

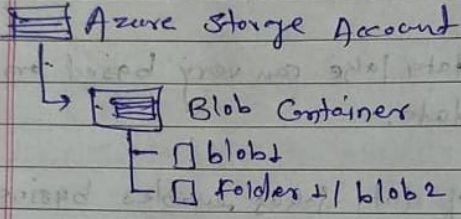
Name: Pradip Bochare



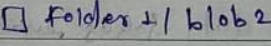
+ Notes

Page No.
 Date

* Azure Data Lake Storage (ADLS)

- In Azure Blob Storage you can ~~can~~ store large amounts of unstructured data in flat namespace within a blob container
- In terms of blob manageability, blobs are stored as a single-level hierarchy in flat namespace.

 Azure Storage Account

- ↳  Blob Container
 - ↳  blob1
 - ↳  folders + blob2

- You can access this data by using HTTP or HTTPS
- ADLS Gen2 builds on blob storage & optimizes it for high volume data by using hierarchical namespace that organizes blob data into directories & stores metadata about each directory & the files within it.
- Hierarchical namespaces keep data organized which yields better storage and retrieval performance for analytical use case & lowers cost of analysis.
- Blob storage can be also used to archive rarely used data or to store website assets such as images and media.

* Stages for processing big data

Data lakes have fundamental role in a wide range of big data architectures. These architecture can involve creation of:

- enterprise data warehouse
- Advanced analytics against big data
- Real-time analytical solution.

* Four stages for processing big data.

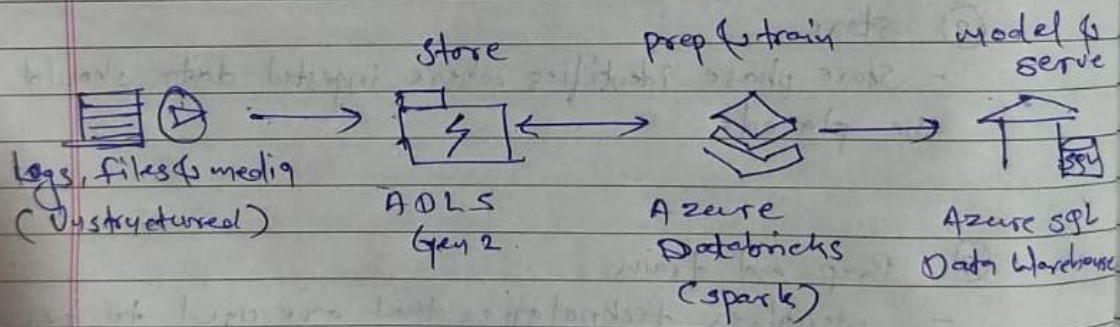
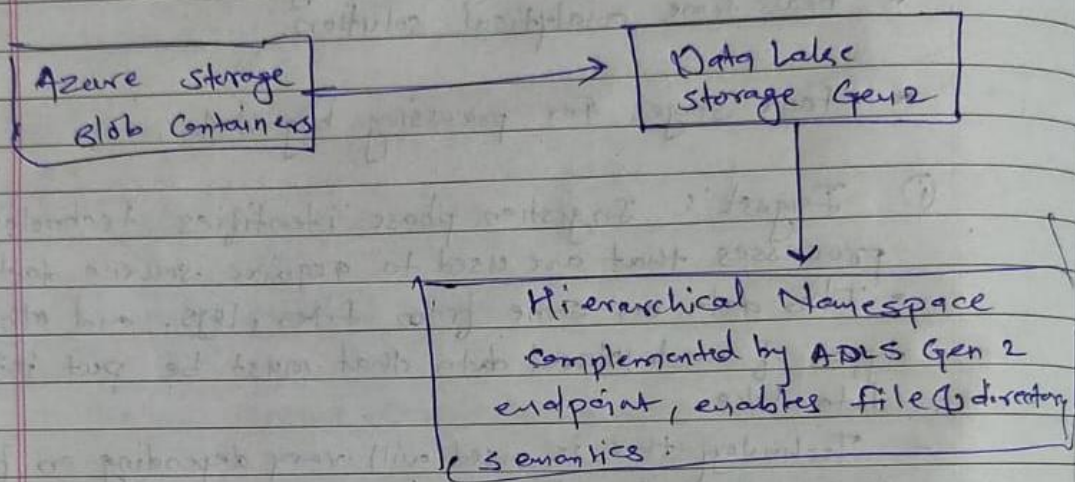
- ① Ingest: Ingestion phase identifies technology and processes that are used to acquire source data.
 - This data can come from files, logs, and other types of unstructured data that must be put into data lake
 - Technology that is used will vary depending on frequency that data is transferred.

- ② Store
 - Store phase identifies where ingested data should be placed.

- ③ Prep and train.
 - identifies technologies that are used to perform data preparation and model training and scoring for machine learning solutions

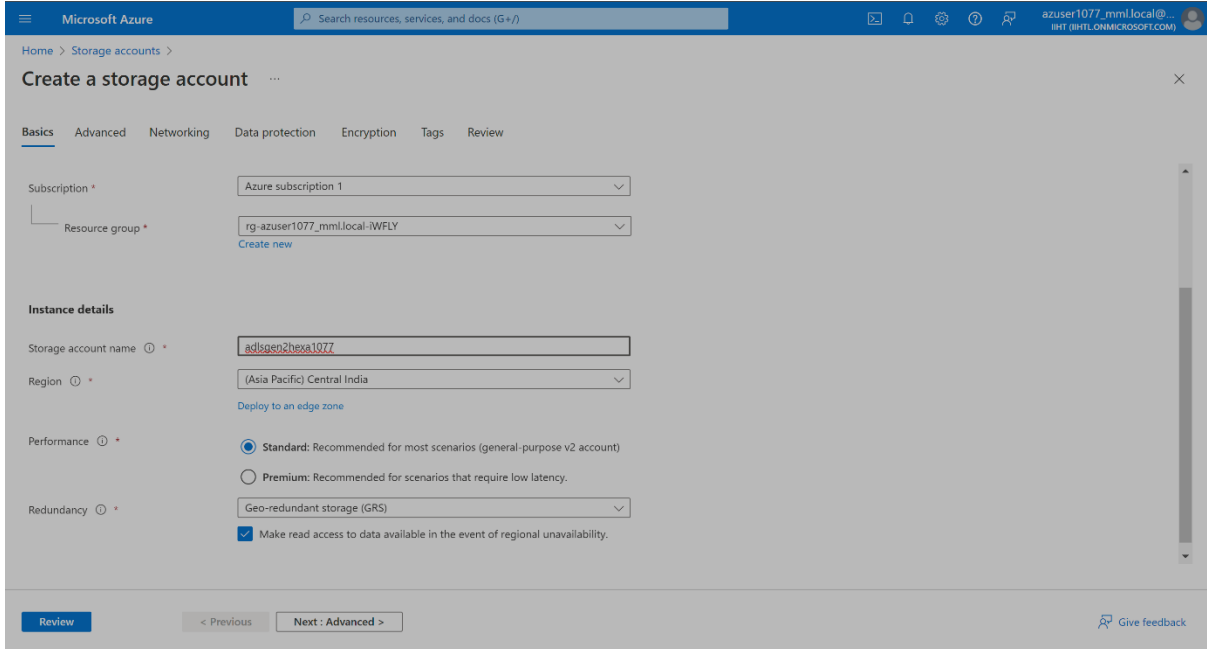
④ Model and serve

- This phase involves technologies that will present data to users.
- This technologies can include visualization tools such as Power BI, or analytical data stores such as Azure Synapse Analytics.



Making Azure Data Lake Storage Gen2

- Selecting Resource group and other parameters for ADLS

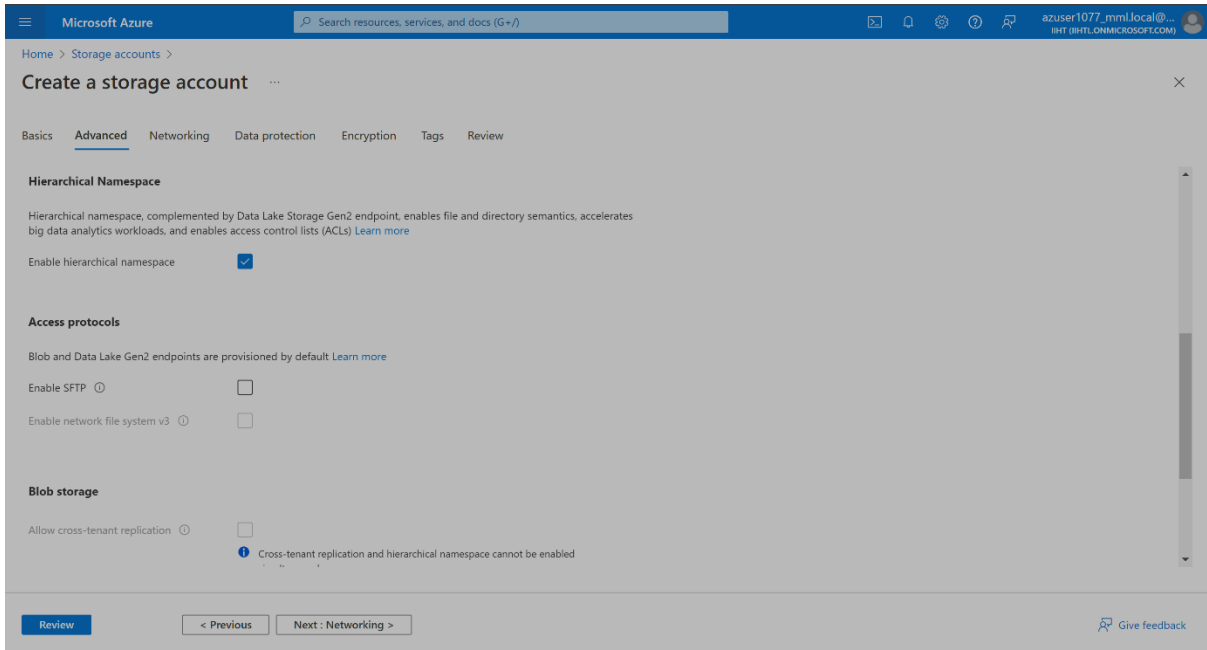


The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The 'Basics' tab is selected, and the following configuration is shown:

- Subscription:** Azure subscription 1
- Resource group:** rg-azuser1077_mml.local-IWFLY (with a 'Create new' link below)
- Instance details:**
 - Storage account name:** adlsgen2hexa1077
 - Region:** (Asia Pacific) Central India (with a 'Deploy to an edge zone' link below)
 - Performance:** Standard: Recommended for most scenarios (general-purpose v2 account) (selected)
 - Redundancy:** Geo-redundant storage (GRS) (selected)
 - ☒ Make read access to data available in the event of regional unavailability.

Navigation buttons at the bottom include 'Review', '< Previous', 'Next: Advanced >', and 'Give feedback'.

- Enabling Hierarchical Namespace

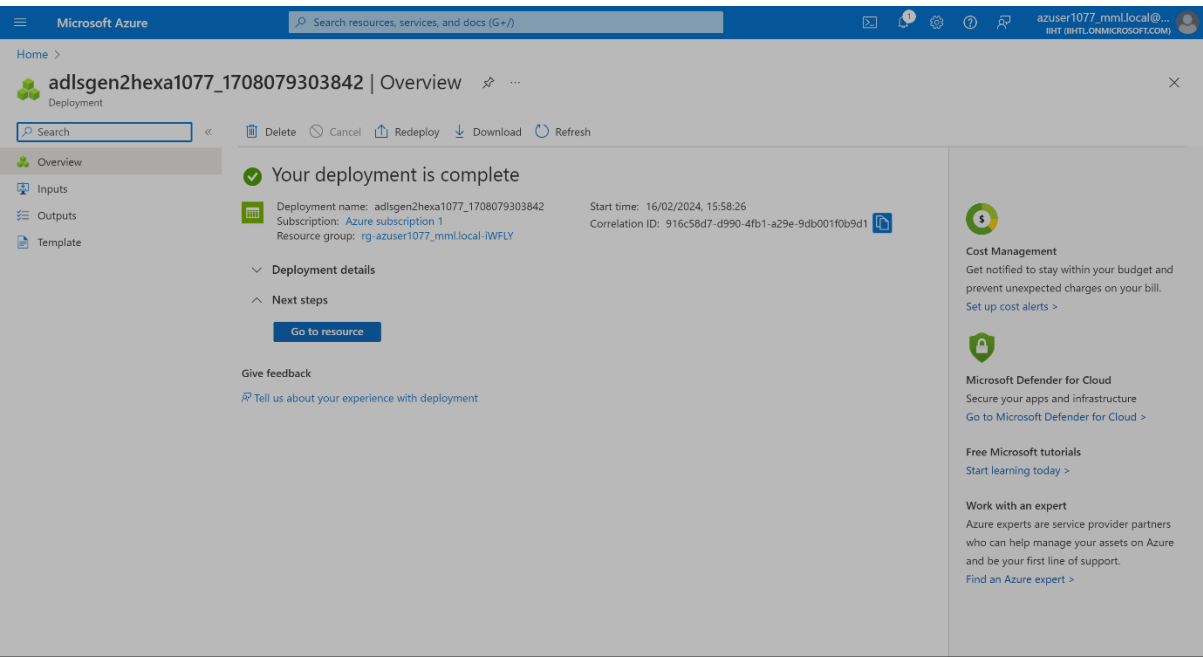


The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal, with the 'Advanced' tab selected. The 'Hierarchical Namespace' section is expanded, showing the following configuration:

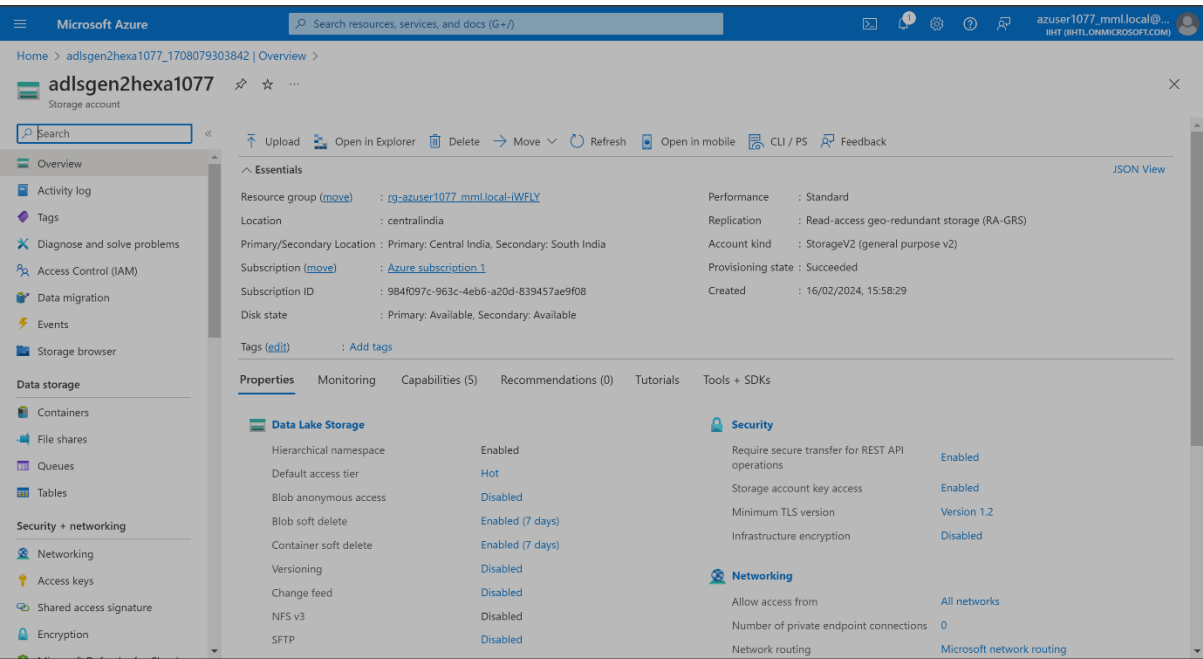
- Hierarchical Namespace:** Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs). [Learn more](#)
- Enable hierarchical namespace:** ☒
- Access protocols:** Blob and Data Lake Gen2 endpoints are provisioned by default. [Learn more](#)
 - Enable SFTP:** ☐
 - Enable network file system v3:** ☐
- Blob storage:**
 - Allow cross-tenant replication:** ☐ (with a warning icon and text: 'Cross-tenant replication and hierarchical namespace cannot be enabled')

Navigation buttons at the bottom include 'Review', '< Previous', 'Next: Networking >', and 'Give feedback'.

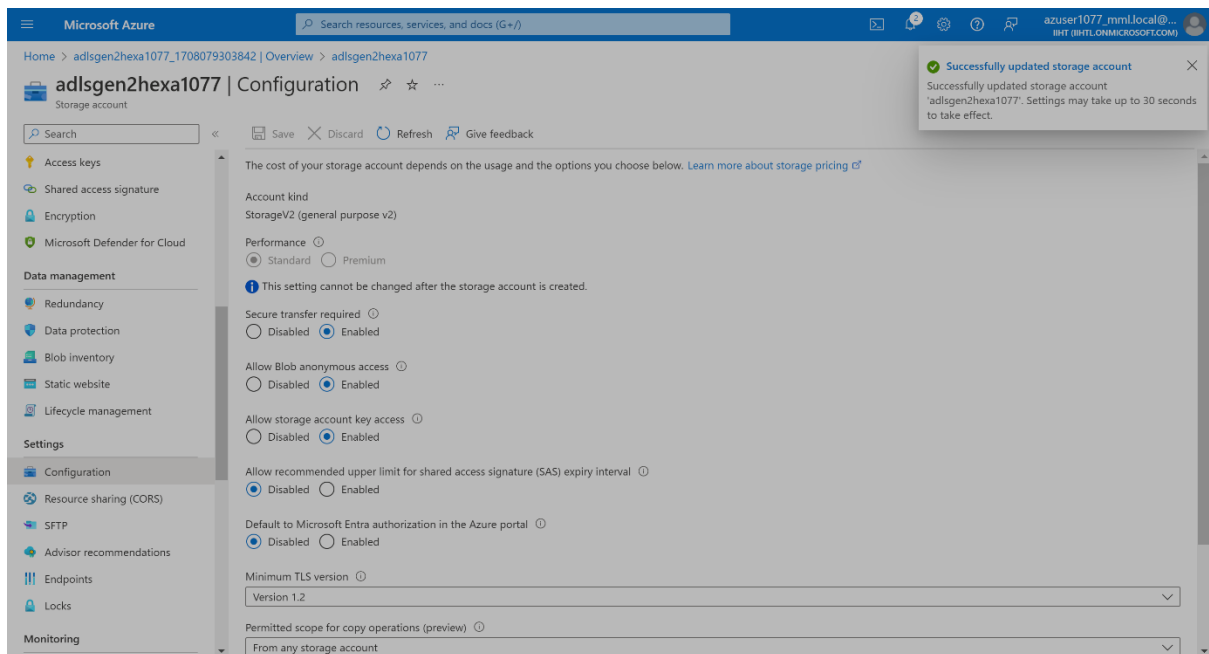
- ADLS is deployed



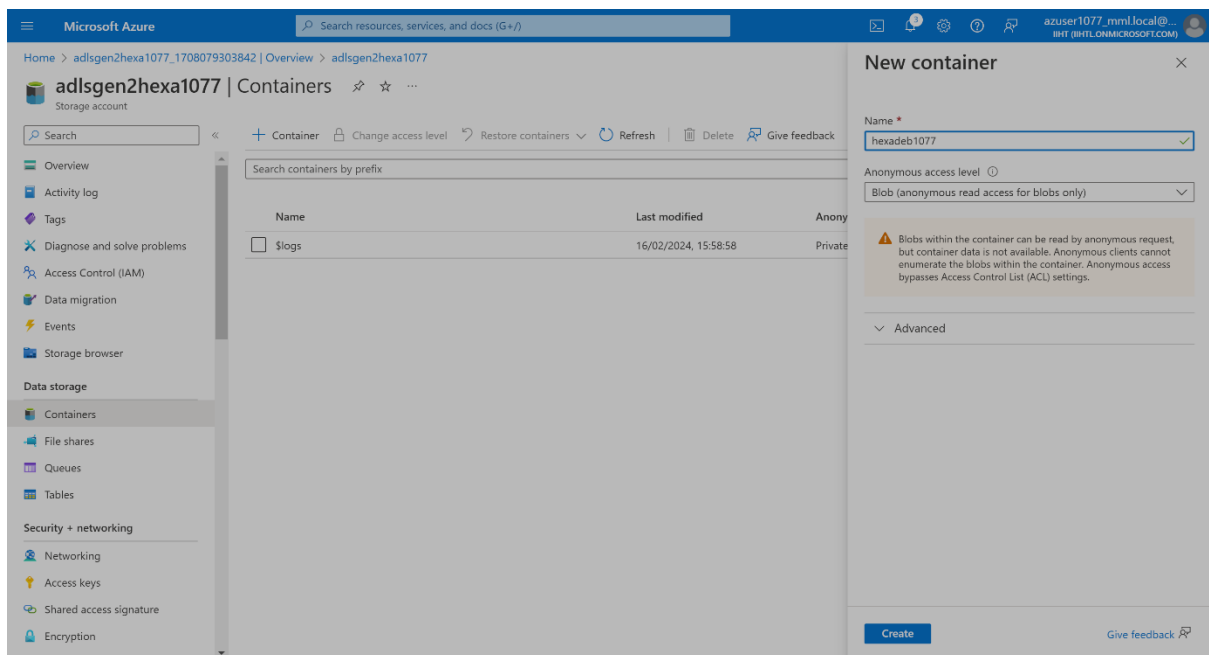
- Overview of ADLS



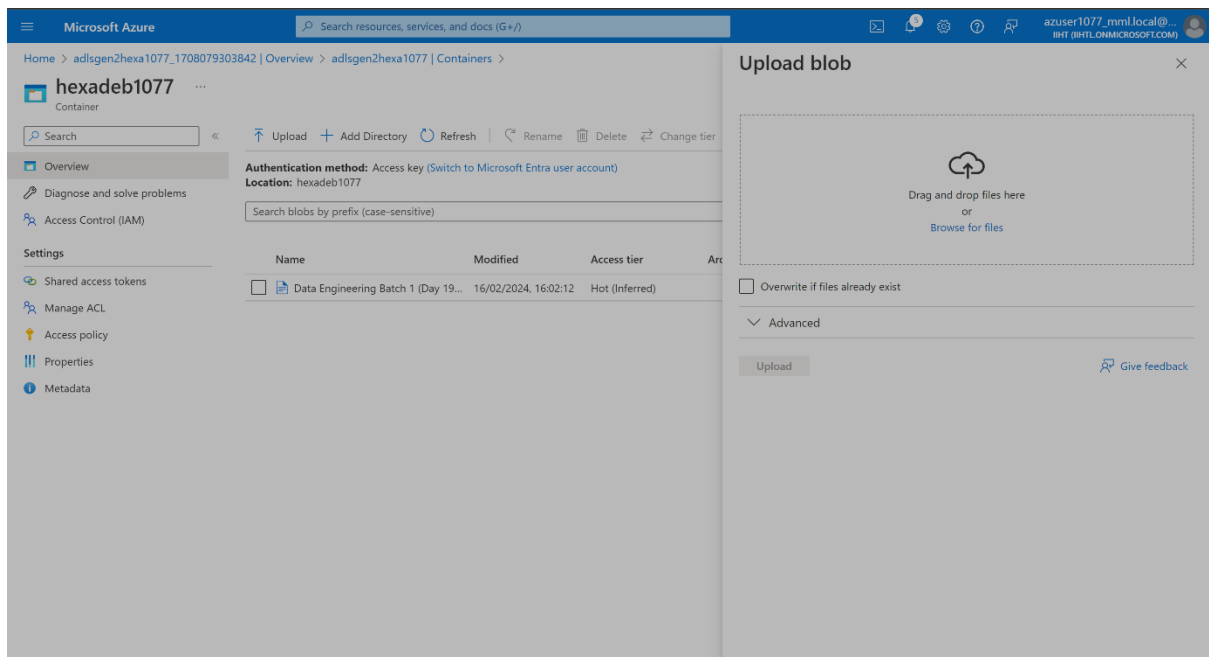
- Changing Configuration to allow Blob access



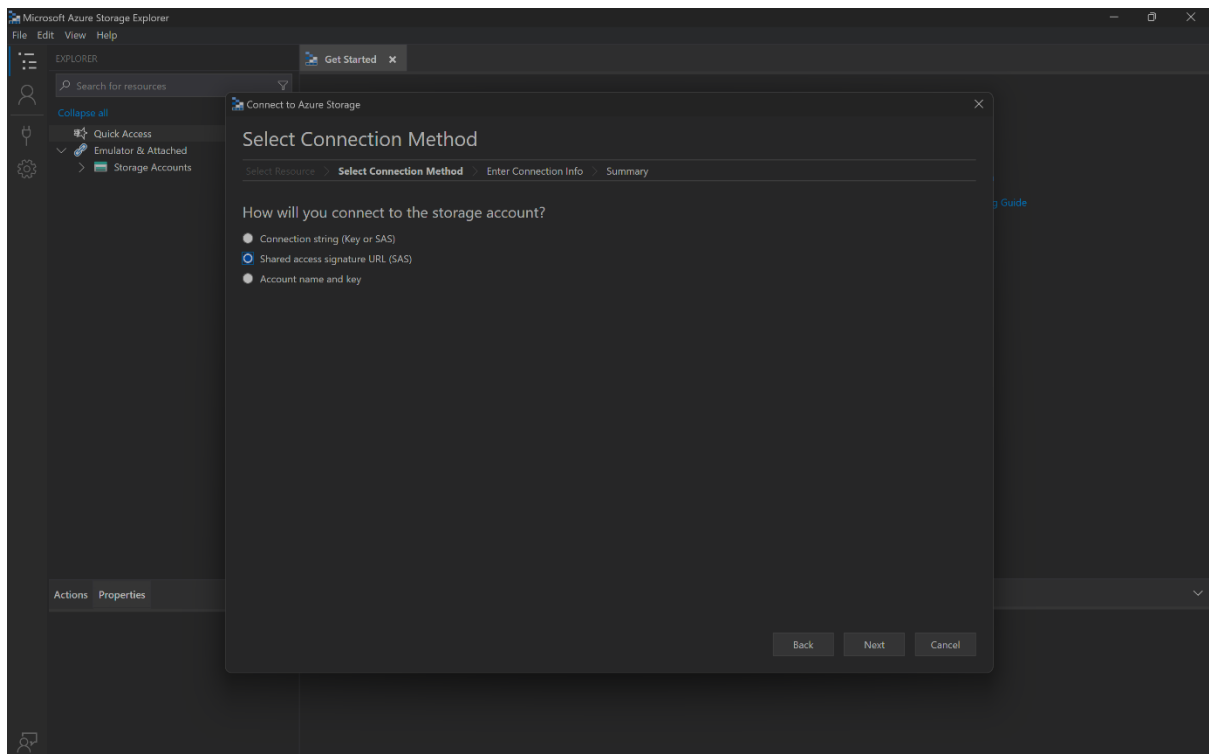
- Creating Container in ADLS



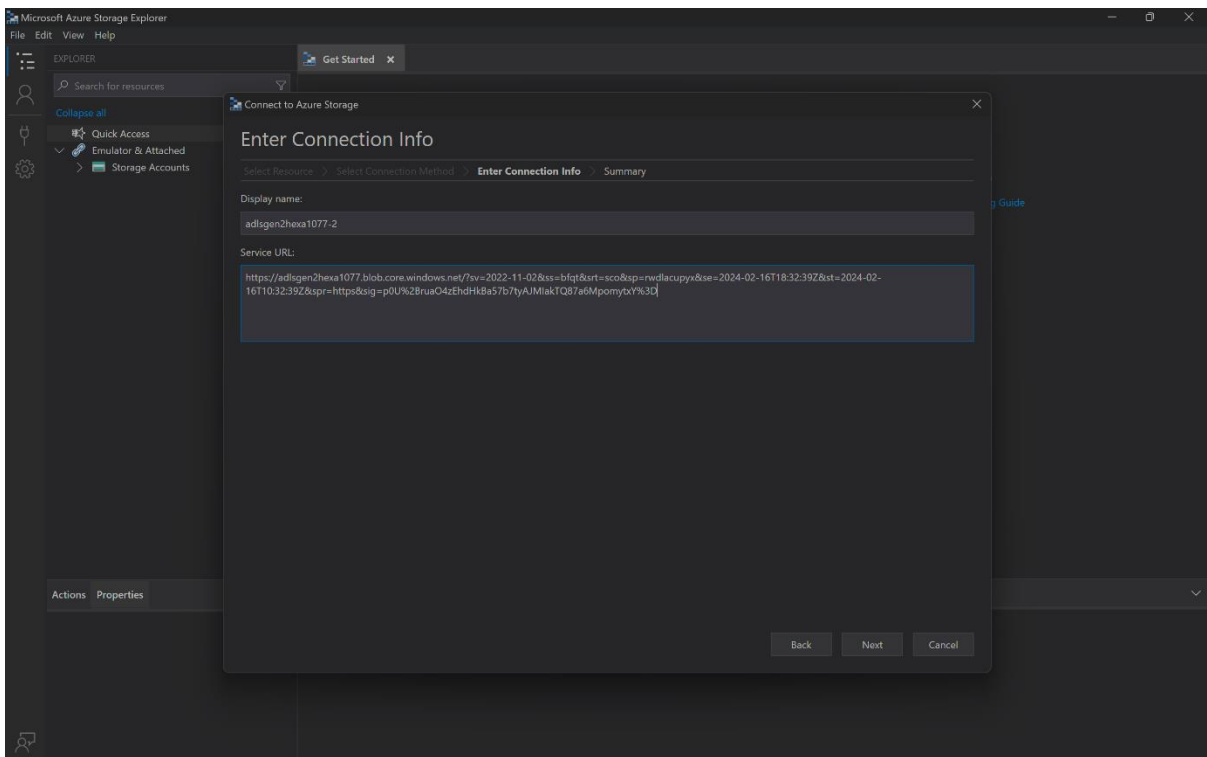
- Uploading files in ADLS



- Connecting ADLS to Microsoft Azure Storage Explorer



- Giving Service URL to connect



- ADLS is connected

