Daily Assessment

Name : Chakradhar Bhogapurapu

Date : 03/01/2024

Mail : chakribhogapurapu@gmail.com

Topic : Azure storage Explorer

**Azure data lake storage generation 2**

A data lake is a single centralized repository to store all data both structured and unstructured. Azure Data Lake Storage is a cloud-based, enterprise data lake solution.

**Benefits if Azure Data Lake Storage,**

🡪Hadoop compatible access.

🡪Security

🡪Performance

🡪Data redundancy

**There are 4 stages in processing big data.**

🡪Ingest - The ingestion phase identifies the technology and processes that are

used to acquire the source data. This data can come from files, logs, and other

types of unstructured data that must be put into the data lake.

🡪Store - The store phase identifies where the ingested data should be placed.

Azure Data Lake Storage Gen2 provides a secure and scalable storage solution

that is compatible with commonly used big data processing technologies.

🡪Prep and train - The prep and train phase identifies the technologies that are

used to perform data preparation and model training and scoring for machine

learning solutions. Common technologies that are used in this phase are Azure

Synapse Analytics, Azure Databricks, Azure HDInsight, and Azure Machine

🡪Learning - Model and serve - Finally, the model and serve phase involves the

technologies that will present the data to users. These technologies can include

visualization tools such as Microsoft Power BI, or analytical data stores such

as Azure Synapse Analytics.

**Data warehousing:**

🡪Data warehousing has evolved in recent years to integrate large volumes of data

stored as files in a data lake with relational tables in a data warehouse. In a typical

example of a data warehousing solution, data is extracted from operational data stores,

such as Azure SQL database or Azure Cosmos DB, and transformed into structures

more suitable for analytical workloads.

🡪There are multiple ways to implement this kind of data warehousing architecture. The

diagram shows a solution in which Azure Synapse Analytics hosts pipelines to

perform extract, transform, and load (ETL) processes using Azure Data Factory

technology.

🡪These processes extract data from operational data sources and load it into a data lake

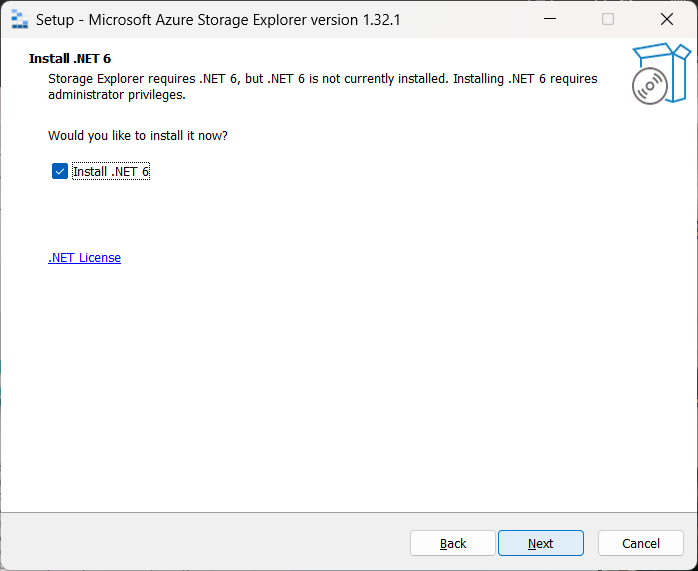
hosted in an Azure Data Lake Storage Gen2 container. The data is then processed and

loaded into a relational data warehouse in an Azure Synapse Analytics dedicated SQL

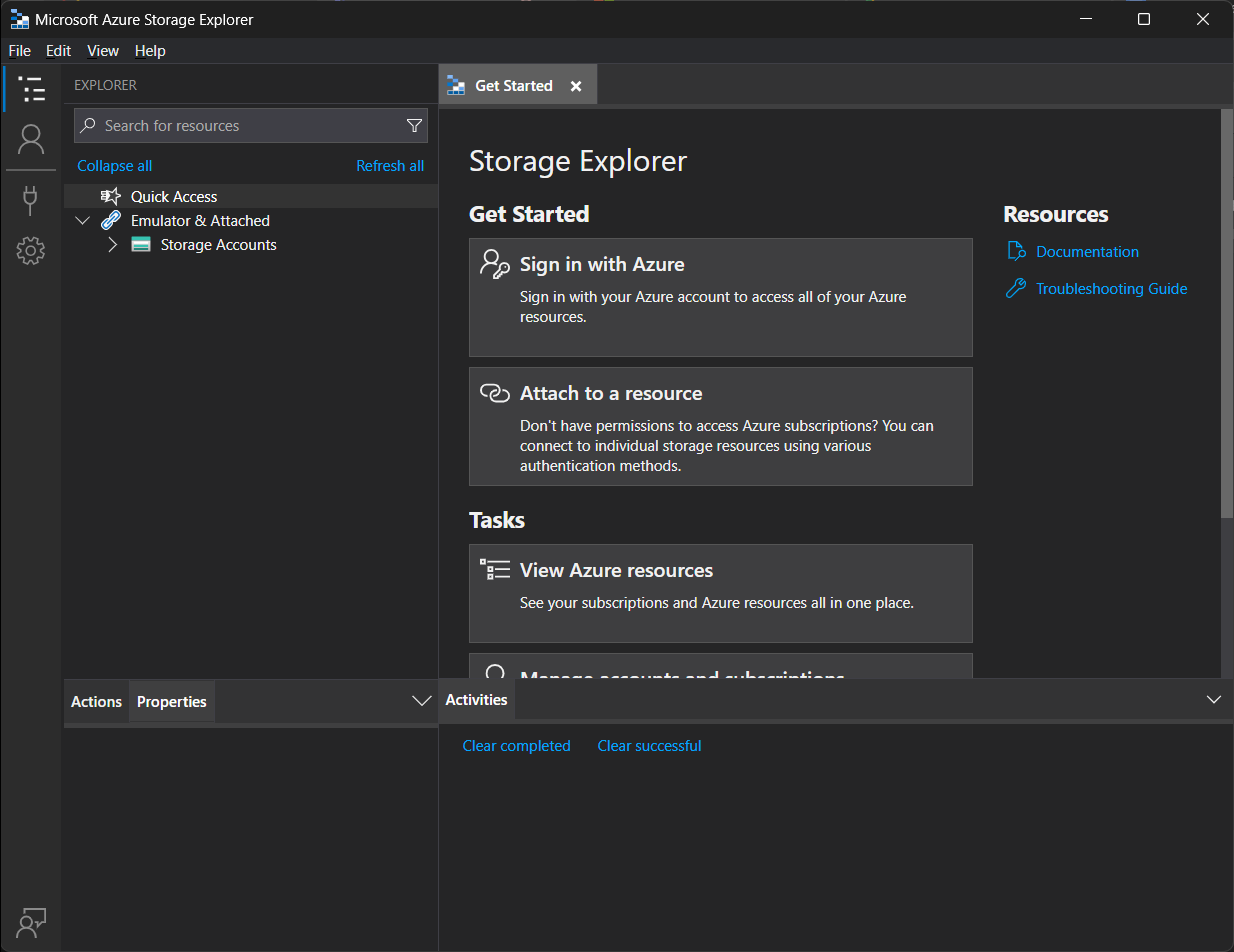
pool, from where it can support data visualization and reporting using Microsoft

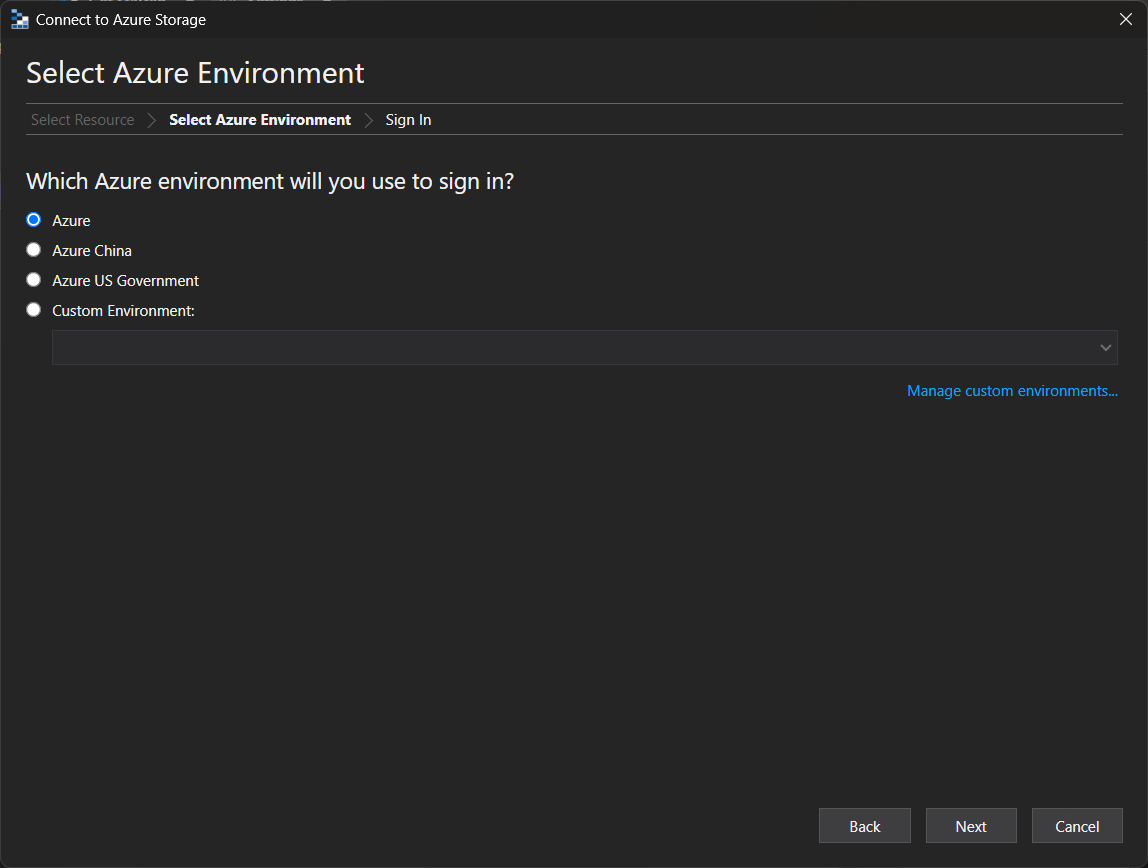
Power BI.

**Creation of Azure Data Lake Storage Access Gen2**

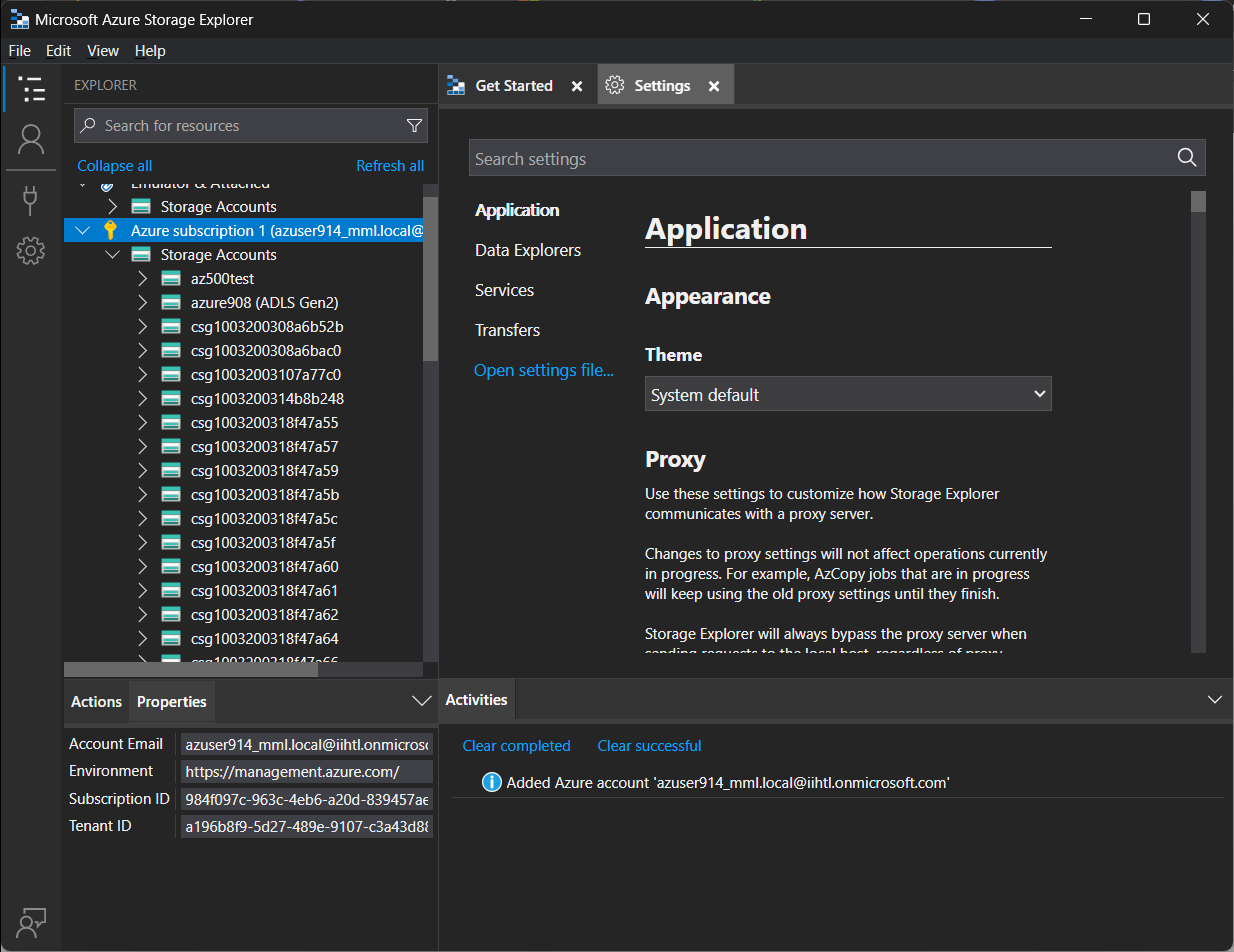
Download the azure storage explorer from the browser. 

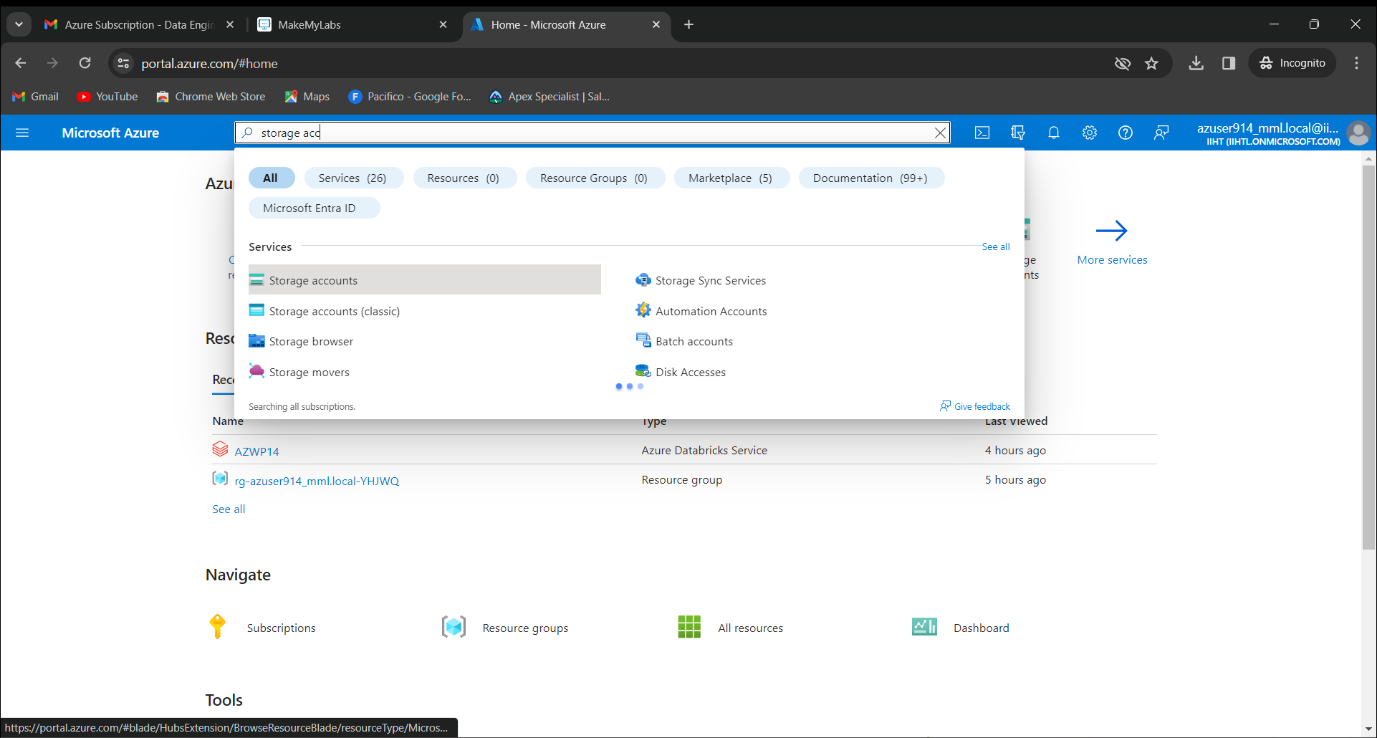
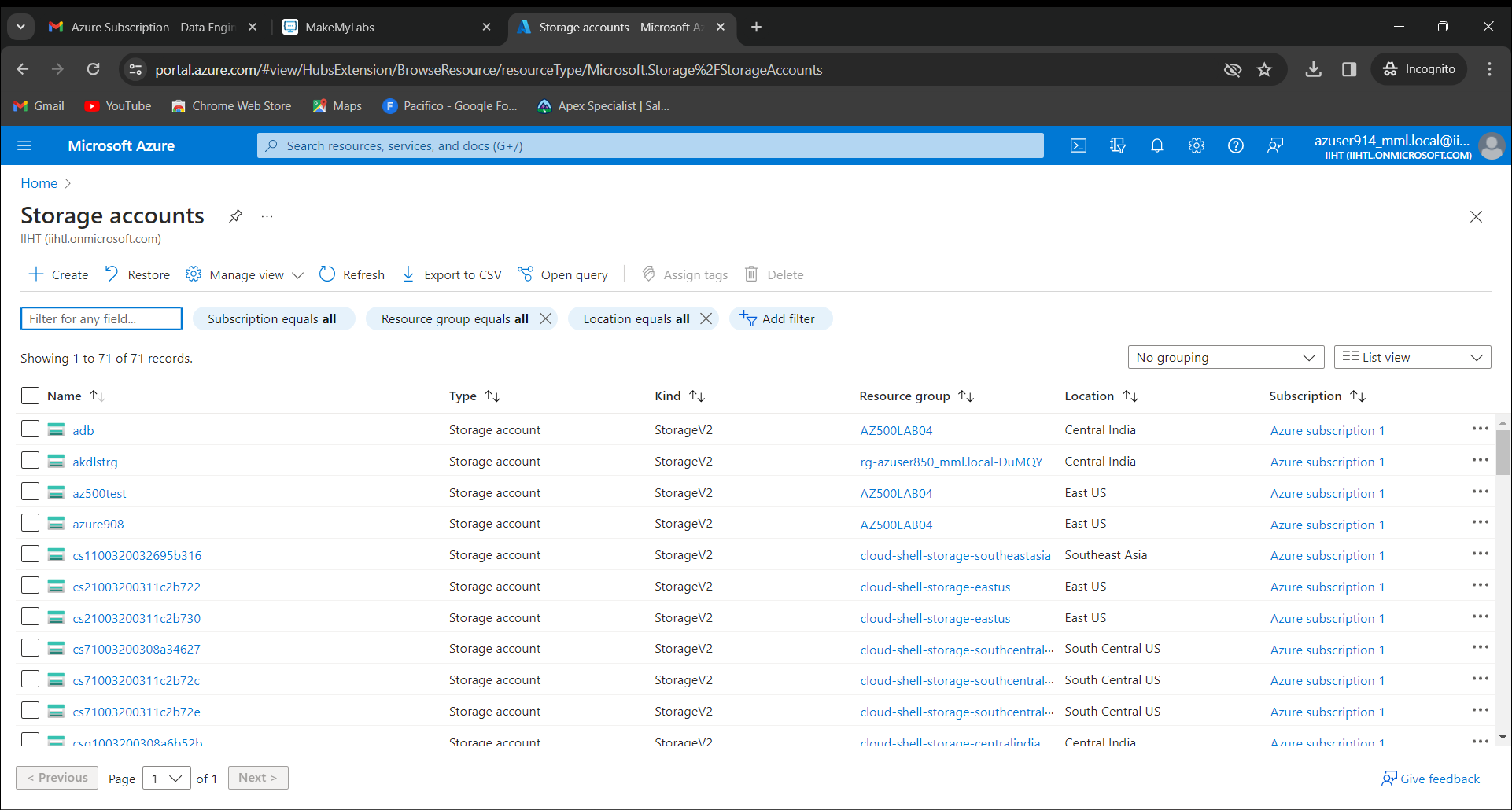
Now sign in with your account

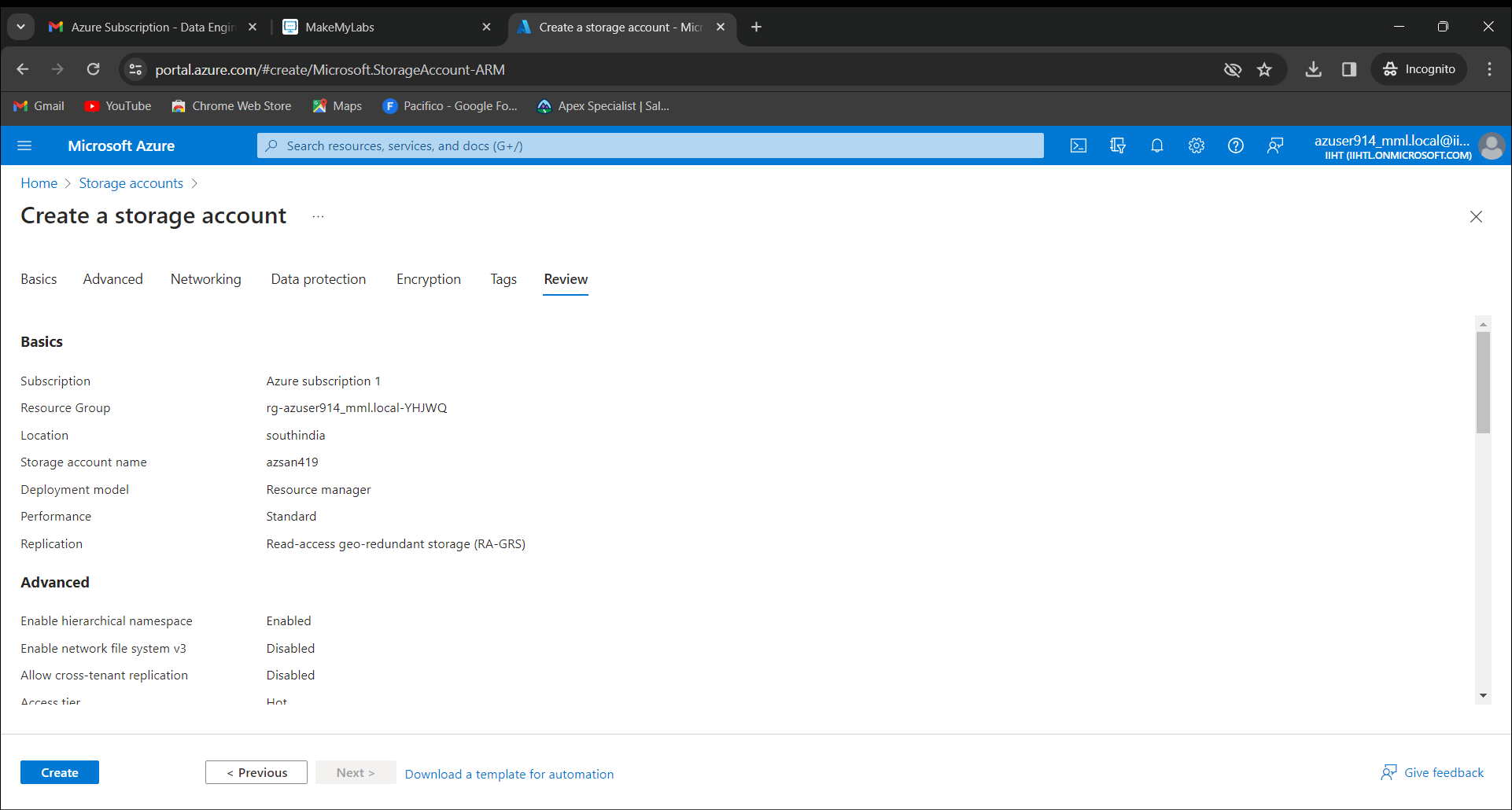
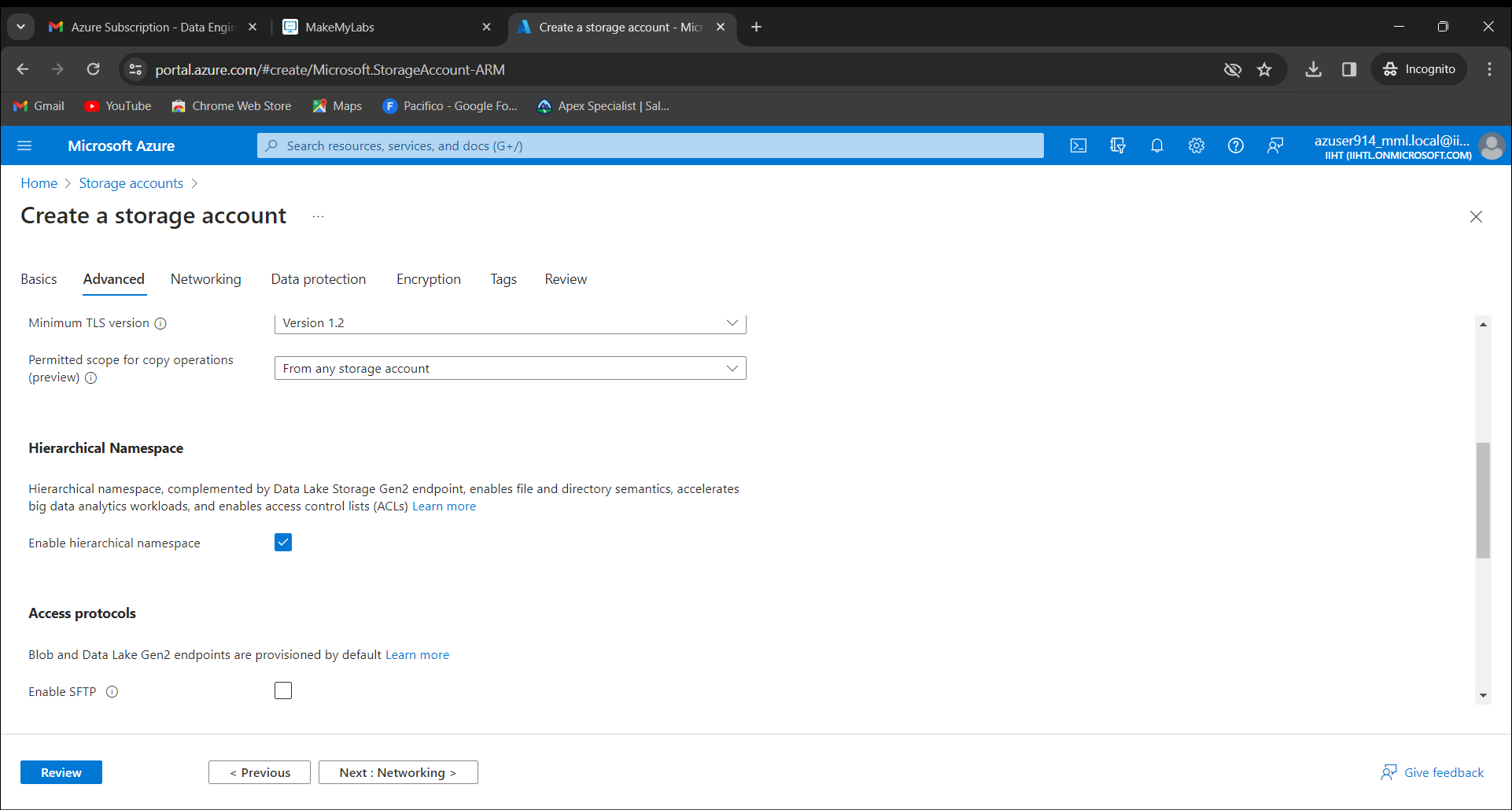
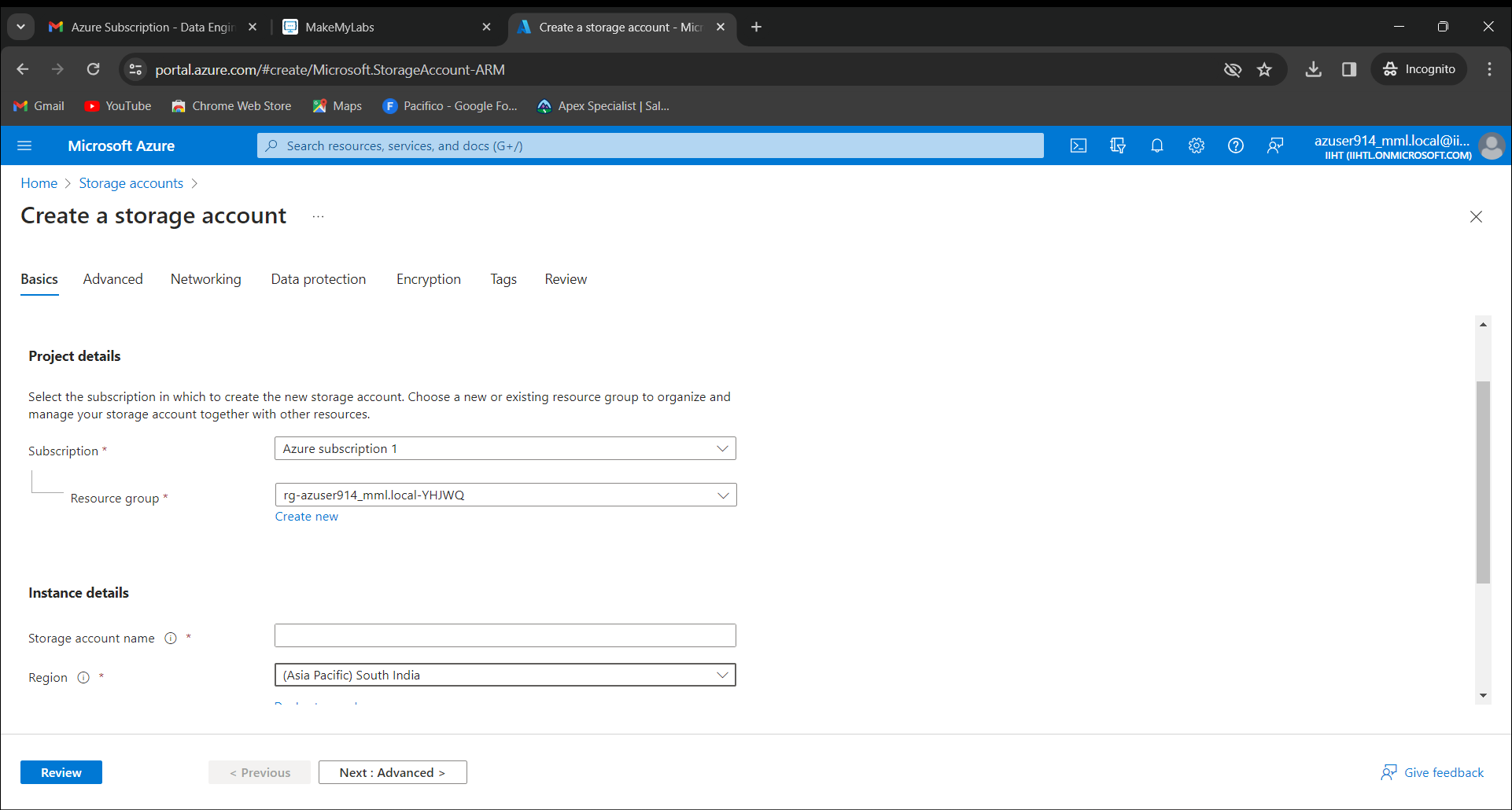


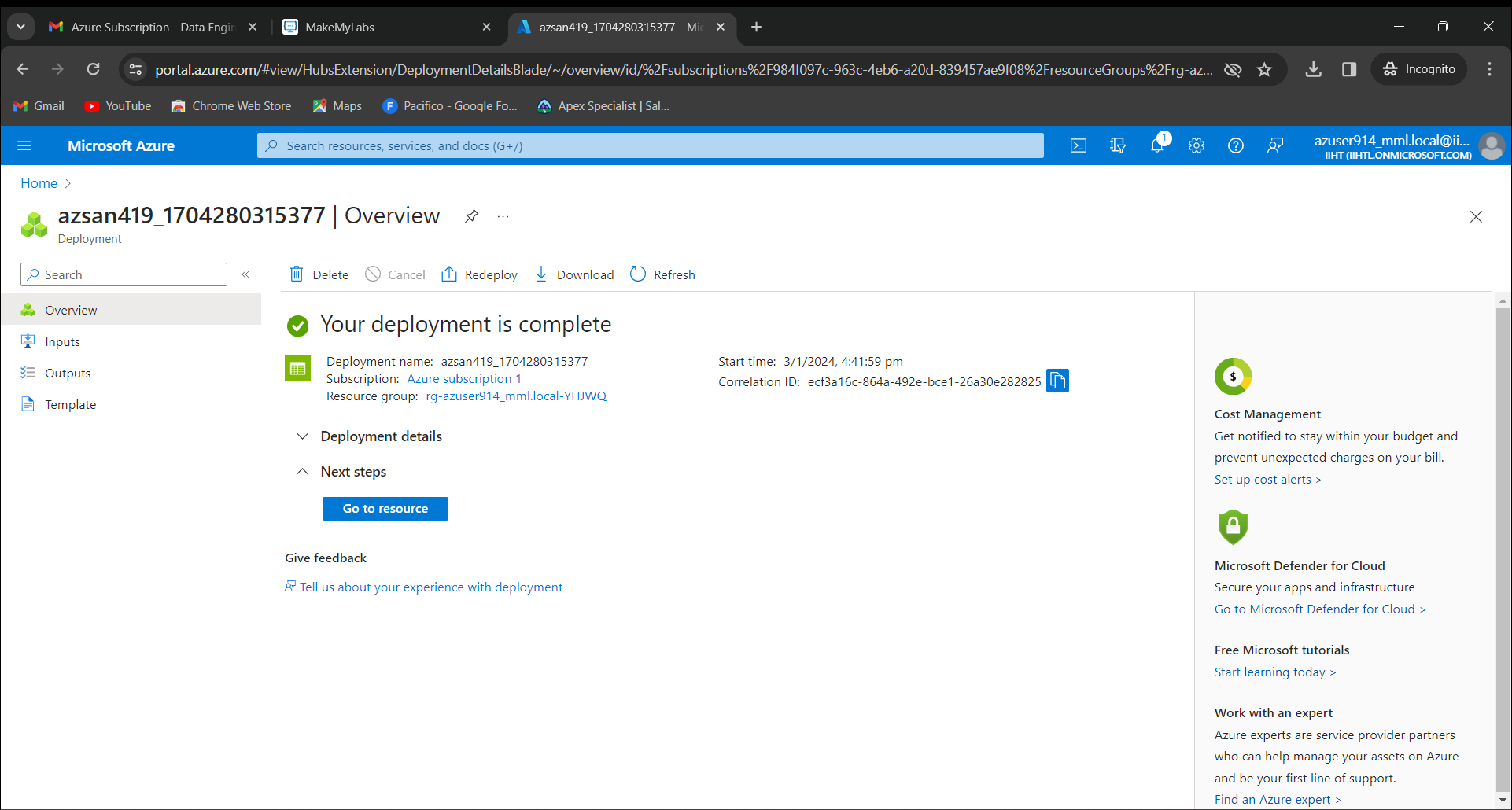
Select Azure and click on next

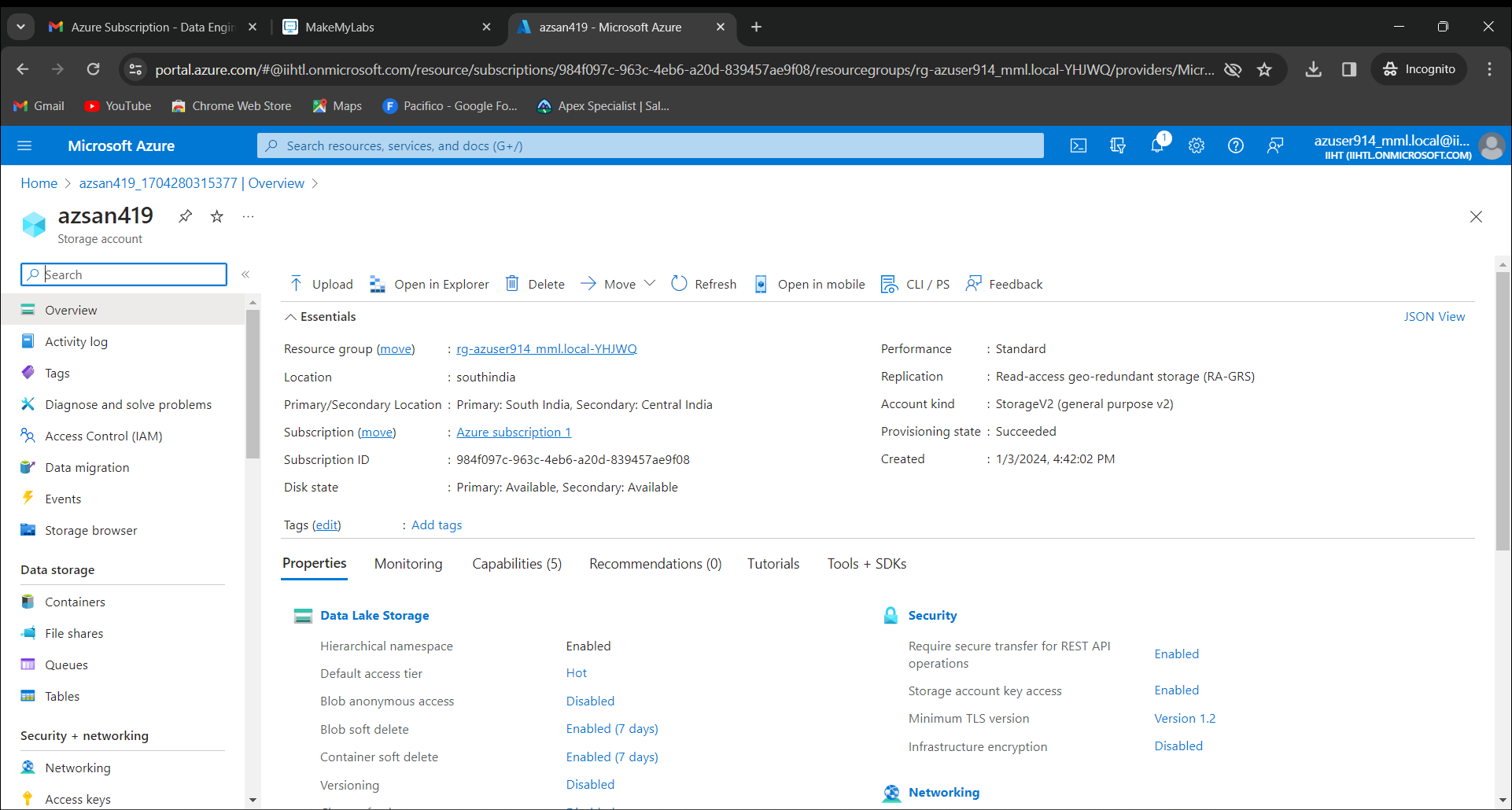
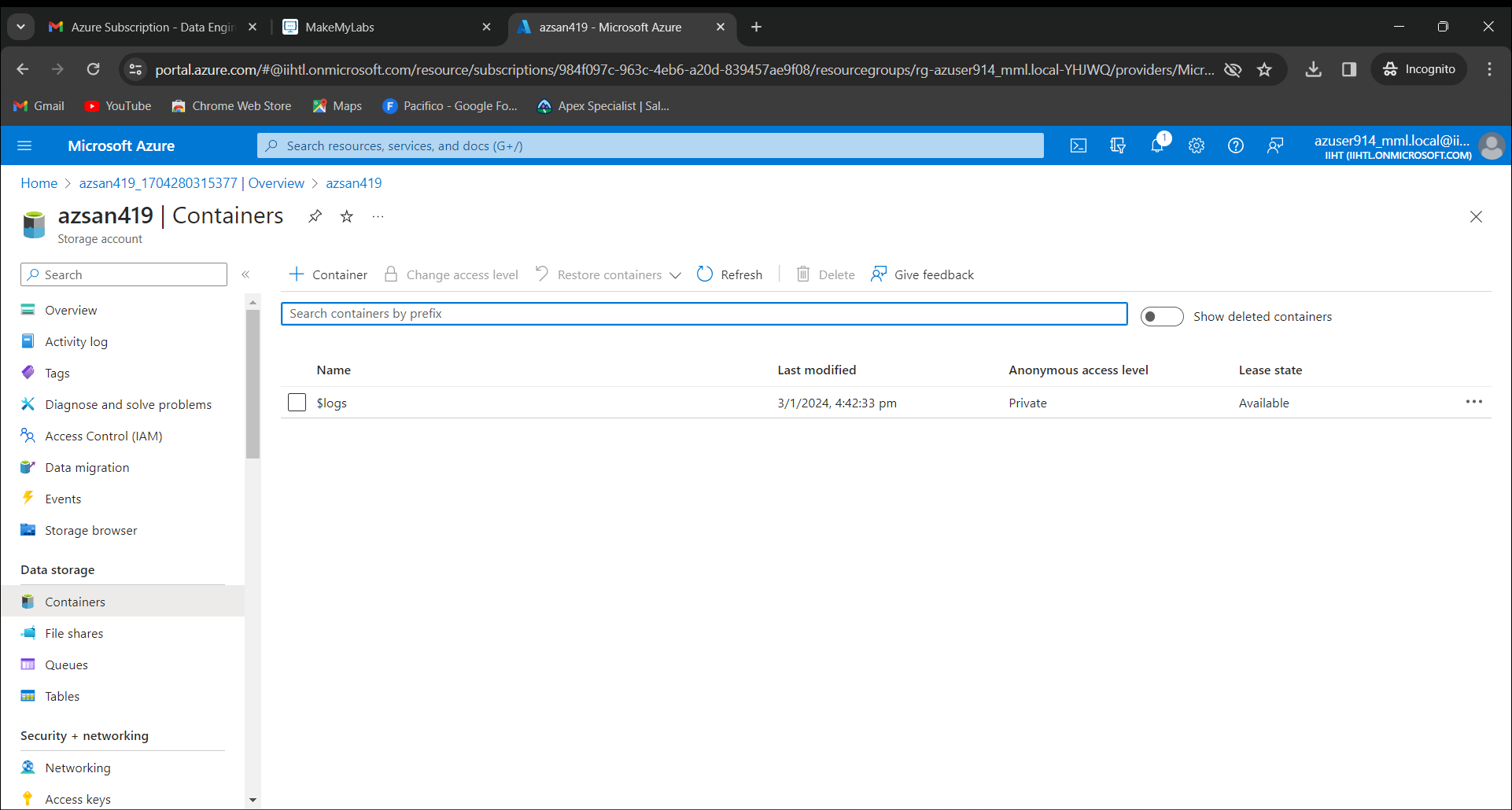
You will be directed to the web browser.

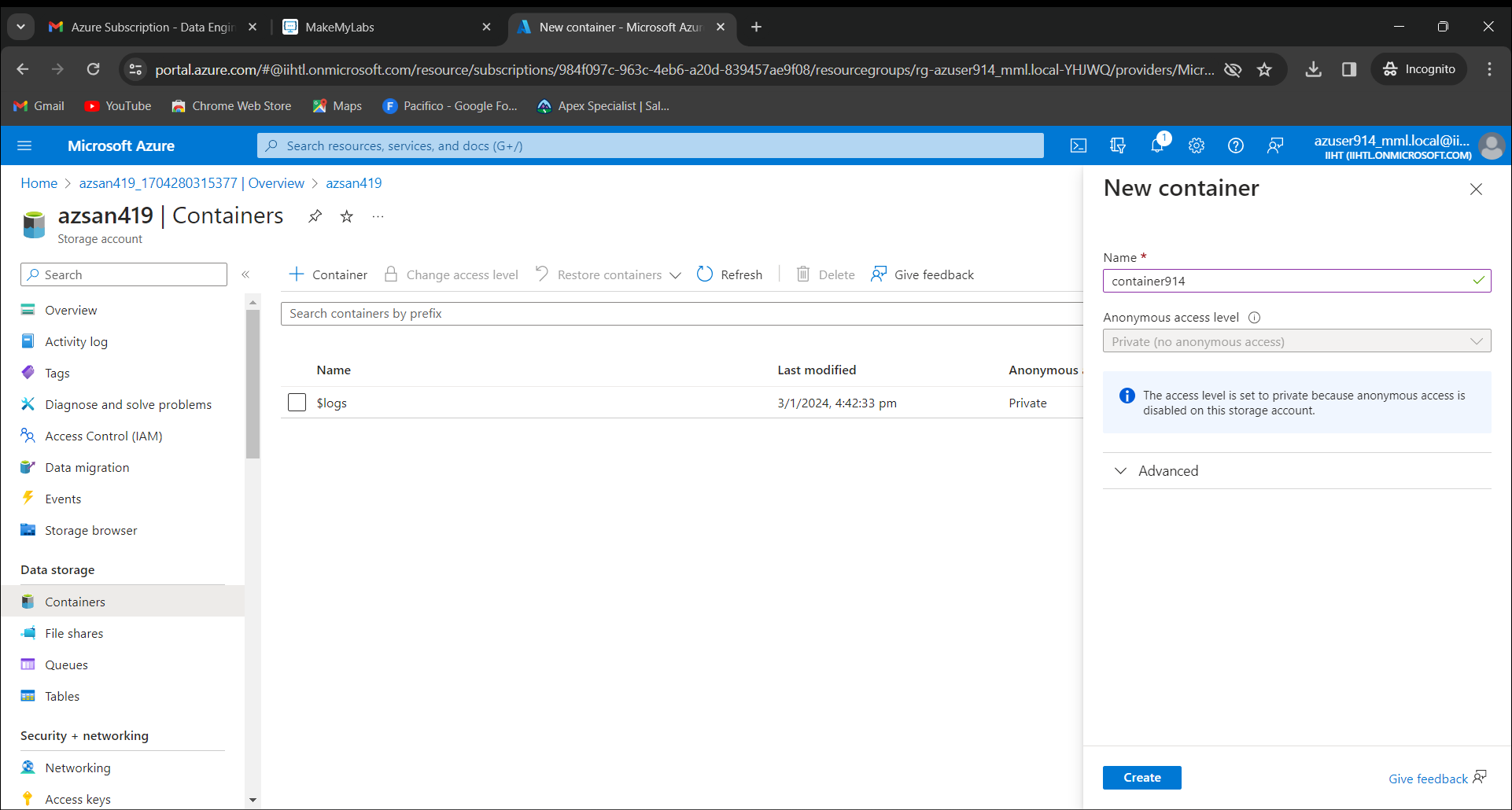
Click on azure subscription 1. In storage account we need to create our storage. 

In Azure, search for storage accounts. Click on craete

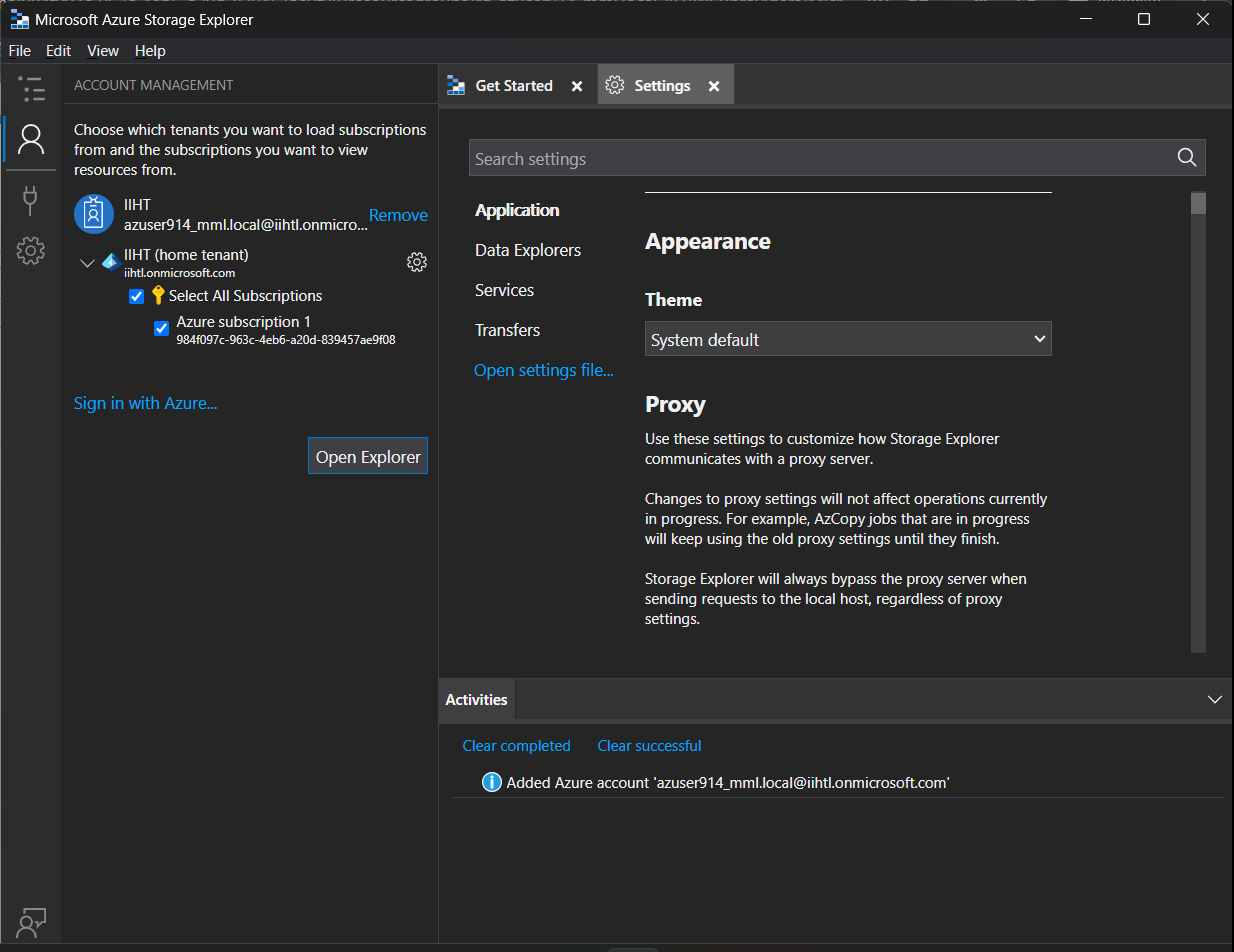
Give the resource group and storage account name



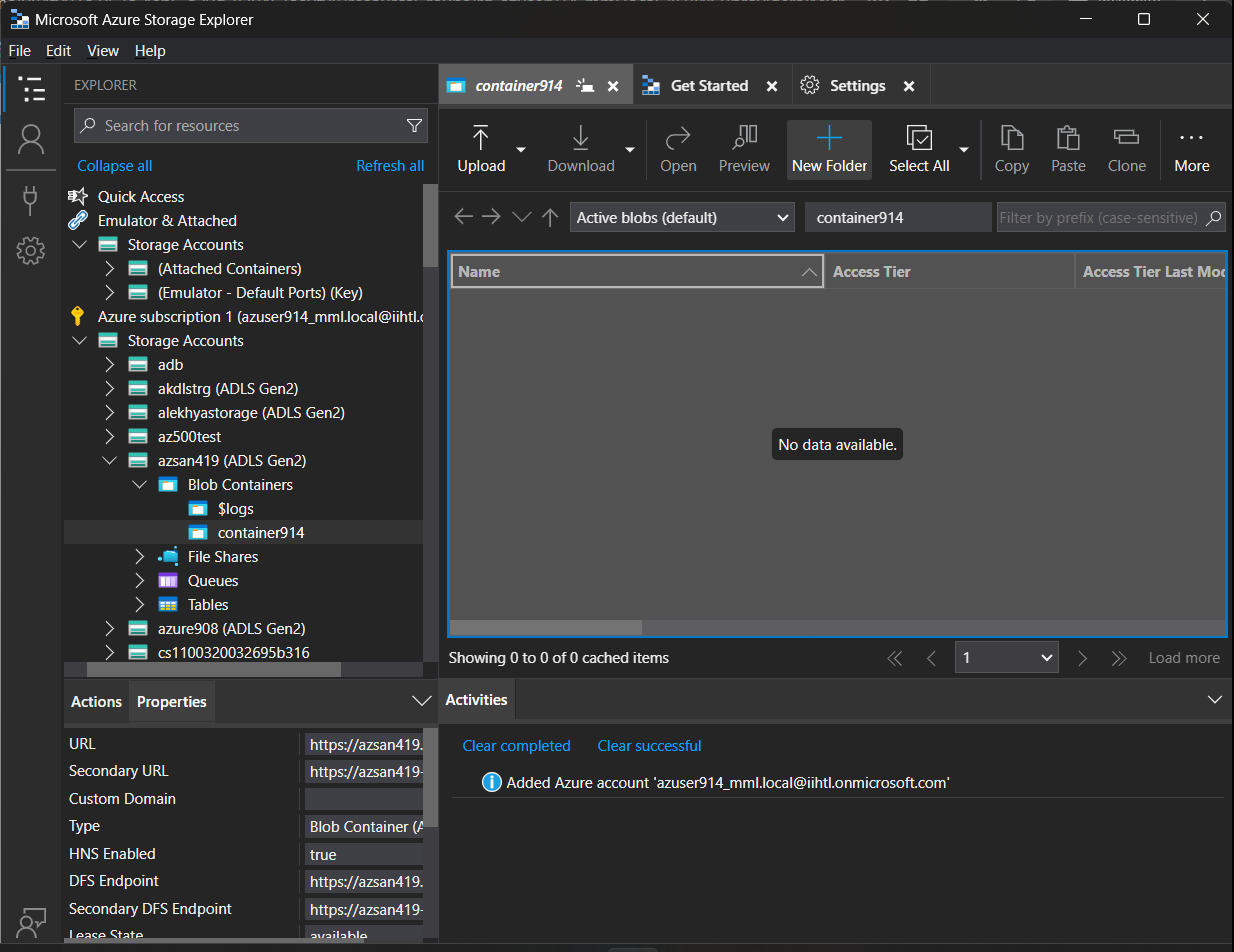
Click on go to resources and now we get the option container in right side menu bar. Click on containers.Click on +container

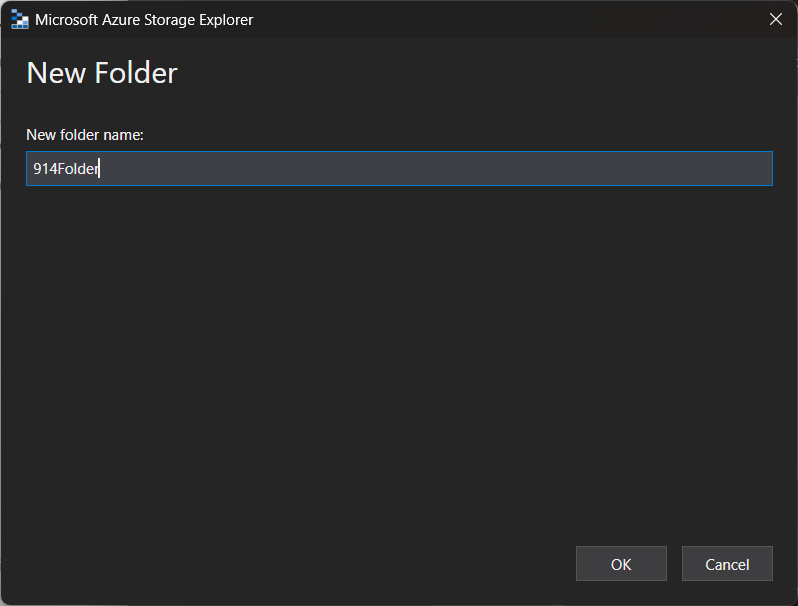
Create a new container

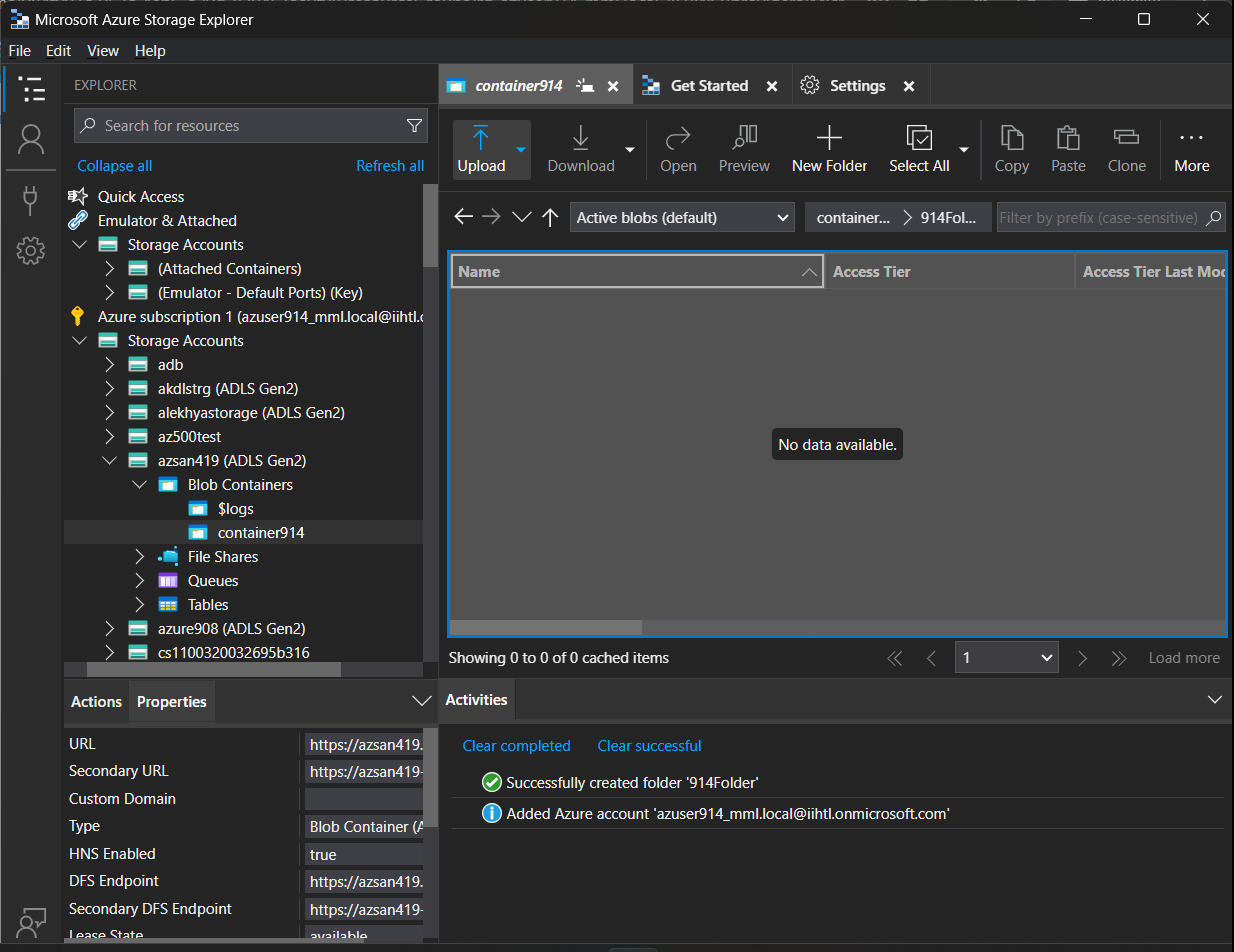
Click on open explorer



Now click on the storage account that was names as “azsan419(ADLS Gen2)”

Open container 914.

Create new folder named as 914Folder

Click on uploadUpload a CSV file.