

Akilesh K

k.akilesh123@gmail.com

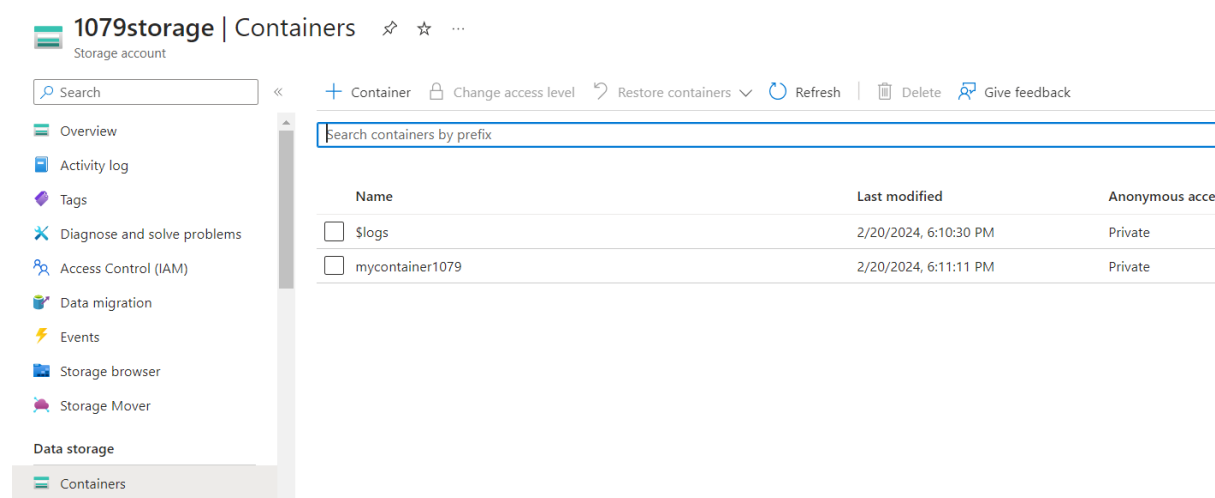
Data engineering - Batch 1

Date: 20-02-24

DAY 22 – Azure DataFactory

Transfer of data from blob storage account to another storage account using Azure datafactory pipeline

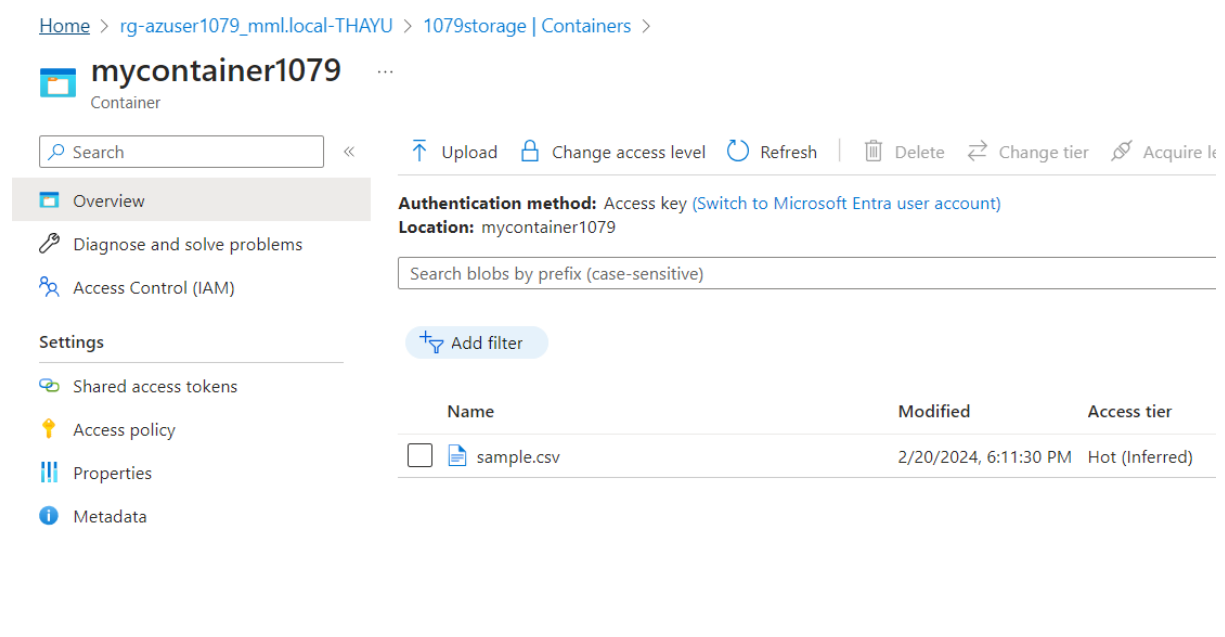
Creating a source storage account



The screenshot shows the Azure portal interface for a storage account named '1079storage'. The left sidebar contains navigation links: Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Storage Mover, and Data storage. The 'Containers' link is selected. The main area shows a list of containers with the following columns: Name, Last modified, and Anonymous access. Two containers are listed: '\$logs' and 'mycontainer1079', both created on 2/20/2024.

Name	Last modified	Anonymous access
\$logs	2/20/2024, 6:10:30 PM	Private
mycontainer1079	2/20/2024, 6:11:11 PM	Private

Upload a sample file in the container



The screenshot shows the Azure portal interface for the 'mycontainer1079' container. The left sidebar contains navigation links: Overview, Diagnose and solve problems, Access Control (IAM), Settings, Shared access tokens, Access policy, Properties, and Metadata. The 'Overview' link is selected. The main area shows the container details, including the authentication method (Access key) and location (mycontainer1079). A search bar for blobs is present. Below the search bar, there is a table of blobs with the following columns: Name, Modified, and Access tier. One blob is listed: 'sample.csv', uploaded on 2/20/2024.

Name	Modified	Access tier
sample.csv	2/20/2024, 6:11:30 PM	Hot (Inferred)

Create a destination storage account

Microsoft Azure

Search resources, services, and docs (G+/)

Home > 1079newstorage

1079newstorage | Containers

Storage account

Search

Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix

Name	Last modified
<input type="checkbox"/> \$logs	2/20/2024, 6:11:09 PM
<input type="checkbox"/> newcontainer1079	2/20/2024, 6:11:50 PM

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Storage Mover

Data storage

Containers

Create a empty container

Home > 1079newstorage | Containers >

newcontainer1079

Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: newcontainer1079

Search blobs by prefix (case-sensitive)

Add filter

Name	Modified	Access tier
No results		

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

Create a azure datafactory

Home > Microsoft.DataFactory-20240220181351 | Overview >

1079hexadf

Data factory (V2)

Search << Delete

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Networking
- Managed identities
- Properties
- Locks

Getting started

- Quick start

Essentials

Resource group ([move](#)) : [rg-azuser1079_mml.local-THAYU](#) Type : Data factory (V2)

Status : Succeeded Getting started : [Quick start](#)

Location : East US

Subscription ([move](#)) : [Azure subscription 1](#)

Subscription ID : 984f097c-963c-4eb6-a20d-839457ae9f08

Azure Data Factory Studio

[Launch studio](#)

Launch the studio and click ingest data

Microsoft Azure | Data Factory | 1079hexadf Search factory and documentation azuser1079_mml.local@mml.local

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration. Learn more

[Set up code repository](#)

Data factory

1079hexadf

New

Ingest
Copy data at scale once or on a schedule.

Orchestrate
Code-free data pipelines.

Transform data
Transform your data using data flows.

Configure SSIS
Manage & run your SSIS packages in the cloud.




Recent resources


Click built in copy task

Microsoft Azure

Data Factory > 1079hexadf


Search factory and documentation








Copy Data tool


>>











1 Properties

2 Source

3 Destination

4 Settings


5 Review and finish

Use Copy Data Tool to perform a one-time or scheduled data load from 90+ data sources. Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool generate the artifacts for you, including pip


Properties

Select copy data task type and configure task schedule

Task type

**Built-in copy task**

You will get single pipeline to copy data from 90+ data source easily.

**Metadata-driven copy task**

You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.

You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

Task cadence or task schedule *

☒ Run once now ☐ Schedule ☐ Tumbling window

Choose azure storage and create a new connection by mentioning the storage name

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a

Source type


Azure Blob Storage

Connection *

Select...

+ New connection

New connection

 Azure Blob Storage [Learn more](#)

Name *

1079storage

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime

Authentication type

Account key

Connection string

Azure Key Vault

Account selection method ⓘ

☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ

Azure subscription 1 (984f097c-963c-4eb6-a20d-839457ae9f08)

Storage account name *

1079storage

↻

Additional connection properties

Specify the source container path

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type Azure Blob Storage

Connection * 1079storage [Edit](#) [+ New connection](#)

if the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse. Append a slash (/) at the end if the path refers to a folder.

[Browse](#)

Options

☐ Binary copy ⓘ

☒ Recursively ⓘ

☐ Enable partitions discovery ⓘ

Max concurrent connections ⓘ

Select the new storage account for destination datastore

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type Azure Blob Storage

Connection * Select... [+ New connection](#)

New connection

[Azure Blob Storage](#) [Learn more](#)

Name *

Description

Connect via integration runtime * ⓘ AutoResolveIntegrationRuntime

Authentication type Account key

[Connection string](#) [Azure Key Vault](#)

Account selection method ⓘ

☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ Azure subscription 1 (984f097c-963c-4eb6-a20d-839457ae9f08)



Storage account name * 1079newstorage [Refresh](#)

Mention the destination container path



Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type

 Azure Blob Storage 

Connection *

 1079newstorage 



Edit



New connection

Folder path *


If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

newcontainer1079/




Browse

File name

Copy behavior 

Select...



Max concurrent connections 

Create a pipeline name


Settings

Enter name and description for the copy data task, more options for data movement

Task name *


CopyPipeline_1079

Task description


Data consistency verification 

☐

Fault tolerance 

Enable logging 

☐

Enable staging 


☐

> Advanced


Summary of the pipeline is provided

Summary

You are running pipeline to copy data from Azure Blob Storage to Azure Blob Storage.

 Azure Blob Storage

→

 Azure Blob Storage

Edit

Properties

Task name

CopyPipeline_1079

Task description

Source

Connection name

1079storage

Dataset name

SourceDataset_h86

Column delimiter

;

Row delimiter

Escape character


\

Quote char


"

Edit

The pipeline gets executed

 Azure Blob Storage

→

 Azure Blob Storage

Deployment complete

Deployment step	Status
Validating copy runtime environment	✔ Succeeded
> Creating datasets	✔ Succeeded
> Creating pipelines	✔ Succeeded
> Running pipelines	✔ Succeeded

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

Finish

Edit pipeline

Monitor

The data gets transferred to destination blob storage

Home > rg-azuser1079_mml.local-THAYU > 1079newstorage | Containers >

newcontainer1079 ...
Container

Search << Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots

Overview
Diagnose and solve problems
Access Control (IAM)

Settings
Shared access tokens
Access policy
Properties
Metadata

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: newcontainer1079

Search blobs by prefix (case-sensitive)

Add filter

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> sample.txt	2/20/2024, 6:21:49 PM	Hot (Inferred)		Block blob

The pipeline is executed successfully

Data Factory Validate all Publish all Preview experience Off

Factory Resources
Filter resources by name +
Pipelines 1
CopyPipeline_1079
Change Data Capture (preview) 0
Datasets 2
Data flows 0
Power Query 0

Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDI Insight
Iteration & conditionals
Machine Learning
Power Query

Copy data
Copy_h86

Parameters Variables Settings **Output**

Pipeline run ID: f966d943-6d7f-4013-b2f5-9fc07752c007 **Pipeline status:** Succeeded [View debug run consumption](#)

All status
Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime
Copy_h86	Succeeded	Copy data	2/20/2024, 6:21:33 PM	18s	AutoResolveIntegrator