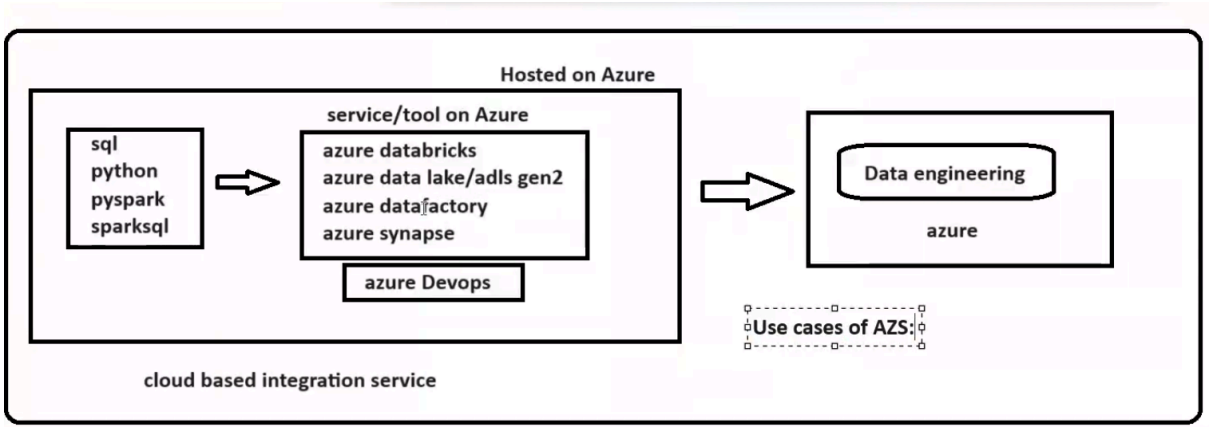
NAME : AKULA SHARATH CHANDRA
BATCH: DATA ENGINEERING
DATE :20-02-2024
TOPIC : AZURE DATA FACTORY

1)DATA FACTORY: Azure Data Factory 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.
ADF does not store any data itself. It allows you to create data-driven workflows to orchestrate the movement of data between supported data stores and then process the data using compute services in other regions or in an on-premise environment. It also allows you to monitor and manage workflows using both programmatic and Ul mechanisms.

2)ADF USE CASES:

ADF can be used for:
—>Supporting data migrations
—>Getting data from a client's server or online data to an Azure Data Lake.
—>Carrying out various data integration processes.
—>Integrating data from different ERP systems and loading it into Azure.
—>Synapse for reporting.



3)HOW DOES ADF WORKS:

The Data Factory service allows you to create data pipelines that move and transform data and then run the pipelines on a specified schedule (hourly, daily, weekly, etc.).
It works on 3 steps

Step 1: Connect and Collect

Connect to all the required sources of data and processing such as SaaS services, file shares, FTP, and web services. Then, move the data as needed to a centralized location for subsequent processing by using the Copy Activity in a data pipeline to move data from both on-premise and cloud source data stores to a centralization data store in the cloud for further analysis.

Step 2: Transform and Enrich

Once data is present in a centralized data store in the cloud, it is transformed using compute services such as HDInsight Hadoop, Spark, Azure Data Lake Analytics, and Machine Learning.

Step 3: Publish

Deliver transformed data from the cloud to on-premise sources like SQL Server or keep it in your cloud storage sources for consumption by BI and analytics tools and other applications.

4) DATA MIGRATION: Data migration occurs between two cloud data stores and between an on-premise data store and a cloud data store.

COPY ACTIVITY:

Copy Activity in Azure Data Factory copies data from a source data store to a sink data store. Azure supports various data stores such as source or sink data stores like Azure Blob storage, Azure Cosmos DB (DocumentDB API), Azure Data Lake Store, Oracle, Cassandra, etc.

→ Azure Data Factory supports transformation activities such as Hive, MapReduce, Spark, etc that can be added to pipelines either individually or chained with other activities.

5)AZURE DATA FACTORY KEY COMPONENTS: Azure Data Factory has four key components that work together to define input and output data, processing events, and the schedule and resources required to execute the desired data flow.

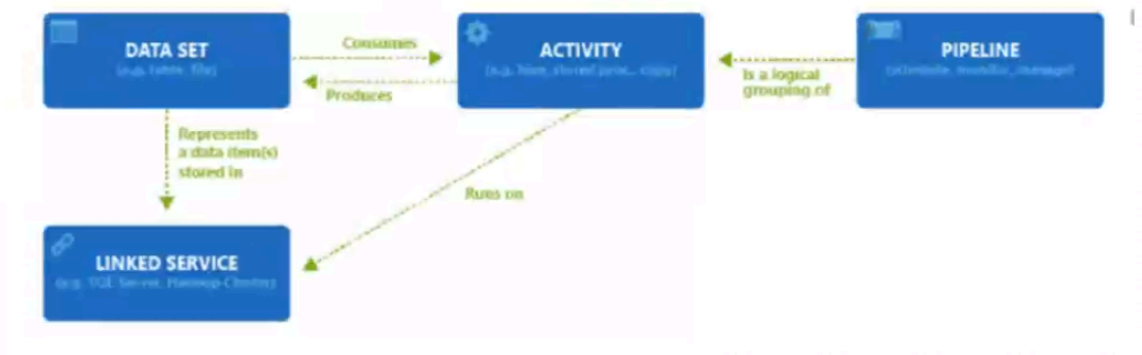
—>**Datasets represent data structures within the data stores:** Input datasets provide activity input in a pipeline, while output datasets represent activity output. For instance, an Azure Blob dataset specifies the source in Blob Storage for reading, and an Azure SQL Table dataset designates the table for writing activity output.

—>**A pipeline is a group of activities:** A pipeline in Azure Data Factory is a collection of activities that collaboratively execute a task. Multiple pipelines can exist in a data factory.

—>**Activities define the actions to perform on your data:**Activities, specifying data actions, come in two types: data movement and data transformation.

—>**Linked services define the information needed for Azure Data Factory to connect to external resources:** Linked services provide the necessary information, like connection strings, for Data Factory to connect to external resources, such as Azure Storage.

6)HOW DOES DATA FACTORY COMPONENTS WORKS



We Can use one of the following tools or APIs to create data pipelines in Azure Data Factory:

—>LINKED SERVICE

—>Azure portal

—>Visual Studio

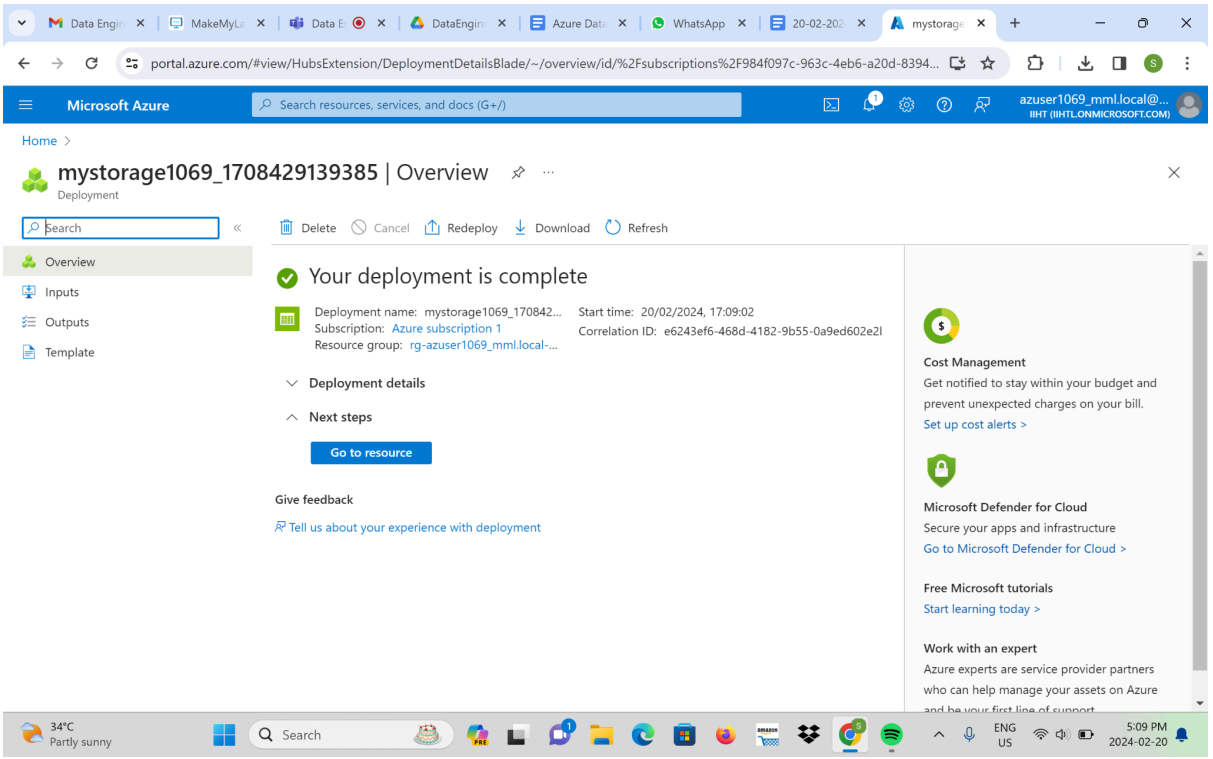
—>PowerShell

—>NET API

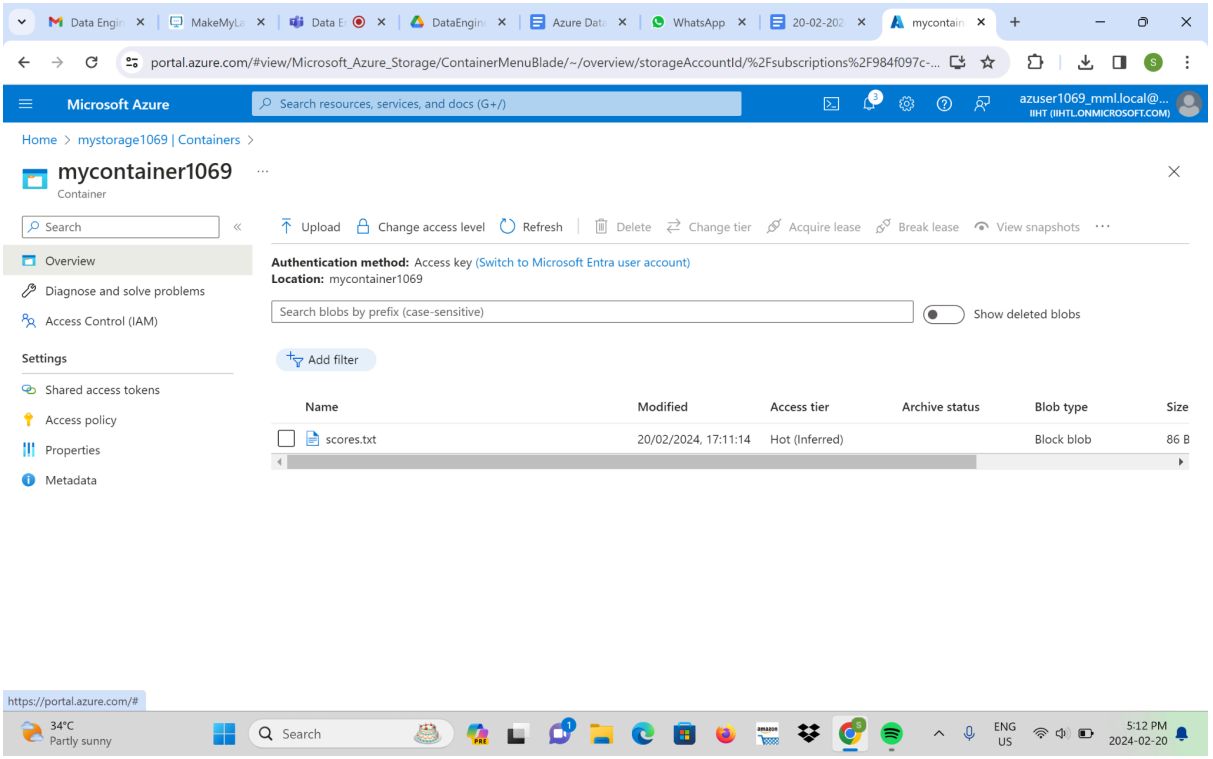
—>Azure Resource Manager template

7)CREATING A DATA FACTORY ACCOUNT:

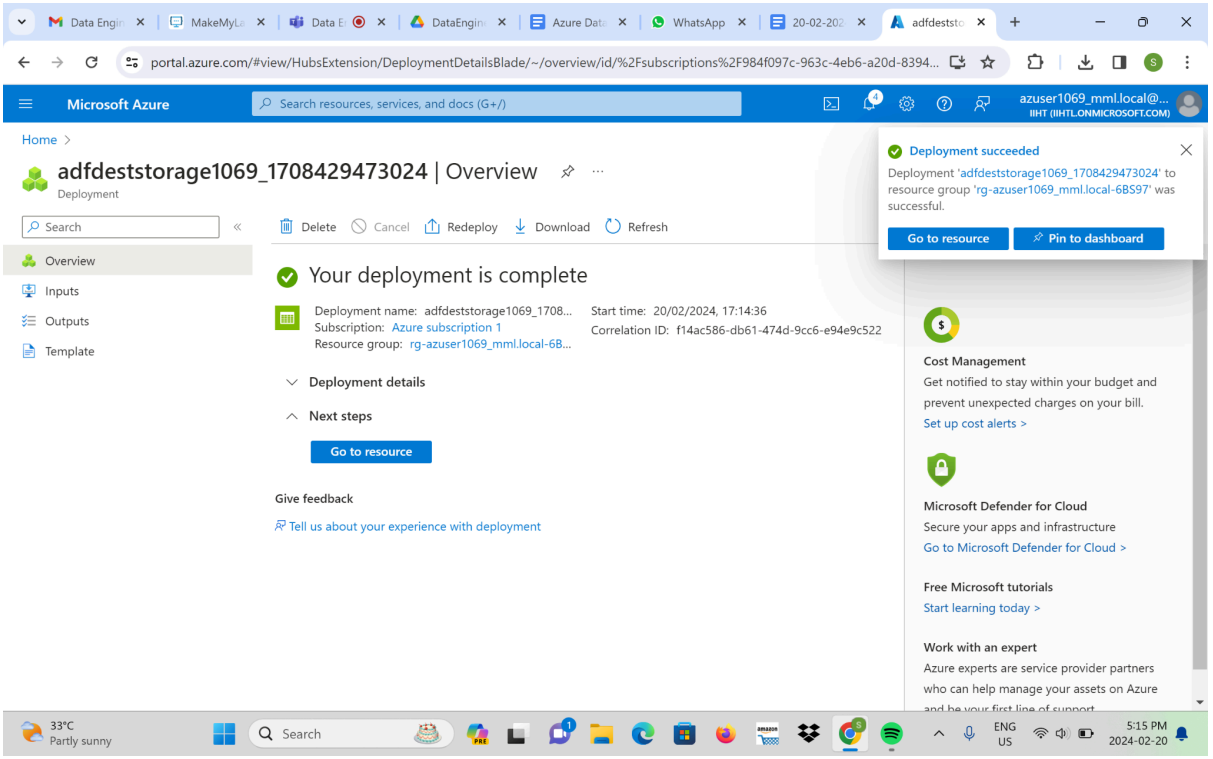
—> to create a data factory account firstly we need to create two storage accounts one is for source group and one is for destination group.

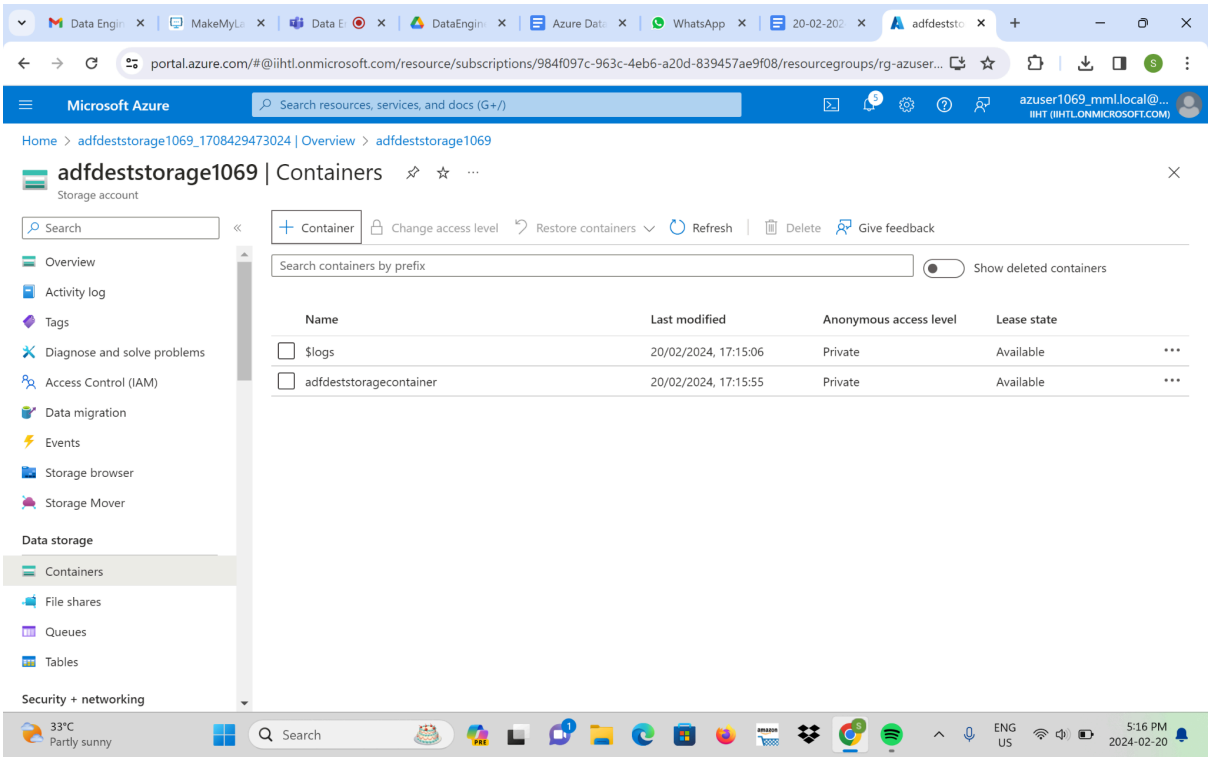


—>Then i have created a container and uploaded a file in it .

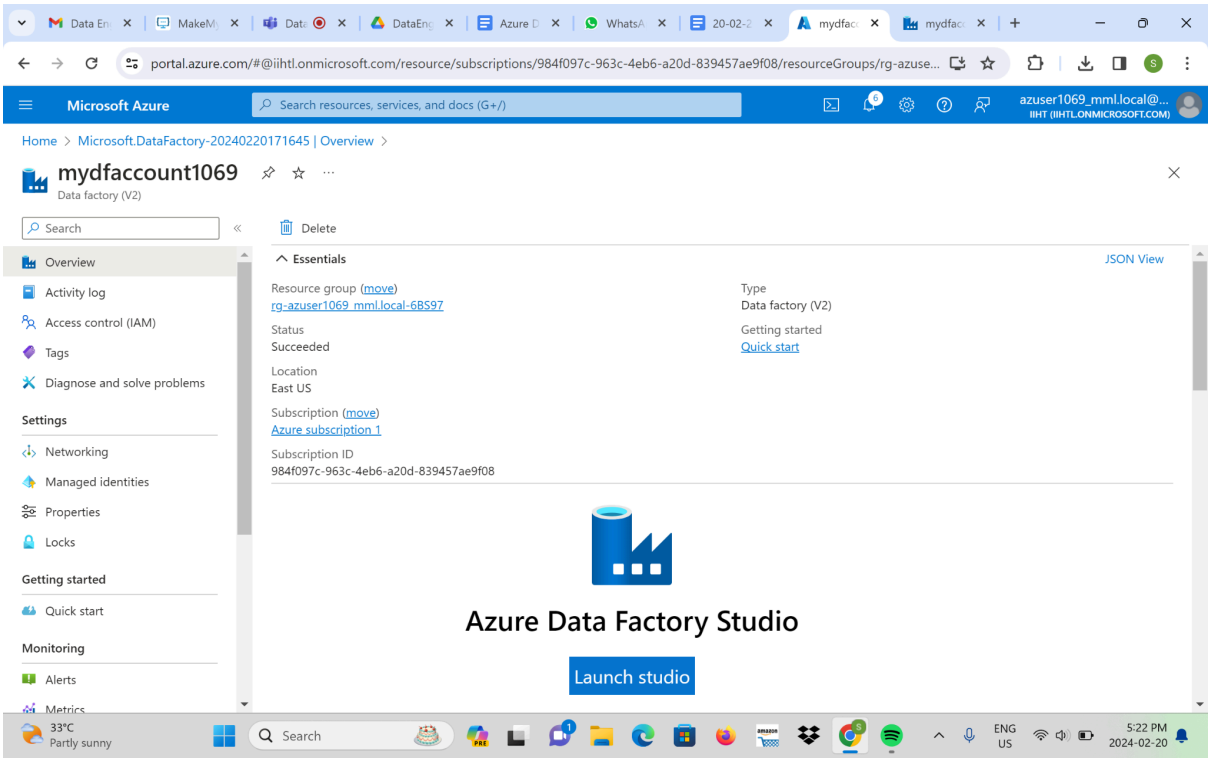


—>Then again, I have created another storage account for the destination type .
—> After that I have created an empty container for my destination storage account .

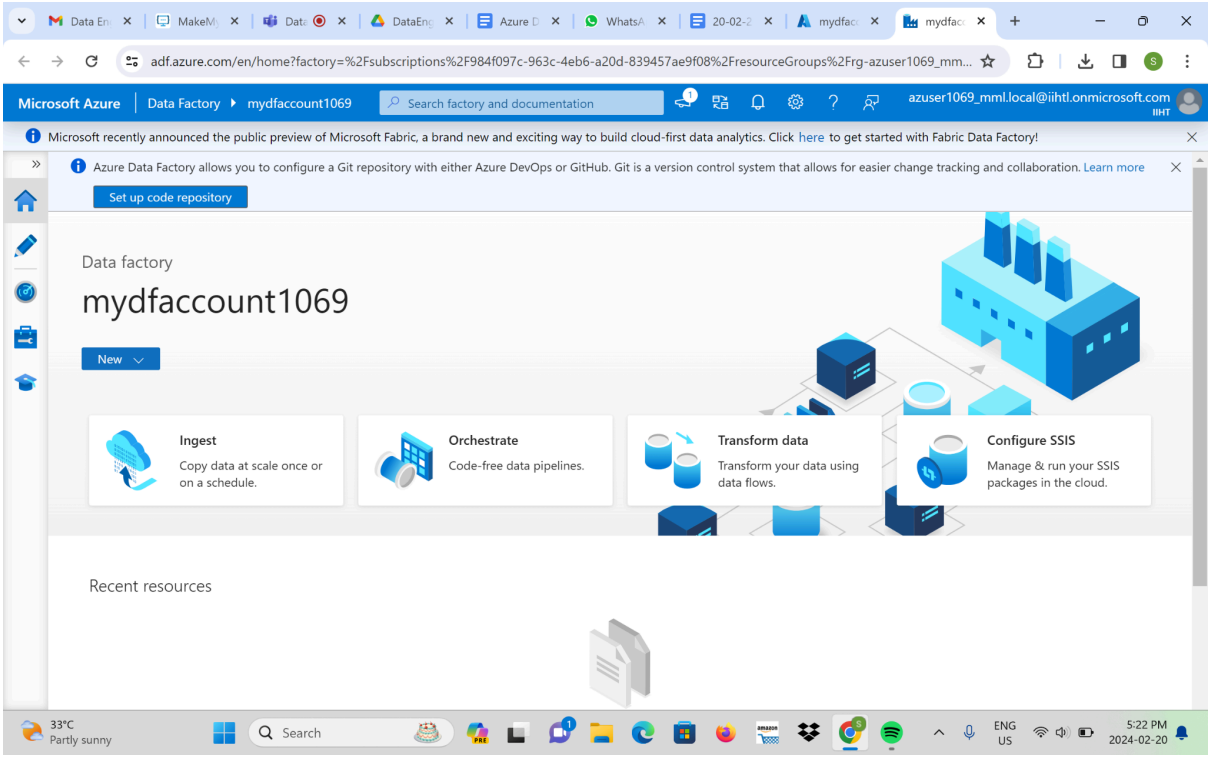

Microsoft Azure
Search resources, services, and docs (G+/)
azuser1069_mml.local@...
IIHT (IIHTL.ONMICROSOFT.COM)
Home
adfdeststorage1069_1708429473024 | Overview
Deployment
Search
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Download
Refresh
Your deployment is complete
Deployment name: adfdeststorage1069_1708...
Subscription: Azure subscription 1
Resource group: rg-azuser1069_mml.local-6B...
Start time: 20/02/2024, 17:14:36
Correlation ID: f14ac586-db61-474d-9cc6-e94e9c522
Deployment details
Next steps
Go to resource
Give feedback
Tell us about your experience with deployment
Deployment succeeded
Deployment 'adfdeststorage1069_1708429473024' to resource group 'rg-azuser1069_mml.local-6BS97' was successful.
Go to resource
Pin to dashboard
Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >
Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >
Free Microsoft tutorials
Start learning today >
Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure
33°C
Partly sunny
Search
ENG
US
5:15 PM
2024-02-20


Microsoft Azure
Search resources, services, and docs (G+/)
azuser1069_mml.local@...
IIHT (IIHTL.ONMICROSOFT.COM)
Home
adfdeststorage1069_1708429473024 | Overview
adfdeststorage1069
adfdeststorage1069 | Containers
Storage account
Search
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Storage Mover
Data storage
Containers
File shares
Queues
Tables
Security + networking
Container
Change access level
Restore containers
Refresh
Delete
Give feedback
Search containers by prefix
Show deleted containers
Name
Last modified
Anonymous access level
Lease state
$logs
20/02/2024, 17:15:06
Private
Available
adfdeststoragecontainer
20/02/2024, 17:15:55
Private
Available
33°C
Partly sunny
Search
ENG
US
5:16 PM
2024-02-20

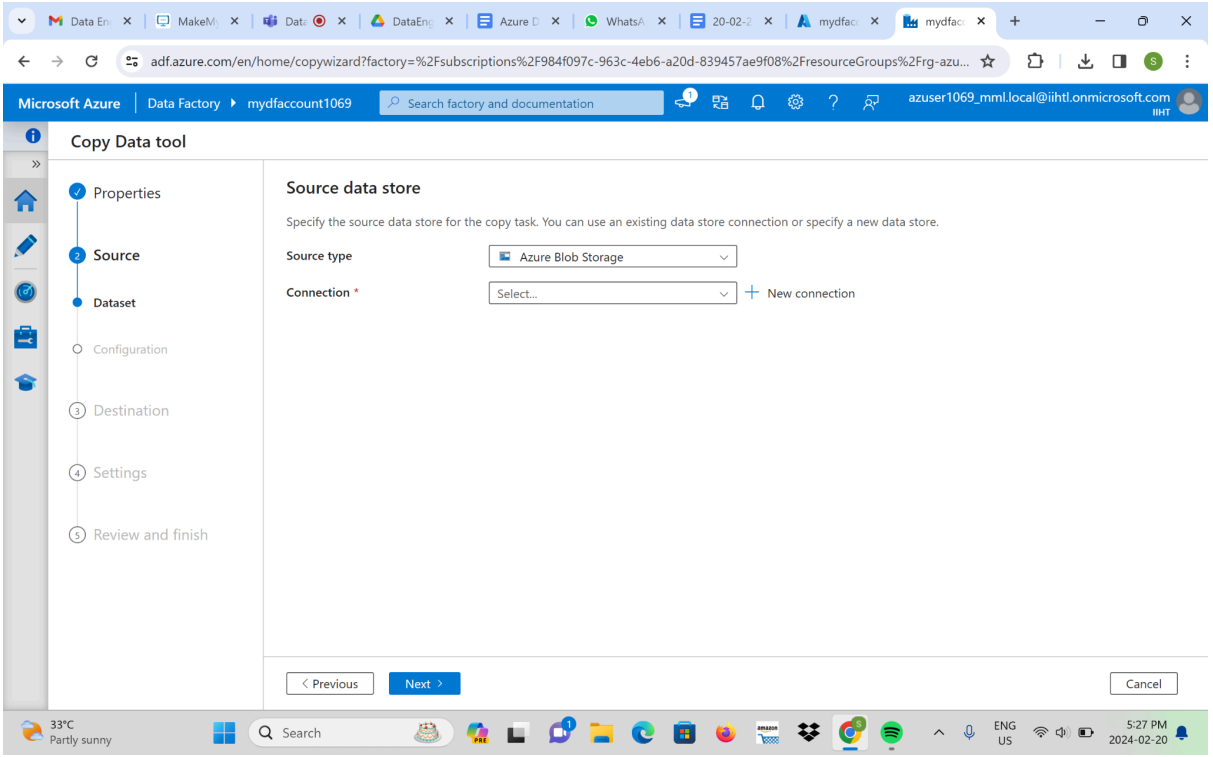
—> Then I created a data factory account and then I clicked on launch studio.



—>Then I have selected an option called ingestion to set the connection .

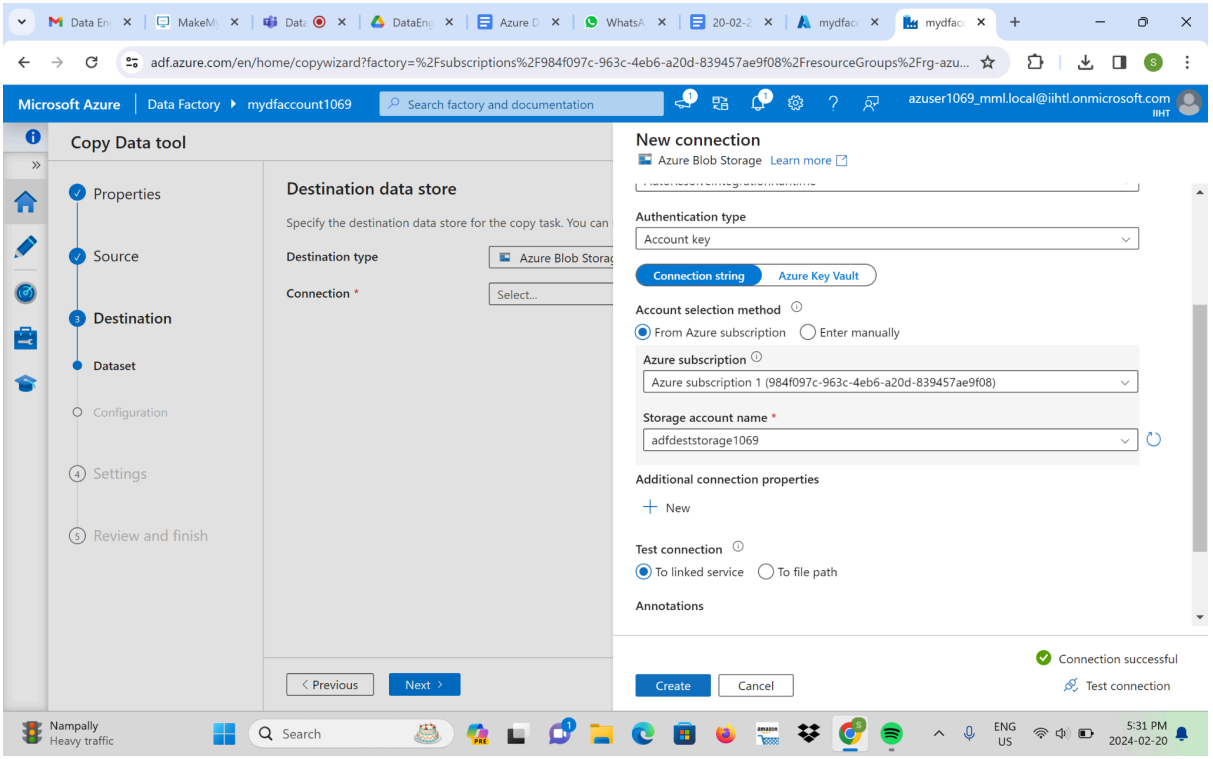


—>Then I have created a source data source and then I have selected azure blob storage as source type and created a connection by giving a source storage name and source container in it.

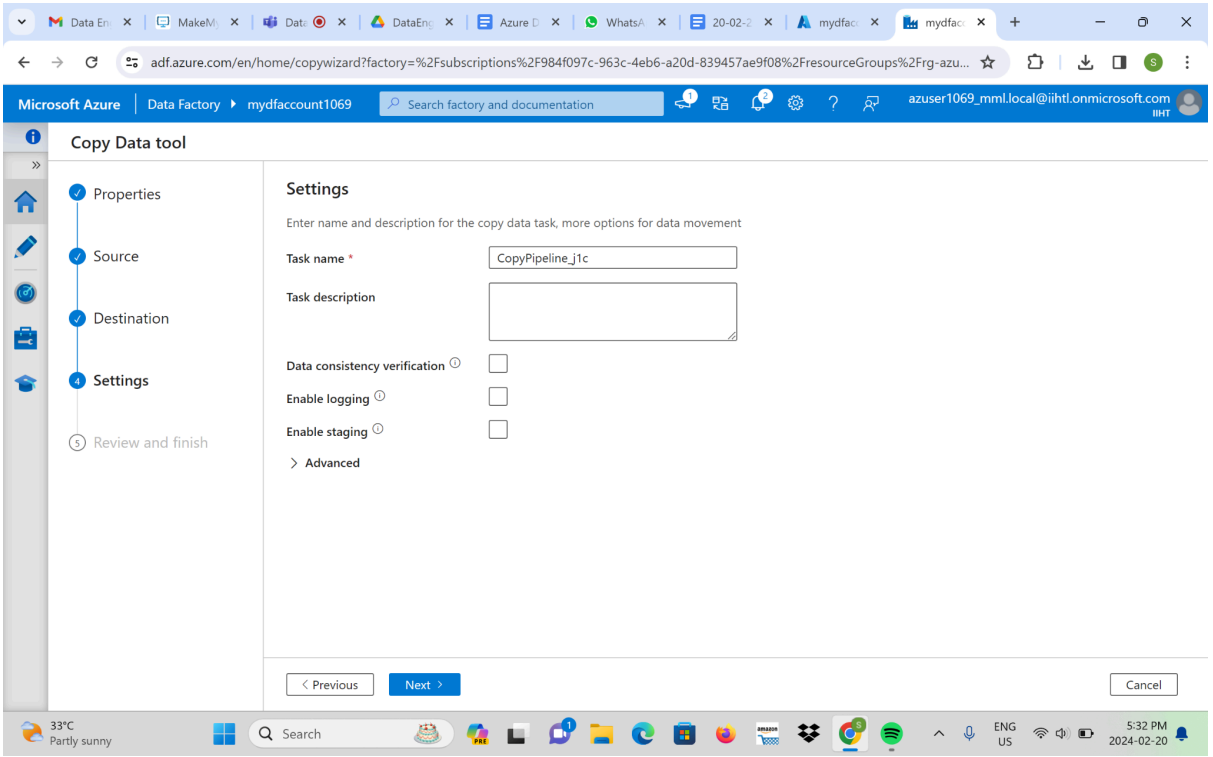


—>Later I have created a destination data source and then I have selected azure blob storage as destination type and created a connection by giving a destination storage name and destination container in it.

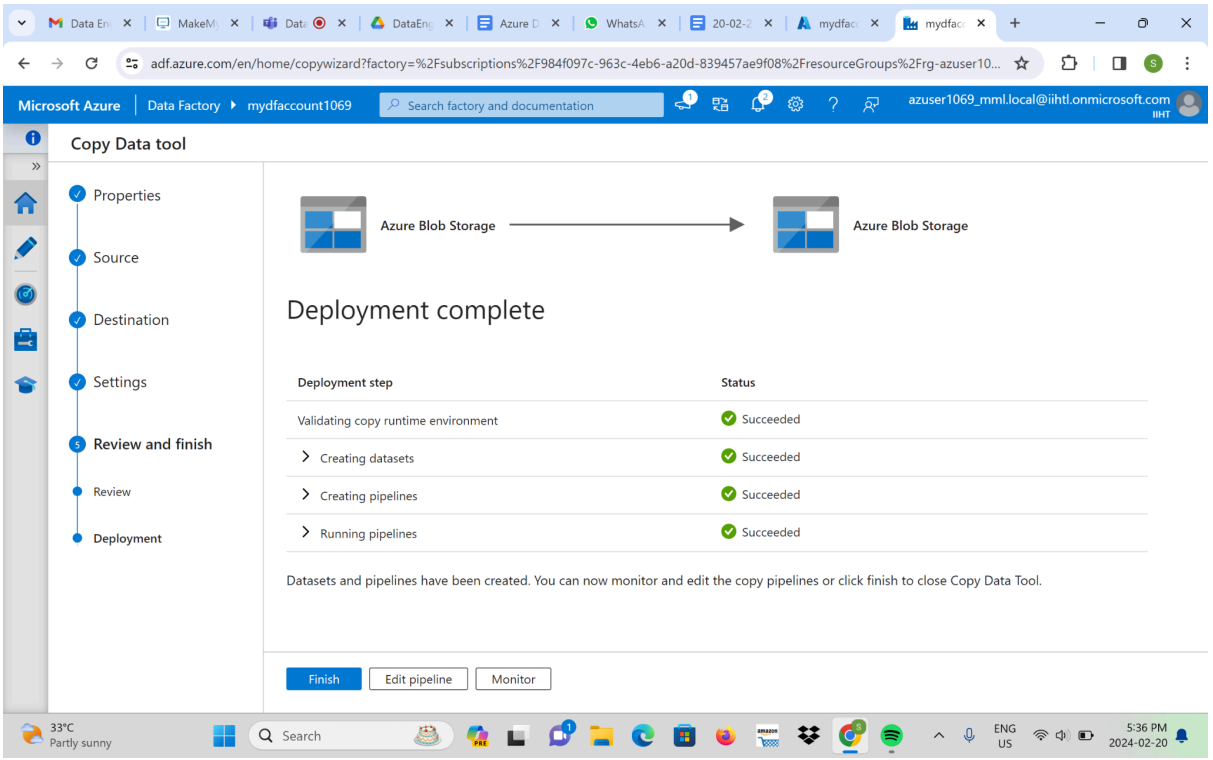
—>In the destination container there is no file present because we are copying source to destination type .So our motive is to copy the data from source and review at destination group . For that we need to give an empty destination container .



—>Then we need to give the task name as “copyActivity” for the above process and click on the next .

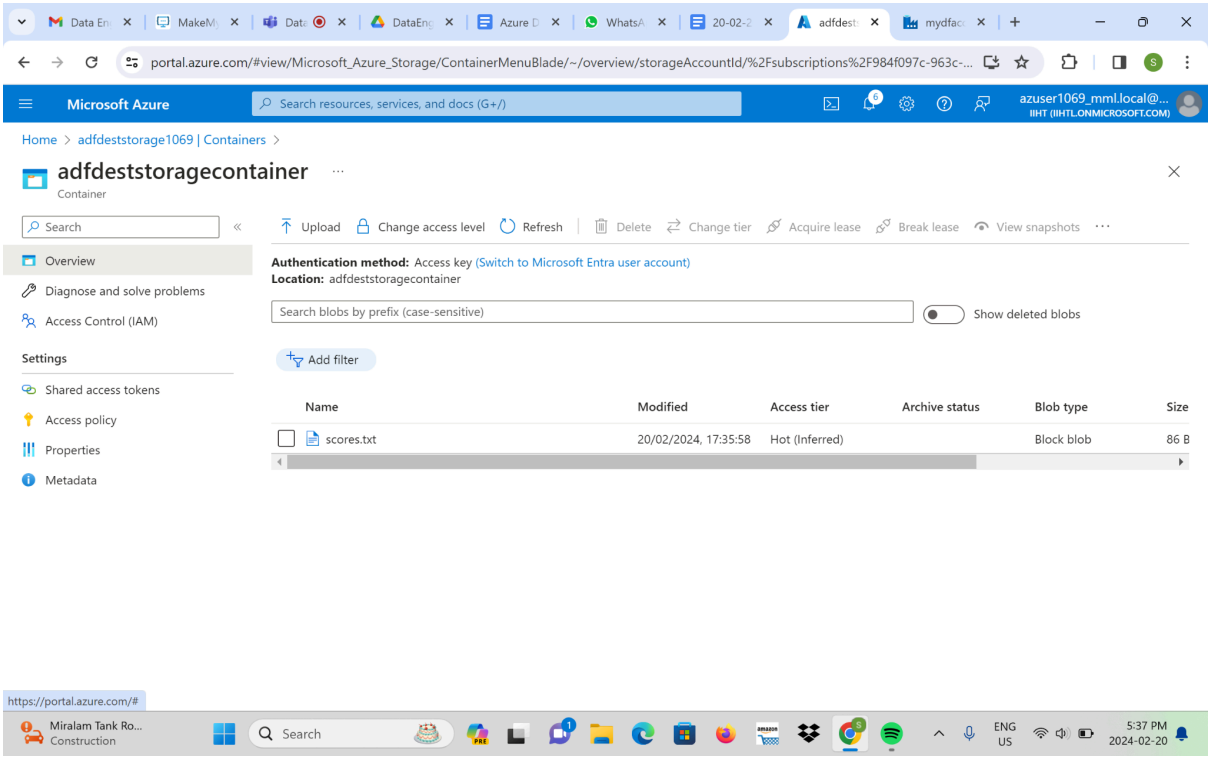


—>After giving the task name we need to review all the settings and click on finish .



—>Then we can see that datasets and pipelines are successfully created and running successfully.

—>Thereby we can see that in the destination container the file has been successfully copied from the source account .


Microsoft Azure
Search resources, services, and docs (G+/)
azuser1069_mml.local@...
IIHT (IIHTL.ONMICROSOFT.COM)
Home > adfdeststorage1069 | Containers >
adfdeststoragecontainer
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Access policy
Properties
Metadata
Upload
Change access level
Refresh
Delete
Change tier
Acquire lease
Break lease
View snapshots
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: adfdeststoragecontainer
Search blobs by prefix (case-sensitive)
Show deleted blobs
Add filter
Name
Modified
Access tier
Archive status
Blob type
Size
scores.txt
20/02/2024, 17:35:58
Hot (Inferred)
Block blob
86 B