

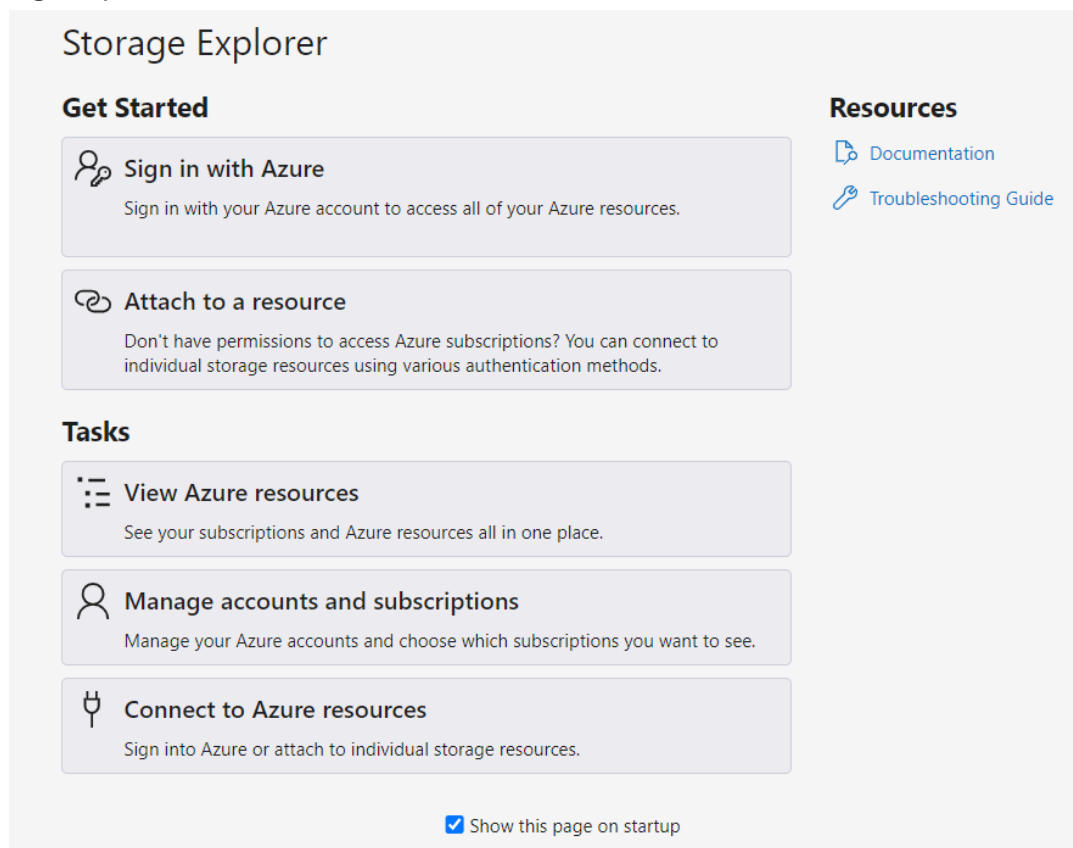
# Azure Storage Container

- Each team has now been given access to a storage container.
- This is where you should store the data you will use for the group project – you should **not** store your group project datasets in your Databricks Workspace directory.
- The follow sections show how you can upload data to the storage container using Storage Explorer, and then how you can read the data from the storage container in your Databricks notebooks.
- We have also added a data/ storage container that you have read-only access to, containing various datasets. See the final section of this document for how to access this.

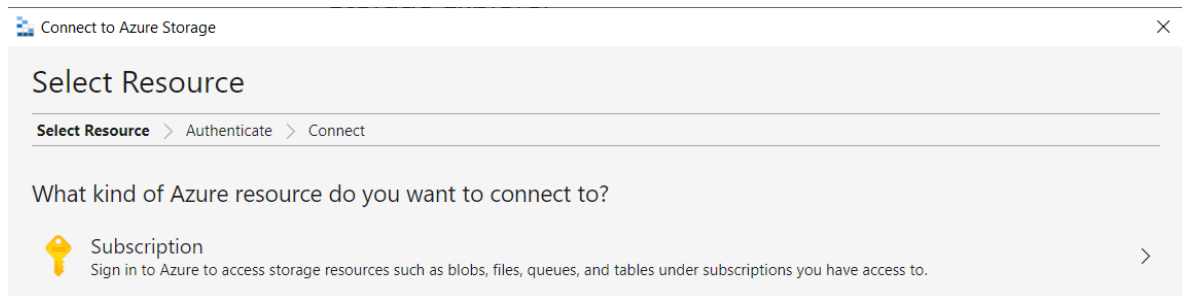
## Uploading data using Storage Explorer

### Setting up Storage Explorer

1. Download Storage Explorer from the following link and install on your device:  
<https://azure.microsoft.com/en-gb/products/storage/storage-explorer>
2. Open Storage Explorer, and select 'Connect to Azure resources'



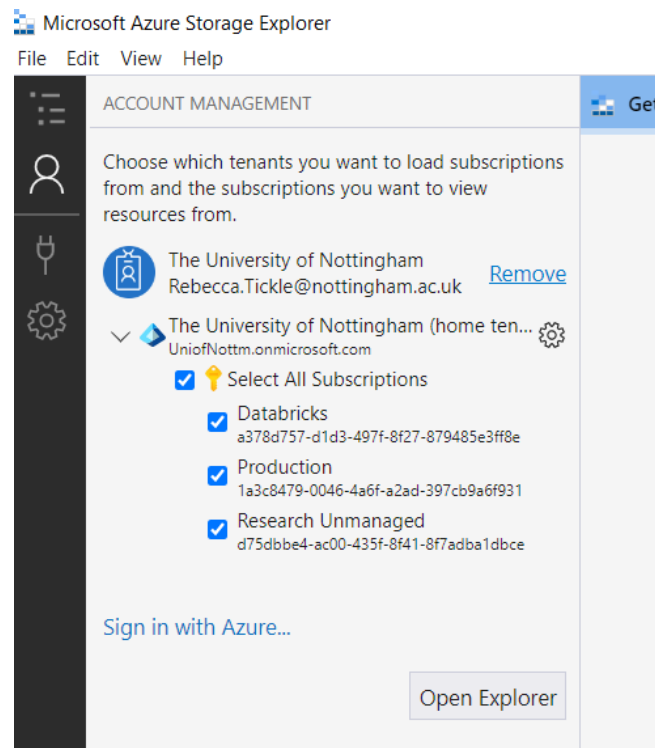
### 3. Select 'Subscription'



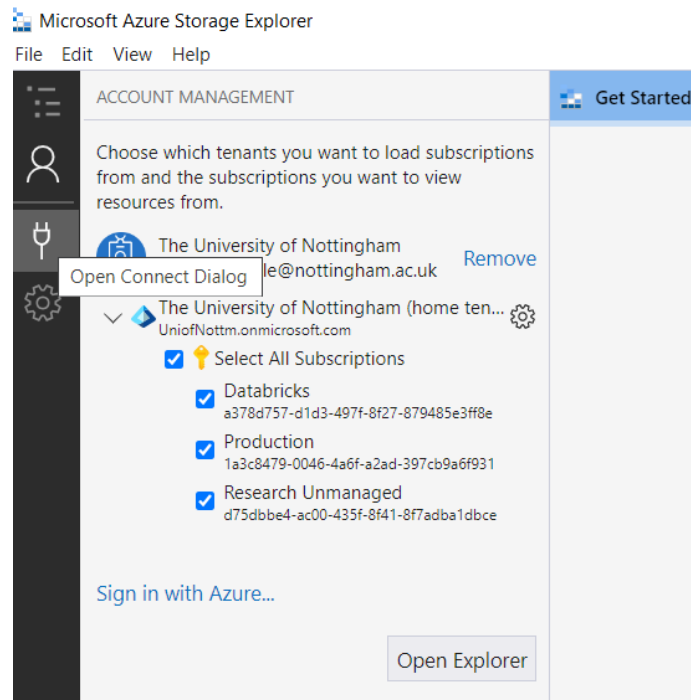
### 4. Select 'Azure'



5. Click through, and it should then take you to a browser to log in with your university credentials
6. When you have logged in, return to Storage Explorer – if your login was successful, you should see your university account in the sidebar



## 7. Open the Connect Dialog







## 8. Select 'ADLS Gen 2 Container'

Connect to Azure Storage

## Select Resource

Select Resource > Authenticate > Connect

What kind of Azure resource do you want to connect to?

-  Subscription  
Sign in to Azure to access storage resources such as blobs, files, queues, and tables under subscriptions you have access to. >
-  Storage account or service  
Attach to one or more services in a Storage account. >
-  Blob container or directory  
Attach to an individual Blob container or directory. >
-  ADLS Gen2 container or directory  
Attach to an individual ADLS Gen2 container or directory. >

### 9. Select 'Sign in using OAuth'

Connect to Azure Storage

## Select Connection Method

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

How will you connect to the blob container?

- ☒ Sign in using OAuth
- ☐ Shared access signature URL (SAS)
- ☐ Anonymously (my blob container allows public access)


### 10. Select your university account and click 'Next'

Connect to Azure Storage

## Select Account & Tenant

Select Resource > Select Connection Method > **Select Account & Tenant** > Enter Connection Info > Summary

Select an Azure account with access to the resource:

-  Rebecca Tickle  
Rebecca.Tickle@nottingham.ac.uk

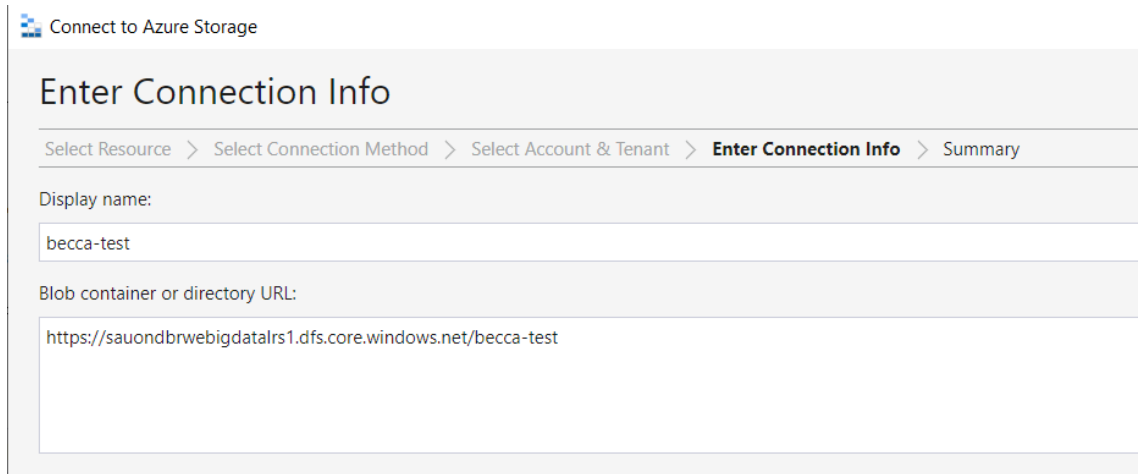
Select the tenant that contains the resource:

The University of Nottingham (67bda7ee-fd80-41ef-ac91-358418290a1e) ▼

Back Next Cancel

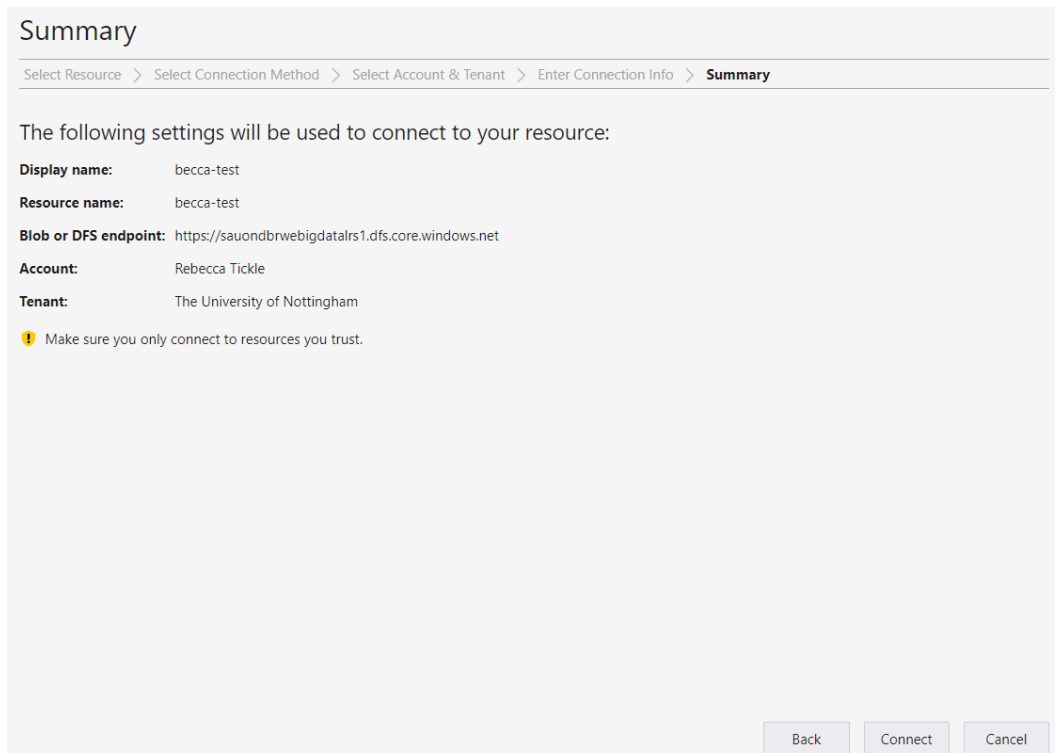
11. Enter the URL of your team's storage container as shown below.

- The storage container URL is:  
**https://sauondbrwebigdatalrs1.dfs.core.windows.net/CONTAINER\_NAME**
- You will need to replace **CONTAINER\_NAME** with the name of your team's container. For all teams, this is **2024-teamX**, where X is the number of your team.



The screenshot shows the 'Enter Connection Info' step of the 'Connect to Azure Storage' wizard. The breadcrumb trail at the top indicates the sequence: Select Resource > Select Connection Method > Select Account & Tenant > **Enter Connection Info** > Summary. The 'Display name' field contains 'becca-test'. The 'Blob container or directory URL' field contains 'https://sauondbrwebigdatalrs1.dfs.core.windows.net/becca-test'.

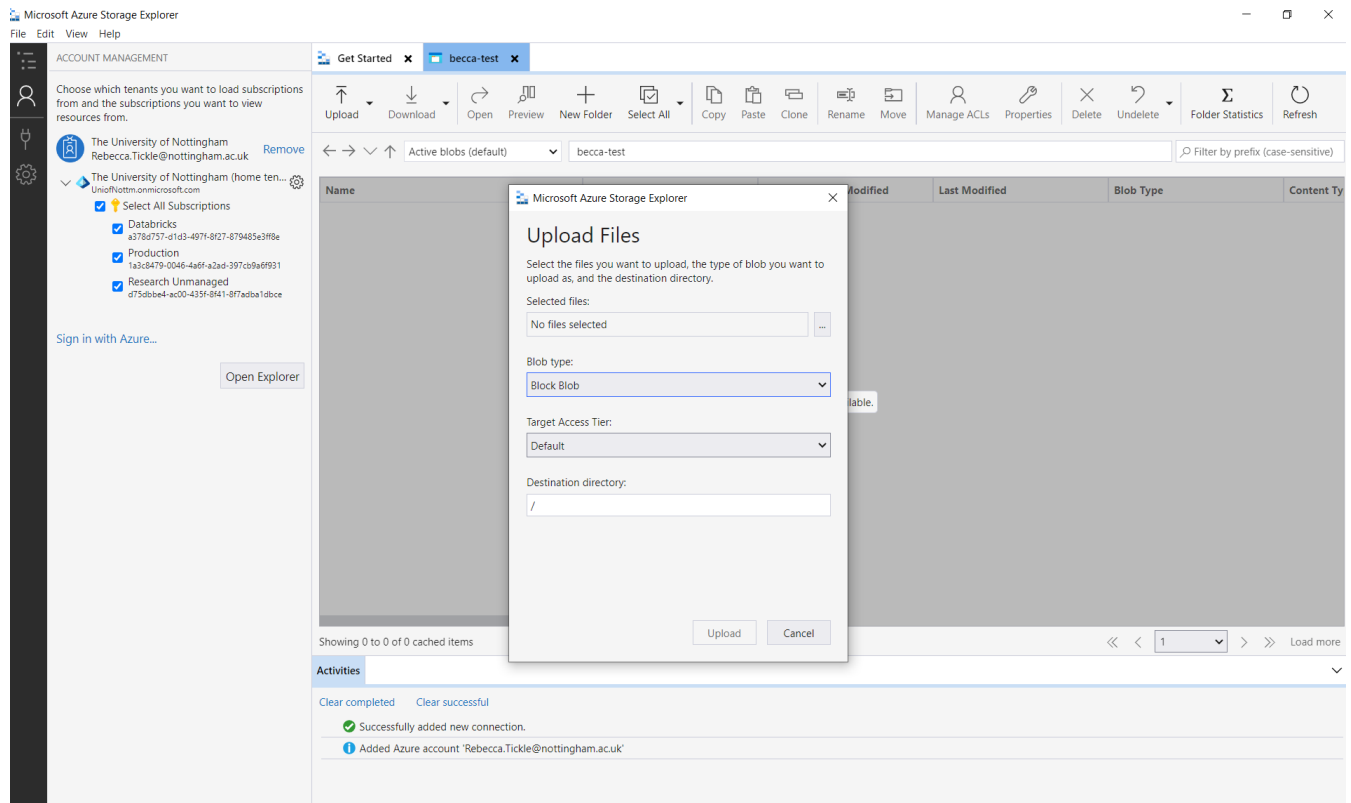
12. On the Summary page, click 'Connect'



The screenshot shows the 'Summary' step of the 'Connect to Azure Storage' wizard. The breadcrumb trail at the top indicates the sequence: Select Resource > Select Connection Method > Select Account & Tenant > Enter Connection Info > **Summary**. The text states: 'The following settings will be used to connect to your resource:'. The settings listed are: Display name: becca-test, Resource name: becca-test, Blob or DFS endpoint: https://sauondbrwebigdatalrs1.dfs.core.windows.net, Account: Rebecca Tickle, and Tenant: The University of Nottingham. A warning icon and text at the bottom state: 'Make sure you only connect to resources you trust.' At the bottom right, there are three buttons: Back, Connect, and Cancel.

## Uploading files

13. Click 'Upload' to open the upload dialog. Choose the files you want to upload.



## Accessing your data from Databricks notebooks

- All storage containers are mounted at /mnt/ in the Databricks.
- To read a file from your storage container, use the path: `/mnt/CONTAINER_NAME/file.csv`

## Read-only data/

- We have given you **read-only** access to a data/ storage container as well.
- This contains some datasets you may wish to use in your projects, to avoid you having to download/upload these datasets yourself. See page 9 of the [Group Project Description document](#) on Moodle for the datasets that are included.
- You can read a file from here as follows: `/mnt/data/<path to file>`
- If you would like to browse the datasets available here, you can use Storage Explorer – just connect to the data container rather than your team container. I.e., the URL you use for connection should be `https://sauondbrwebigdata1rs1.dfs.core.windows.net/data`