

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.

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

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

Project details

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 *

Azure subscription 1



Resource group *

my-rg

[Create new](#)

Instance details

Storage account name * ⓘ

shreestorage11

Region * ⓘ

(US) Central US

[Deploy to an Azure Extended Zone](#)

Primary service ⓘ

Select a primary service

Performance * ⓘ



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



Premium: Recommended for scenarios that require low latency.

Redundancy * ⓘ

Locally-redundant storage (LRS)

* Below screenshot represent “Advanced” options:

Basics Advanced Networking Data protection Encryption Tags Review + create

Security

Configure security settings that impact your storage account.

Require secure transfer for REST API operations ⓘ	<input checked="" type="checkbox"/>
Allow enabling anonymous access on individual containers ⓘ	<input type="checkbox"/>
Enable storage account key access ⓘ	<input checked="" type="checkbox"/>
Default to Microsoft Entra authorization in the Azure portal ⓘ	<input type="checkbox"/>
Minimum TLS version ⓘ	<div>Version 1.2 ▾</div>
Permitted scope for copy operations (preview) ⓘ	<div>From any storage account ▾</div>

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 ⓘ	<input type="checkbox"/>
---------------------------------	--------------------------

Access protocols

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

Enable SFTP ⓘ	<input type="checkbox"/>
i SFTP can only be enabled for hierarchical namespace accounts	
Enable network file system v3 ⓘ	<input type="checkbox"/>
i To enable NFS v3 'hierarchical namespace' must be enabled. Learn more about NFS v3 ↗	

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



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

Basics Advanced Networking Data protection Encryption Tags Review + create

Network connectivity

You can connect to your storage account either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Network access *


- ☒ Enable public access from all networks
- ☐ Enable public access from selected virtual networks and IP addresses
- ☐ Disable public access and use private access
-  Enabling public access from all networks might make this resource available publicly. Unless public access is required, we recommend using a more restricted access type. [Learn more](#) 


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

Basics Advanced Networking Data protection Encryption Tags Review + create

Recovery


Protect your data from accidental or erroneous deletion or modification.

☐ Enable point-in-time restore for containers
Use point-in-time restore to restore one or more containers to an earlier state. If point-in-time restore is enabled, then versioning, change feed, and blob soft delete must also be enabled. [Learn more](#) 

☒ Enable soft delete for blobs
Soft delete enables you to recover blobs that were previously marked for deletion, including blobs that were overwritten. [Learn more](#) 


Days to retain deleted blobs ⓘ

7

☒ Enable soft delete for containers
Soft delete enables you to recover containers that were previously marked for deletion. [Learn more](#) 

Days to retain deleted containers ⓘ

7

☒ Enable soft delete for file shares
Soft delete enables you to recover file shares that were previously marked for deletion. [Learn more](#) 

Days to retain deleted file shares ⓘ

7

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

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

Encryption type * ⓘ

☒ Microsoft-managed keys (MMK)

☐ Customer-managed keys (CMK)

Enable support for customer-managed keys ⓘ

☒ Blobs and files only

☐ All service types (blobs, files, tables, and queues)

⚠ This option cannot be changed after this storage account is created.

Enable infrastructure encryption ⓘ ☐

*** Below screenshot represent Tags options:**

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

Name	Value	Resource
<input type="text"/>	:	<input type="text"/>
<input type="text"/>	:	<input type="text" value="All resources selected"/>

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

[View automation template](#)

Basics

Subscription	Azure subscription 1
Resource group	my-rg
Location	Central US
Storage account name	shreestorage2
Primary service	Azure Blob Storage or Azure Data Lake Storage Gen 2
Primary workload	
Performance	Standard
Replication	Read-access geo-redundant storage (RA-GRS)

Advanced

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

Security

Secure transfer	Enabled
Blob anonymous access	Disabled
Allow storage account key access	Enabled
Default to Microsoft Entra authorization in the Azure portal	Disabled
Minimum TLS version	Version 1.2

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.**

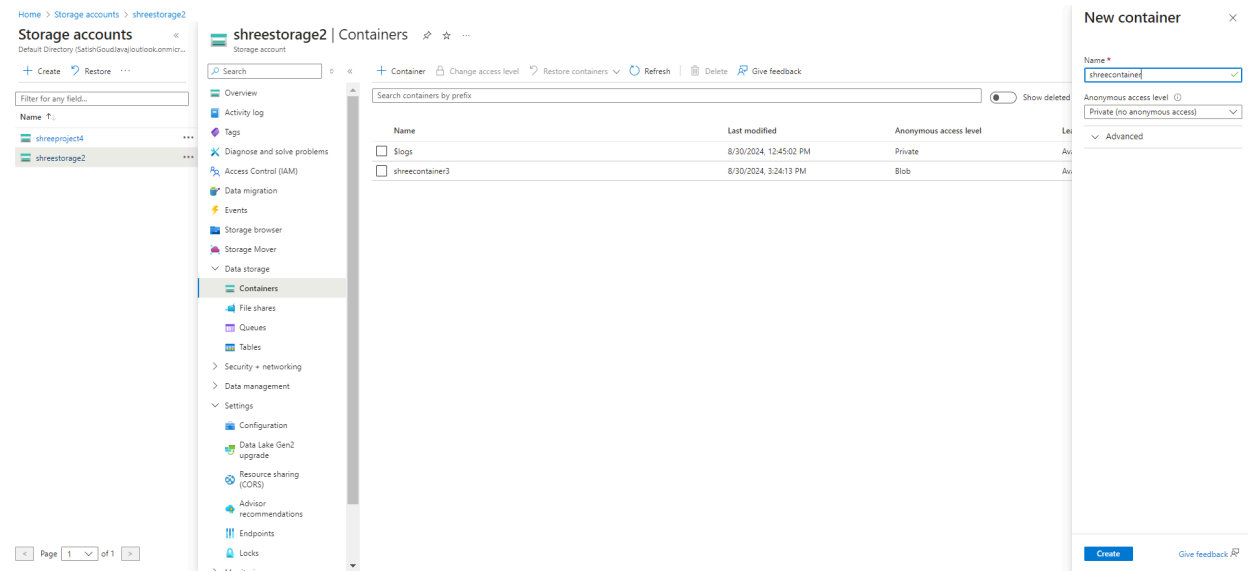
Step 1 : Open a storage account, the one which is created.

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.

Home > Storage accounts > shreestorage2 | Containers >

shreecontainer3

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

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

Location: shreecontainer3

Search blobs by prefix (case-sensitive)

Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
Raw-Data						

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.

shreecontainer3

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

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

Location: shreecontainer3 / Raw-Data

Search blobs by prefix (case-sensitive)

Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[.]						
demo_mlexpec.tsv	8/30/2024, 3:25:06 PM	Hot (Inferred)		Block blob	4.06 MiB	Available
iris.json	8/30/2024, 3:25:05 PM	Hot (Inferred)		Block blob	15.43 KiB	Available

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

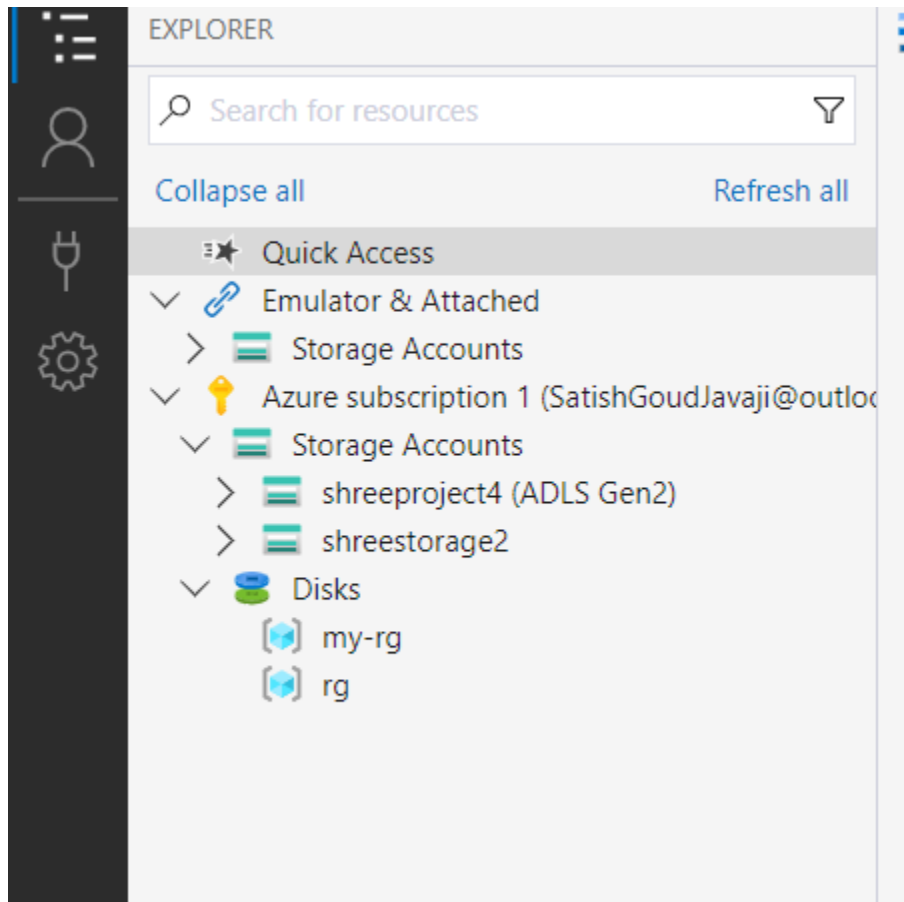
- **processed-data/** (empty for now)
- **metadata/** (empty for now)

- 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.

The screenshot shows the Microsoft Azure Storage Explorer interface. On the left, the Explorer pane shows the hierarchy: Quick Access > Azure subscription 1 (SatishGoudJavaji@outlook) > Storage Accounts > shreeproject4 (ADLS Gen2) > shreestorage2 > Blob Containers > shreecontainer3. The main pane shows the 'shreecontainer3' folder with a table of items:

Name	Last Modified	Public Access
\$logs	2024-08-30 12:45 PM	off
shreecontainer3	2024-08-30 3:24 PM	blob

Below this, the 'Raw-Data' folder is expanded, showing a table of items:

Name	Access Tier	Access Tier L
demo_mlexpec.tsv	Hot (inferred)	
iris.json	Hot (inferred)	

The bottom pane shows the properties of the selected container 'shreecontainer3'.

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

File		Edit	View
unit,sex,age,geo,time	2021	2020	2019
2000	1999	1998	1997
1977	1976	1975	1974
YR,F,Y1,AL	79.4	80.4	80.2
79.7	79.8	79.2	79.8
79.2	79.8	79.6	79.6
79.6	79.6	79.6	79.6
79.6	79.6	79.6	79.6
YR,F,Y1,AM	79.1	79.2	78.5
79.1	79.2	78.0	77.9
79.1	79.2	78.0	77.9
79.1	79.2	78.0	77.9
YR,F,Y1,AT	82.9	83.5	83.3
80.3	80.3	80.1	79.6
75.3	75.1	75.2	75.1
75.3	75.1	75.2	75.1
YR,F,Y1,AZ	78.6	78.1	77.7
78.6	78.1	77.7	77.5
78.6	78.1	77.7	77.5
78.6	78.1	77.7	77.5
YR,F,Y1,BE	82.3	83.6	83.2
80.3	80.1	80.1	79.8
75.5	75.3	75.4	75.1
75.5	75.3	75.4	75.1
YR,F,Y1,BG	76.9	78.2	78.0
75.0	74.6	73.9	74.5
74.3	73.9	74.2	74.4
74.3	73.9	74.2	74.4
YR,F,Y1,BY	78.7	78.6	78.4
78.7	78.6	78.4	78.2
78.7	78.6	78.4	78.2
78.7	78.6	78.4	78.2
YR,F,Y1,CH	84.4	85.1	85.0
82.0	82.0	81.6	81.6
78.0	77.8	77.5	77.0
78.0	77.8	77.5	77.0
YR,F,Y1,CY	83.5	83.6	83.9
79.3	79.3	79.6	79.3
79.3	79.3	79.6	79.3
79.3	79.3	79.6	79.3
YR,F,Y1,CZ	80.5	81.4	81.2
77.6	77.5	77.0	76.2
74.5	74.3	73.9	74.0
74.5	74.3	73.9	74.0
YR,F,Y1,DE	82.8	82.9	82.5
80.3	80.1	79.8	79.4
80.3	80.1	79.8	79.4
80.3	80.1	79.8	79.4
YR,F,Y1,DE_TOT	82.8	82.9	82.5
80.3	80.1	79.8	79.4
75.1	74.8	74.9	74.7
75.1	74.8	74.9	74.7
YR,F,Y1,DK	82.8	82.7	82.2
82.8	82.7	82.2	82.4
82.8	82.7	82.2	82.4
82.8	82.7	82.2	82.4

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

```
File Edit View

[
  {"sepal.length": 5.1, "sepal.width": 3.5, "petal.length": 1.4, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.9, "sepal.width": 3.0, "petal.length": 1.4, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.7, "sepal.width": 3.2, "petal.length": 1.3, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.6, "sepal.width": 3.1, "petal.length": 1.5, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.0, "sepal.width": 3.6, "petal.length": 1.4, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.4, "sepal.width": 3.9, "petal.length": 1.7, "petal.width": 0.4, "species": "setosa"},
  {"sepal.length": 4.6, "sepal.width": 3.4, "petal.length": 1.4, "petal.width": 0.3, "species": "setosa"},
  {"sepal.length": 5.0, "sepal.width": 3.4, "petal.length": 1.5, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.4, "sepal.width": 2.9, "petal.length": 1.4, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.9, "sepal.width": 3.1, "petal.length": 1.5, "petal.width": 0.1, "species": "setosa"},
  {"sepal.length": 5.4, "sepal.width": 3.7, "petal.length": 1.5, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.8, "sepal.width": 3.4, "petal.length": 1.6, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.8, "sepal.width": 3.0, "petal.length": 1.4, "petal.width": 0.1, "species": "setosa"},
  {"sepal.length": 4.3, "sepal.width": 3.0, "petal.length": 1.1, "petal.width": 0.1, "species": "setosa"},
  {"sepal.length": 5.8, "sepal.width": 4.0, "petal.length": 1.2, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.7, "sepal.width": 4.4, "petal.length": 1.5, "petal.width": 0.4, "species": "setosa"},
  {"sepal.length": 5.4, "sepal.width": 3.9, "petal.length": 1.3, "petal.width": 0.4, "species": "setosa"},
  {"sepal.length": 5.1, "sepal.width": 3.5, "petal.length": 1.4, "petal.width": 0.3, "species": "setosa"},
  {"sepal.length": 5.7, "sepal.width": 3.8, "petal.length": 1.7, "petal.width": 0.3, "species": "setosa"},
  {"sepal.length": 5.1, "sepal.width": 3.8, "petal.length": 1.5, "petal.width": 0.3, "species": "setosa"},
  {"sepal.length": 5.4, "sepal.width": 3.4, "petal.length": 1.7, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.1, "sepal.width": 3.7, "petal.length": 1.5, "petal.width": 0.4, "species": "setosa"},
  {"sepal.length": 4.6, "sepal.width": 3.6, "petal.length": 1.0, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.1, "sepal.width": 3.3, "petal.length": 1.7, "petal.width": 0.5, "species": "setosa"},
  {"sepal.length": 4.8, "sepal.width": 3.4, "petal.length": 1.9, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.0, "sepal.width": 3.0, "petal.length": 1.6, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.0, "sepal.width": 3.4, "petal.length": 1.6, "petal.width": 0.4, "species": "setosa"},
  {"sepal.length": 5.2, "sepal.width": 3.5, "petal.length": 1.5, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.2, "sepal.width": 3.4, "petal.length": 1.4, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.7, "sepal.width": 3.2, "petal.length": 1.6, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.8, "sepal.width": 3.1, "petal.length": 1.6, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.4, "sepal.width": 3.4, "petal.length": 1.5, "petal.width": 0.4, "species": "setosa"},
  {"sepal.length": 5.2, "sepal.width": 4.1, "petal.length": 1.5, "petal.width": 0.1, "species": "setosa"},
  {"sepal.length": 5.5, "sepal.width": 4.2, "petal.length": 1.4, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.9, "sepal.width": 3.1, "petal.length": 1.5, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.0, "sepal.width": 3.2, "petal.length": 1.2, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 5.5, "sepal.width": 3.5, "petal.length": 1.3, "petal.width": 0.2, "species": "setosa"},
  {"sepal.length": 4.9, "sepal.width": 3.6, "petal.length": 1.4, "petal.width": 0.1, "species": "setosa"},
  {"sepal.length": 4.4, "sepal.width": 3.0, "petal.length": 1.3, "petal.width": 0.2, "species": "setosa"}
]
```