-access	
	S3 bucket path (optional) ⑦	External Id	
	o?ulldatabrieke unitu estaler som	35/48C32-eC41-4884-901e-C118/430e820 t	
	Optional location for storing managed t		
	optional location for storing managed t	Skip (not recommended) IAM role configured	
	IAM role ARN (optional) ⑦		
	arn:aws:jam::654654235321:role/	databricke_unity_catalon_accase	
	Enter the IAM role that Databricks will u	ise to access the S3 bucket. Learn more 🖸	
	Create Cancel		

□ Select the workspace to be associated with the new metadata store.

≡ 🥪 Account						•	С
😹 Workspaces	Catalog > Create metastore >						
👌 Catalog	Create metastore						
🗠 Usage							
🐣 User management	Create metastore —	Assign to workspaces					
양 Cloud resources 毌 Previews	Assign acme-us-east-2 to	workspaces					
Settings	Filter workspaces	Q					
	Name	Status	Pricing tier	Region	Created	Metastore	
	Acme	Running	Enterprise	us-east-2	today at 12:45 PM	-	
	Vishal3	Running	Enterprise	us-west-2 🕐	09/06/2024	us-west-2 🗹 🛆	
	Vishal2	Failed	Enterprise	us-west-2 🕐	09/06/2024	-	
	Vishal	Failed	Enterprise	us-west-2 🕕	09/05/2024	-	
	Andy	Running	Enterprise	us-west-2 🕕	06/05/2024	us-west-2 🖪 🛆	
	PM Team	Running	Enterprise	us-east-1 🕛	07/16/2023	metastore-us-east-1 🗹 🛆	
	Onehouse Benchmarks	Running	Enterprise	us-east-2	10/17/2022	-	
	Ethan performance benchmar	Running	Enterprise	us-west-1 🕛	05/19/2022	-	
	*						
	1 workspace selected						
	Assign Skip						





Configure the Unity Catalog for CORS

Refer to the Databricks documentation to configure the metastore storage for CORS here: <u>https://docs.databricks.com/en/data-governance/unity-catalog/storage-cors.html</u>

Connect the Workspace to the Onehouse data storage location

To sync your Onehouse tables with Databricks Unity Catalog, you must grant Databricks access to the storage location where your Onehouse tables are stored. This involves creating a Databricks Storage Credential and External Location.

Refer to the Databricks documentation: Create a storage credential for connecting to AWS S3 at: <u>https://docs.databricks.com/en/connect/unity-catalog/cloud-storage/storage-credentials.html</u>

Create and attach IAM policy for Onehouse customer storage access

Create a new AWS IAM Policy. This will be to allow access to the Onehouse customer (data lake) storage. Replace the bucket name with your Onehouse S3 customer bucket location, and the 'Resource' with the ARN for the role you created earlier.

```
"Version": "2012-10-17",
"Statement": [
    {
        "Effect": "Allow",
        "Action": [
            "s3:PutObject",
            "s3:GetObject",
            "s3:ListBucket",
            "s3:DeleteObject",
            "s3:GetBucketLocation"
        ],
        "Resource": [
            "arn:aws:s3:::BUCKET_NAME",
            "arn:aws:s3:::BUCKET_NAME/*"
        1
    },
             "Effect": "Allow",
             "Action": "sts:AssumeRole",
             "Resource": "<DATABRICKS_ACCESS_ROLE_ARN"
      }
1
```

□ Click 'Next'. Name the new policy something like: databricks-onehouse-S3-access-ajax

Click 'Create Policy'.

Note: We will reuse the AWS Role created earlier for Unity Catalog access, and simply attach the new policy for Onehouse S3 customer bucket access.

- Back in the databricks-unity-catalog-access role, click 'Add permissions' and 'Attach policies'.
- Select the databricks-onehouse-S3-access-ajax policy that was just created and click 'Add permissions'.

Create a Databricks Storage Credential

□ In the Databricks workspace console, click 'Catalog', then the settings gear icon, then 'Credentials'.

📃 😂 date	abricks		Q Search data	a, notek
+ New	e Catalog Ø Serverl Type to	ess Starter Warehou search	se Serverless s	Delta SI J uick :
\equiv 😂 databricks		Q Search data, notebooks, n	ecents, and more \$\$ + P	Acme v 🛛 🔶 😮
+ New Workspace Workspace Calog Workflows Compute SoL SoL SoL Editor Dashboards Click 'Create	Catalog Ø Serverless Starter Warehouse Serverless Type to search ✓ My organization > ☐ system > ☐ main ✓ Shared > ☐ samples ✓ Legacy > ☐ hive_metastore Credential'	Image: Second State	 Clean Rooms >> External Data >> Add data 	Send feedback Filter Last viewed Type Ig to see your recent table here.
atabricks 😂		Q Search data, notebooks,	recents, and more 第 + P	Acme 🗸 🔶 😋
+ New Workspace Recents Catalog	Catalog Explorer > External Data External Locations O Credentials	D Connections ①		Create credential
Workflows Compute	Name Purpos	credential Type	Properties	Owner Comment
SQL SQL Editor Cueries	on 99bcc61d-fadd-4364-£ STORA	GE IAM Role	IAM role (ARN): arr::aws:iam::654654235321:role/databricks-u External ID: 35748c32-ec4f-4aa4-9dfe-c1f87430ea2d	' cameron@onehouse.ai

□ For the storage credential name, choose something like onehouse-s3-access. Paste in the AWS ARN of the role you created earlier. Click 'Create'.

atabricks 😂		Q Search data, notebooks, recents, and more # + P		Acme v 🛛 💠 😮
+ New	Catalog Explorer >	Create a new storage credential	×	
Uvrkspace	External Data	A storage credential represents an authentication and authorization mechanism for accessing		
C Recents	External Locations ① Creden			
Catalog	Q Filter credentials	Storage Credential Service Credential Preview		Create credential
🗞 Workflows		* Credential Type		
Compute	Name	AWS IAM Pole	~	Owner Comment
SQL	0- 99bcc61d-fadd-4364-8			cameron@onehouse.ai
SQL Editor	_	Copy from instance profile \vee		
D Queries				
E Dashboards		* Credential name		
📑 Genie		onehouse-s3-access-ajax		
💭 Alerts		* IAM role (ARN) Learn more		
Query History			_	
G SQL Warehouses		arn:aws:iam::654654235321:role/databricks-unity-catalog-access		
Data Engineering		Enter the ARN of the IAM role that has access to the S3 bucket		
≝⊟ Job Runs		Comment		
🕅 Data Ingestion				
No. Delta Live Tables			10	
Machine Learning		> Advanced Options		
Playground				
L Experiments		Cancel Croat	e	
Features				
% Models				

□ You can confirm that the role trust policy matches what is offered in the next dialog, but you should have already set this up.

\equiv 📦 databricks		O Search data, notebooks, recents, and more \$			Acme 🗸 🛛 💠 💽
+ New	Catalog Explorer >	Storage credential created	\times		
	External Data	Please navigate to the AWS console and update your IAM role with the custom trust policy			
C Recents	External Locations ① Creden	provided below to establish a cross-account trust relationship. Learn more			
🔓 Catalog	O Filter credentials	IAM role (ARN): arn:aws:lam::654654235321:role/databricks-unity-catalog-access			Create credential
🗞 Workflows		External ID			
Compute	Name	25749c22_cc4f_4cc4.0dfc_c1f97420cc2d	B	Owner	Comment
SQL	• 99bcc61d-fadd-4364-8	30/40032-6041-4884-3016-0110/4306820	-	cameron@onehouse.ai	
SQL Editor		Trust policy			
Cueries		{ "Version": "2012-10-17",	С		
Dashboards		"Statement": [{			
□ Alerts		"Effect": "Allow", "Principal": {			
Query History		"AWS": ["arn:aws:iam::414351767826:role/unity-catalog-prod-UCMasterRole-14S5ZJ\	KOTY		
G SQL Warehouses		"arn:aws:iam::654654235321:role/databricks-unity-catalog-access"]			
Data Engineering		}, "Action": "sts:AssumeRole",			
🚈 Job Runs		"Condition": { "StringEquals": {			
3 Data Ingestion		"sts:ExternalId": "35748c32-ec4f-4aa4-9dfe-c1f87430ea2d" }			
Number 2018 Delta Live Tables		}			
Machine Learning		1			
Playground					
L Experiments		2	ine		
☐ ¹ Features					
% Models					

□ Click 'Validate Configuration' and ensure that all the permissions are confirmed good.

\equiv 💚 databricks		Q Search data, notebooks, recents, and more	3€ + P	Acme - 🔶 📀
+ New	Catalog Explorer > Credentials > on onehouse-s3-access Overview Usage Permissions	S-ajax Workspaces		: Validate Configuration
Catalog & Workflows Compute	Description Add description			About this credential Owner cameron@onehouse.ai 🖉
SQL SQL Editor SQL Editor Dashboards SGN Genie Alerts Query History G SQL Warehouses Data Engineering Ed Dob Runs Data Insection	Credential Type Purpose IAM role (ARN) External ID	Test connection Success - Assume Role Success - Self Assume Role Success - External ID Condition All Permissions Confirmed The storage credential grants permission to perform all necessary operations	X Done	

Create a Databricks External Location

Since we are referencing an existing Onehouse S3 bucket, refer to the Databricks documentation: 'Creating an external location manually using Catalog Explorer' at: <u>https://docs.databricks.com/en/connect/unity-catalog/cloud-storage/external-locations.html#c</u>reate-an-external-location-manually-using-catalog-explorer

□ In the Databricks workspace console, click 'Catalog', then the settings gear icon, then 'External Locations'.

📃 📦 datab	oricks				Q Sea	rch data, no	tek	
+ New		Catalog Ø Serverle	ss Starter Wareł	ouse Serverless	⊛ ⊂ +	Delta	ı SI	
Uvrkspace		Type to se	earch		۷ -	Quic	k	
= 😂 databricks			Q Search data, notebo	ooks, recents, and more	₩ + P			Acme 🗸 💠 📀
+ New Ca	atalog © Serverless Starter W	arehouse Serverless	log + Delta Sh Delta Sharing	aring > Clean Rooms >	External Data >	Add data 🗸		💬 Send feedback
Workspace	Type to search		Clean Rooms					
A Catalog	✓ My organization		External Locations	🖈 Favorites 📋 0	Catalogs			Q Filter
➢ Workflows△ Compute	> 🗋 main > Shared		Credentials Connections				Last viewed	Туре
SQL	> 🛃 samples Y Legacy > 📮 hive_metasto	re	Metastore acme-us-east-2	٥				
 ▶ Queries 				No re	ecents table available ye	t, start exploring to s	ee your recent table here	

Choose 'Manual', then click 'Next'.



Provide a name such as onehouse-s3-projectname. Paste in the S3 URL for the existing Onehouse customer data bucket. Choose the storage credential that we created earlier. Click 'Create'.

\equiv 🥯 databricks	Q Search data, notebooks, recents, and more ೫ + P	Acme 🗸 🔶	• ©
+ New	External Locations >		
U Workspace	Create a new external location manually		
() Recents	An external location allows you to access your data stored in S3. You will need the S3 nath and the ability to create an		
👌 Catalog	IAM role which gives access to that path Learn more 🖆		
S Workflows			
Compute	* External location name		
	onehouse-s3-ajax		
SQL			
SQL Editor	• URL ©		
Queries	Enter the bucket path that you want to use as the external location		
Dashboards	s3://oh-project-ajax Copy from DBFS ~		
📑 Genie	* Storage credential Learn more 🖸		
💭 Alerts	Provide a storage credential capable of accessing the URL		
① Query History	onehouse-s3-access-ajax (IAM Role)		
G SQL Warehouses	IAM roje (APN): arr::ws:iam::654654235321:roje/databricks-unitv-catalog-access		
Data Engineering	External ID: 35748c32-ec4f-4aa4-9dfe-c1f87430ea2d		
🚈 Job Runs			
🕅 Data Ingestion	Comment		
🖫 Delta Live Tables			
	•		
Machine Learning			
Playground	> Advanced Options		
Experiments			
Features			
S Models	Back		
0.000100			

Click 'Test connection' and ensure that all permissions are confirmed good.

\equiv 📦 databricks		Q Search data, notebooks, recents, and more 3€ + P	Acme v 🛛 💠
+ New Workspace Recents Catalog	Catalog Explorer > External Locations > Catalog Explorer > External Locations > Coerview Permissions Browse V Description	Norkspaces	: Test cortyset About this external location
Workflows Compute SOL SQL Editor Queries	Add description Te Credential URL	st connection ⊘ Success - Read ⊘ Success - List ⊙ Success - Write ⊗ Success - Delete	X Owner cameron@onehouse.ai Ø
 Dashboards Genie Alerts Query History SQL Warehouses 	Limit to read-only use	Success - Path Exists Success - Assume Role Success - Sett Assume Role Success - External ID Condition I Permissions Confirmed The associated Storage Credential grants permission to perform all necessary operations.	
Data Engineering 또 Job Runs 웴 Data Ingestion % Delta Live Tables		Done	
Machine Learning 한 Playground 문 Experiments 라 Features 양 Models			

Gather the external catalog information in Databricks

There are two ways that Onehouse can authenticate to Databricks to synchronize the metadata catalog: OAuth, and Access Token.

- For the Access Token Auth Type, refer to the Databricks documentation for personal access token generation at: <u>https://docs.databricks.com/en/integrations/jdbc/authentication.html#databricks-personal-access-token</u>
- For the OAuth access type, refer to the Databricks documentation for service principal creation and retrieving credentials at:
 https://docs.databricks.com/en/integrations/jdbc/authentication.html#oauth-machine-t_o-machine-m2m-authentication
- □ In the Databricks workspace console, select 'SQL Warehouses', and then select the 'Serverless Starter Warehouse' that should have been created with the new workspace, or select the SQL Warehouse that you plan to use for catalog synchronization.

□ Make note of the Server hostname and HTTP path.



Create the Metadata Catalog Sync in Onehouse

Finally! We are ready to create the metadata sync in Onehouse and attach it to a data lakehouse table.

Create the Databricks metadata catalog in Onehouse

- □ In Onehouse, click 'Catalogs' in the left navigation panel.
- □ Scroll over to the Databricks Unity Catalog icon and choose 'Add new Catalog'.
- □ Fill in the new catalog form:
 - Enter a Onehouse catalog name such as databricks-catalog
 - The Unity Catalog Name is the location in the Unity catalog where the tables will be synced. If the catalog location doesn't already exist, it will be created.
 - Paste in the Server Hostname from Databricks.
 - Paste in the server HTTP Path from Databricks.

C ONEHOUSE	Ajax Company us-west-2 (~	4 ³⁹⁺ C ∽
	Catalogs	New catalog	×
🛆 Home	Catalog	Create a data catalog to register metadata of your databases and tables	
🖯 Data	Create and		
K Stream Captures		Name*	
Table Services		uatabilities-catalog	
පී Sources	NAME	Type*	
Catalogs		Databricks Unity Catalog 🔹	Detabalate Units October
ılı Usage		Unity Catalog Name*	
Settings ^	(+) Add new Ca	Ajax-retail	Add new Catalog
Users & Access		Server Hostname*	
User Preferences	🕀 Add Ne	dbc-6a470af3-aced.cloud.databricks.com	Actions :
Integrations			
API Settings	🗌 Nai	HTTP Path*	Created On 个
Cloud accounts	aiax	/sql/1.0/wateriouses/000/c04040102200	Aug 14, 2024 11:03 AM
Project Settings		Credential Type	
♀ Notifications	sno	Secret Manager ()	Aug 23, 2024 1:08 PM
G Support		 Unehouse will use secret arn to get credentials. 	
		Credentials () Stored securely by Onehouse	
All rights reserved. © 2025			Version: release-v0.139.1 About Docs

- Choose the Credential type.
 - For OAuth type, enter the service principal credentials 'client-id' and 'client-secret' from your workspace.
 - For Access Token type, enter the 'personal-access-token' for your workspace user.

C ONEHOUSE	Ajax Company	~	60° C ∨
		Unity Catalog Name*	
	Catalogs	Ajax-retail	
🛈 Home	Catalog	Server Hostname*	
😫 Data	Create and	dbc-6a470af3-aced.cloud.databricks.com	
💐 Stream Captures		HTTP Path*	
🖬 Table Services		/sql/1.0/warehouses/8887c0454516e2c5	
皆 Sources			
Catalogs	Hive Metastore	Credential Type	Databricks Unity Catalog
III Usage	Add new Ca	O Secret Manager (j) Onehouse will use secret arn to get credentials.	talog
Settings ^		- Credentials ()	
Users & Access		Stored securely by Onehouse	
User Preferences	🕀 Add No	Type*	Actions :
Integrations		Access Token	
API Settings	🗌 Nai		Created On 个
Cloud accounts		Personal Access Token*	Aug 14, 2024 14:03 AM
Project Settings	ajax	Personal Accoss Token	Aug 14, 2024 Theo Aut
Q Notifications	sno	Create	Aug 23, 2024 1:08 PM
G Support			
			1 00 000E 0.E0 MM
All rights reserved. © 2025			Version: release-v0.139.1 About Docs

Attach the Onehouse Databricks catalog to one or more Stream Captures.

Edit or create a stream capture that uses the Databricks catalog sync.

ONEHOUSE	Ajax Company us-west-2 (Oregon)	~	С ⁹⁹⁺ С ~
	Capture > PostgresRetailCDC:re	rtail_item_master	
HomeData	Stream Capture Name:	PostgresRetailCDC:retail_item_master 25 2:32 PM etail_item_master	Running Actions :
✓ Stream Captures□ Table Services			
B Ources	Destination		
la Catalogs ₁ ₁ Usage	Data lake	retail	
Settings	Database	retail_bronze	
NotificationsSupport	Catalogs	snowflake_catalog, ajax, xtable	Cancel
		Catalogs	
	×	snowflake_catalog ajax xtable databricks-catalog	*
		Save	
	Table name	retail_item_master	
All rights reserved. © 2025			Version: release-v0.139.1 About Docs

After a time, the table metadata will be synchronized to the Databricks Unity Catalog. It can then be referenced in queries.

\equiv \gtrless databricks			Qs	Search data, notebooks, rec	cents, and more		ж + Р
+ New	≙	Catalog C X	Ø Nev	w Query 2025-01-10 3:02p	m • +		
 Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries 	\$	For you All For you All My organization \$\[\[-\]_\]estimation \$\[-\]_\]estimation \$\[-\]_\]estimation_schema	► R	un selected (1000) SELECT category_code, from acme.retail_bronz LIMIT 10	<pre></pre>	e. 🖯 default 🗸 rice, item_upc Ister	. ☆
 Dashboards Genie Alerts Query History SQL Warehouses 		 iii retail_scans iii retail_stores iii main Shared iii samples 	Raw r	esults V +	x		
Data Engineering		✓ Legacy		1 ² 3 category_code	1 ² 3 item_id	1.2 item_price	^{A^B_C item_}
		> 📮 hive_metastore	1	20100	152967	1	2855988
Job Runs			2	20100	152957	1.06	2134552

###