

Azure Databricks Assignment - 4

Mitushi Vishwakarma

Week - 4

Azure Databricks
Day - 4

Database :-

- collection of data or information.
- accessed electronically
- support OLTP.
- DBMS store data in database

Characteristics

- RDB store data in tables fixed rows and columns
- NoSQL / Non-RDB store in JSON, key-value pairs, tables with rows and dynamic columns
- Security features
- ACID transactions to ensure data integrity
- APIs to easily interact
- flexible deployment topologies to isolate workloads to a specific set of resources

Types of data in DB :-

- patient medical records
- items in online store
- financial records
- IoT device readings
- mobile application information

Examples :-

RDB :- MS-SQL server, Oracle, MySQL

27-11-2023

27

NOVEMBER
WK 48 • DAY 31-034
MONDAY
2023

DECEMBER
2023
S S F T W T S
48 49 50 51 52
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Document database :- MongoDB

Key-value :- DynamoDB

Wide-column → Cassandra

Graph :- Amazon Neptune

Data warehouses and data lakes
Support OLAP. OLAP systems are typically
used to collect data from a variety of sources.

Data Warehouse :-

- a system that stores highly structured information. (collection of databases)
- typically store current and historical data
- goal is to combine data in order to analyze, look for insights and create business intelligence in form of reports and dashboards.

Characteristics

- contain range of data
 - ↳ raw ingested data
 - ↳ highly curated
 - ↳ cleansed, filtered and aggregated data

48	27	28	29	30
47	20	21	22	23
46	13	14	15	16
45	6	7	8	9
44	0	1	2	3
43	4	5	6	7
42	10	11	12	13
41	17	18	19	20
40	24	25	26	27
39	31	32	33	34

2023
SATURDAY
WK 47 • DAY 329-036
NOVEMBER

25-11-2023

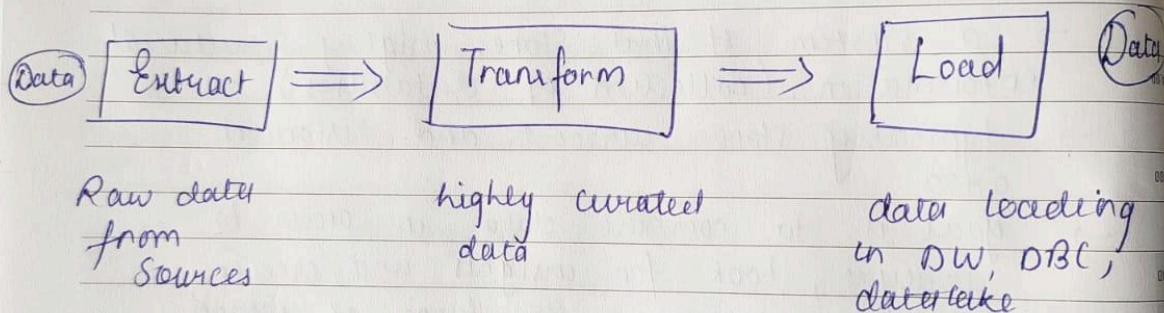
25

ETL move data from original source to data warehouse.

ETL move data on regular schedule so

- pre-defined and fixed relational schema.
- Some also support semi-structured data

ETL Processes.



ETL pipelines -

Data warehouse examples.

- Amazon Redshift
- Snowflake.
- IBM Db2 Warehouse
- Google Bigquery.
- Azure Synapse

24-11-2023

24

NOVEMBER

WK 47 • DAY 328-037

FRIDAY

49 50 51 52
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Data Lake

- repository of data from disparate sources stored in original raw format.
- primary purpose is to analyse data to gain insights

Data Lake is usually used in ELT

Characteristics

- large amounts of data
 - structured, semi, unstructured
 - audio, pdf, JSON.
- use of data lake can vary based on structure of data.

Data in data lakes can be processed with a variety of OLAP systems. (BI)

Examples:-

Azure Data Lake Storage Gen2

AWS S3

Google Cloud Storage.

44	1	2	3	4	5
45	6	7	8	9	10
46	11	12	13	14	15
47	16	17	18	19	20
48	21	22	23	24	25
	26	27	28	29	30

2023
THURSDAY
WK 47 • DAY 327-038
NOVEMBER

23-11-2023

23

Created a blob storage account in azure :

The image shows two screenshots from the Microsoft Azure portal. The top screenshot is the 'Create a storage account' page. The bottom screenshot is the 'Overview' page for the created storage account.

Create a storage account

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

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Resource group * [Create new](#)

Instance details

Storage account name

Region [Deploy to an edge zone](#)

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

[Review](#) [< Previous](#) [Next: Advanced >](#) [Give feedback](#)

hexastorage1046_1708057719326 | Overview

Deployment

Search [Delete](#) [Cancel](#) [Redeploy](#) [Download](#) [Refresh](#)

Deployment is in progress

Deployment name: hexastorage1046_1708057719326 Start time: 16/2/2024, 9:58:42 am
Subscription: Azure subscription 1 Correlation ID: 36dd09b9-2c54-4e95-934f-31d426cf16d9
Resource group: rg-azuser1046_mml.local-xASSG

Deployment details

Resource	Type	Status	Operation details
No results.			

Give feedback

[Tell us about your experience with deployment](#)

Microsoft Defender for Cloud
Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

Free Microsoft tutorials
[Start learning today >](#)

Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
[Find an Azure expert >](#)

hexastorage1046_1708057719326 | Overview

Deployment

Search

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name: hexastorage1046_1708057719326
Subscription: Azure subscription 1
Resource group: rg-azuser1046_mmllocal-xASSG

Start time: 16/2/2024, 9:58:42 am
Correlation ID: 36dd09b9-2c54-4e95-934f-31d426cf16d9

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

Cost Management

Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >

Microsoft Defender for Cloud

Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

Start learning today >

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >

hexastorage1046 - Microsoft Azure

Search resources, services, and docs (G+)

Home > hexastorage1046_1708057719326 | Overview

hexastorage1046

Storage account

Search

Upload

Open in Explorer

Delete

Move

Refresh

Open in mobile

CLI / PS

Feedback

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Storage Mover

Data storage

Containers

File shares

Queues

Tables

Security + networking

Essentials

Resource group (move) : rg-azuser1046_mmllocal-xASSG

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

Tags (edit) : Add tags

Performance : Standard

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

Account kind : StorageV2 (general purpose v2)

Provisioning state : Succeeded

Created : 2/16/2024, 9:58:44 AM

Properties

Monitoring

Capabilities (7)

Recommendations (0)

Tutorials

Tools + SDKs

Blob service

Hierarchical namespace : Disabled

Default access tier : Hot

Blob anonymous access : Disabled

Blob soft delete : Enabled (7 days)

Container soft delete : Enabled (7 days)

Security

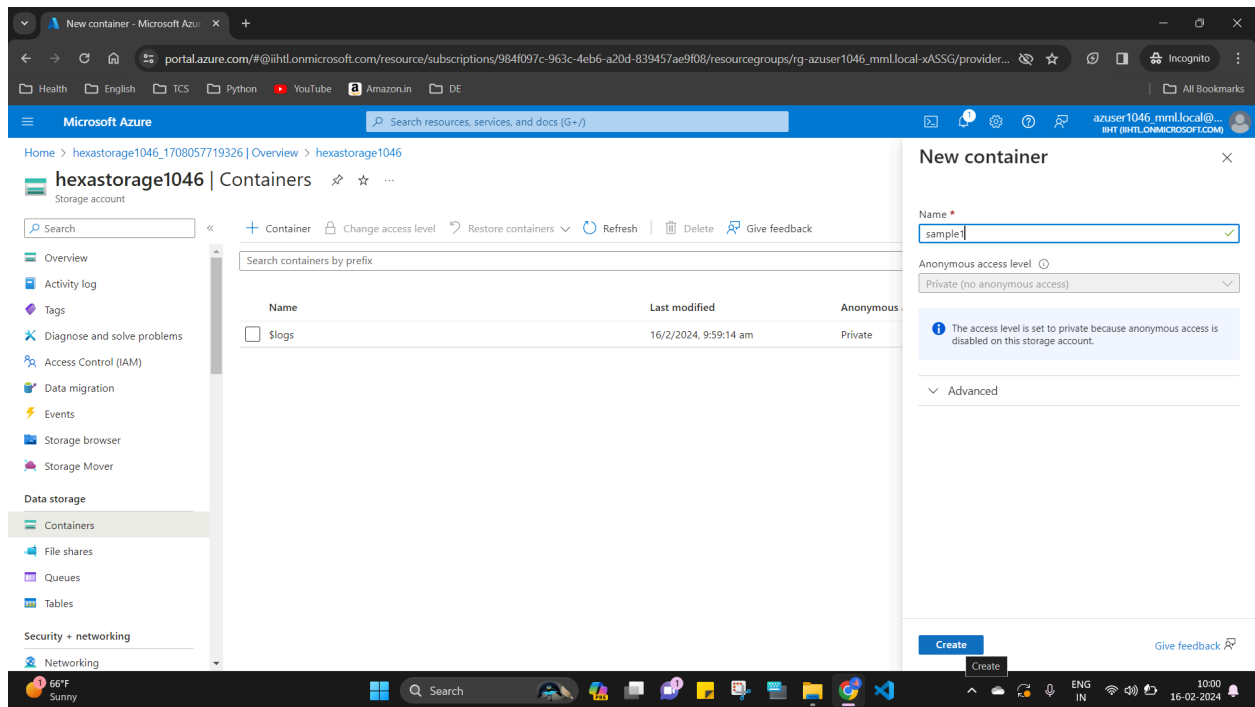
Require secure transfer for REST API operations : Enabled

Storage account key access : Enabled

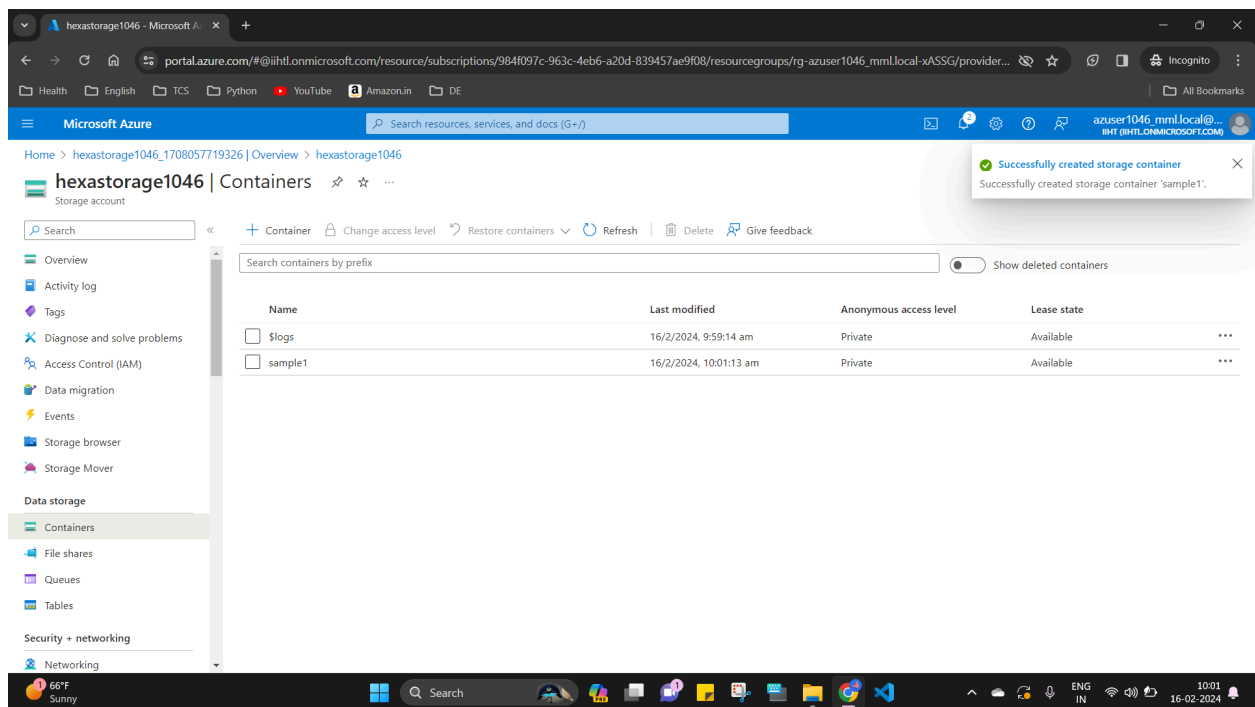
Minimum TLS version : Version 1.2

Infrastructure encryption : Disabled

Created container in blob storage account :



The screenshot shows the Microsoft Azure portal interface. On the left, the navigation pane is open, showing the 'Containers' section under 'Data storage'. The main area displays the 'hexastorage1046' storage account. A 'New container' dialog box is open on the right. The dialog has a 'Name' field with the value 'sample1', an 'Anonymous access level' dropdown set to 'Private (no anonymous access)', and an 'Advanced' section. A 'Create' button is at the bottom of the dialog. A notification at the bottom of the dialog states: 'The access level is set to private because anonymous access is disabled on this storage account.'



The screenshot shows the Microsoft Azure portal interface. On the left, the navigation pane is open, showing the 'Containers' section under 'Data storage'. The main area displays the 'hexastorage1046' storage account. A table lists the containers. A success message is displayed at the top right of the main area.

Name	Last modified	Anonymous access level	Lease state
\$logs	16/2/2024, 9:59:14 am	Private	Available
sample1	16/2/2024, 10:01:13 am	Private	Available

Uploaded file in the container :

sample1
Container

Search

«

Upload

Change access level

Refresh

Delete

Change tier

Acquire lease

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: sample1

Search blobs by prefix (case-sensitive)

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
No results						

Upload blob

1 file(s) selected: DE(Day-16)_Azure_Databricks_Assignment_1_Mitushi Vi...

Drag and drop files here or [Browse for files](#)

☐ Overwrite if files already exist

Advanced

Upload

[Give feedback](#)

sample1 - Microsoft Azure

portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2Fhexastorage1046_1708057719326/Containers/hexastorage1046

Microsoft Azure

Search resources, services, and docs (G+)

azuser1046_mml.local@...
BHT (BHTLONMICROSOFT.COM)

Home > hexastorage1046_1708057719326 | Overview > hexastorage1046 | Containers >

sample1
Container

Search

«

Upload

Change access level

Refresh

Delete

Change tier

Acquire lease

Break lease

View snapshots

Create snapshot

[Give feedback](#)

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: sample1

Search blobs by prefix (case-sensitive)

Show deleted blobs

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> DE(Day-16)_Azure_Databricks_Assignment_1_Mitush...	16/2/2024, 10:02:14 ...	Hot (Inferred)		Block blob	2.76 MiB	Available

Successfully uploaded blob(s)

Successfully uploaded 1 blob(s).

66°F
Sunny

Search

ENG
IN

10:01
16-02-2024

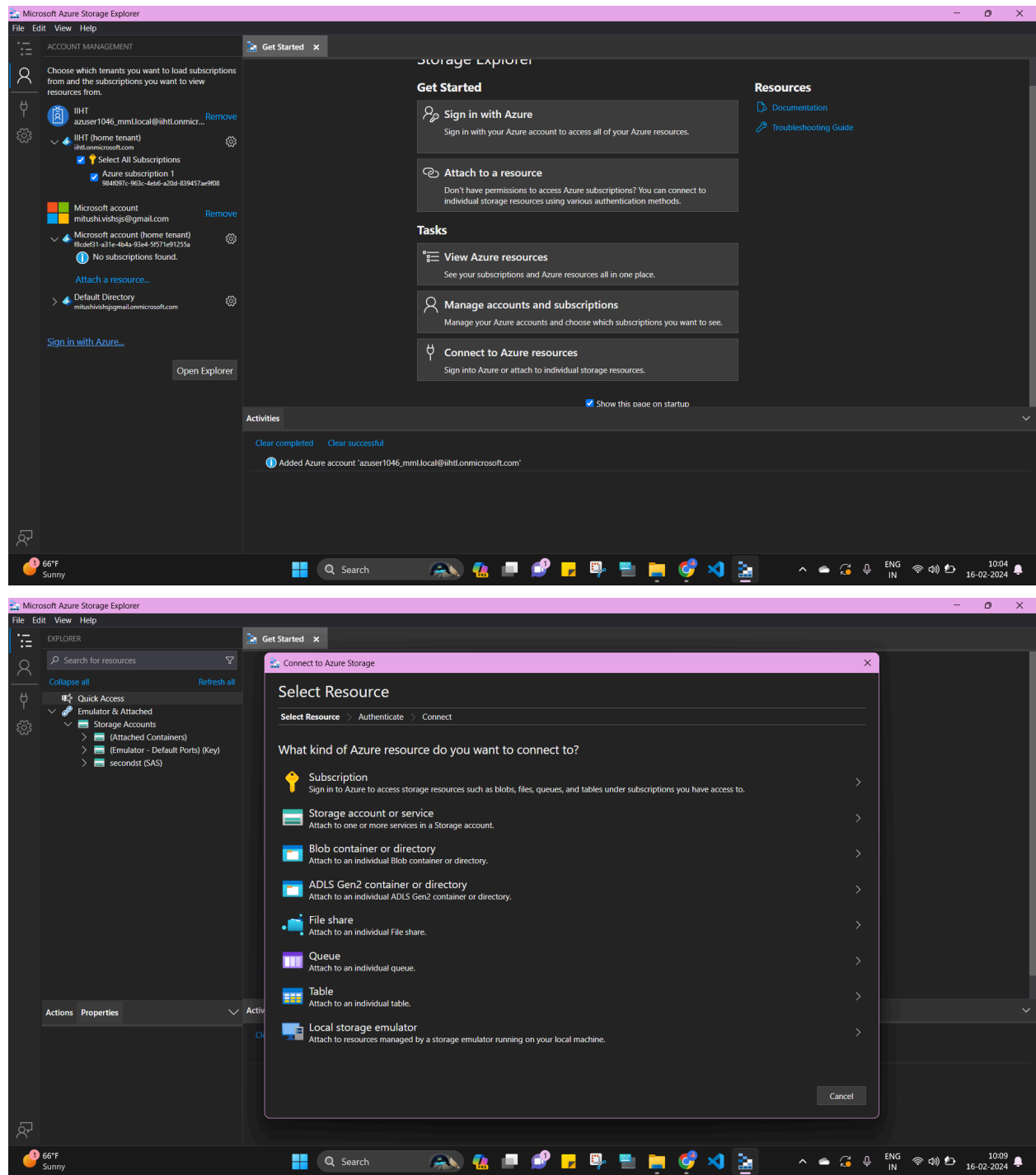
66°F
Sunny

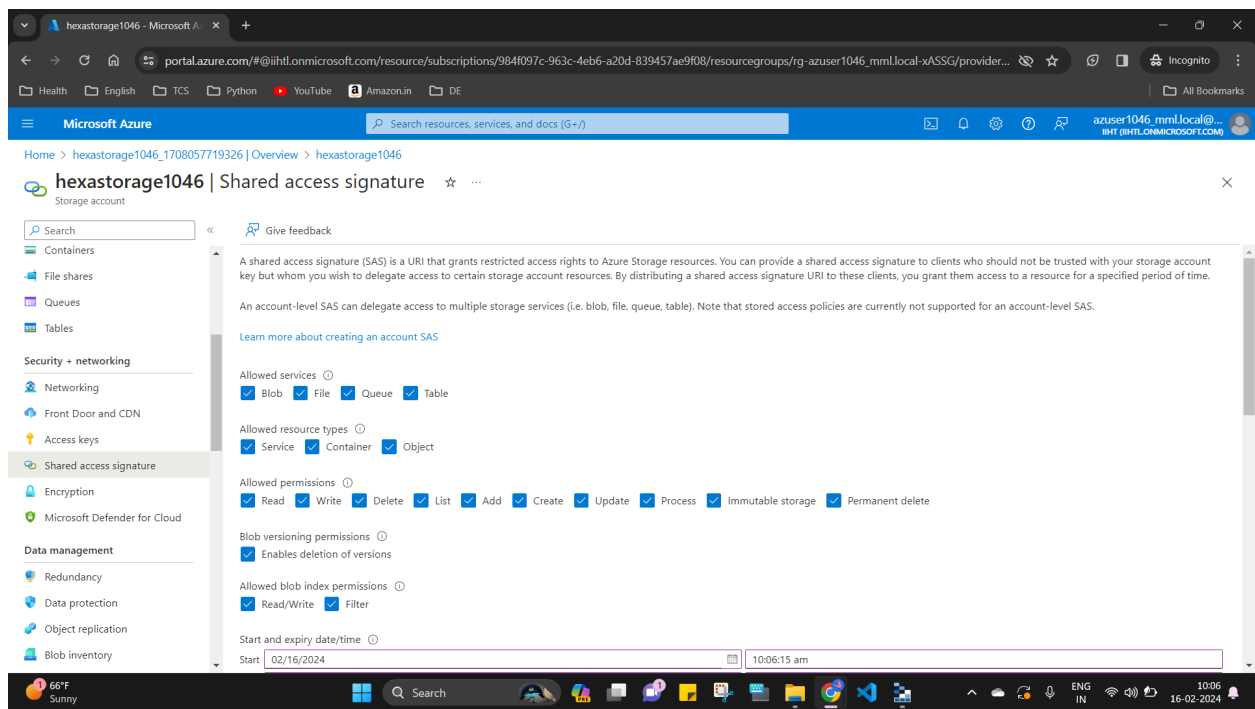
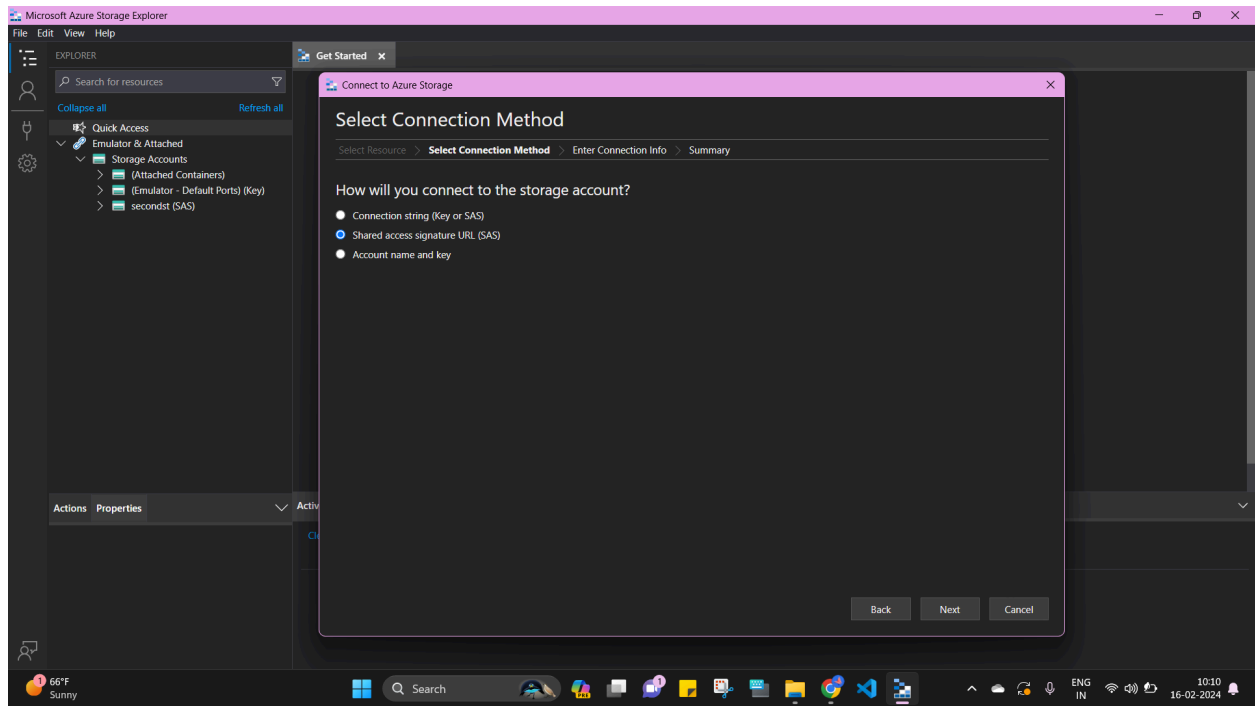
Search

ENG
IN

10:02
16-02-2024

Attaching Storage account into storage explorer in local :





hexastorage1046 - Microsoft A x +

portal.azure.com/#@ihlt.onmicrosoft.com/resource/subscriptions/984f097c-963c-4eb6-a20d-839457ae9f08/resourcegroups/rg-azuser1046_mmlLocal-xASSG/provider...

Health English TCS Python YouTube Amazon.in DE azuser1046_mmlLocal@... WHT (WHT.ONMICROSOFT.COM)

Microsoft Azure Search resources, services, and docs (G+)

Home > hexastorage1046_1708057719326 | Overview > hexastorage1046

hexastorage1046 | Shared access signature ☆ ...

Storage account

Search Containers File shares Queues Tables Security & networking Networking Front Door and CDN Access keys Shared access signature Encryption Microsoft Defender for Cloud Data management Redundancy Data protection Object replication Blob inventory

Give feedback key1

Generate SAS and connection string

Connection string
BlobEndpoint=https://hexastorage1046.blob.core.windows.net/QueueEndpoint=https://hexastorage1046.queue.core.windows.net/FileEndpoint=https://hexastorage1046.file.core.windows.net/TableEndpo...

SAS token
?sv=2022-11-02&ss=bfqt&srt=sco&sp=rwdlacupiytf&se=2024-02-16T12:36:15Z&st=2024-02-16T04:36:15Z&sig=QkdG%2FleZMgU74Hi7SEsedywXVlp06sNIPCywgrRRk%3D

Blob service SAS URL
https://hexastorage1046.blob.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=sco&sp=rwdlacupiytf&se=2024-02-16T12:36:15Z&st=2024-02-16T04:36:15Z&sig=QkdG%2FleZMgU74Hi7SELS... Copied

File service SAS URL
https://hexastorage1046.file.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=sco&sp=rwdlacupiytf&se=2024-02-16T12:36:15Z&st=2024-02-16T04:36:15Z&sig=QkdG%2FleZMgU74Hi7SELS...

Queue service SAS URL
https://hexastorage1046.queue.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=sco&sp=rwdlacupiytf&se=2024-02-16T12:36:15Z&st=2024-02-16T04:36:15Z&sig=QkdG%2FleZMgU74Hi7SELS...

Table service SAS URL
https://hexastorage1046.table.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=sco&sp=rwdlacupiytf&se=2024-02-16T12:36:15Z&st=2024-02-16T04:36:15Z&sig=QkdG%2FleZMgU74Hi7SELS...

66°F Sunny 10:10 16-02-2024

Microsoft Azure Storage Explorer

File Edit View Help

EXPLORER Search for resources Collapse all Refresh all Quick Access Emulator & Attached Storage Accounts (Attached Containers) (Emulator - Default Ports) (Key) secondst (SAS)

Get Started x

Connect to Azure Storage

Enter Connection Info

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

Display name:
hexastorage1046

Service URL:
https://hexastorage1046.blob.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=sco&sp=rwdlacupiytf&se=2024-02-16T12:36:15Z&st=2024-02-16T04:36:15Z&sig=QkdG%2FleZMgU74Hi7SEsedywXVlp06sNIPCywgrRRk%3D

Back Next Cancel

66°F Sunny 10:10 16-02-2024

