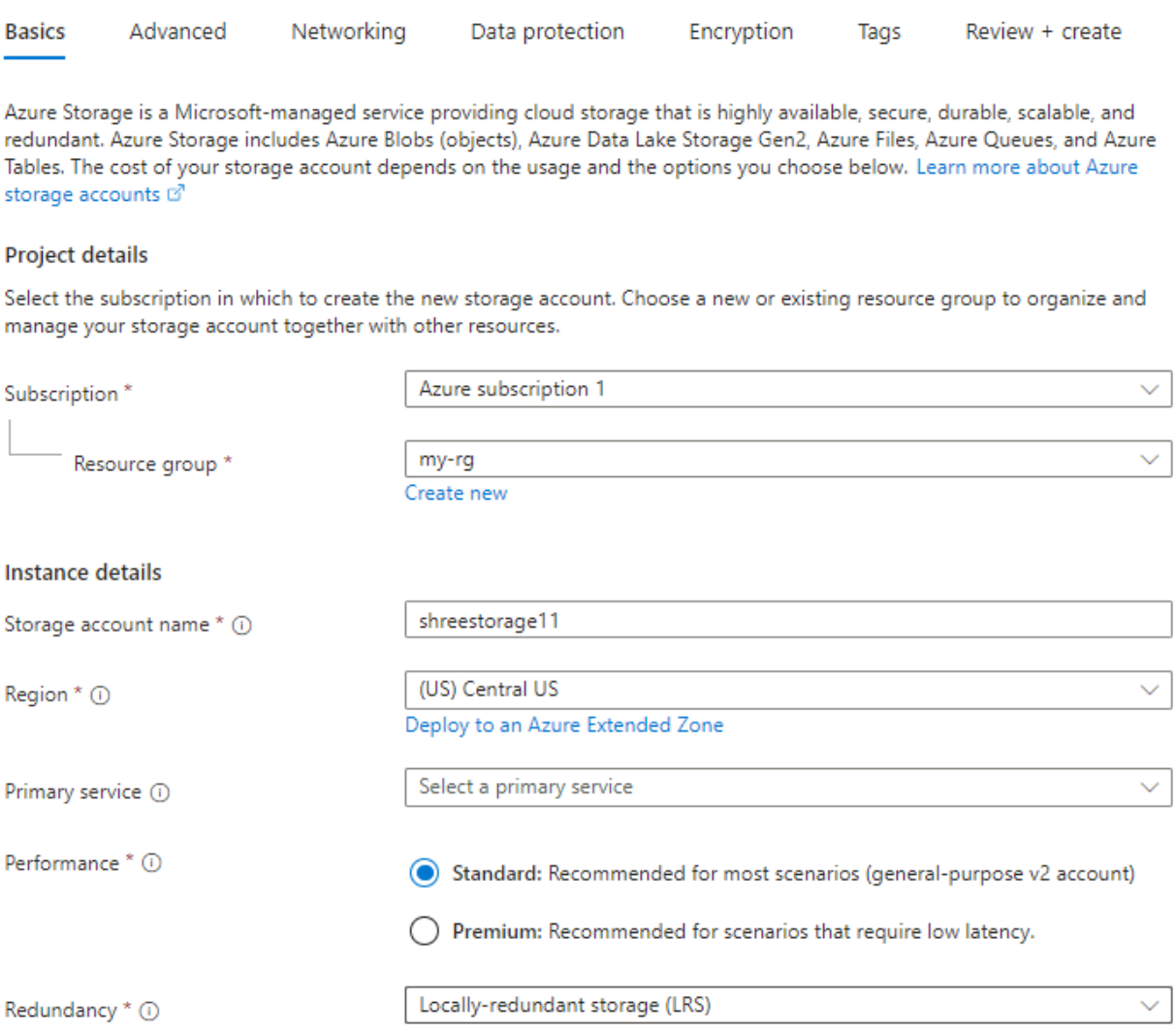
Data Organization and Storage with Azure

1. **Set Up Azure Storage Account**
    - **Create an Azure Storage Account if you don't already have one.**
    - **Choose between Azure Blob Storage or Azure Data Lake Storage Gen2 based on your preference.**

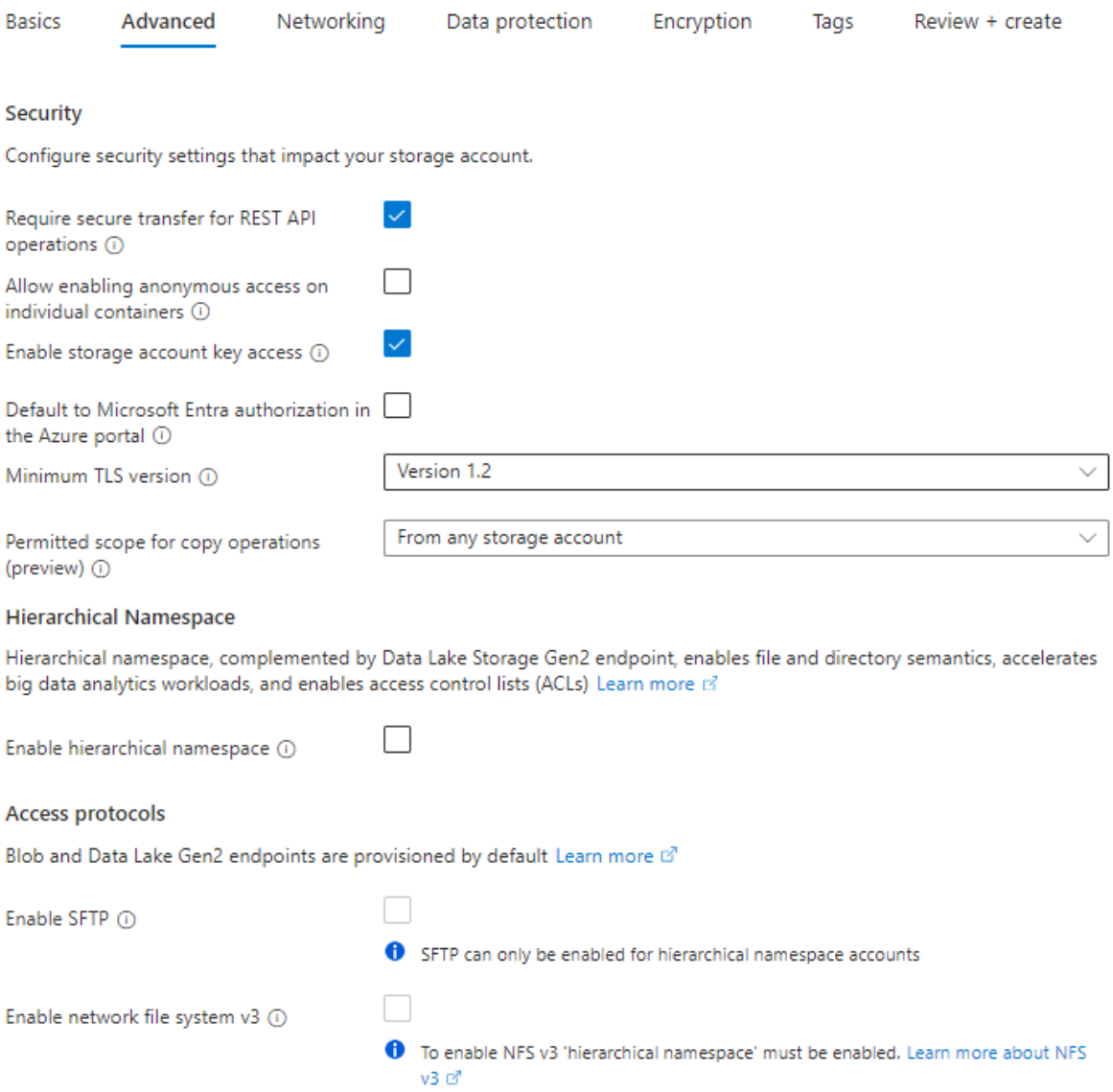
Step 1: Open Azure portal, Search Bar : Storage accounts, Open storage account
Step 2: Click on "Create" to create a storage account.
Step 3: Follow the below steps what needs to be given in the respective fields:

Basics	
Subscription :	Default (Azure Subscription 1)
Resource group :	my-rg [If already created resource group mention the name of the resource, if not, you can click on "create new" to create a new one and give the name for resource group - "my-rg" (can give any name]
Storage account name :	shreestorage2 (only contains lower case letters and numbers).
Region :	(Canada) Canada Central
Primary Service :	Azure Blob Storage or Azure Data Lake Storage Gen 2
Performance :	Standard: Recommended for most scenarios (general-purpose V2 accounts)
Redundancy :	Locally-redundant storage (LRS)

- Follow the below screenshot for reference.

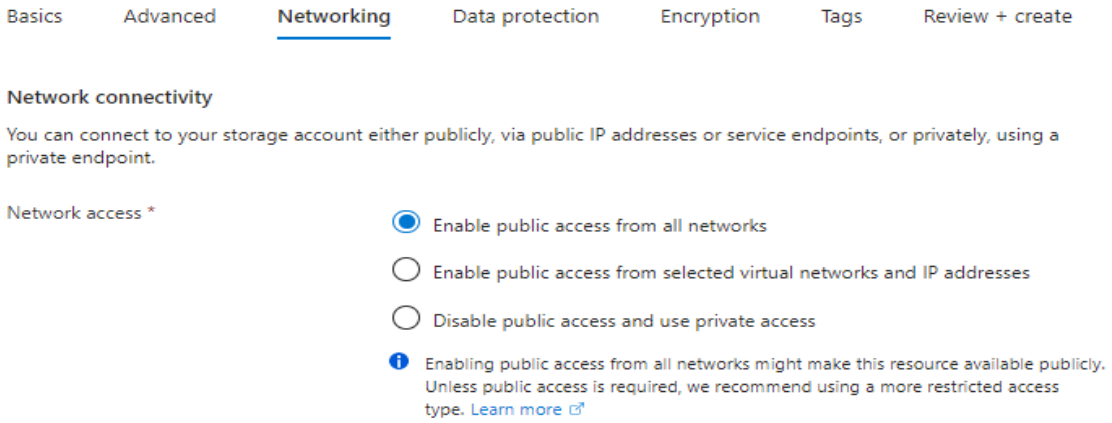


*** Below screenshot represent “Advanced” options:**

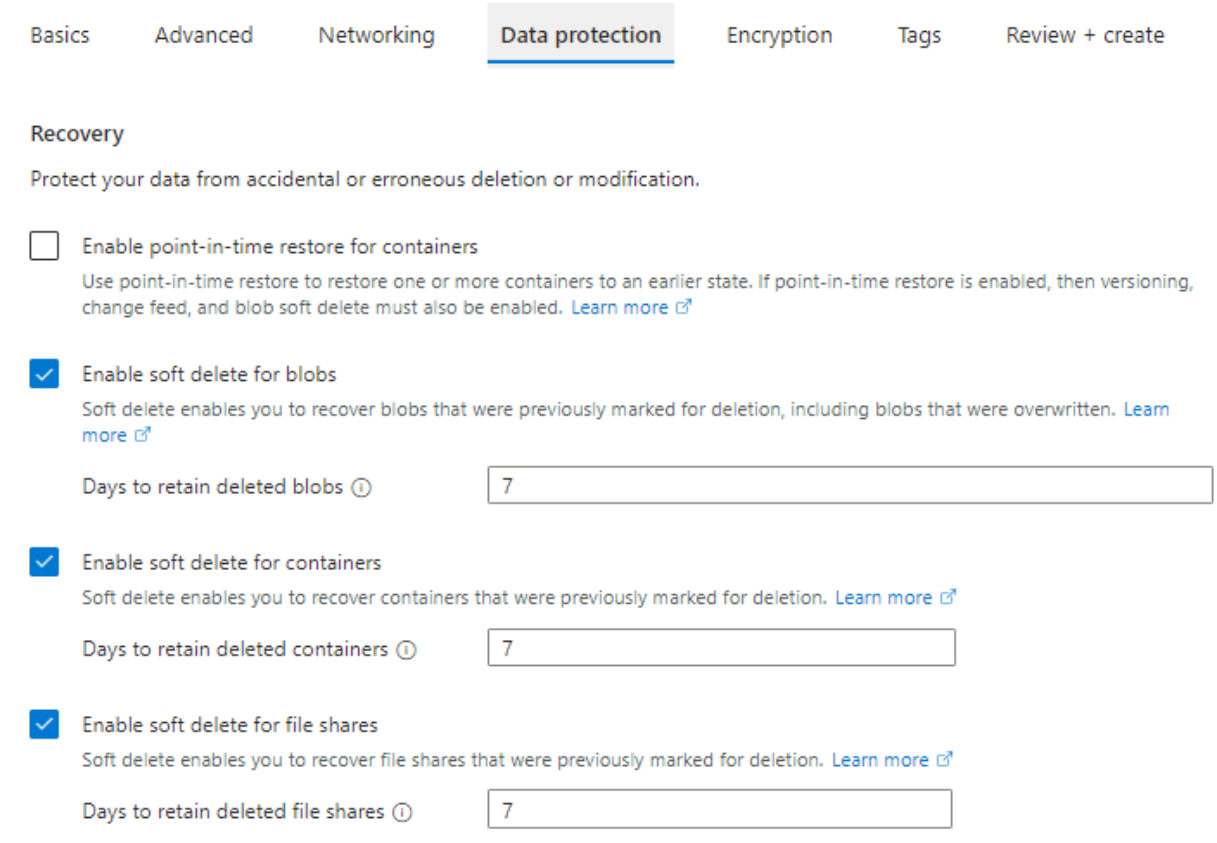


- For Networking, Data Protection, Encryption, Tags – Keeping the Options and fields as “Default”.

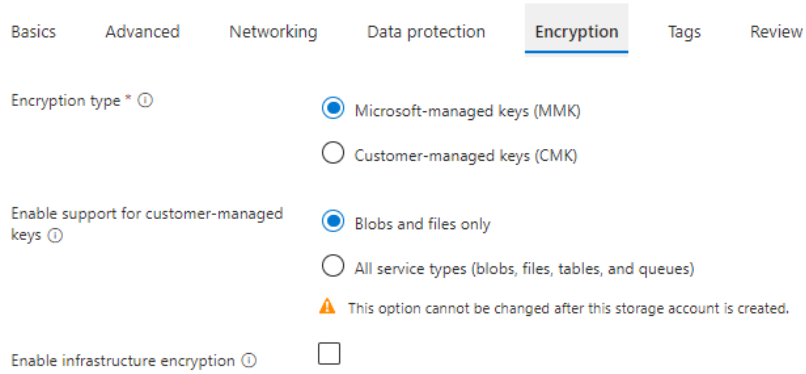
*** Below screenshot represent “Networking” options:**



*** Below screenshot represent “Data Protection” options:**

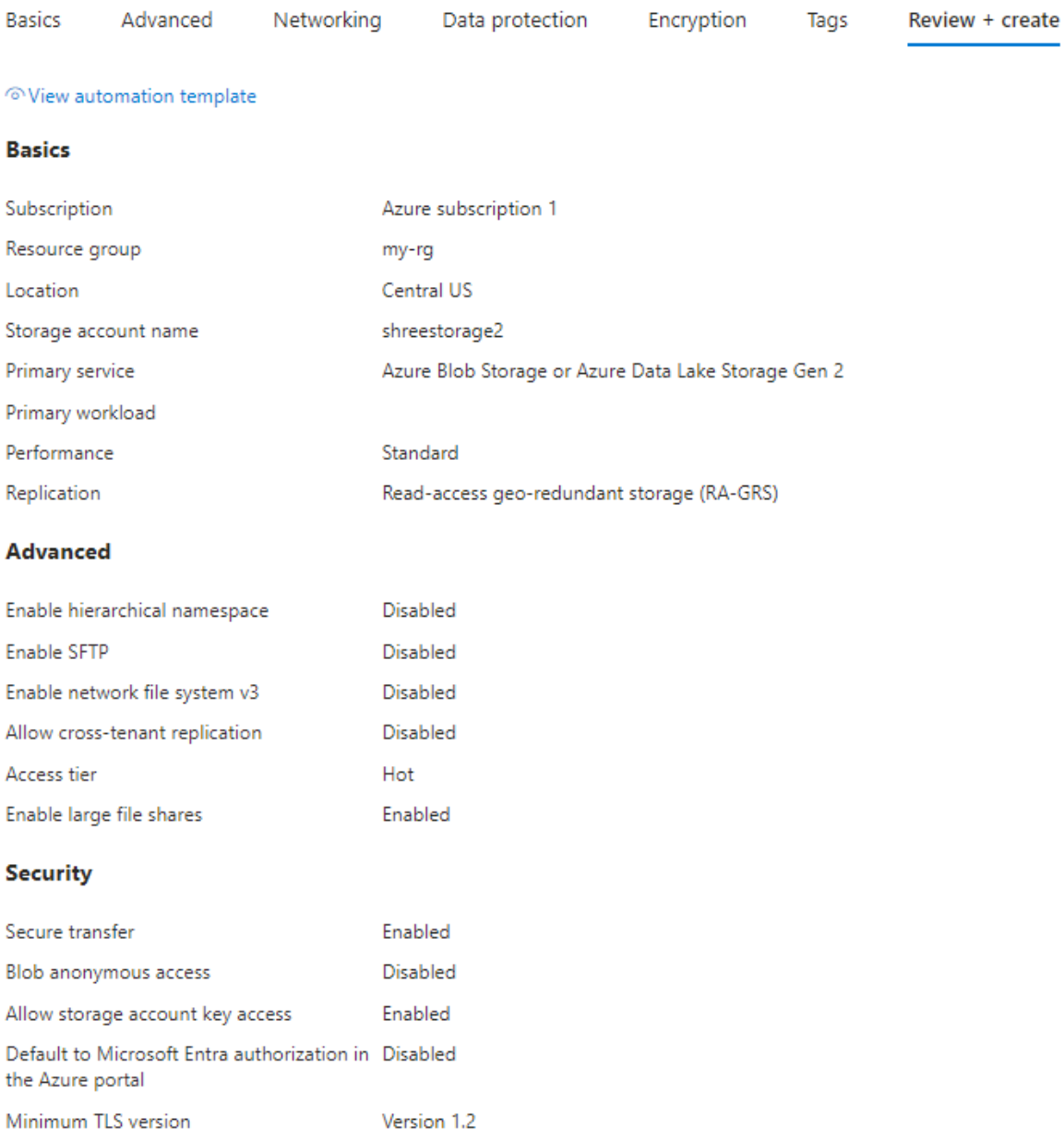


*** Below screenshot represent “Encryption” options:**



*** Below screenshot represent Tags options:**

*** Below screenshot represent “Review + Create” options:**



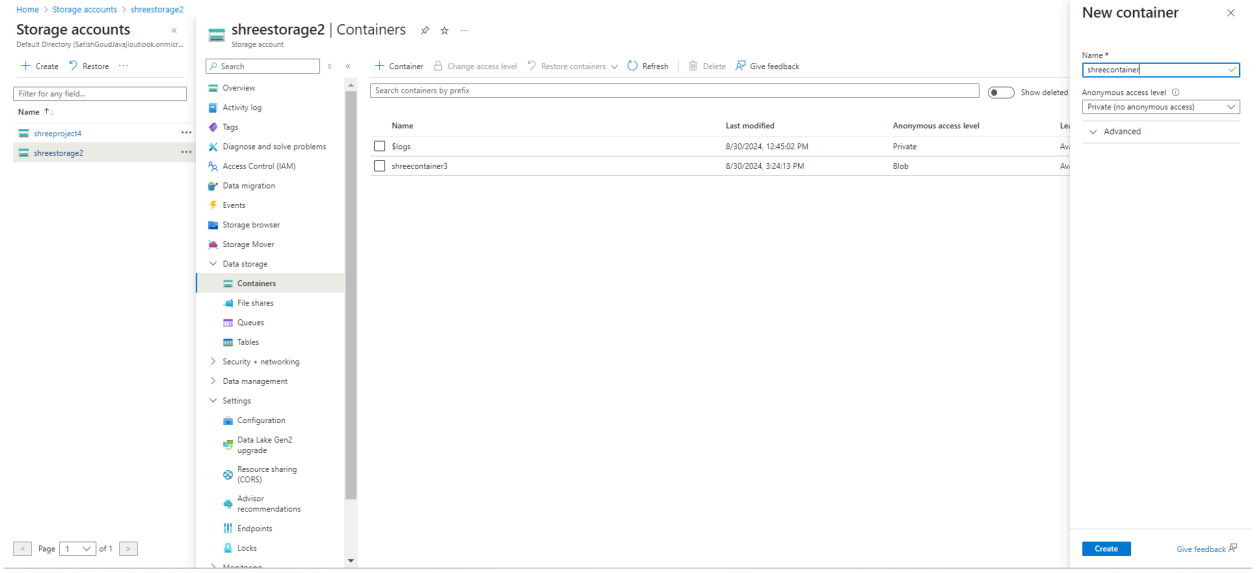
Step 4 : Click on create.
Step 5 : It will show as "**Your deployment is complete**"

2. **Create Containers/Folders**
    - **Create a container (for Blob Storage) or a file system (for Data Lake Storage Gen2).**
    - **Within the container/file system, create folders to organize your data. For example:**
    - `raw-data/`
    - `processed-data/`
    - `metadata/`
3. **Upload Sample Data Sets**
    - **Download the following sample data sets:**

- `sample_data.json` **(JSON format)**
- `sample_data.tsv` **(TSV format)**

- **Upload these files to the appropriate folders in your Azure Storage.**

Step 1 : Open a storage account, the one which is created.
Step 2 : Go to Option “Data Storage”
Step 3: Select “Containers”.
Step 4: Go to New Container - Name field - “shreecontainer3”, select create.
Step 5: Follow below screenshot for reference.



Step 6: Create a folder and name the folder (Naming the folder as raw-data for reference).
Step 7: To create a folder - select “Upload”.
Step 8: In “Browse the file” select the file which needs to be uploaded in the folder. (using JSON file).
Step 9: In “Advanced” field - go to “Upload a folder” and give the name of the folder (Giving raw-data for reference).

Upload blob

1 file(s) selected: iris.json

Drag and drop files here or Browse for files

Overwrite if files already exist

Advanced

Blob type

Block blob

Upload .vhd files as page blobs (recommended)

Block size

4 MiB

Access tier

Hot (Inferred)

Upload to folder

Raw-Data1

Blob index tags

Key	Value

Encryption scope

Use existing default container scope

Choose an existing scope

Retention policy

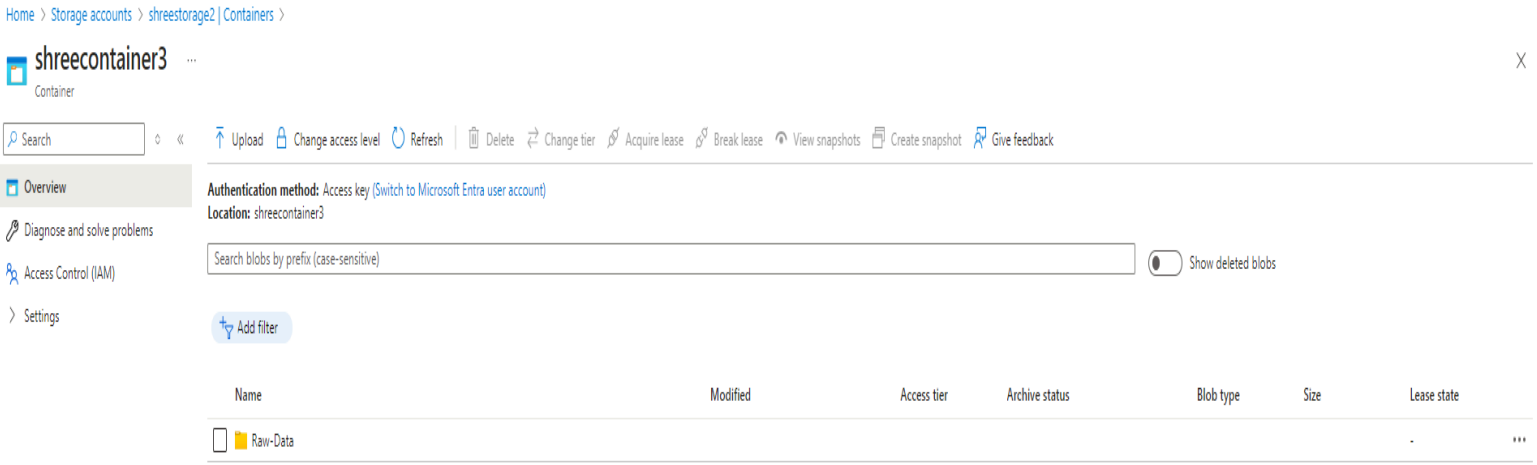
No retention

Choose custom retention period

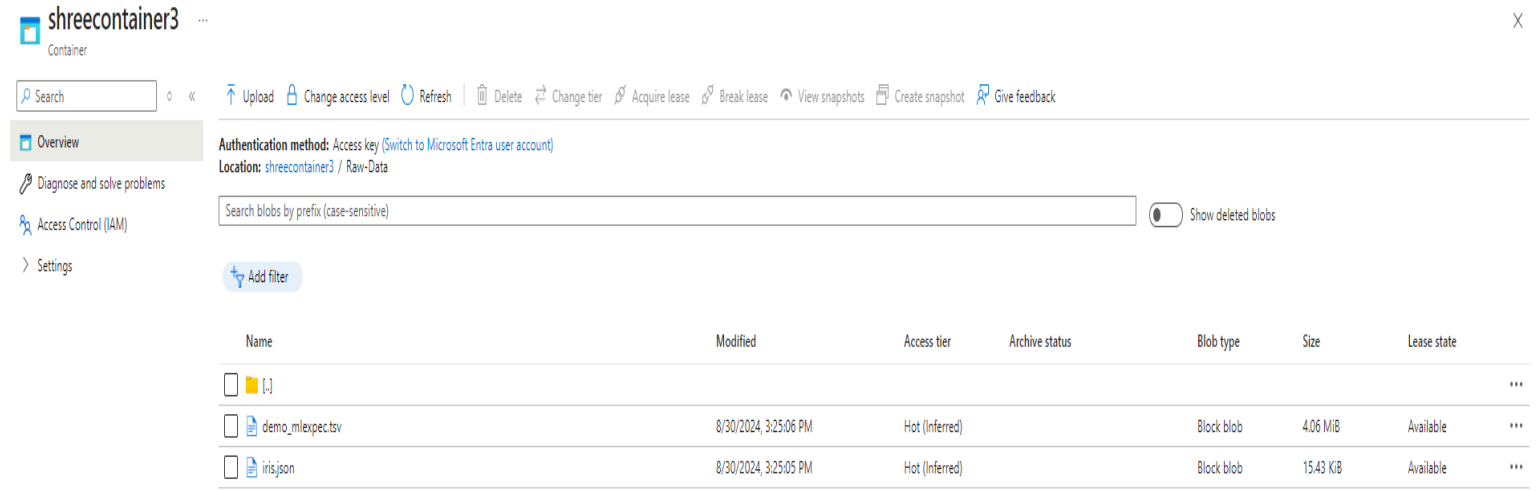
Upload

Give feedback

Step 10: Click on “Upload”. A folder “Raw-Data” will be created.



Step 11: To attach a TSV file to the existing folder “Raw-Data” - Follow step 9, in “Upload a folder” field give the existing folder name “Raw-Data”.
Step 12: A new TSV file is attached to the folder.
Step 13: Open folder “Raw-Data” within the folder both TSV and JSON files are uploaded.
Step 14: Follow screenshot for reference.



4. Organize Data:

- **Ensure that the JSON and TSV files are stored in a structured manner. For example:**
  - `raw-data/sample_data.json`
  - `raw-data/sample_data.tsv`

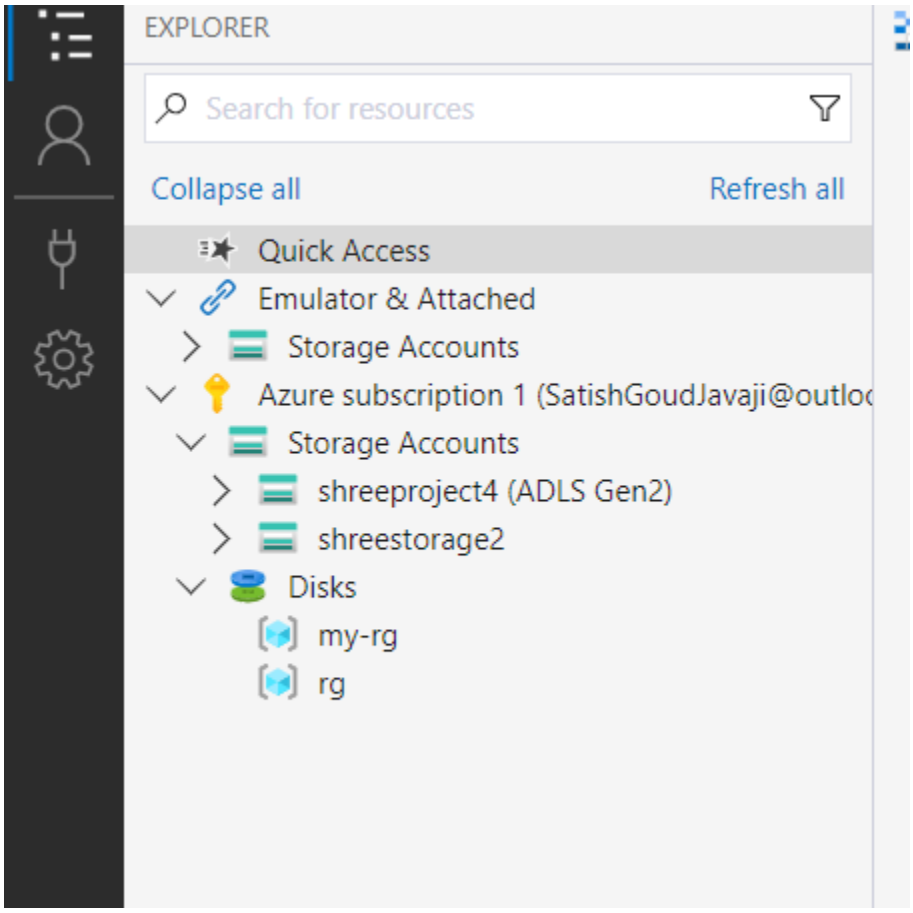
- `processed-data/` **(empty for now)**
- `metadata/` **(empty for now)**

➔ Downloaded JSON and TSV file from kaggle.
➔ JSON and TSV files are stored in a structured manner.

5. Access and Manage Data:

- **Use Azure Storage Explorer or Azure Portal to access and manage your data.**
- **Verify that the data is correctly uploaded and organized.**

Step 1: Open Azure Storage Explorer, you will be able to see the containers which have been created with folders. As shown in the screenshot below.

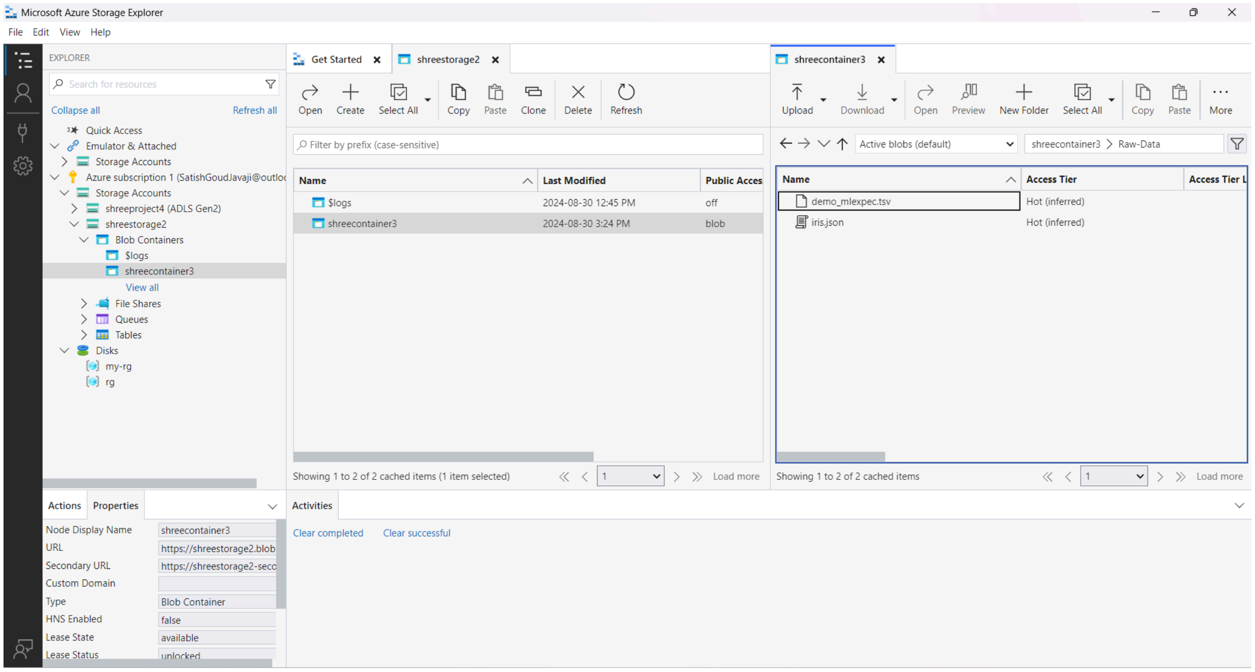


Step 2: Click on “shreestorage2” the storage account which is already created.

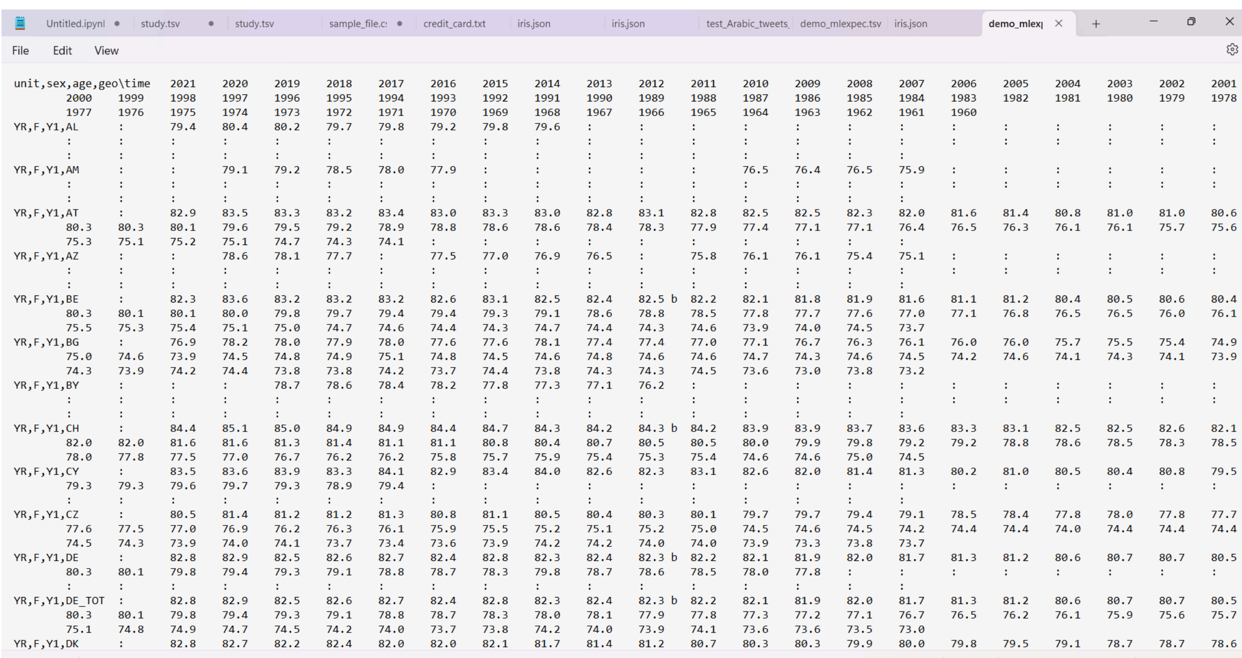
Step 3: Click on “Blob container” and select the container which is already created “shreecontainer3”.

Step 4: Click on “Raw-Data” the folder which is already created.

Step 5: You will be able to see the JSON and TSV file which is uploaded.



Step 6: Select the TSV file, you will be able to see as shown in the below screenshot.



Step 7: Select the JSON file, you will be able to see as shown in the below screenshot.