| Name | Manoj Mani |
| --- | --- |
| Date | 05/01/2024 |
| Email Id | jonam1012@gmail.com |
| Type | Daily Assignment |
| Topic | ETL with Azure Databricks and Azure Synapse and Copy activity-ADF |

ETL with Azure Databricks and Azure Synapse

**Steps:**

**Create Azure Databricks Service:**

* Go to Azure portal, select "Create a resource," then choose "Analytics > Azure Databricks."
* Provide workspace details (name, subscription, resource group, location, pricing tier).
* Create the Databricks service.

**Create Spark Cluster:**

* Access Azure Databricks portal, select the Databricks service, and launch the workspace.
* In the workspace, create a new Spark cluster with specified configurations.

**Configure Data Lake Storage Gen2:**

* Create a notebook in Azure Databricks workspace.
* Run code snippets to set up service principal credentials and configure the storage account.

**Ingest Sample Data:**

* Use notebook to download sample data into Databricks from a public repository.
* Upload the sample data to Azure Data Lake Storage Gen2.

**Extract Data:**

* Load the sample data into a DataFrame in Azure Databricks.

**Transform Data:**

* Perform data transformations on the DataFrame, selecting specific columns and renaming one of them.

**Load Data into Azure Synapse:**

* Configure access to Azure Blob Storage and specify a temporary folder.
* Provide Azure Synapse connection details.
* Load the transformed DataFrame into Azure Synapse as a table.

**Verify and Clean Up:**

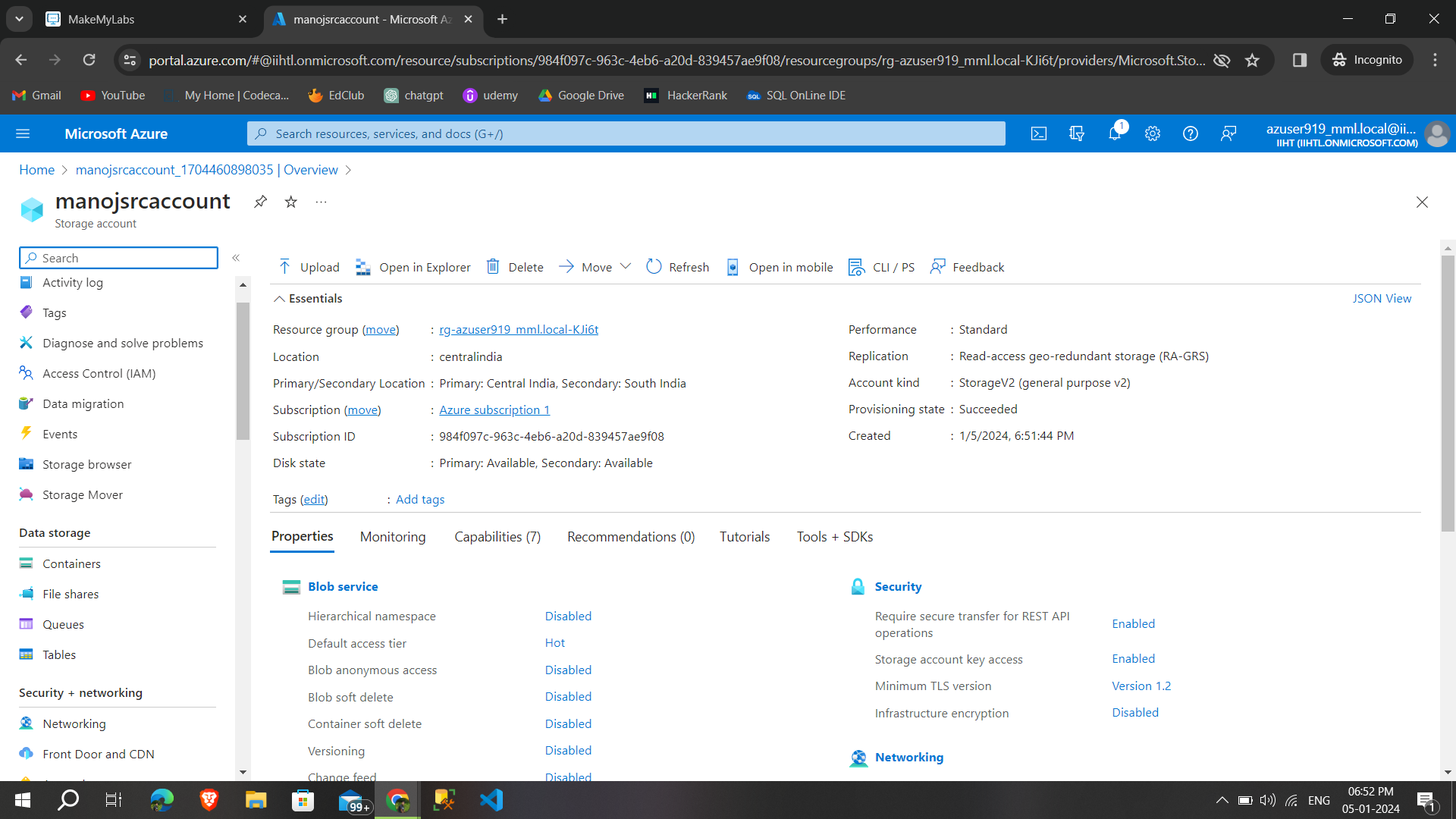
* Connect to the SQL database in Azure Synapse to verify the created table.
* Optionally, terminate the Azure Databricks cluster to avoid unnecessary costs.
* Clean up resources after completing the tutorial to avoid ongoing costs.

Copy activity-ADF

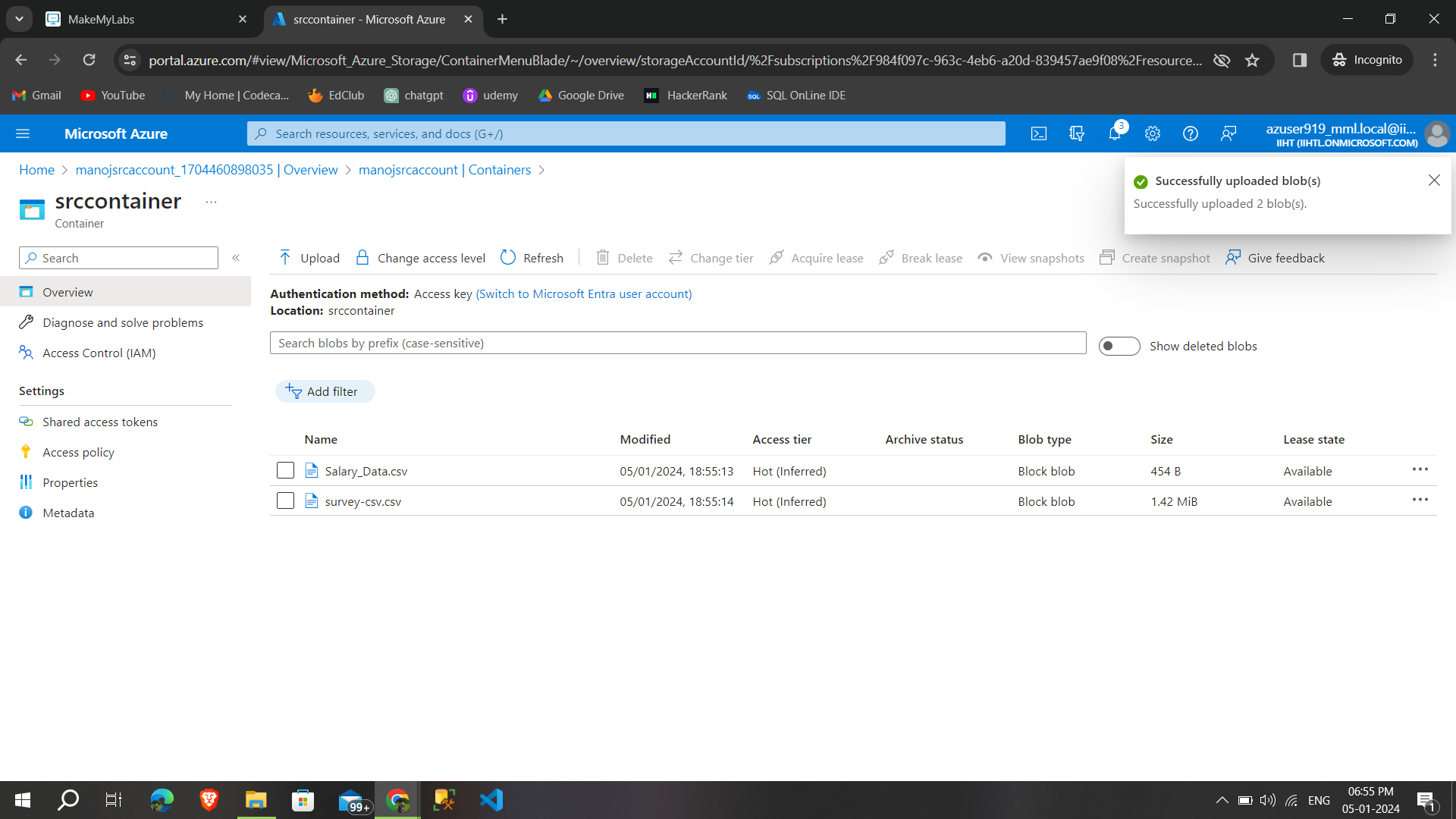
**Azure Data Factory**

* It is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

**Copy Activity** is a key component within ADF that enables the movement of data from source to destination.

**Step 1:** I created a one blob storage account called **manojsrcaccount** and this account will act as a **source** 

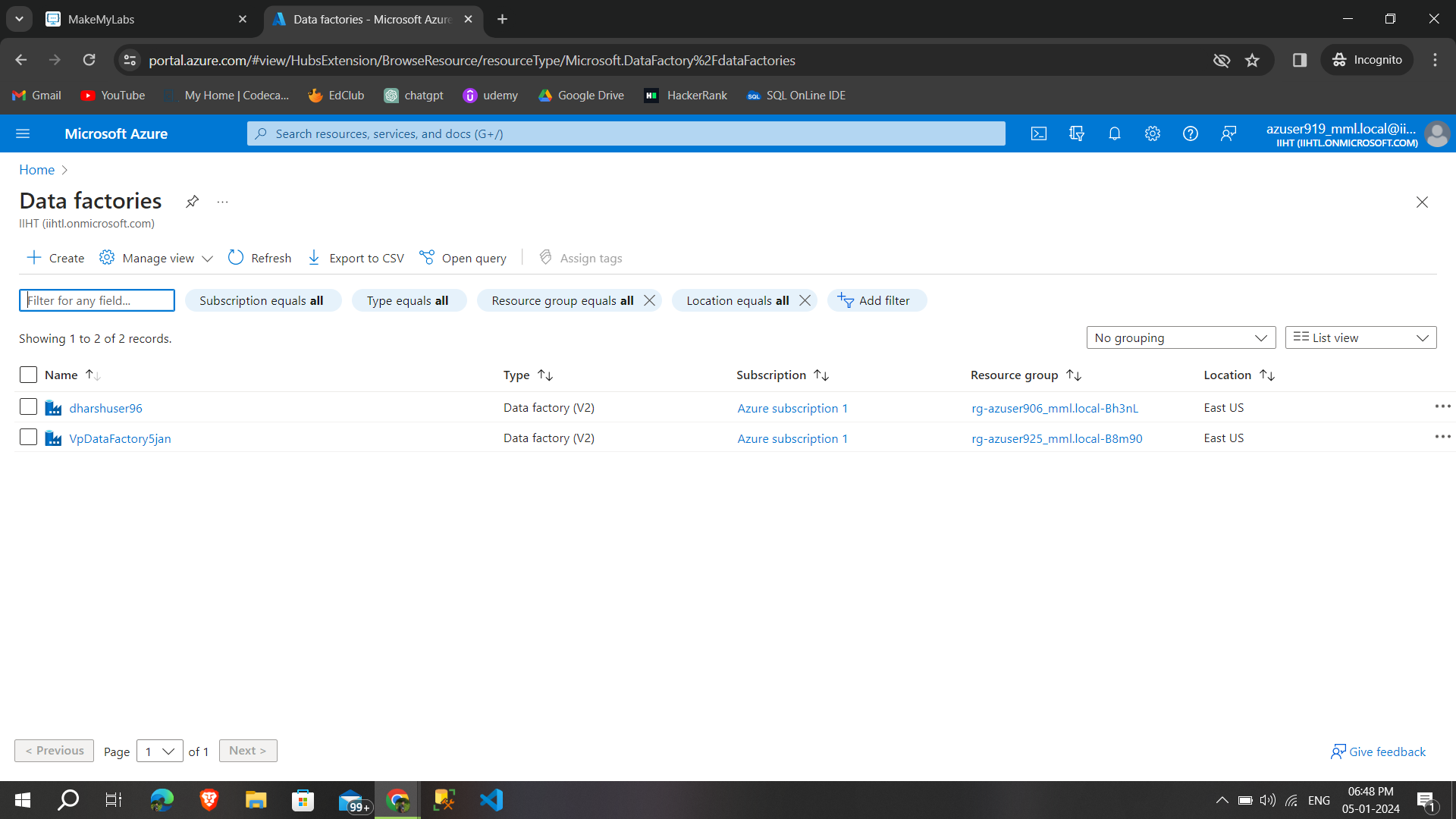
**Step 2** :under that i created a container and **upload two sample files** and these files are to be copied from **source to destination**



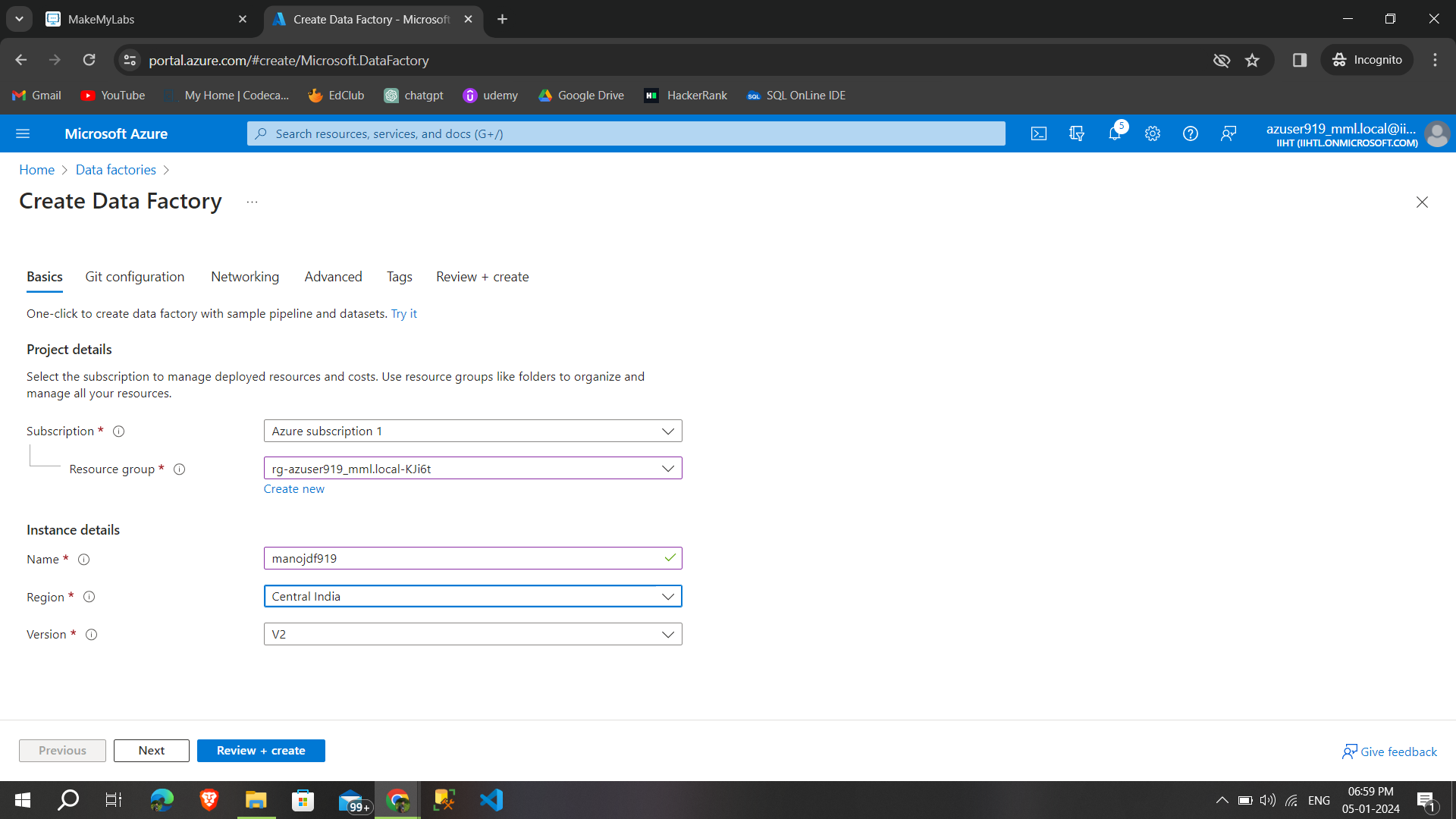
**Step 3:** then i created a storage account named **manojdestaccount** and this act like as a destination account and **created a container**



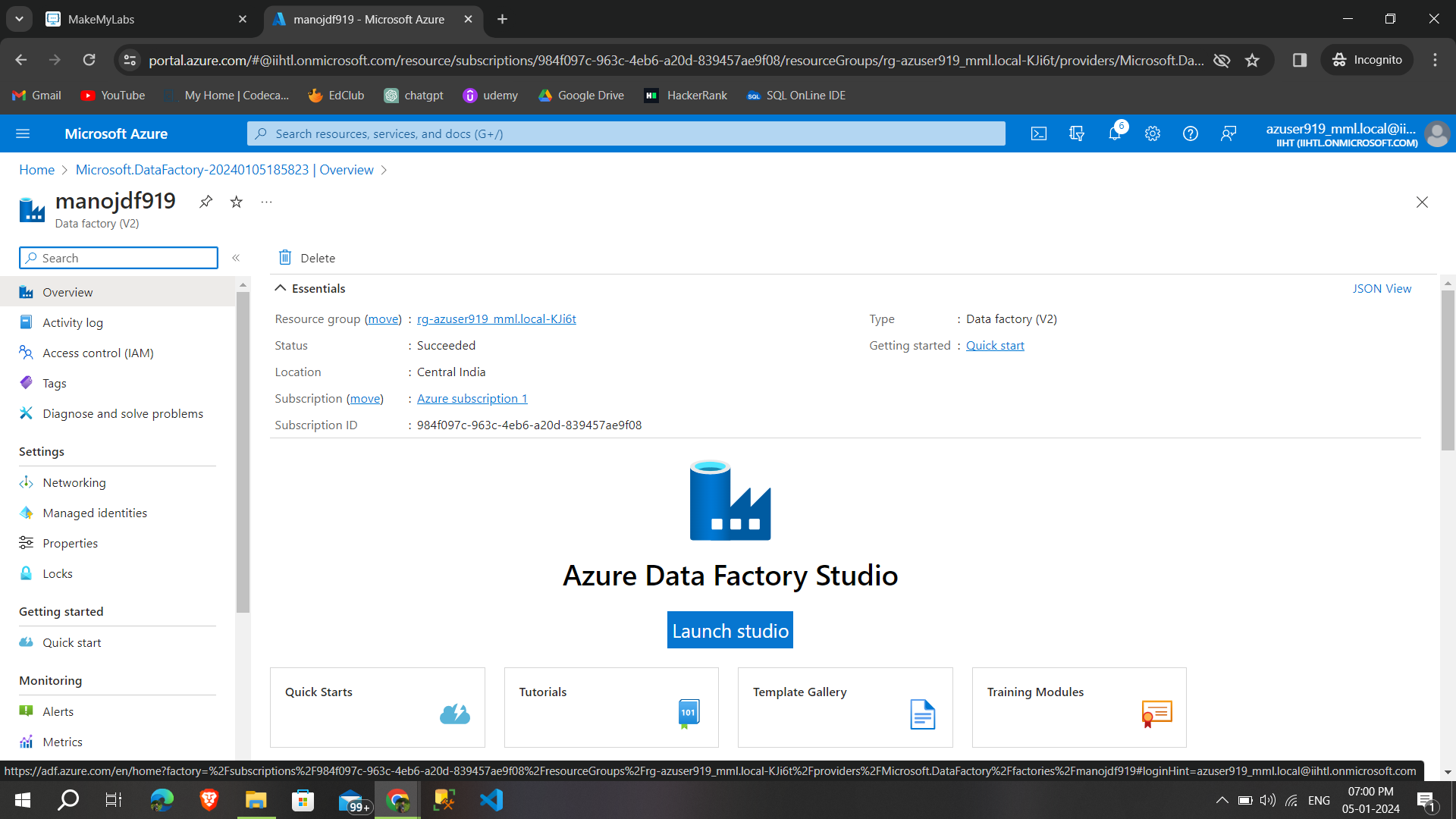
**Step 4:** Then search for **data factory** in azure portal



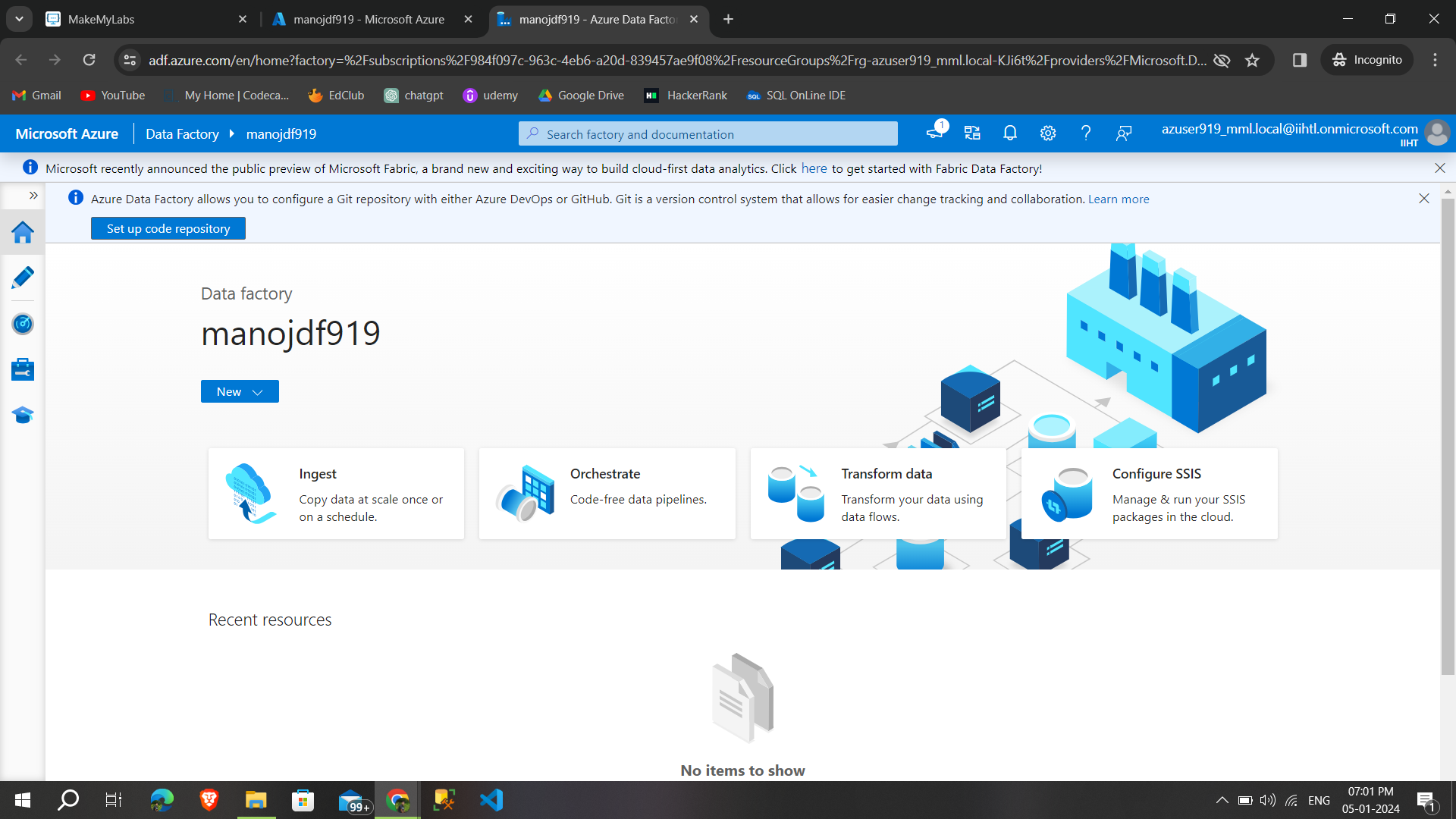
**Step 5:**select create and **give the necessary options** to create the data factory



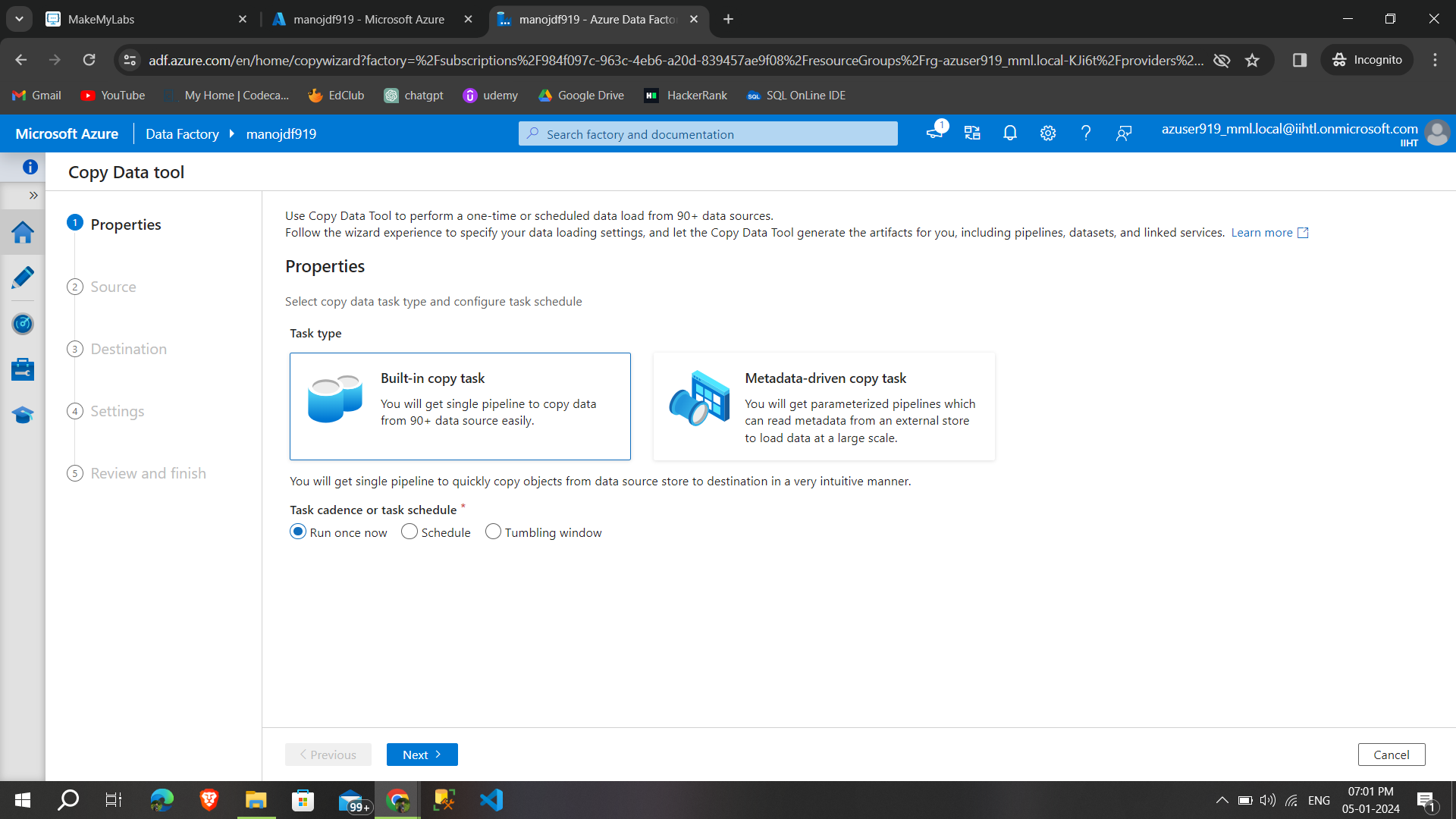
**Step 6:** successfully created a **data factory then click launch factory** to use the service



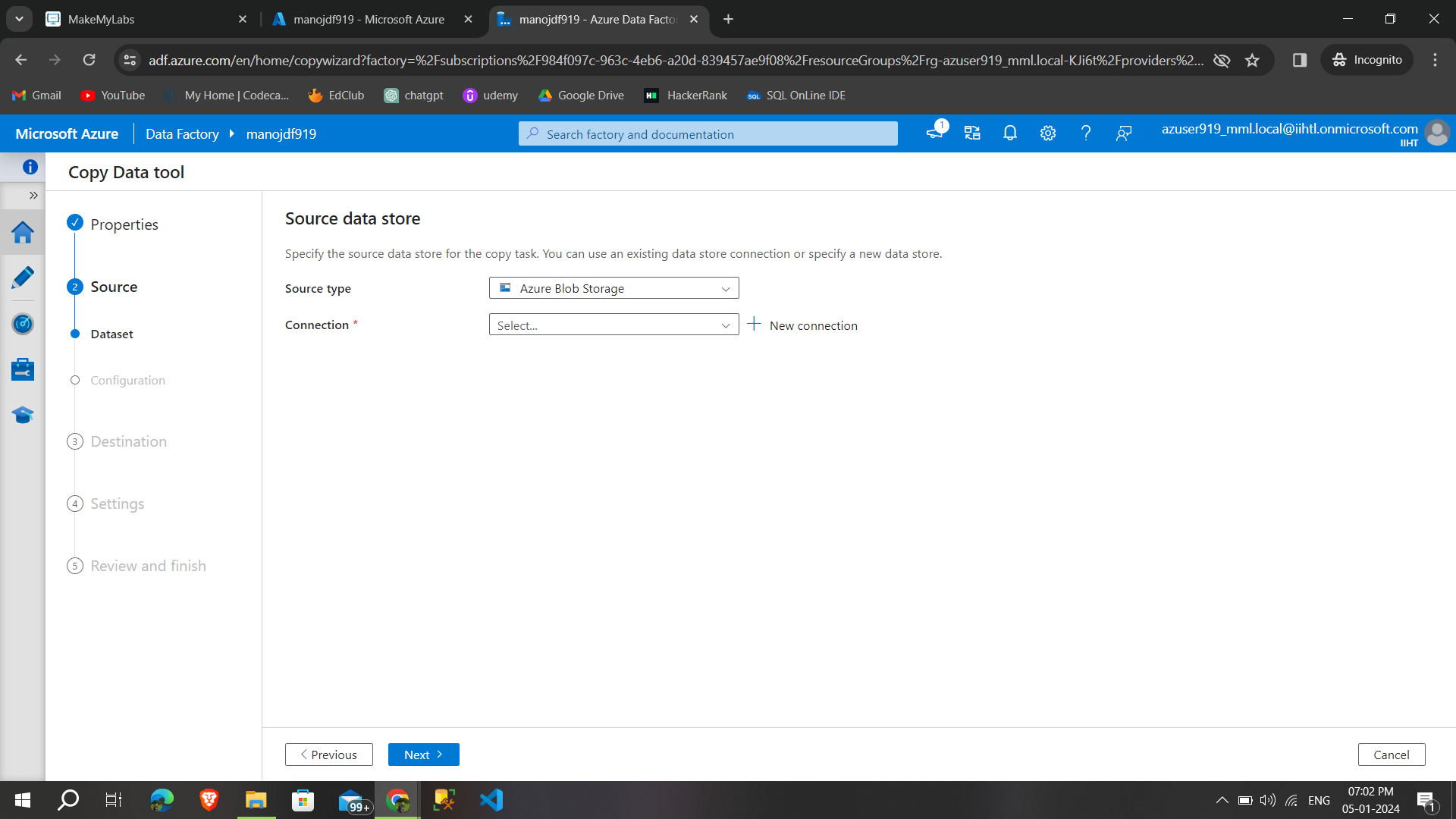
**Step 7:** this is the page of **azure data factory** and by selecting **ingest to perform the process**



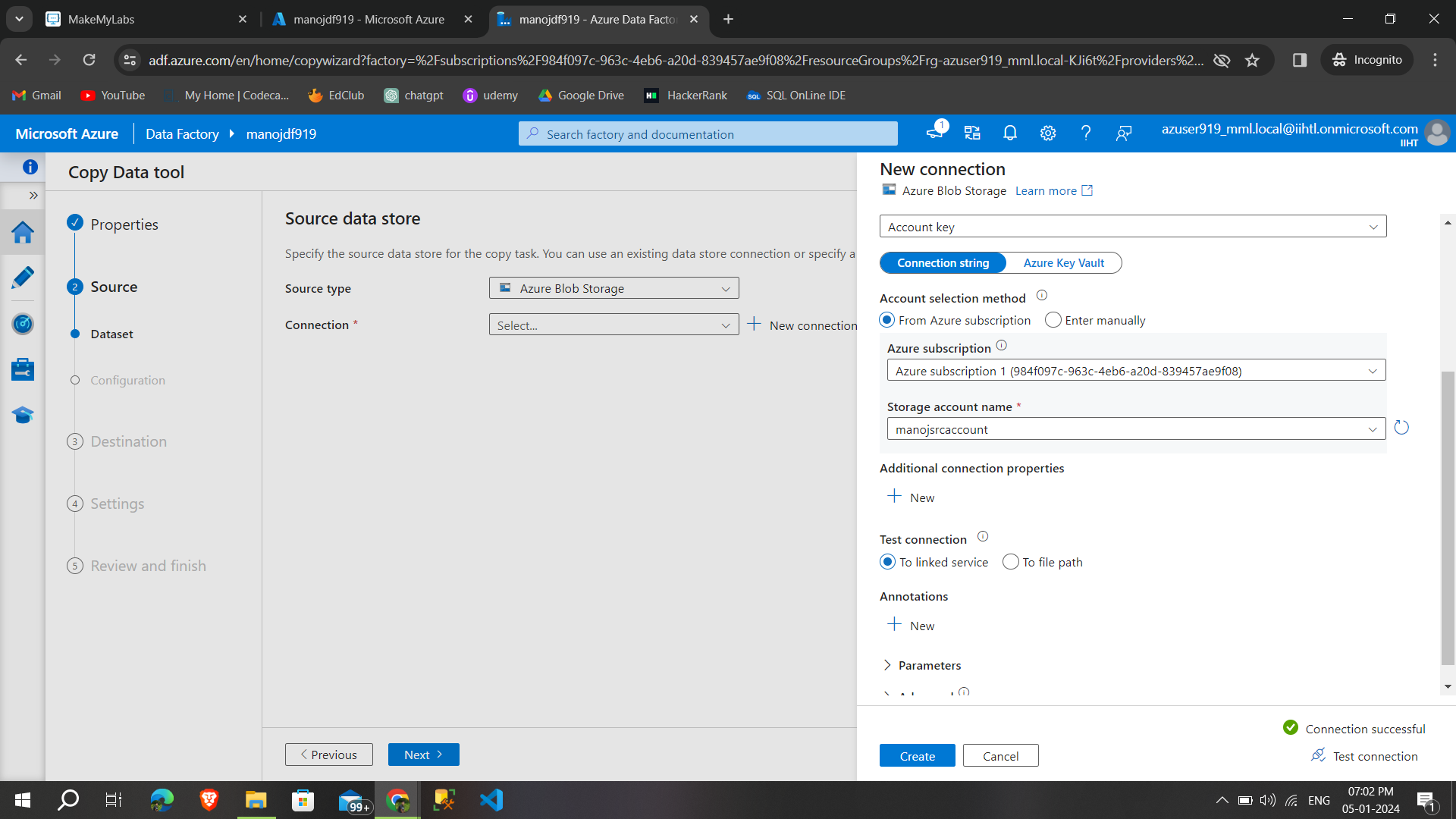
**Step 8:** by selecting the built in **copy task and run once now,** starting the process



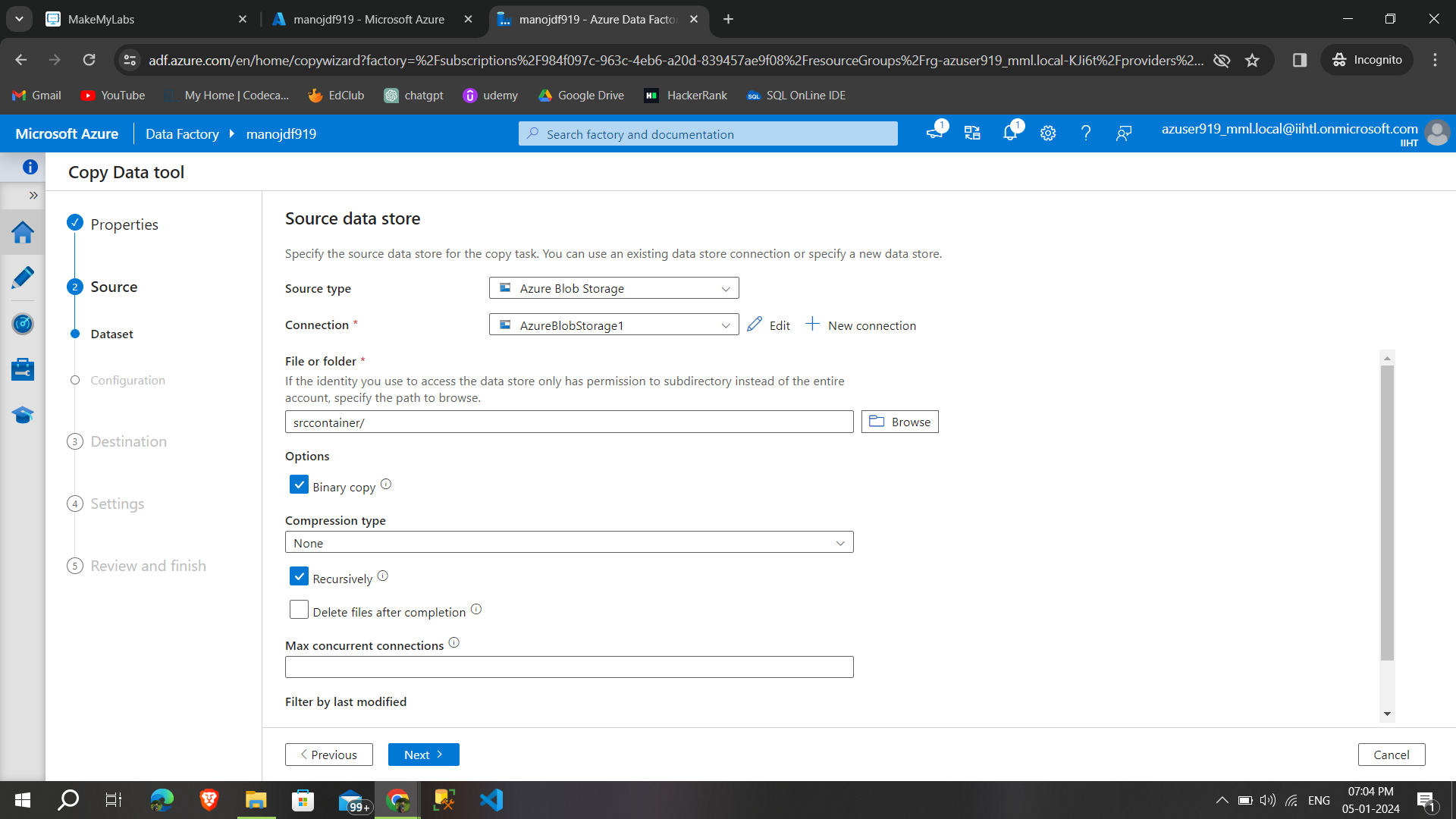
**Step 9:** we have to select the **source data store** that i’ve previously created ,click new connection



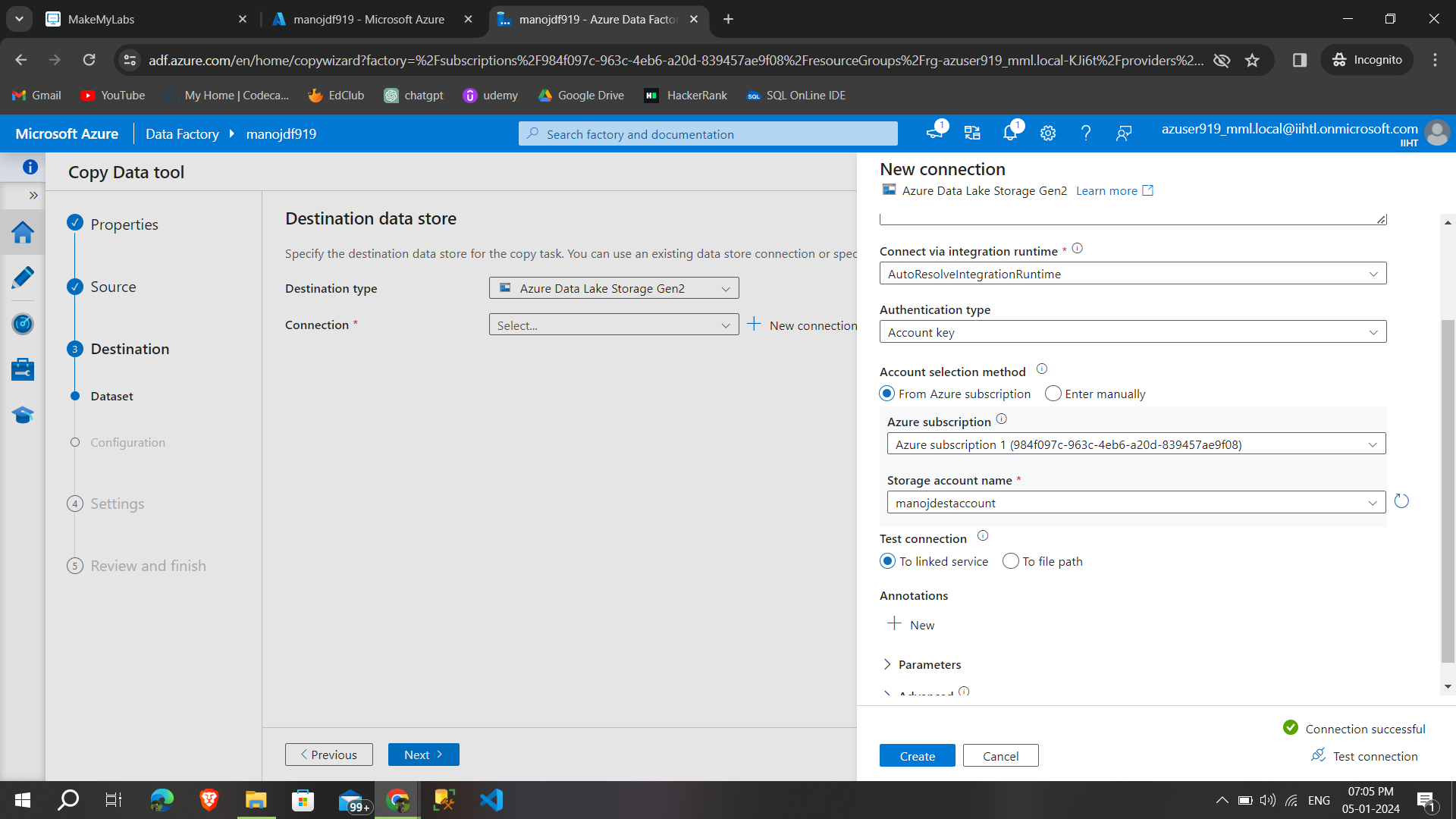
**Step 10:**after selecting the azure subscription,storage account name,**click test connection and the connection was successful**



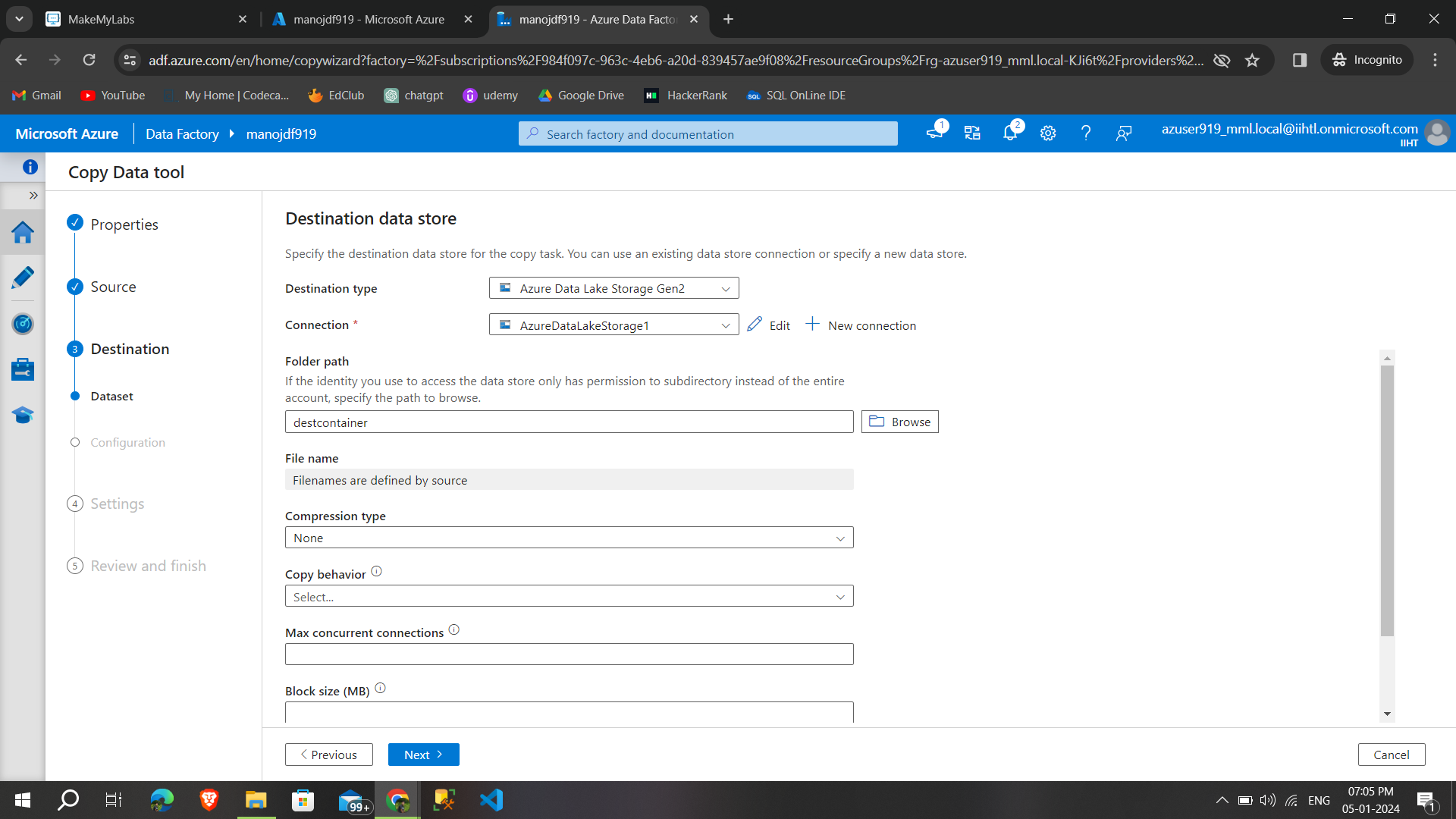
**Step 11:** select the source folder in that source storage account **then click next**



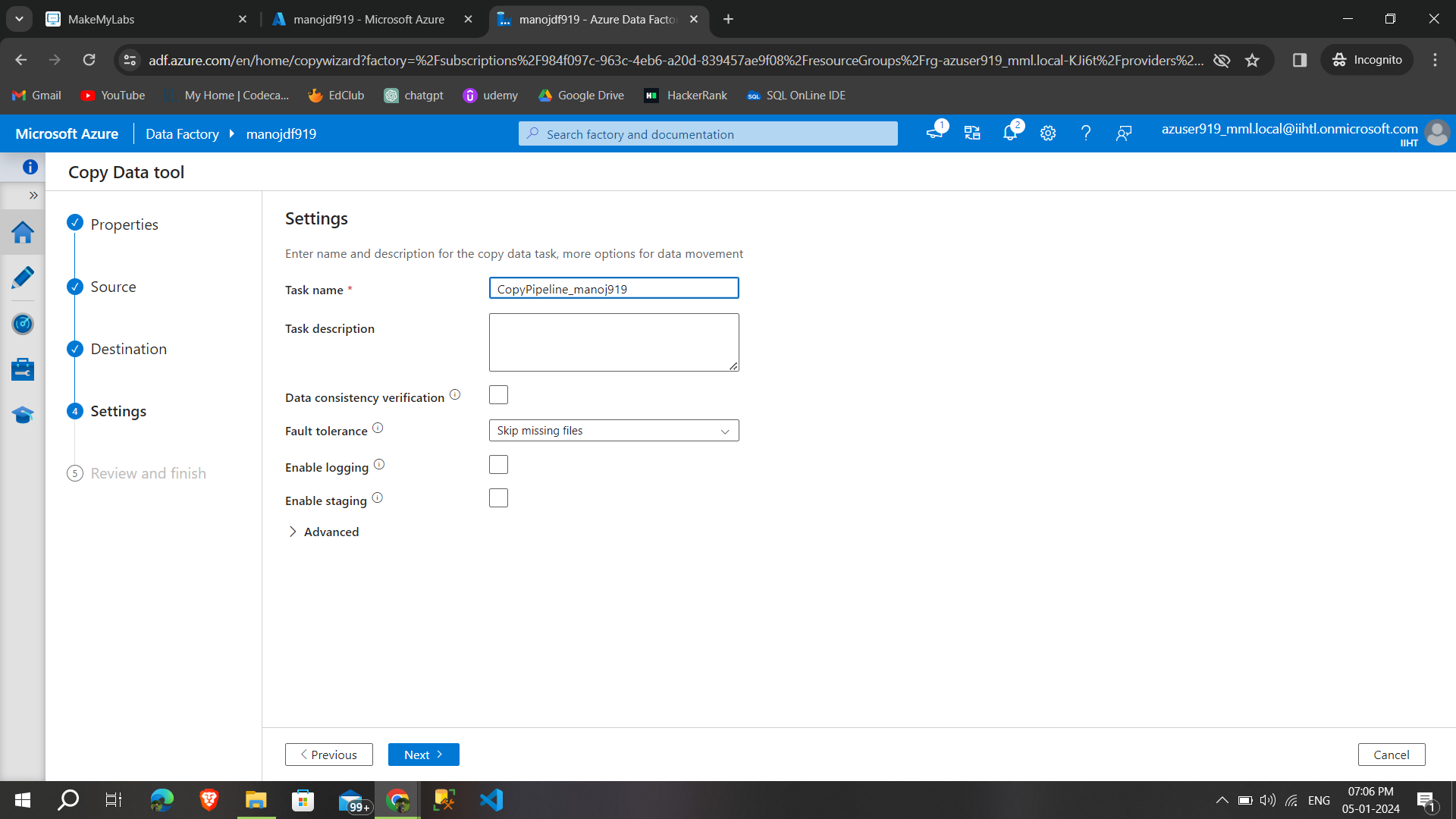
**Step 12:**then need to select **destination storage account and test the connection**



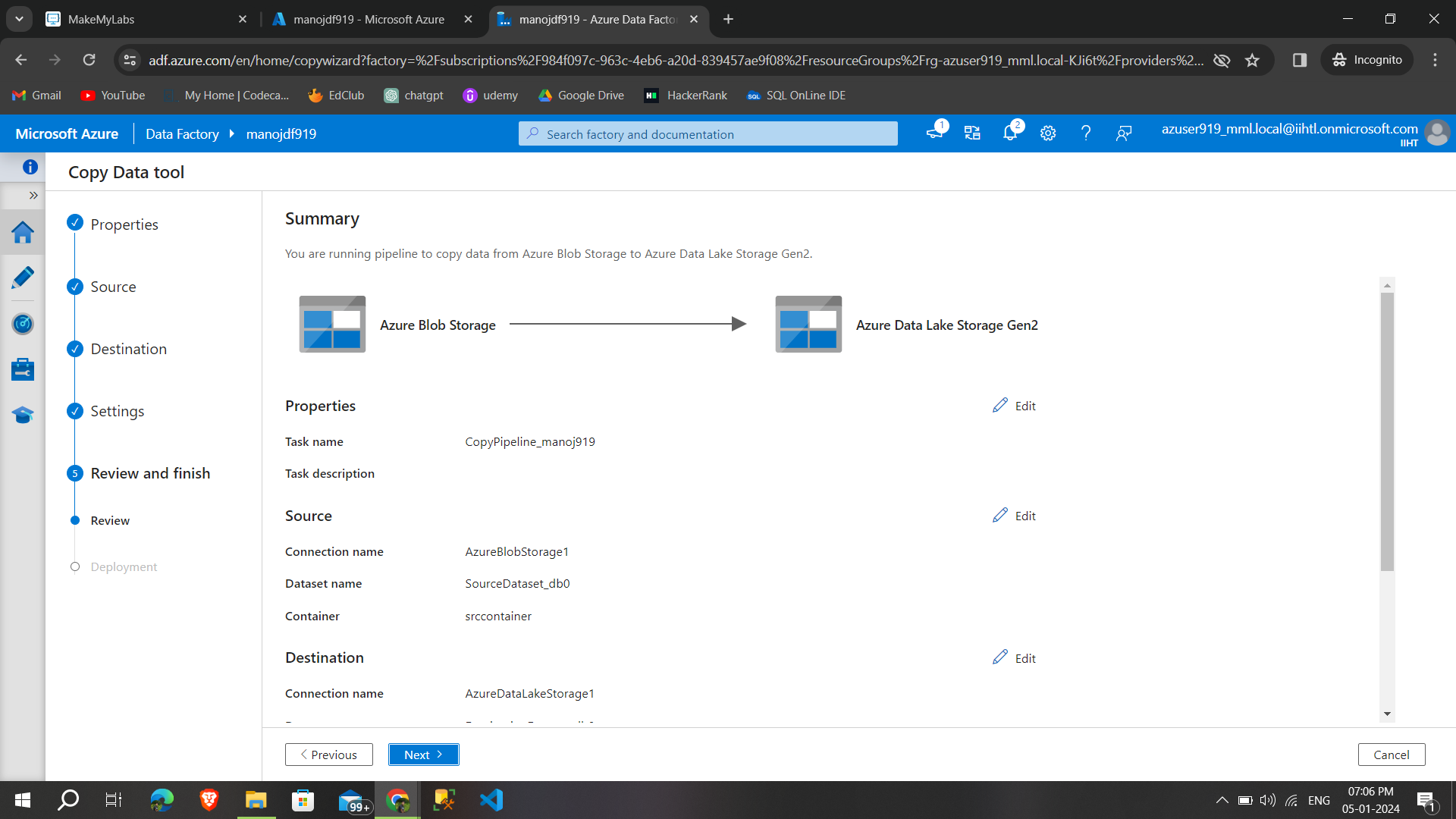
**Step 13:**then select the **destination folder that copy the files,then click next**



**Step 14:** give the taskname and **click next**



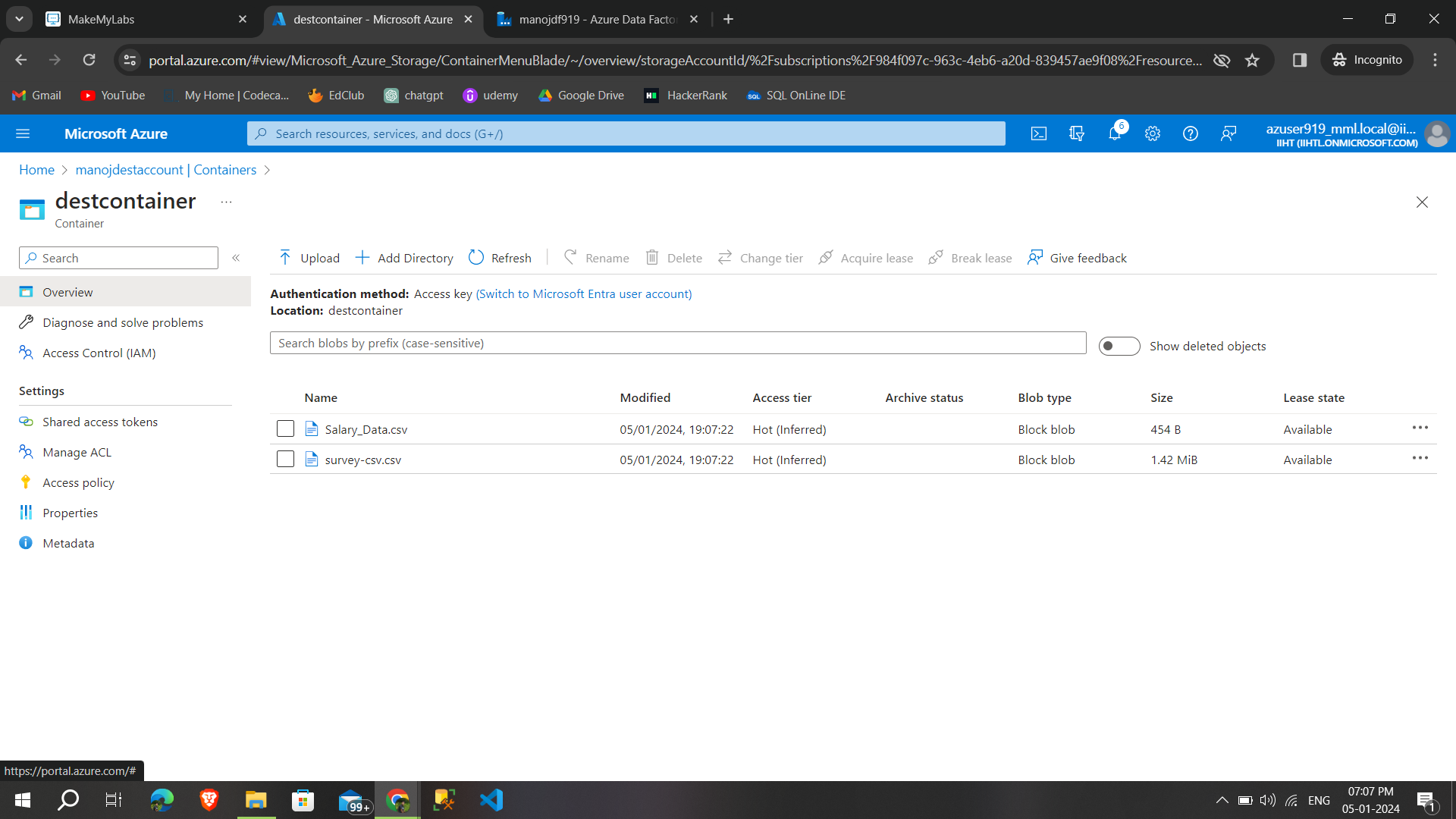
**Step 15:**In summary **we can once again check the details** that have selected after checking **click next to start process**



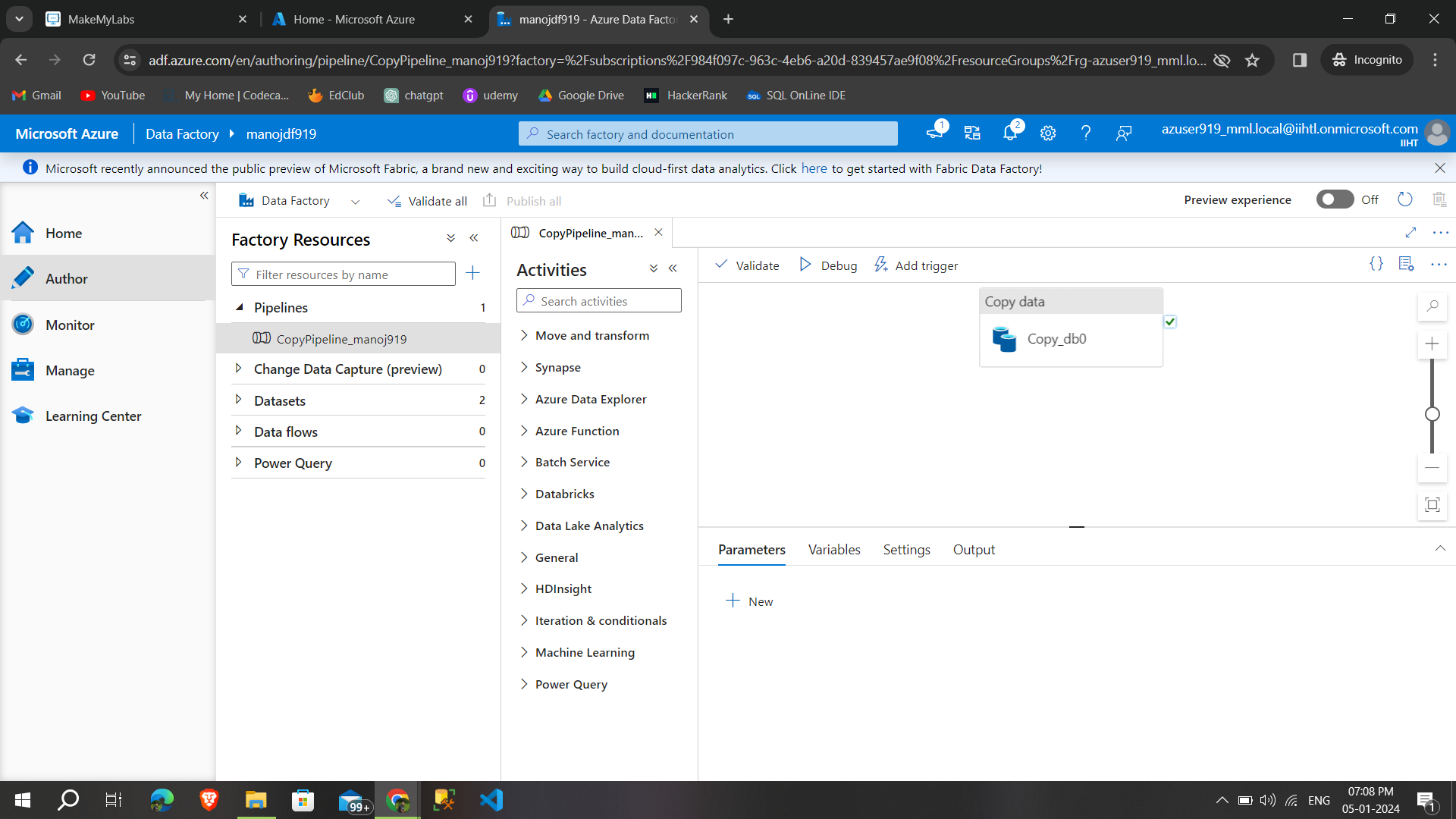
**Step 16:**once the deployment is **complete we can able to see the copied files in destination** account



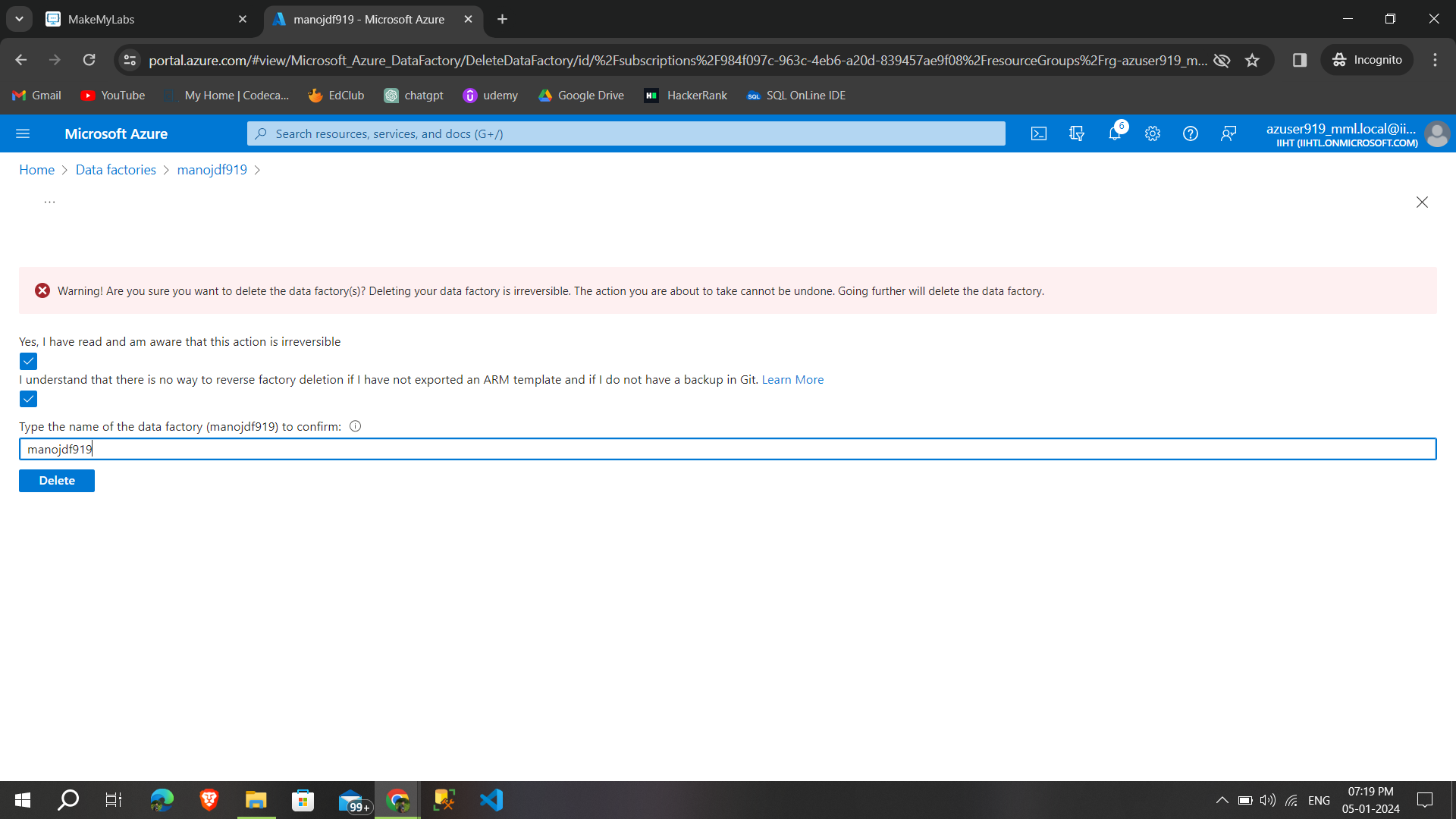
**Step 17:**we can able to **see the copied the files from source account to destination account,**

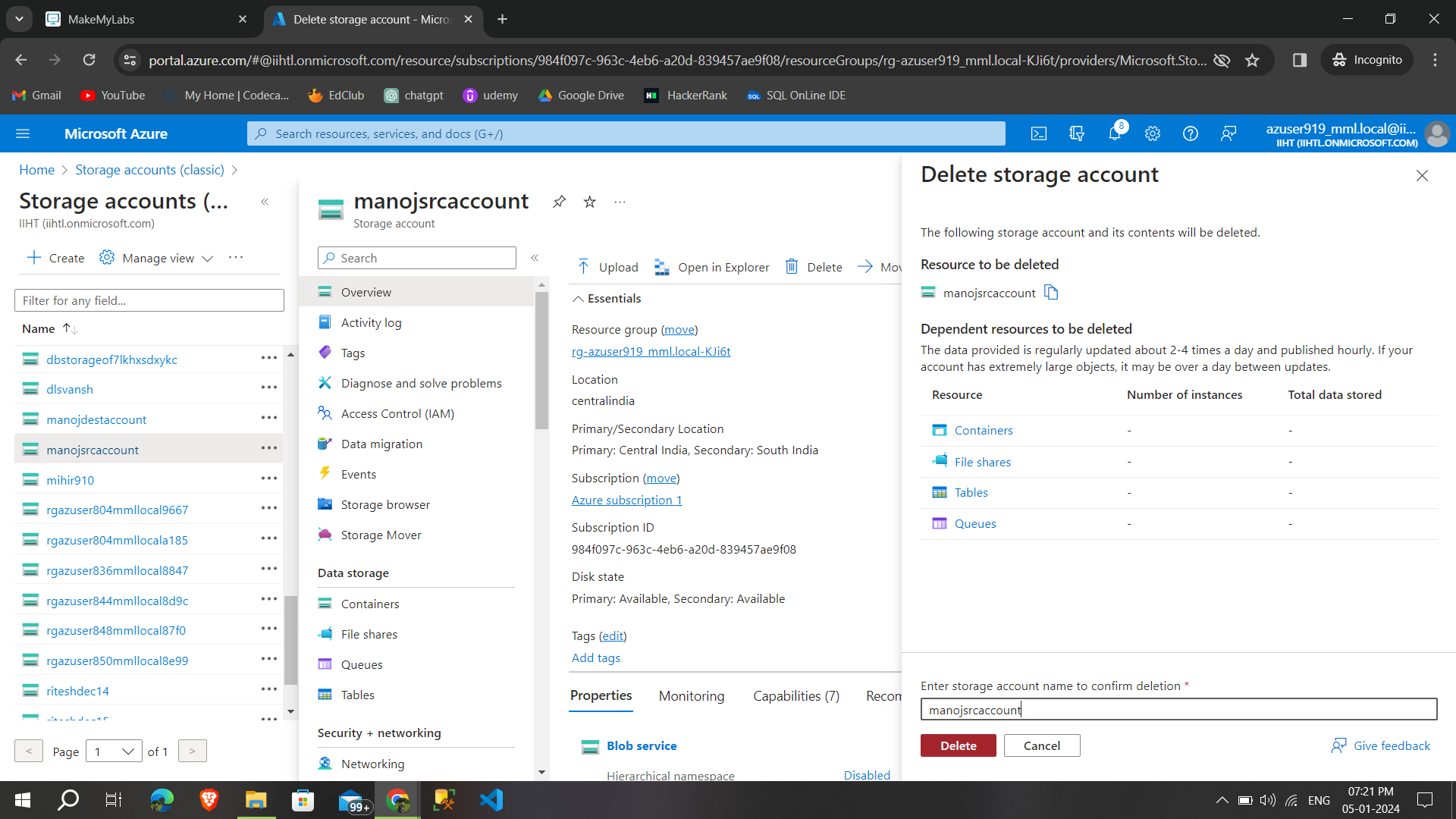


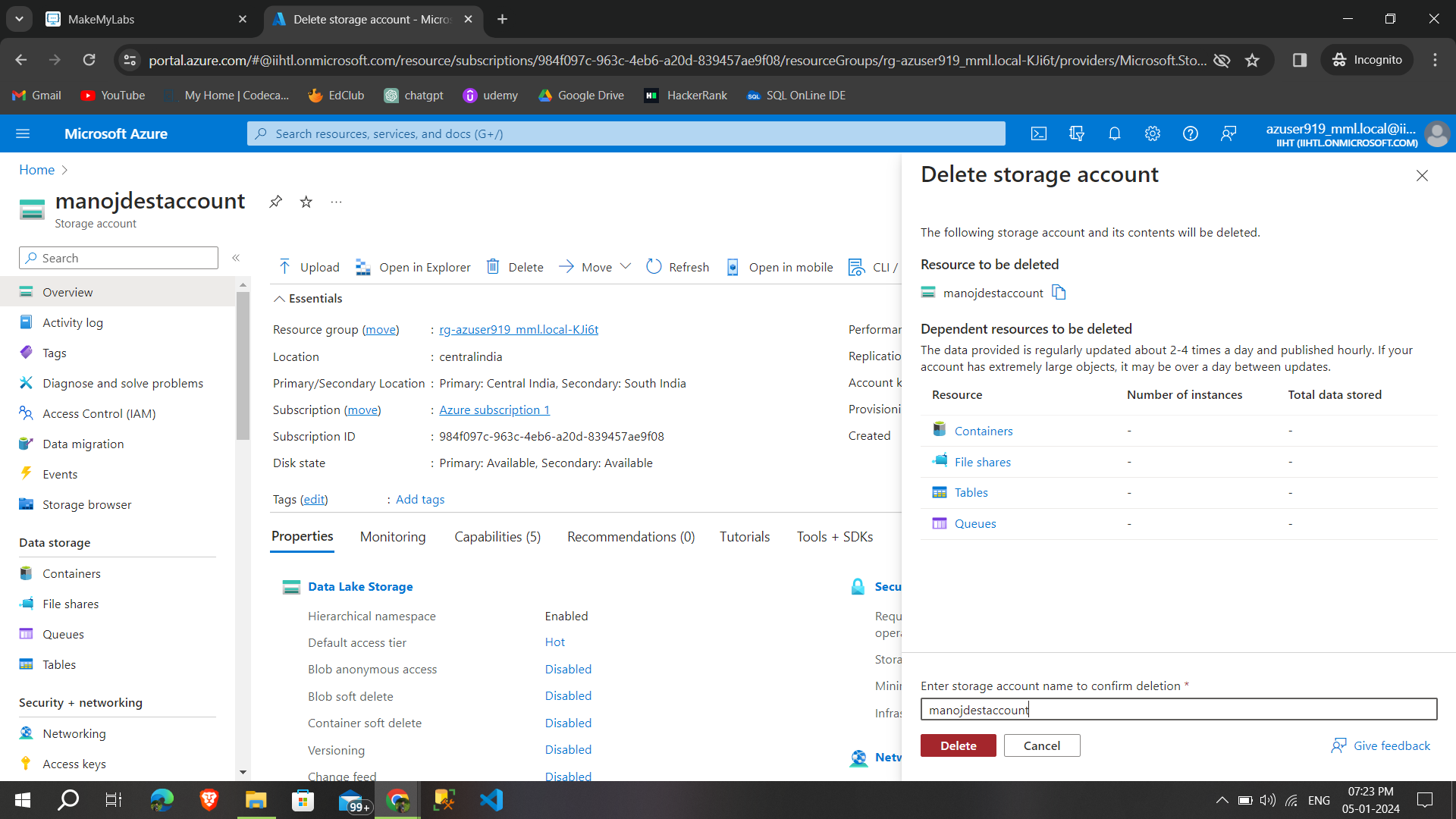
We can able to **perform the pipeline again by validate and debug the pipeline**



**Then I deleted the data factory service**



Also **deleted the storage accounts that created previously** 



These are the steps for Copy activity-ADF