DATABRICKS Assessment 3

Q) Execute & explain, Azure datafactory and its copy activity.

AZURE DATA FACTORY: Azure Data Factory (ADF) is a cloud-based tool from Microsoft Azure that helps you manage data movement and transformation tasks. It allows you to create, schedule, and oversee workflows for moving data from one place to another.

A key feature of ADF is the Copy Activity, which handles the actual data movement between different storage locations. It supports a wide range of sources and destinations, including Azure Blob Storage, SQL Database, and other cloud platforms like AWS S3 and Google Cloud Storage.

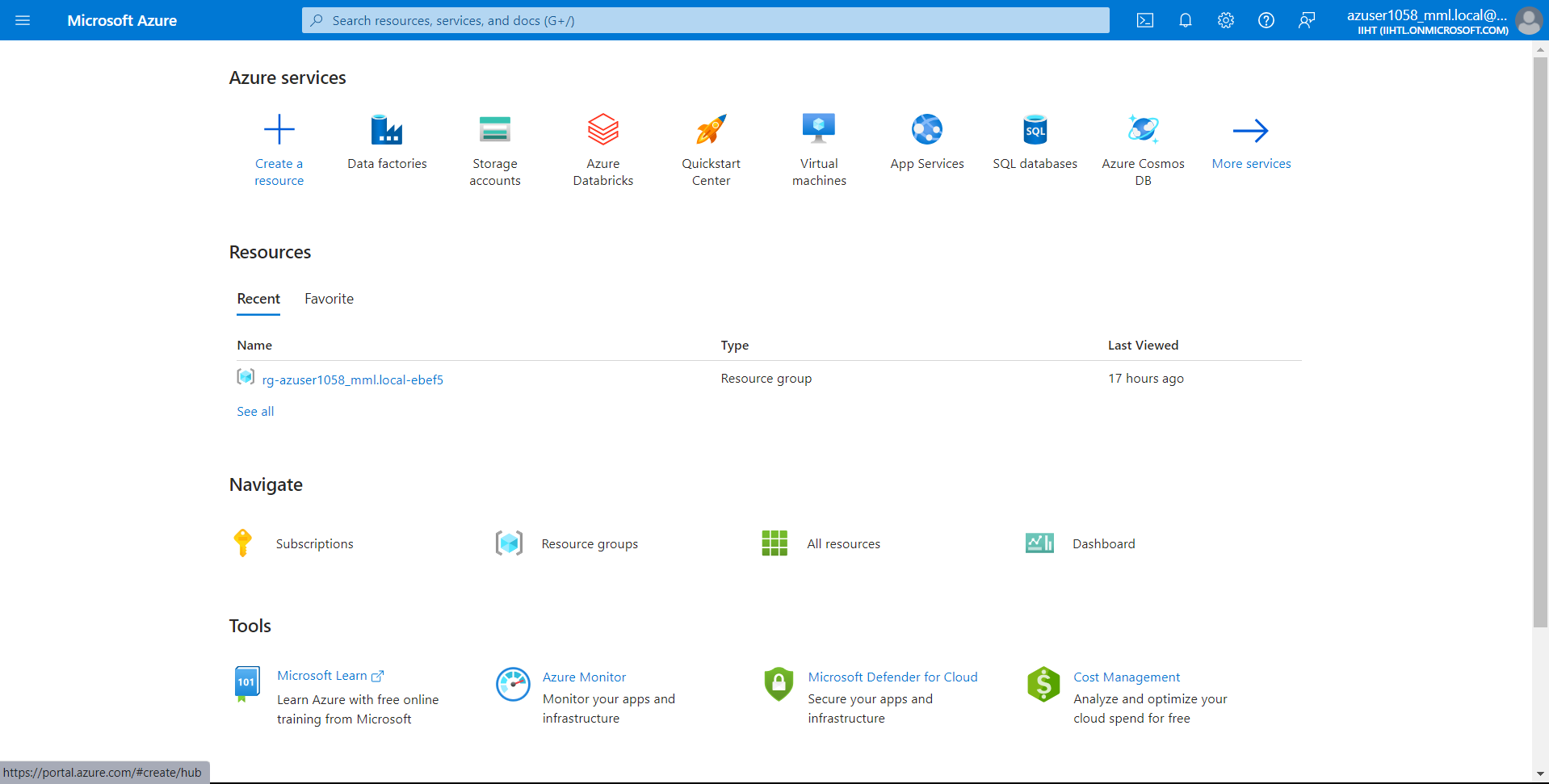
With Copy Activity, you can do things like bringing data into Azure for analysis, loading data into databases or data lakes, replicating data to keep it consistent across environments, and migrating data from on-premises systems to the cloud.

We can perform various data movement operations, such as:

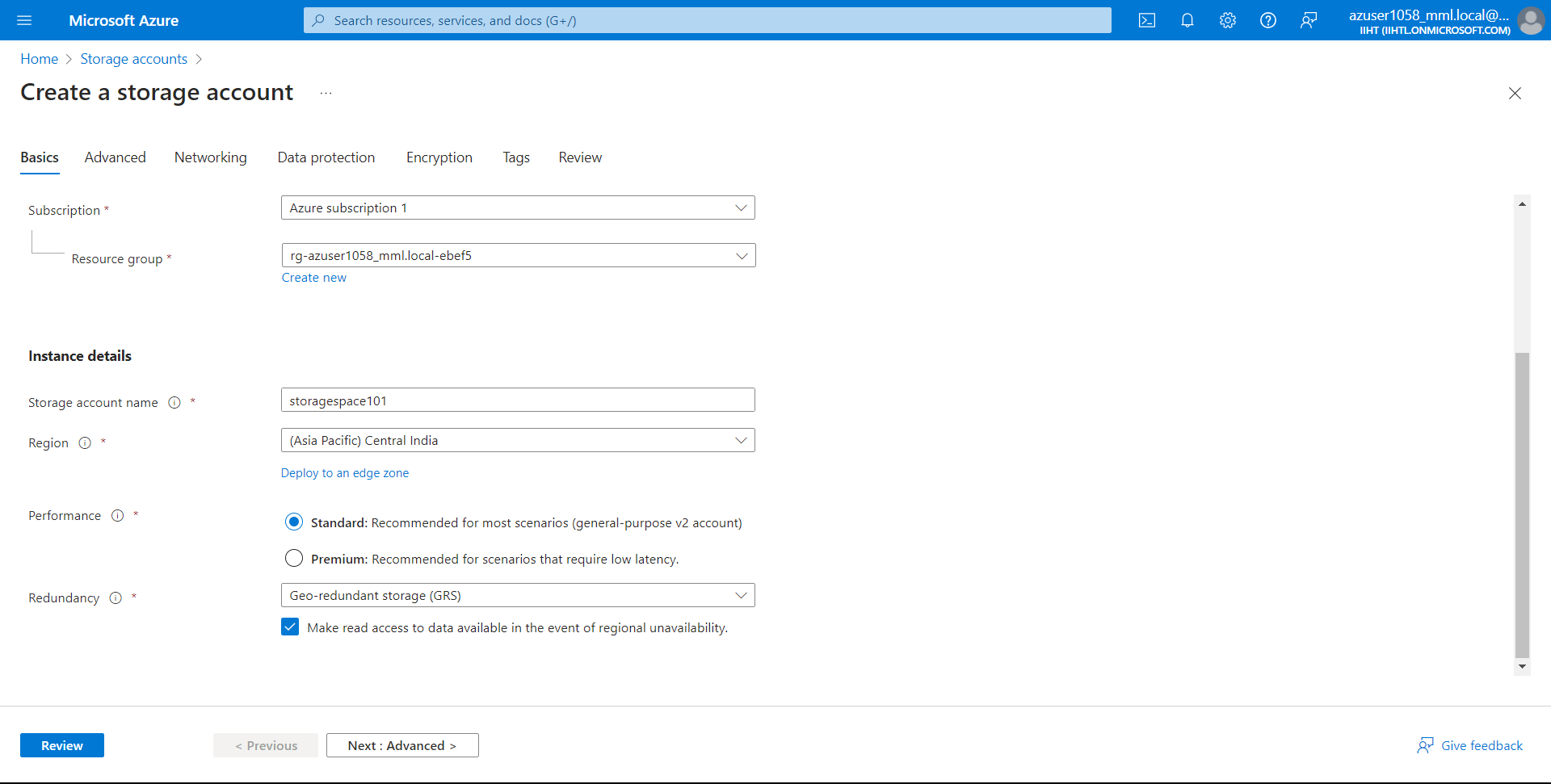
* Data Ingestion: Copy data from source data stores to Azure or other cloud platforms for further processing and analysis. This could involve moving data from on-premises databases, files, or cloud-based applications.
* Data Loading: Load data into data warehouses or data lakes for storage and analysis. This could involve loading data into Azure Synapse Analytics, Azure SQL Data Warehouse, or Azure Data Lake Storage.

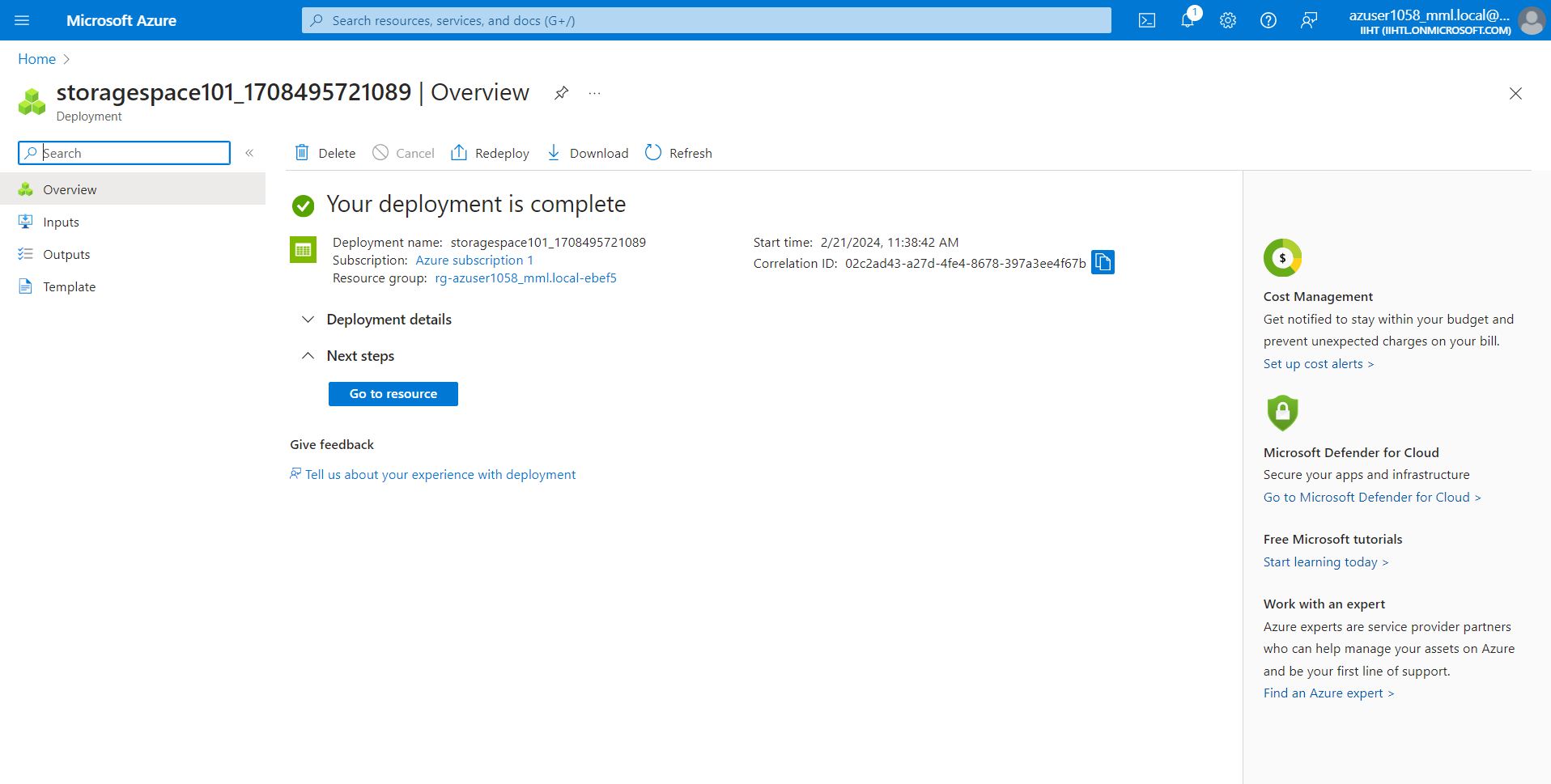
Here are steps for data factory and copy activity :

Go to Azure Portal and sign in

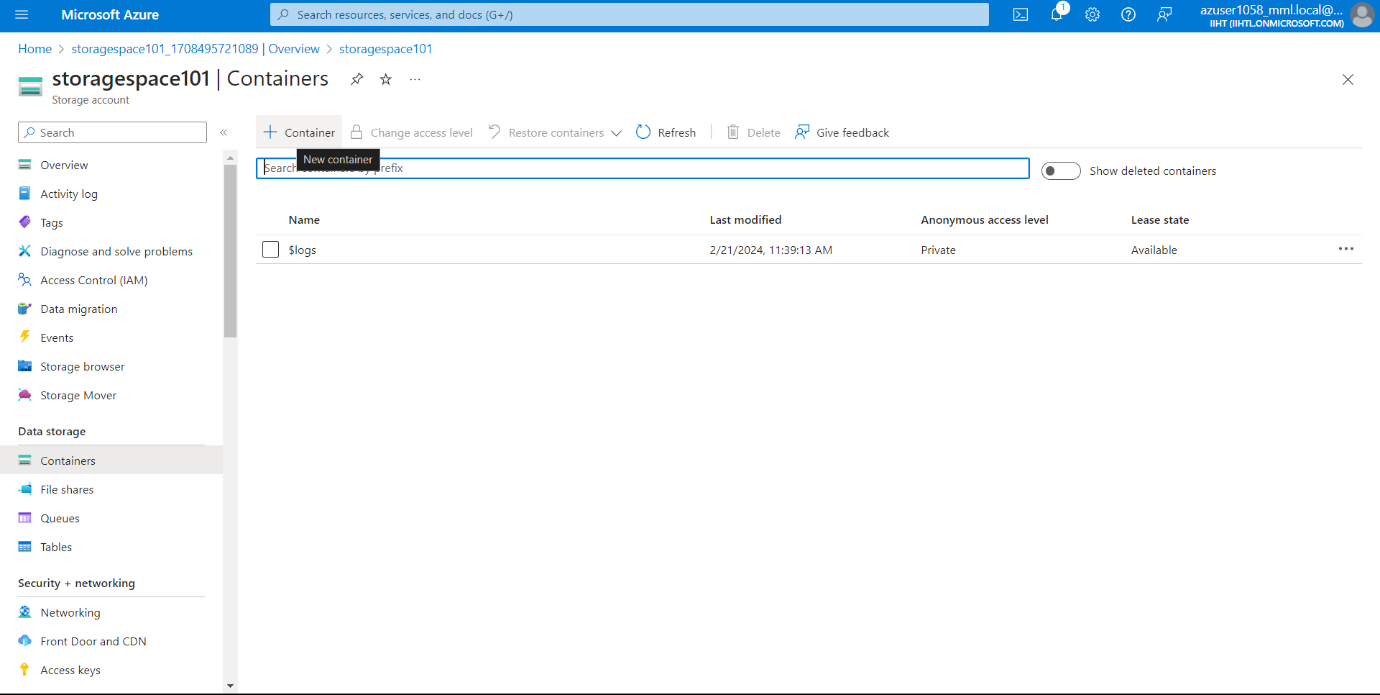


Now create a storage “Storagespace101” ie., “Source” and create the storage and wait for deployment

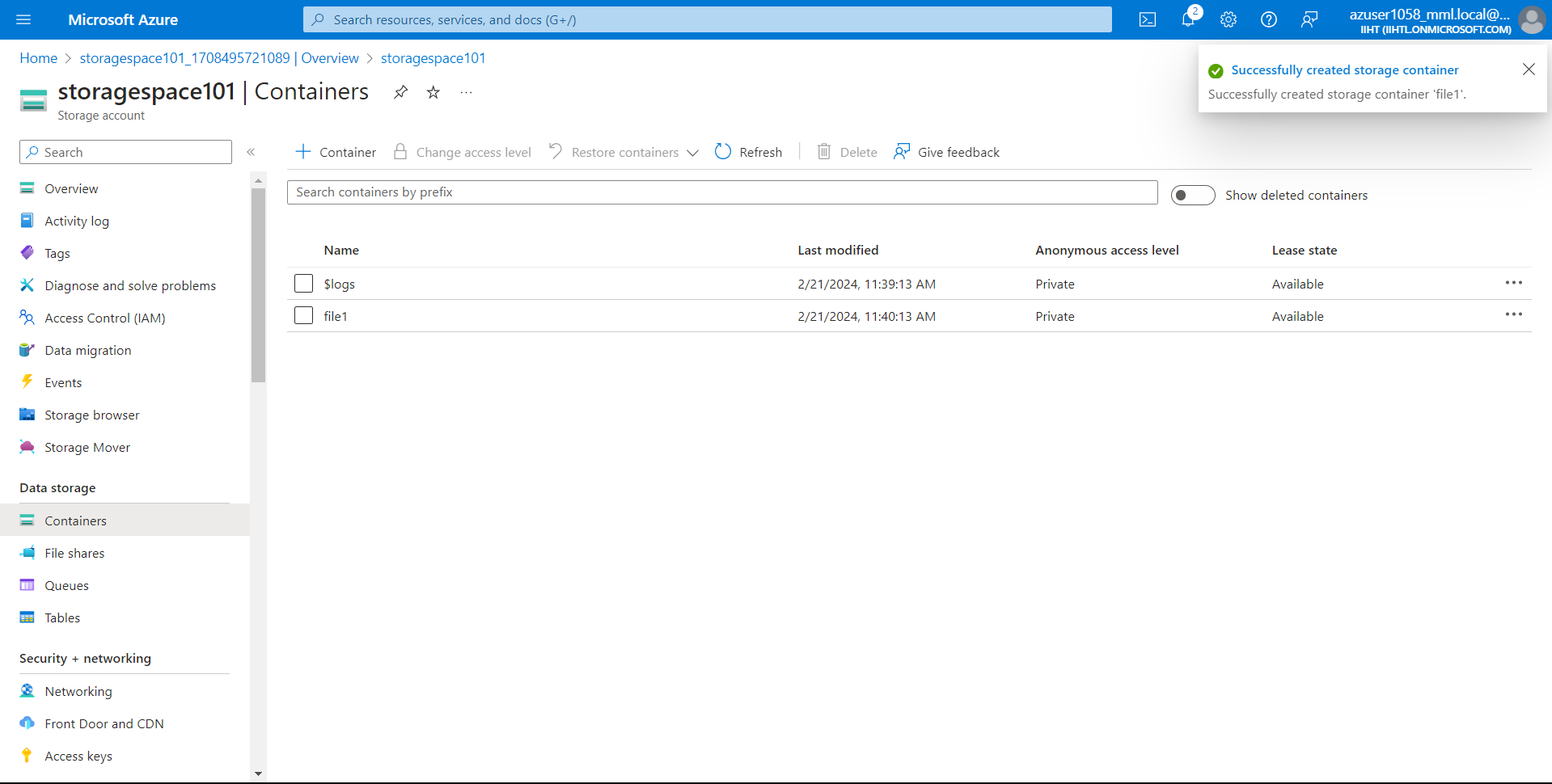


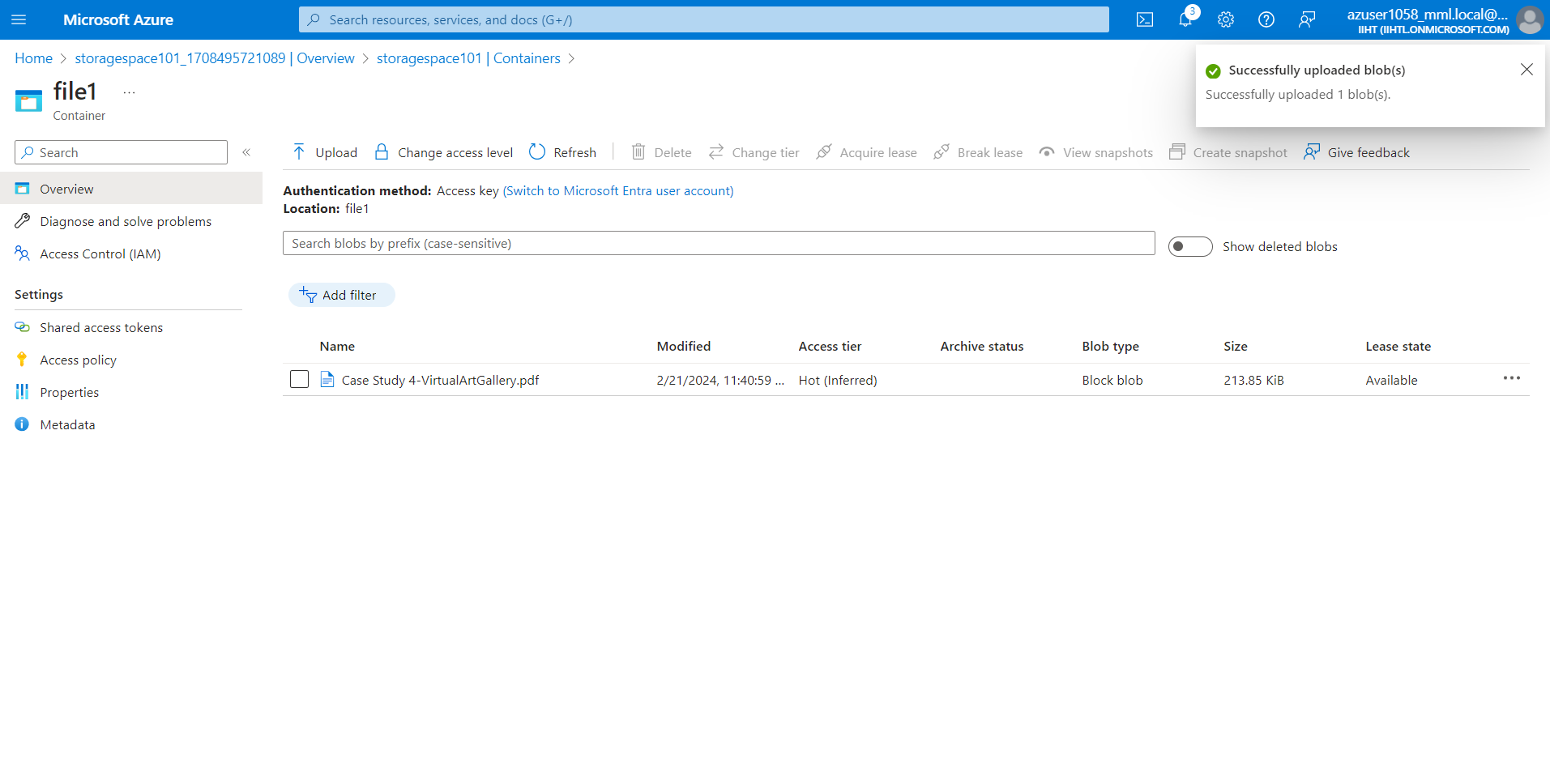


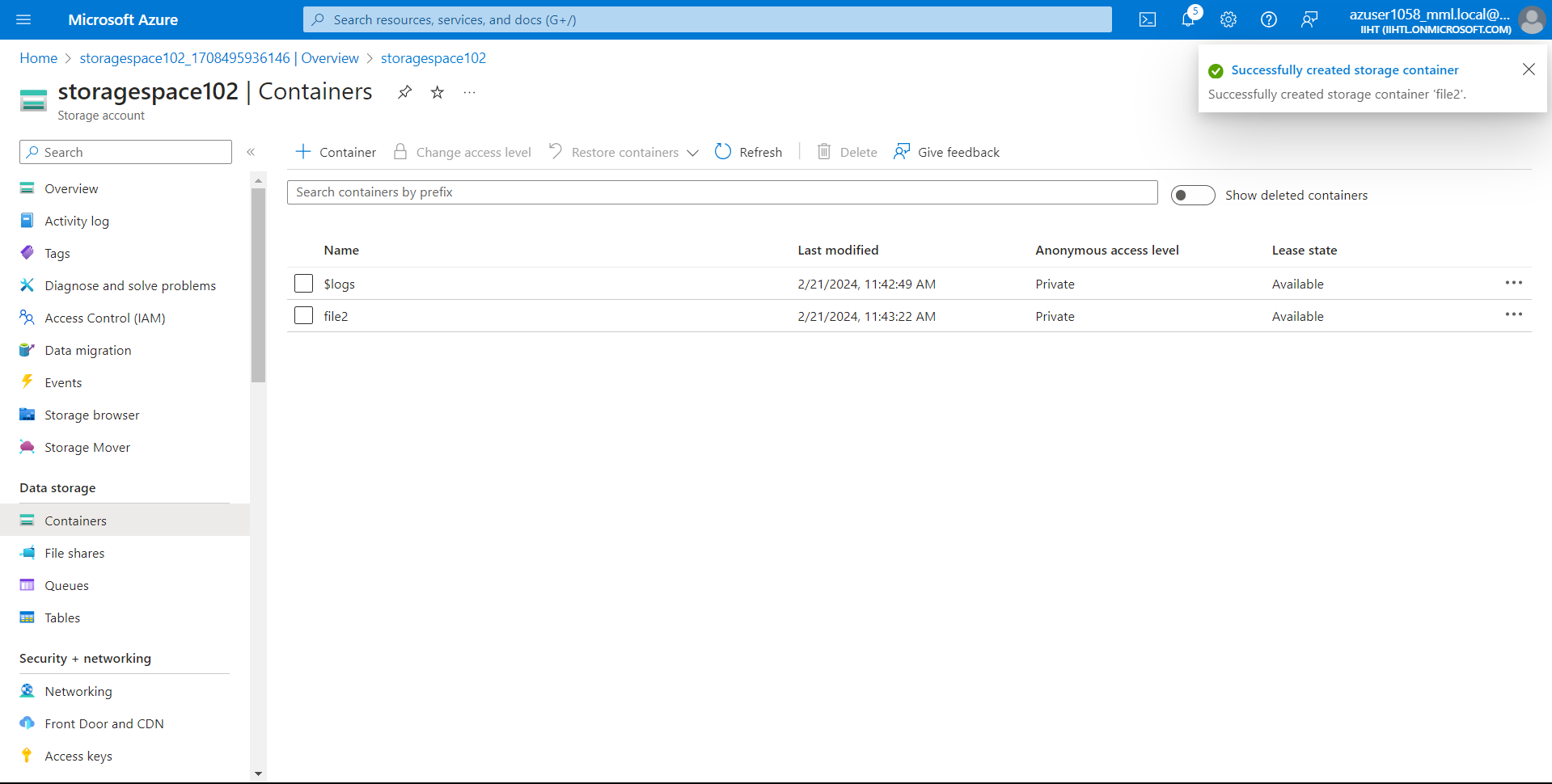
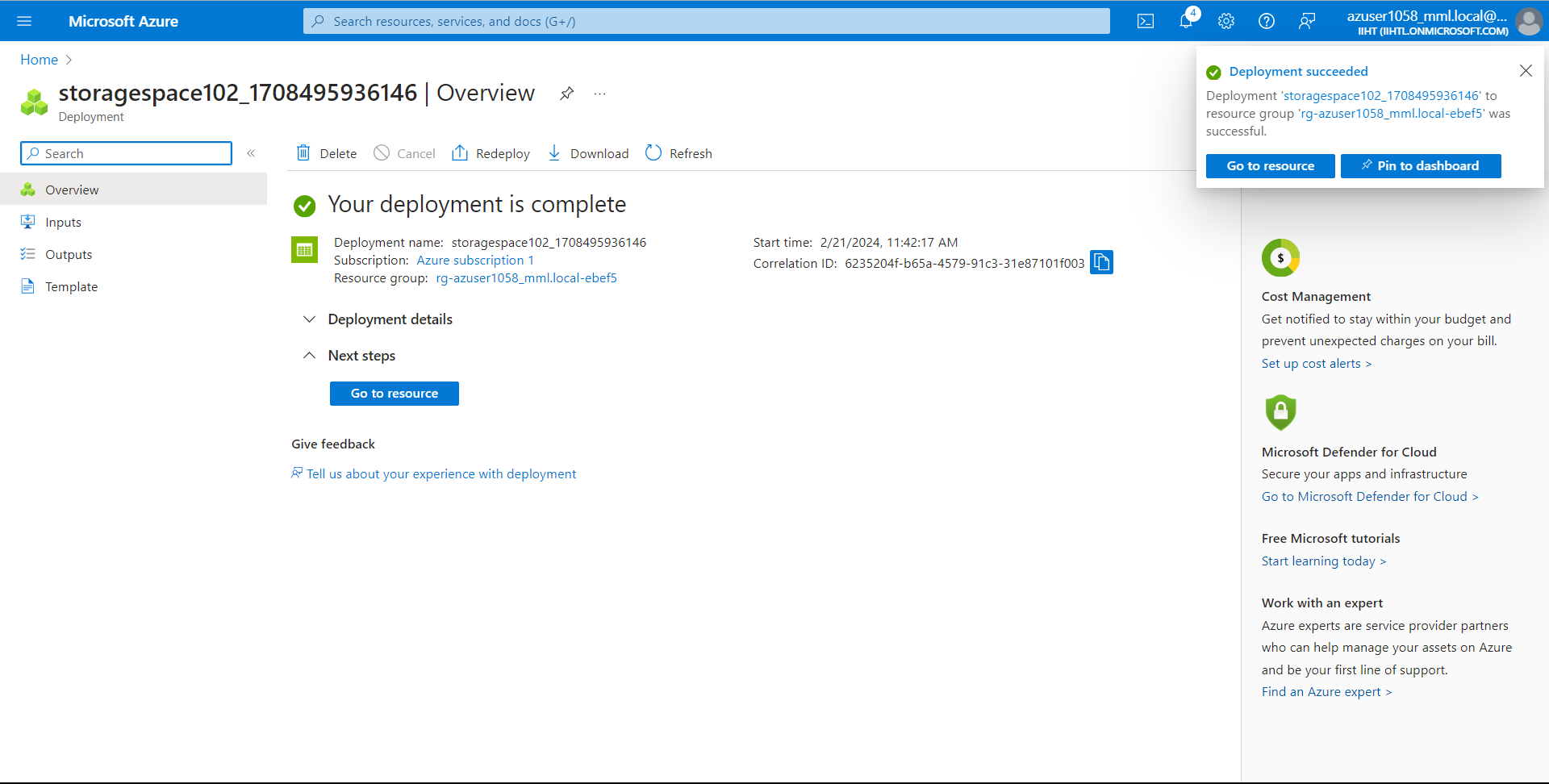
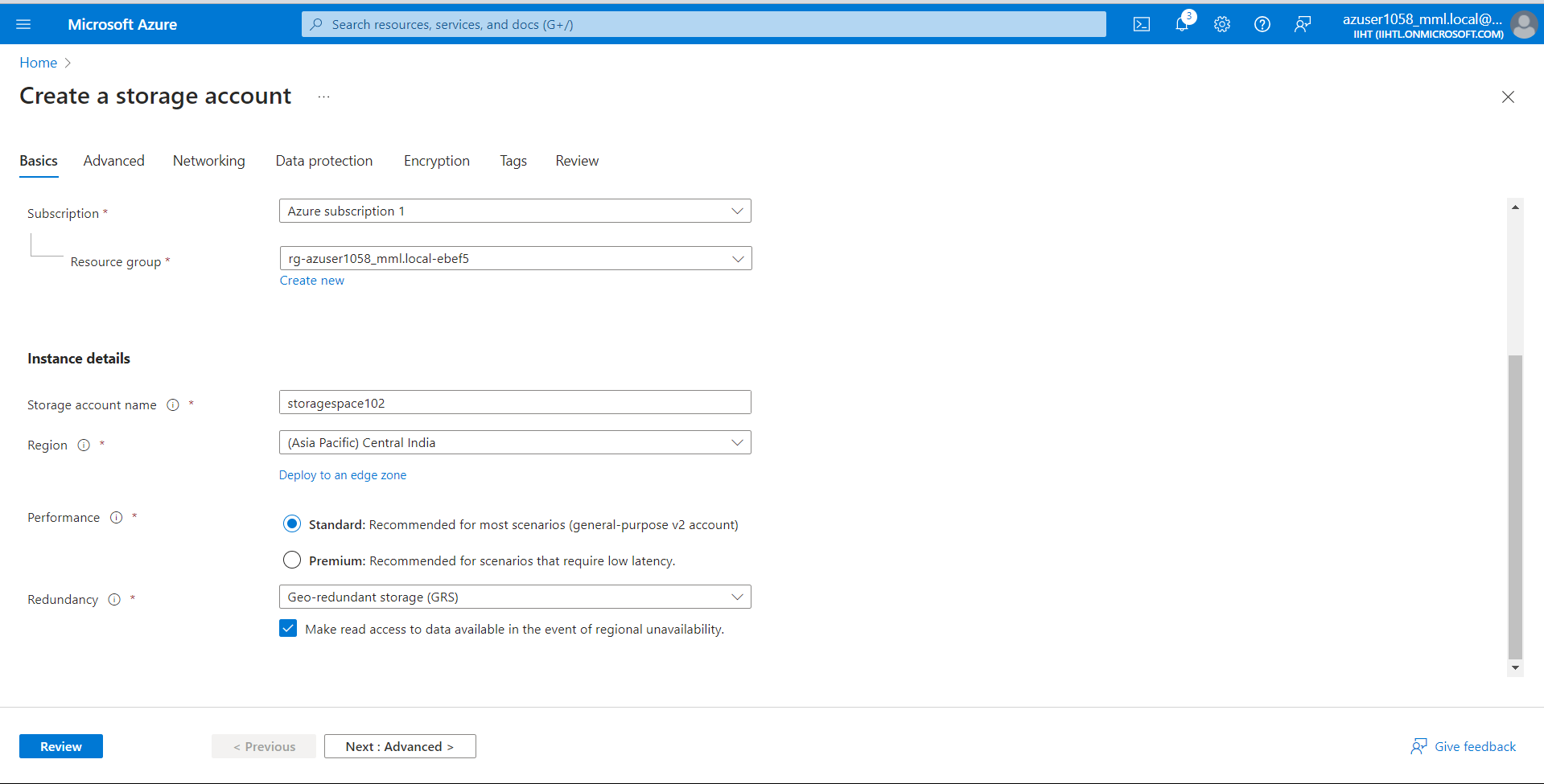
After deployment, click on “Go to resource” and Select “Container” and open it

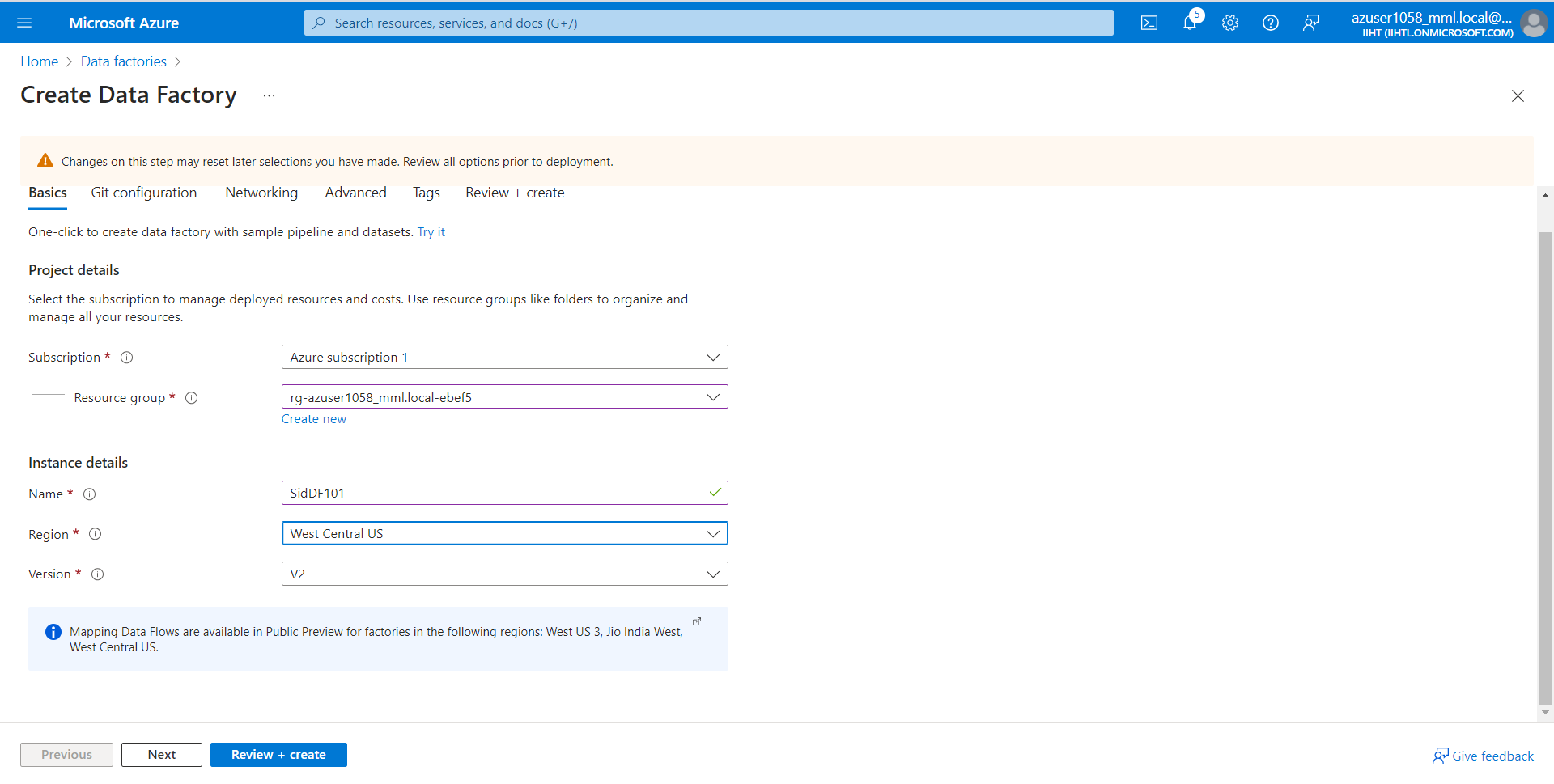


Now lets create container and name it as “file1” and upload a random file to the file1 which we have created.

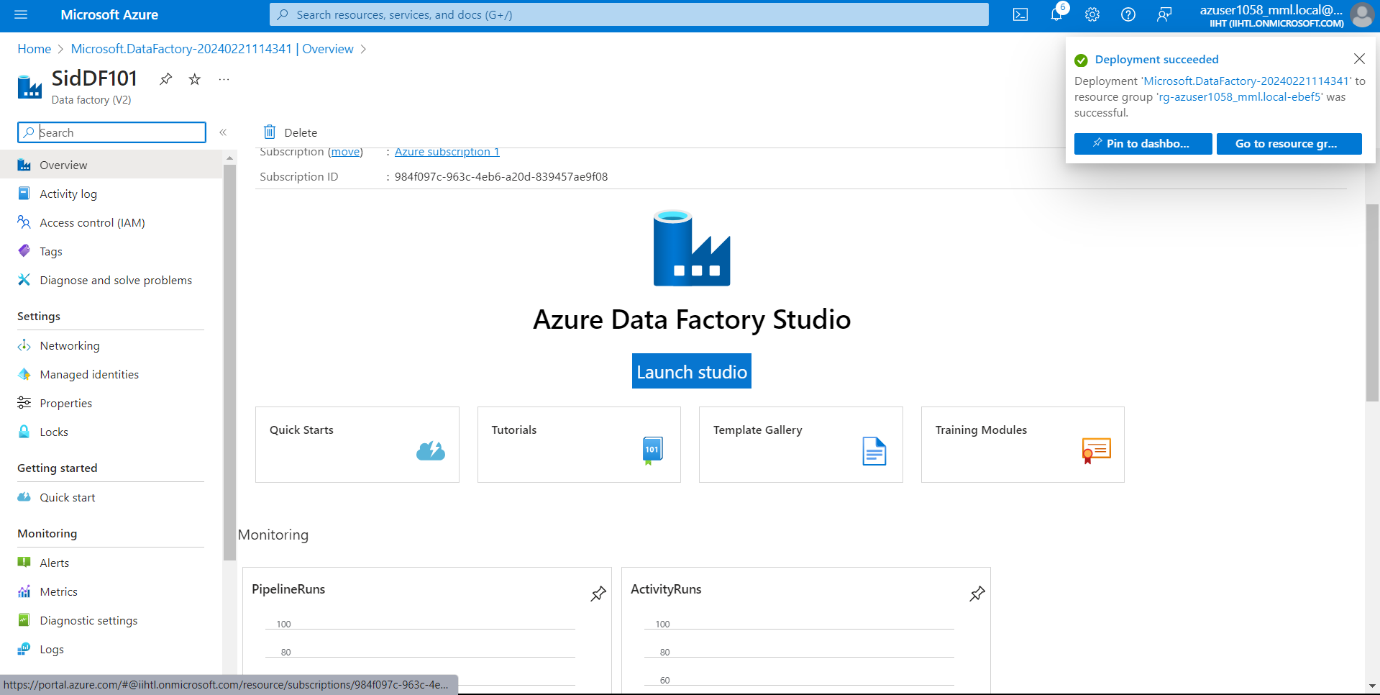




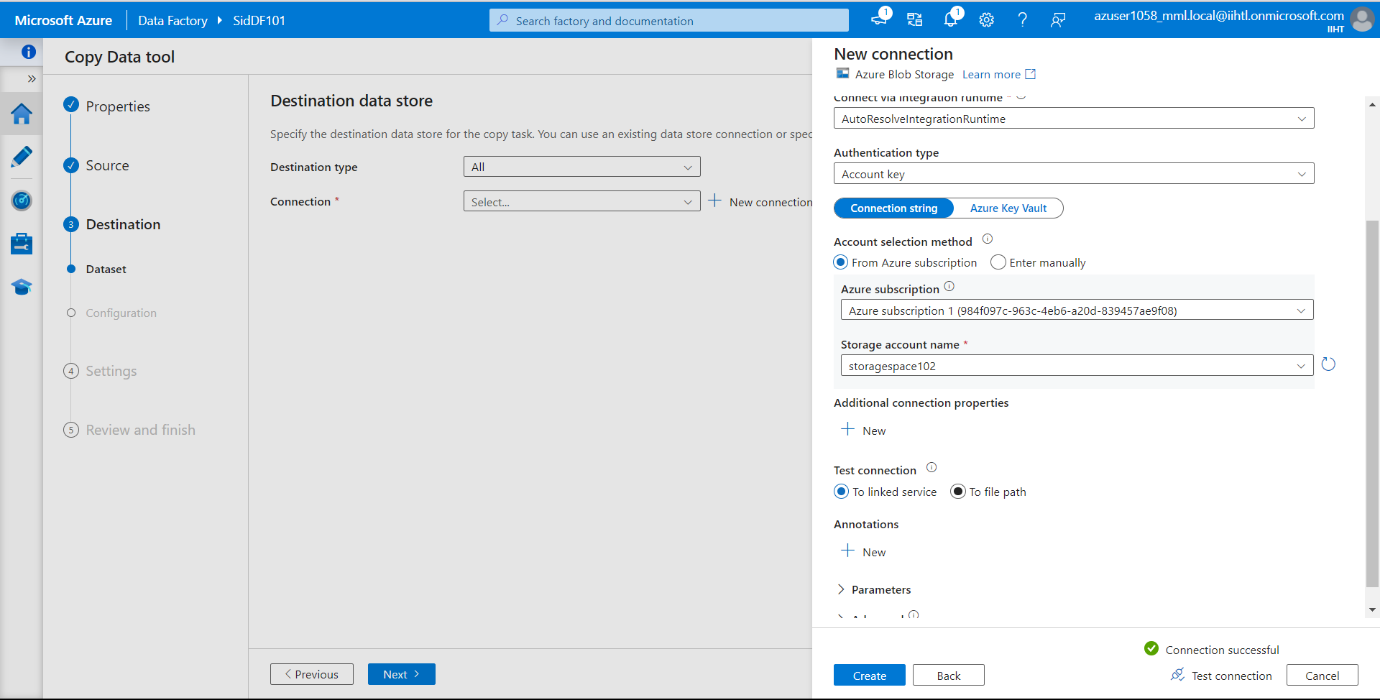
Now lets create another storage ie, “StorageSpace102” like previous steps and lets create container and name it as “file2”.

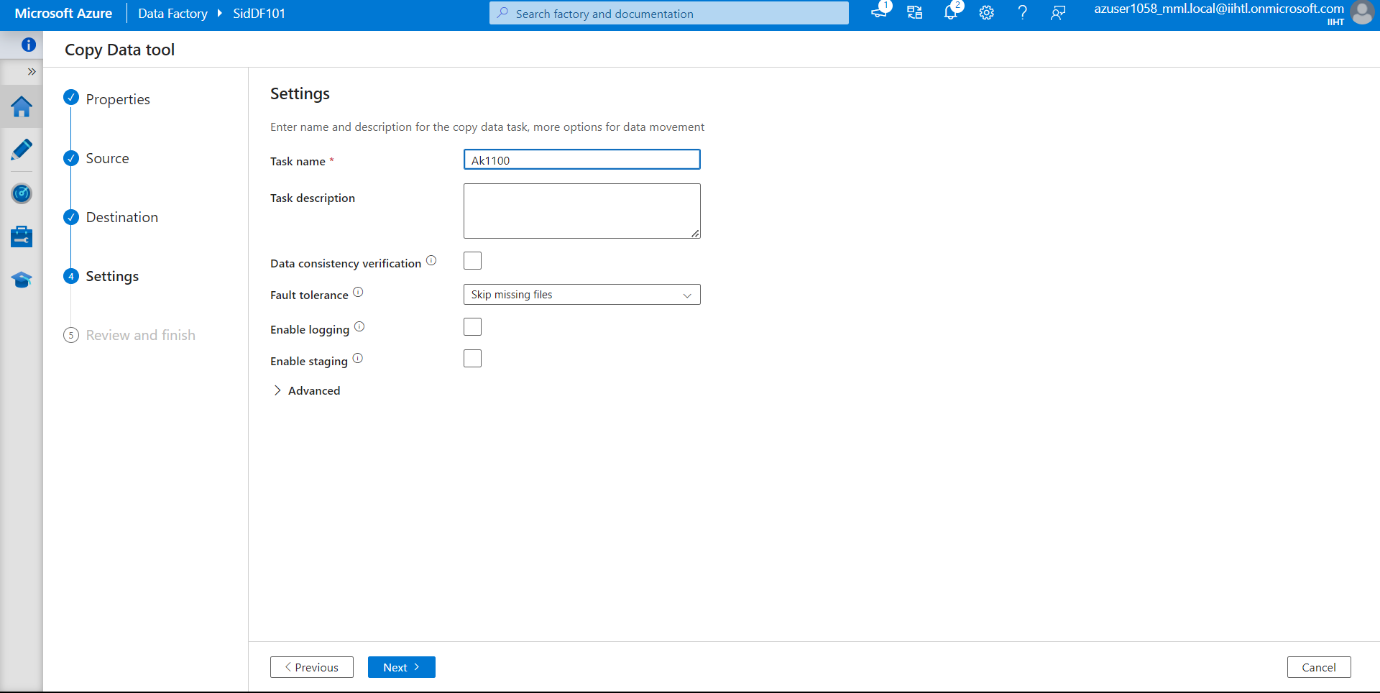
Now lets create a data factory and name it as “SidDF101” and select any region and click on review+create 

After creating we will see an image shown below

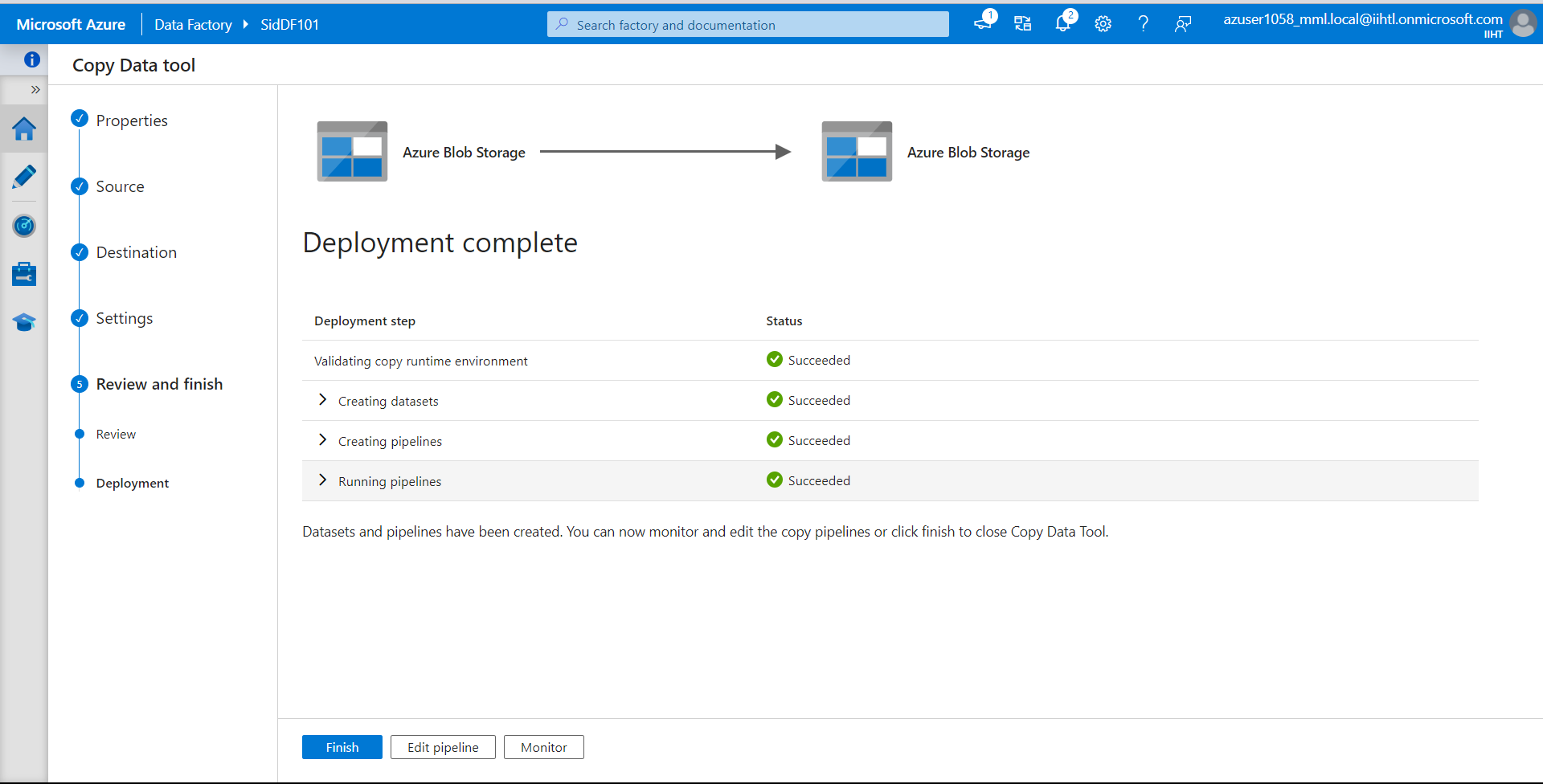


Now we need to launch studio and we need to create connection between source and destination and then we need to check the file is copied from “file1” to “file2”

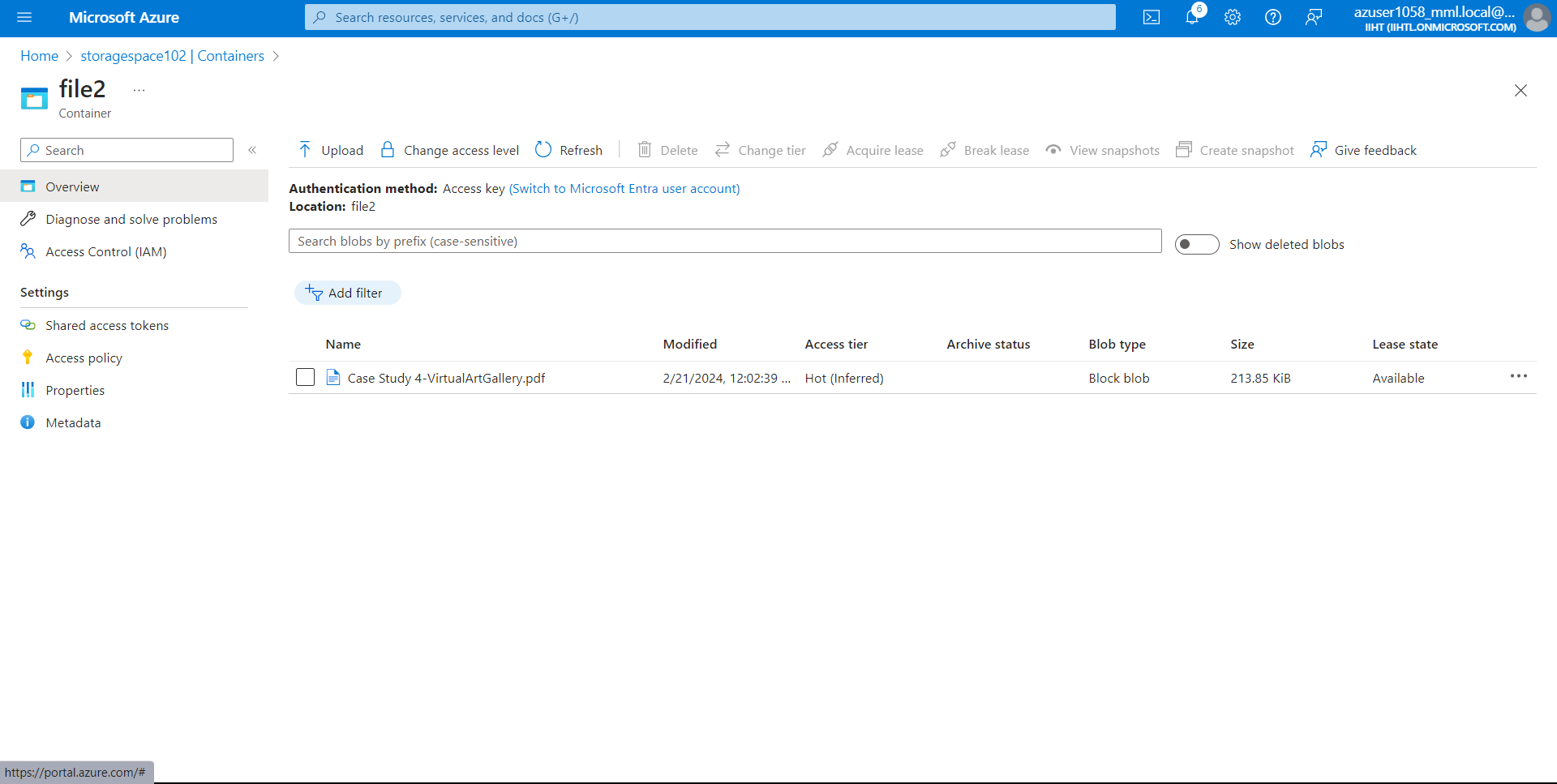
And name the pipeline on random.



Now here shown below the connection is successful and we have copied activity from source to destination



Now lets check In the “Storagespace102” in the container “file2” so that the file is successfully copied or not.



Here we can see that we have successfully copied file from one storage to another storage.