Working with Azure Data Lake Gen2 account

Exercise 1 – Create Azure Data Lake Gen2 Account

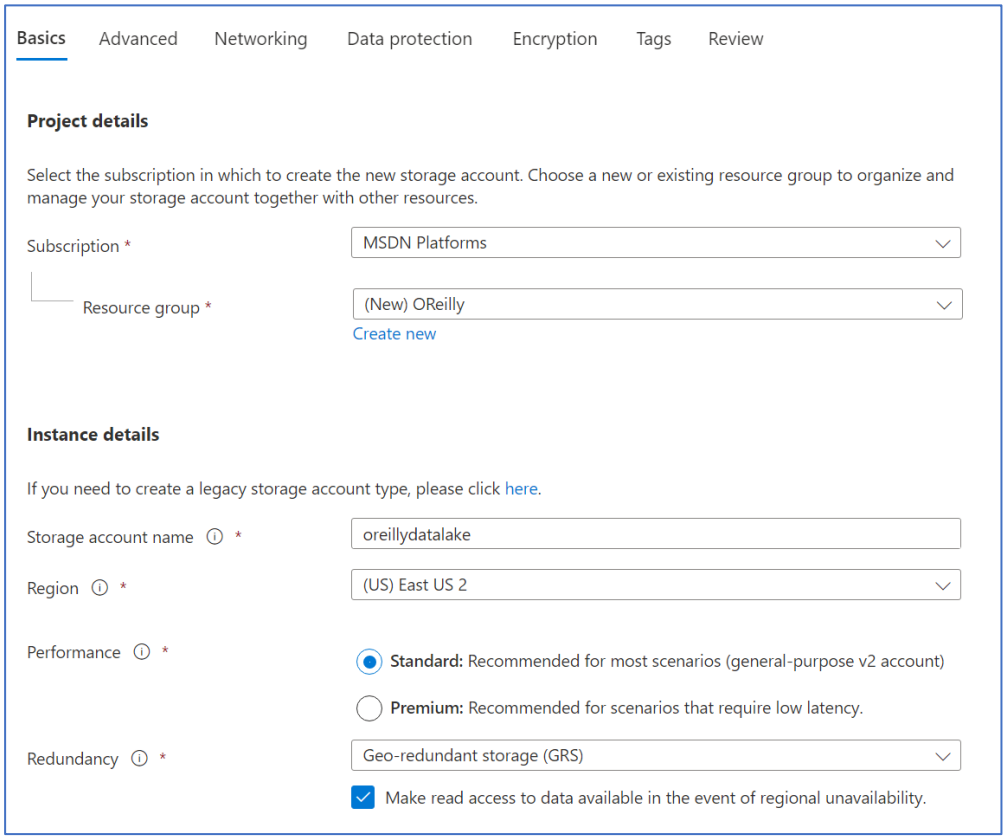
1. Go to Azure portal (portal.azure.com)

2. In the search bar, search for Storage Accounts. And select it.

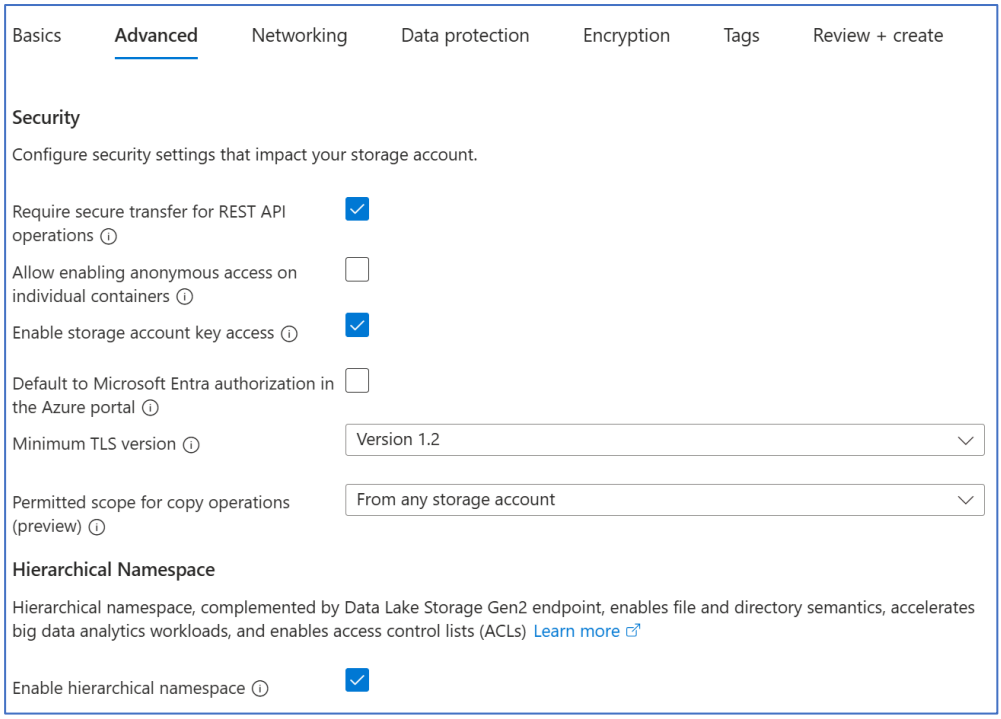
3. Click on **Create**.

4. Fill up the properties to create account:

    a. [Basics Tab]
        i. Select subscription
        ii. On Resource Group, click Create new and provide a name
        iii. Provide a unique storage account name
        iv. Select region of your choice (example – East US 2)
        v. You have options for selecting performance & redundancy options – keep it as is.
        vi. Click Next

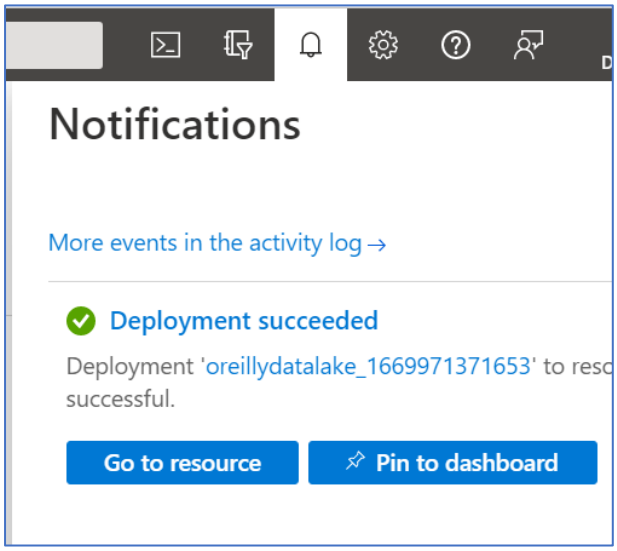


    b. [Advanced Tab]
        i. In **Hierarchical Namespace** section -> Select **Enable Hierarchical Namespace** checkbox.
        ii. Click **Review + create**.



c. Click **Create**.

5. This will create a new Azure Data Lake Gen2 account.



Exercise 2 – Upload Files to Azure Data Lake Gen2 Account

0. In a separate browser window, go to following URL and download sample files:
   https://tinyurl.com/data-lake-bootcamp-2022

1. Open Data Lake account created in the previous step.

2. From left pane, under **Data storage**, go to **Containers**.

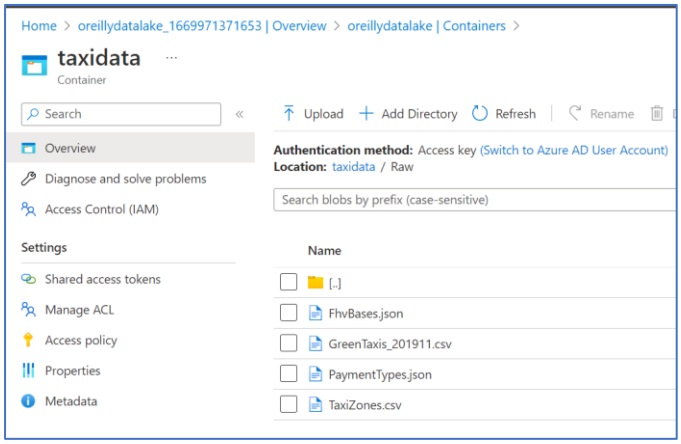
3. Click on **Add container**, fill up the properties, and click **Create**:
   a. Name: taxidata
   b. Public access level: Private

4. Once created, open **taxidata** container.

5. To store raw data, click on **Add Directory**, fill up the properties, and click **Save**:
    a. Name: Raw

6. Click on **Raw** directory to open it.

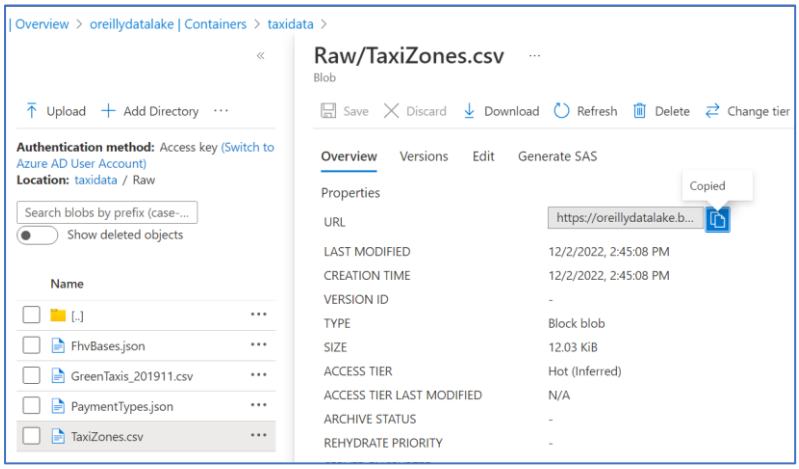
7. In Raw directory, upload the files downloaded in Step 0.



Exercise 3 – Generate and use SAS Token

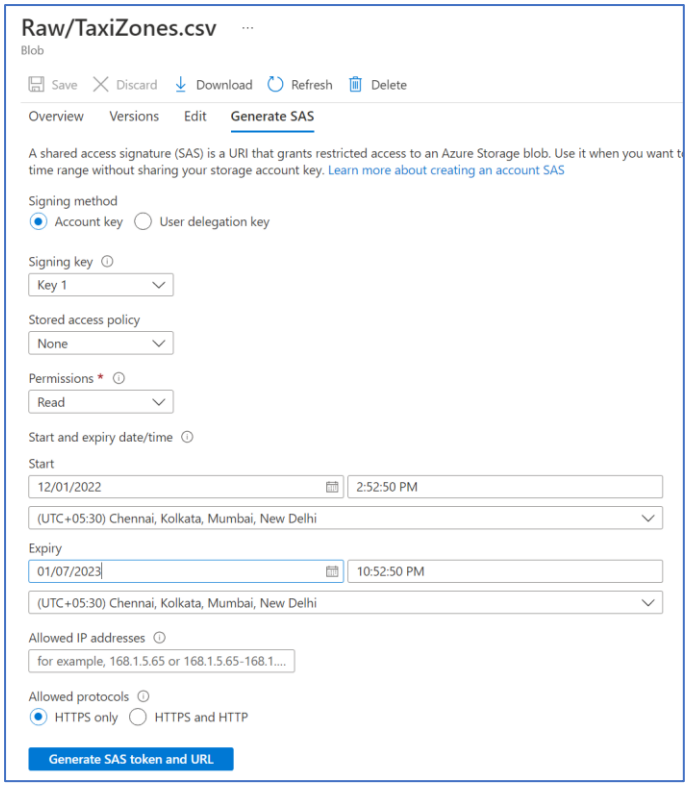
1. In taxidata container, click on TaxiZones.csv file.

2. Copy file URL for TaxiZones.csv.



3. Paste URL in browser and see if its accessible.
    - This should not work, since no access is available.

4. Click on Generate SAS, and define the permissions:
    a. Permissions: Read
    b. Start: Keep one day before to avoid Timezone issues
    c. End: One month ahead in future
    d. Click on Generate SAS token and URL

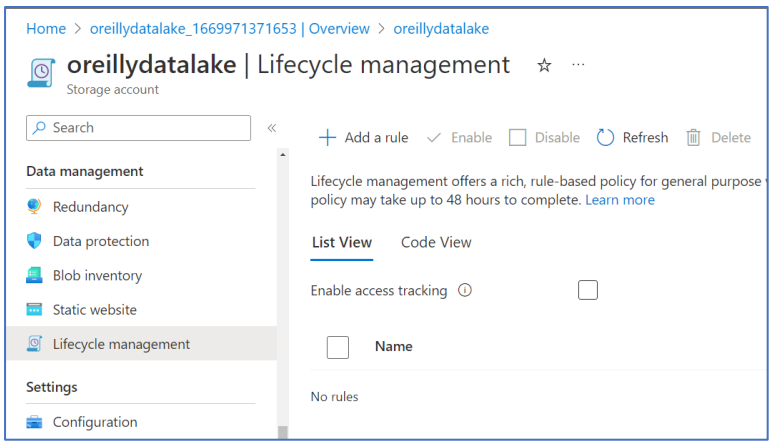


5. Copy Blob SAS URL.
   - This will include file URL and SAS token.

6. Paste URL in browser and see if its accessible.
   - This should work, since access is provided via SAS token.

Exercise 4 – Configure Data Lifecycle

1. From left pane of storage account, under **Data Management**, navigate to **Lifecycle management** tab.



2. Click on **Add a rule**.

3. Provide the details:
   a. Rule name: Rule1
   b. Rule scope: Apply rule to all blobs in your storage account
   c. Blob type: Block blobs
   d. Blob subtype: Base blobs
   e. Click Next

4. Define rule conditions, and click on **Save**:
    a. If Base blobs haven't been modified in 60 days, then move to Cool storage
    b. If Base blobs haven't been modified in 180 days, then move to Archive storage
    c. If Base blobs haven't been modified in 365 days, then Delete the blob