

Akilesh K

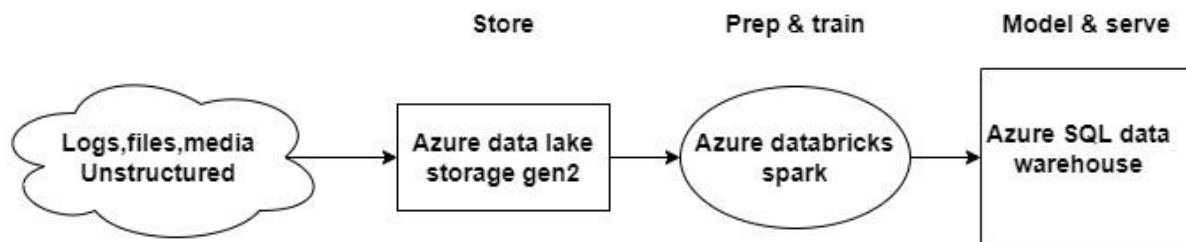
[k.akilesh123@gmail.com](mailto:k.akilesh123@gmail.com)

Data engineering - Batch 1

Date: 16-02-24

## DAY 20 – Azure -Data Lake Storage

### ETL Model



The data lifecycle typically involves several key phases:

- **Ingestion:** This phase involves acquiring source data from various sources such as files, logs, and other unstructured data. Technologies like Azure Synapse Analytics, Azure Data Factory, Apache Kafka, or Stream Analytics may be used for batch or real-time data movement.
- **Storage:** In the storage phase, ingested data is placed into a secure and scalable storage solution. Azure Data Lake Storage Gen2 is commonly used for this purpose due to its compatibility with big data processing technologies.
- **Preparation and Training:** In this phase, technologies such as Azure Synapse Analytics, Azure Databricks, Azure HDInsight, and Azure Machine Learning are used for data preparation and model training in machine learning solutions.
- **Model and Serving:** Finally, the model and serve phase involves presenting the data to users. This may involve visualization tools like Microsoft Power BI or analytical data stores such as Azure Synapse Analytics. Often, a combination of multiple technologies is used based on business requirements.

Overall, the data lifecycle encompasses processes for acquiring, storing, preparing, analyzing, and presenting data, with various technologies utilized at each stage to enable efficient data management and decision-making.

## Data lake storage

- If you want to store data without performing analysis on the data, set the Hierarchical Namespace option to Disabled to set up the storage account as an Azure Blob storage account. You can also use blob storage to archive rarely used data or to store website assets such as images and media.
- If you are performing analytics on the data, set up the storage account as an Azure Data Lake Storage Gen2 account by setting the Hierarchical Namespace option to Enabled. Because Azure Data Lake Storage Gen2 is integrated into the Azure Storage platform, applications can use either the Blob APIs or the Azure Data Lake Storage Gen2 file system APIs to access data.

## Login into Azure portal

The screenshot shows the Microsoft Azure portal interface. At the top is a blue header bar with the Microsoft Azure logo, a search bar, and several utility icons. Below the header, the page is divided into several sections:

- Azure services:** A row of icons for various services: Create a resource, Storage accounts, Data Lake Storage Gen1, Data Lake Analytics, Azure Databricks, Resource groups, Quickstart Center, Virtual machines, App Services, and a link to More services.
- Resources:** A section with tabs for Recent and Favorite. It contains a table with one resource entry.
- Navigate:** A row of icons for Subscriptions, Resource groups, All resources, and Dashboard.
- Tools:** A section that is currently empty.

Name	Type	Last Viewed
rg-azuser1079_mmmllocal-THAYU	Resource group	a few seconds ago

## Go to a storage account

Microsoft Azure

Storage accounts

Home >

IIHT (iiht.onmicrosoft.com)

+ Create

↶ Restore

⚙ Manage view

↺ Refresh

⬇ Export to CSV

🔗 Open query

🏷 Assign tags

🗑 Delete

Filter for any field...

Subscription equals all

Resource group equals all

Location equals all

Add filter

Showing 1 to 100 of 207 records.

<input type="checkbox"/> Name ↑↓	Type ↑↓	Kind ↑↓	Resource group ↑↓
<input type="checkbox"/> adlsgen2hexa1077	Storage account	StorageV2	rg-azuser1077_mml.lo
<input type="checkbox"/> anm123	Storage account	StorageV2	anm123
<input type="checkbox"/> blobstorage1176	Storage account	StorageV2	Rpower
<input type="checkbox"/> blobstorage2319825	Storage account	StorageV2	powershell2319825
<input type="checkbox"/> blobstorage2319933	Storage account	StorageV2	powershell2319933
<input type="checkbox"/> csg10032003540d0576	Storage account	StorageV2	cloud-shell-storage-ce
<input type="checkbox"/> csg10032003540d0c6b	Storage account	StorageV2	cloud-shell-storage-ce
<input type="checkbox"/> csg10032003540d0c71	Storage account	StorageV2	cloud-shell-storage-ce

## Create a storage account

[Home](#) > [Storage accounts](#) >

## Create a storage account

[Basics](#) [Advanced](#) [Networking](#) [Data protection](#) [Encryption](#) [Tags](#) [Review](#)

[Create new](#)

Instance details

Storage account name ⓘ \*

adlshexa1079

Region ⓘ \*

(Asia Pacific) Central India

Deploy to an edge zone

Performance ⓘ \*

☒ Standard: Recommended for most scenarios (general-purpose v2 account)

☐ Premium: Recommended for scenarios that require low latency.

Redundancy ⓘ \*

Geo-redundant storage (GRS)

☒ Make read access to data available in the event of regional unavailability.

Review

< Previous

Next : Advanced >

## In the Advanced option ,check enable hierachial Namespace

Basics **Advanced** Networking Data protection Encryption Tags Review

Permitted scope for copy operations  
(preview) ⓘ

From any storage account



### Hierarchical Namespace

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs) [Learn more](#)

Enable hierarchical namespace



### Access protocols

Blob and Data Lake Gen2 endpoints are provisioned by default [Learn more](#)

Enable SFTP ⓘ



Enable REST API ⓘ



Review

< Previous

Next : Networking >

Review the storage account

Create a storage account ...

- Basics
- Advanced
- Networking
- Data protection
- Encryption
- Tags
- Review

Basics

Subscription	Azure subscription 1
Resource Group	rg-azuser1079_mml.local-THAYU
Location	centralindia
Storage account name	adlshexa1079
Deployment model	Resource manager
Performance	Standard
Replication	Read-access geo-redundant storage (RA-GRS)

Advanced

Enable hierarchical namespace	Enabled
Enable network file system v3	Disabled
Allow cross-tenant replication	Disabled
Access tier	Hot

Create

< Previous

Next >

Download a template for automation

Storage account is deployed

 adlshexa1079\_1708079770218 | Overview ...

Search

«

Delete

Cancel

Redeploy

Download

Refresh

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name: adlshexa1079\_1708079770218

Subscription: Azure subscription 1

Resource group: rg-azuser1079\_mml.local-THAYU

Start time: 2/16/2024, 4:06:14 PM

Correlation ID: 5de3ee81-f71d-40b1-8d6f-f6f8fe575e25

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

Open the storage account

adlshexa1079

Storage account

Search

«

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Data storage

Containers

File shares

Queues

Tables

Upload

Open in Explorer

Delete

Move

Refresh

Open in mobile

CLI / PS

Feedback

Essentials

Resource group (move) : rg-azuser1079\_mmjlocal-THAYU

Location : centralindia

Primary/Secondary Location : Primary: Central India, Secondary: South India

Subscription (move) : Azure subscription 1

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

Disk state : Primary: Available, Secondary: Available

Performance : Standard

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

Account kind : StorageV2 (general purpose v2)

Provisioning state : Succeeded

Created : 2/16/2024, 4:06:15 PM

Tags (edit) : Add tags

Properties

Monitoring

Capabilities (5)

Recommendations (0)

Tutorials

Tools + SDKs

Data Lake Storage

Hierarchical namespace : Enabled

Default access tier : Hot

Blob anonymous access : Disabled

Security

Require secure transfer for REST API operations : Enabled

Storage account key access : Enabled

Create a container

adlshexa1079 | Containers

Storage account

Search

«

Events

Storage browser

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

Access keys

+ Container

Change access level

Restore containers

Refresh

Delete

Give feedback

Search containers by prefix

Name	Last modified	Anonymous acc
<input type="checkbox"/> \$logs	2/16/2024, 4:06:43 PM	Private
<input type="checkbox"/> mycontainer	2/16/2024, 4:08:07 PM	Private

Add sample files

mycontainer

Container

Search

«

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Upload

Add Directory

Refresh

Rename

Delete

Change tier

Acquire lease


Break lease

Give fee


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

Location: mycontainer

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/>  mark.csv	2/16/2024, 4:08:39 PM	Hot (Inferred)		Block blob

## Copy the access key-connection string

 **adlshexa1079** | Access keys ☆ ...  
Storage account

<<

[Diagnose and solve problems](#)  
[Access Control \(IAM\)](#)  
[Data migration](#)  
[Events](#)  
[Storage browser](#)

**Data storage**  
[Containers](#)  
[File shares](#)  
[Queues](#)  
[Tables](#)

**Security + networking**  
[Networking](#)  
**[Access keys](#)**

[Set rotation reminder](#) [Refresh](#) [Give feedback](#)

Access keys authenticate your applications' requests to this storage account. Keep your keys in a secure location Key Vault, and replace them often with new keys. The two keys allow you to replace one while still using the other.

Remember to update the keys with any Azure resources and apps that use this storage account.  
[Learn more about managing storage account access keys](#)

Storage account name  
adlshexa1079

**key1** [Rotate key](#)  
Last rotated: 2/16/2024 (0 days ago)  
Key  
..... [Show](#)

Connection string  
DefaultEndpointsProtocol=https;AccountName=adlshexa1079;AccountKey=WTn... [Hide](#)

**key2** [Rotate key](#)  
Last rotated: 2/16/2024 (0 days ago)

## Open azure storage explorer


<>


[Collapse all](#) [Refresh all](#)

[Quick Access](#)  
Emulator & Attached  
Storage Accounts  
Azure subscription 1 (azuser1079\_mml.local@i...)  
Storage Accounts  
Disks


## Storage Explorer


### Get Started

 **Sign in with Azure**  
Sign in with your Azure account to access all of your Azure resources.

 **Attach to a resource**  
Don't have permissions to access Azure subscriptions? You can connect to individual storage resources using various authentication methods.

### Tasks

 **View Azure resources**  
See your subscriptions and Azure resources all in one place.

 **Manage accounts and subscriptions**

**Resources**  
[Documentation](#)  
[Troubleshooting](#)

## Connect it by pasting the connection string

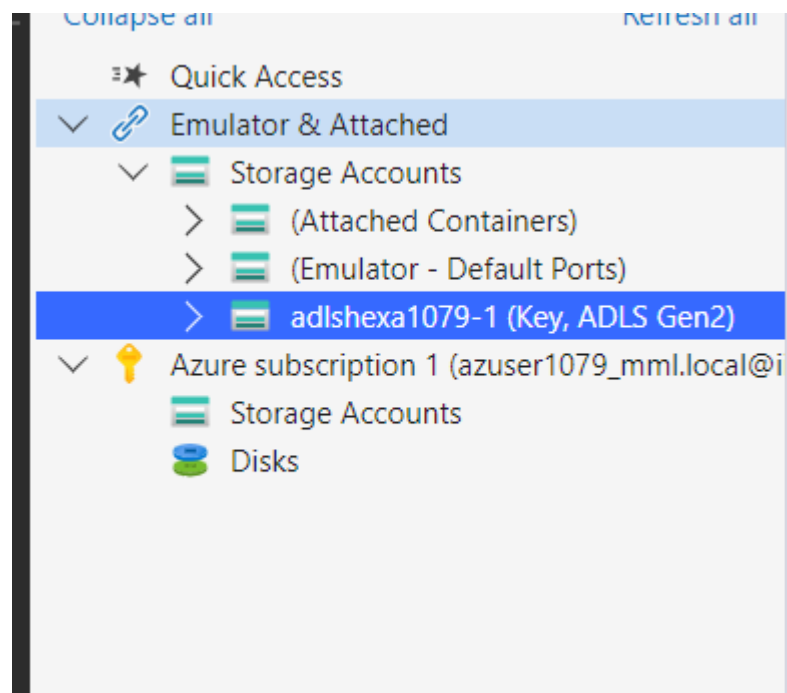
### Enter Connection Info

Select Resource > Select Connection Method > **Enter Connection Info** > Summary

Display name:  
adlshexa1079-1

Connection string:  
DefaultEndpointsProtocol=https;AccountName=adlshexa1079;AccountKey=WTnFjgDSK0mHhPFg3RUD1mOANiF+m2oWkdMt+FwoxSfxJRweLW6BS1Mr7AsJlgyZV  
D/oui/BRx4a+AStHbbmEQ==;EndpointSuffix=core.windows.net

## ADLS Gen2 is added to the file explorer





Sample file is present here in the dashboard

Collapse all

Refresh all

Quick Access

Emulator & Attached

Storage Accounts

(Attached Containers)

(Emulator - Default Ports) (Key)

adlshexa1079-1 (Key, ADLS Gen2)

Blob Containers

\$logs

mycontainer

File Shares

Queues

Tables

Azure subscription 1 (azuser1079\_mml.local@i)

Storage Accounts

Disks

Upload

Download

Open

Preview

New Folder

Select All

Copy

Paste

Clone

Active blobs (default)

mycontainer

Filter by prefix (case-

Name	Access Tier	Access Tie
mark.csv	Hot (inferred)	

Showing 1 to 1 of 1 cached items

Actions

Properties

Activities

URL

Custom Domain

Type

HNS Enabled

DFS Endpoint

Lease State

Lease Status

Public Read Access

https://adlshexa1079.blob.core

Blob Container (ADLS Gen2)

true

https://adlshexa1079.dfs.core.v

available

unlocked

off

Clear completed

Clear successful

Successfully added new connection.