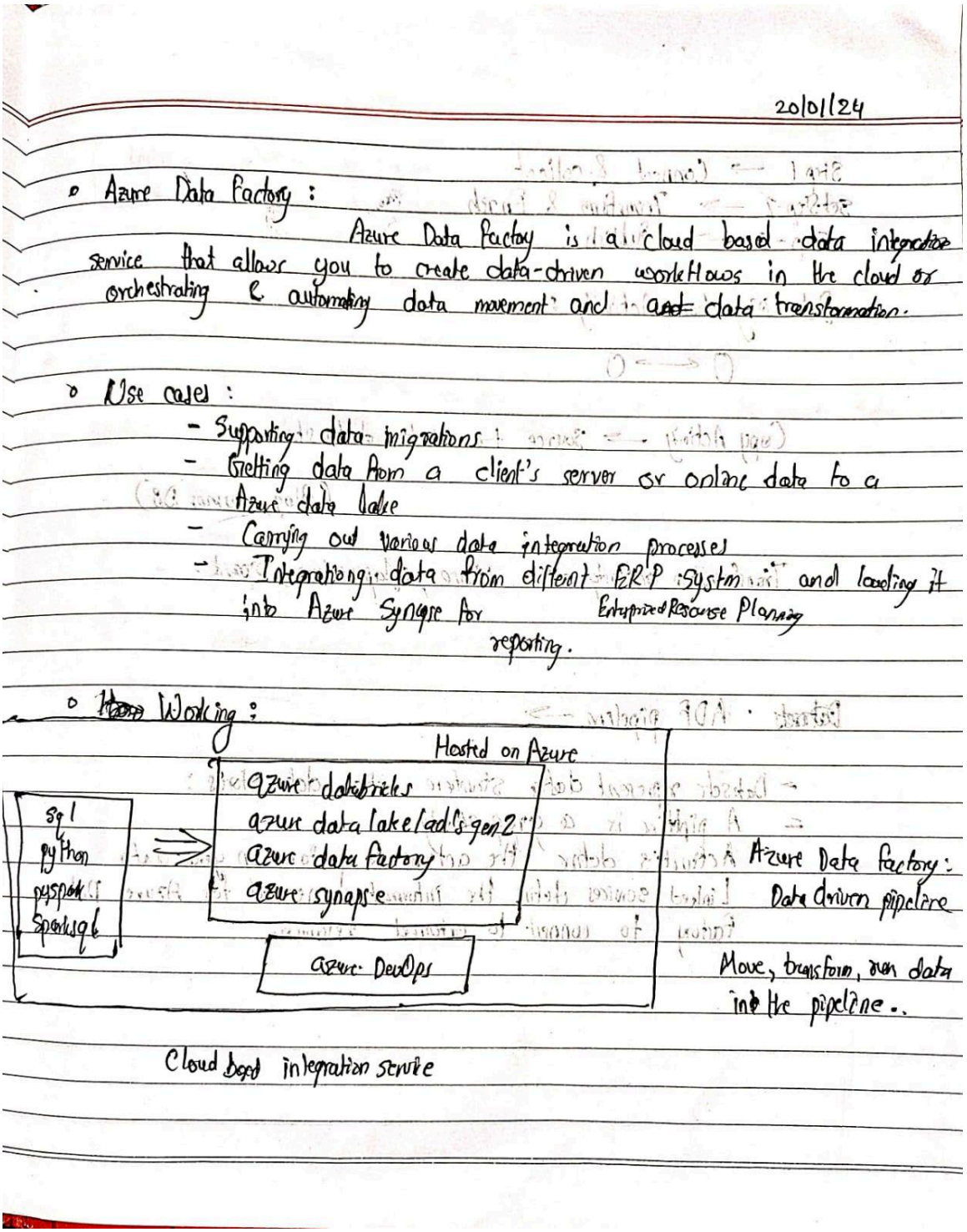
Name : Parth Nandedkar
Date : 20 Feb 2024
Topics : Azure Data Factory
Batch : Data Engineering Batch-1

Handwritten Notes :



Step 1 → Connect & collect
Step-2 → Transform & Enrich
Step-3 → Publish

Data Migration Activity :

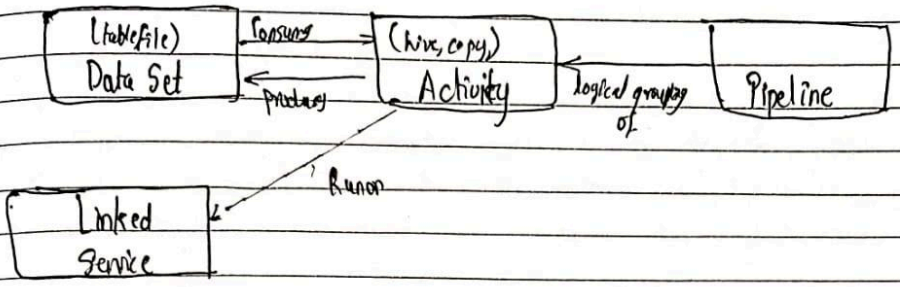
O ⟷ O

Copy Activity → Source to sink i data store -
(Blog, cosmos DB)

Transformation activity → Hive, Map reduce, Spark.

Datasets · ADF pipeline →

- Datasets represent data structure within data stores :
- A pipeline is a group activity
- Activities define the actions to perform on your data
- Linked services define the information needed for Azure Data Factory to connect to external resources.



You can use one of the following tools or APIs to create data pipelines in Azure Data Factory.

- Azure Portal
- Visual Studio
- PowerShell
- .NET API
- REST API
- Azure resource manager template.

Azure Data Factory :

Azure Data Factory (ADF) is a cloud-based data integration service provided by Microsoft Azure. It enables you to create, schedule, and manage data pipelines for ingesting, transforming, and processing data from various sources. Here's an overview of Azure Data Factory and its key features:

Data Orchestration: Azure Data Factory allows you to orchestrate and automate the movement and transformation of data across on-premises and cloud-based data stores. You can create pipelines to orchestrate complex data workflows involving data ingestion, data transformation, and data movement tasks.

Data Integration: ADF provides built-in connectors for integrating with various data sources and destinations, including Azure services (such as Azure Blob Storage, Azure SQL Database, Azure Data Lake Storage, Azure Synapse Analytics), on-premises data sources (using self-hosted integration runtimes), and external sources (such as SQL Server, Oracle, Salesforce, Amazon S3, and more).

Data Transformation: With Azure Data Factory, you can perform data transformations using Azure Data Flows, which provide a visual, code-free environment for building data transformation logic. Data Flows allow you to clean, enrich, aggregate, and transform data at scale, using familiar data manipulation techniques.

Data Movement: ADF enables efficient and scalable data movement between various data stores. You can use Copy Activity to move data between supported data sources, with support for parallelism, fault tolerance, and data compression to optimize data transfer performance.

Data Orchestration and Scheduling: Azure Data Factory allows you to schedule and orchestrate data pipelines using triggers. You can define triggers based on schedules (e.g., recurring intervals, specific times) or events (e.g., data arrival, external events) to automate the execution of your data pipelines.

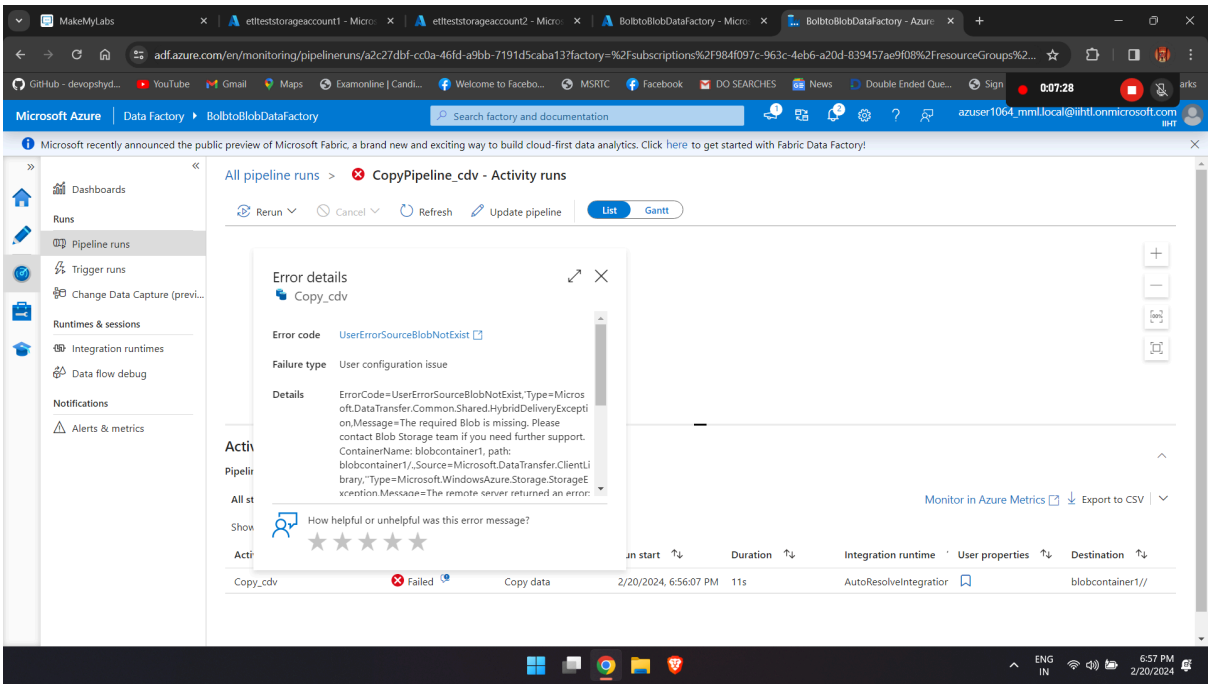
Monitoring and Management: ADF provides monitoring and management capabilities to track the performance and health of your data pipelines. You can monitor pipeline runs, track data lineage, view execution logs, and set up alerts to be notified of pipeline failures or performance issues.

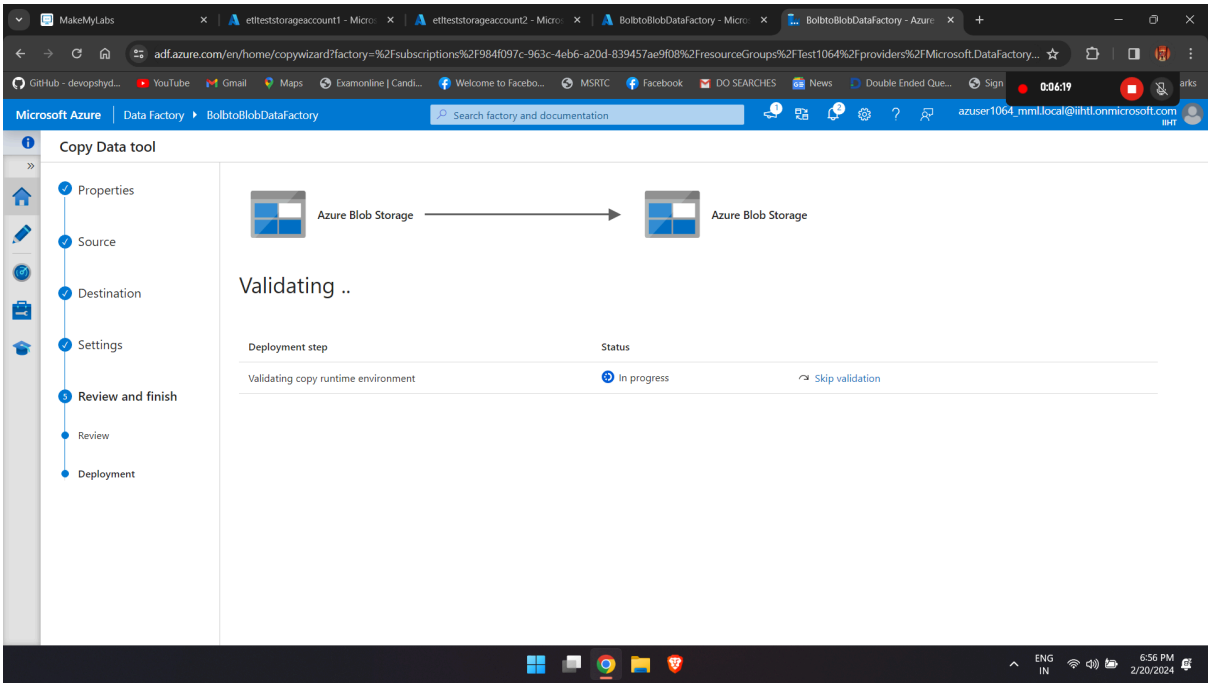
Integration with Azure Services: Azure Data Factory integrates seamlessly with other Azure services, such as Azure Monitor, Azure Active Directory, Azure Key Vault, and Azure DevOps, enabling you to leverage additional capabilities for monitoring, security, and governance.

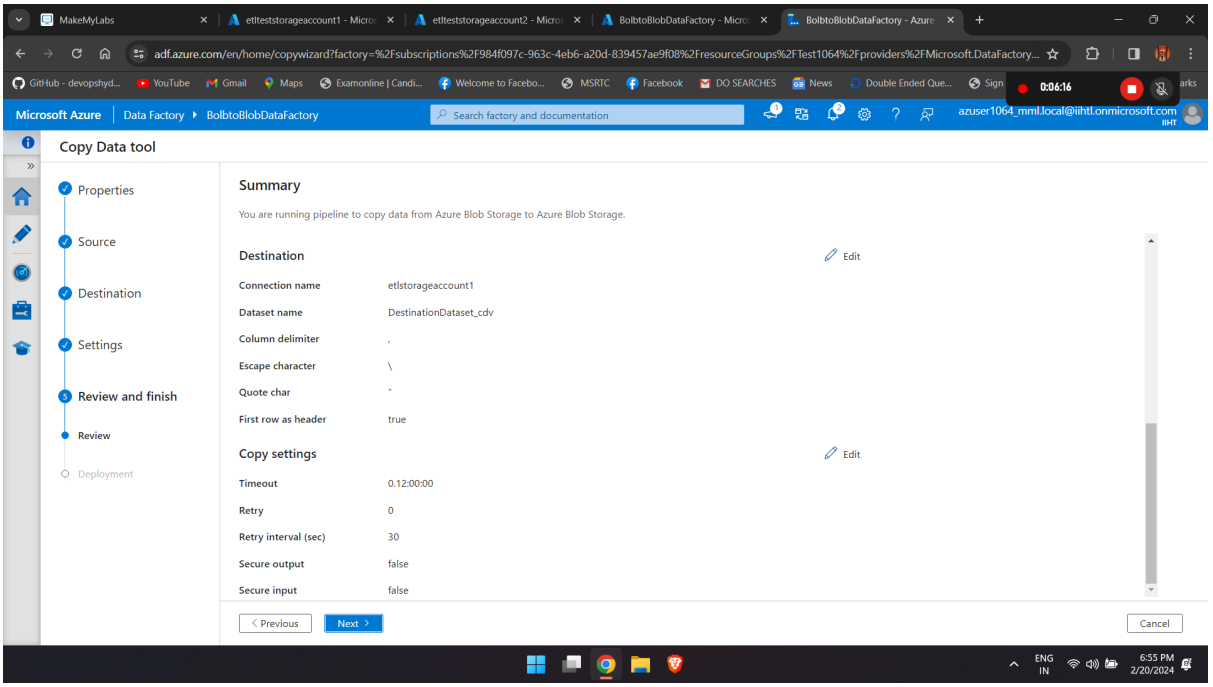
Security and Compliance: ADF provides features for securing your data pipelines and complying with data governance requirements. It supports encryption at rest and in transit, role-based access control (RBAC), audit logging, and data masking to protect sensitive data and ensure compliance with regulatory standards.

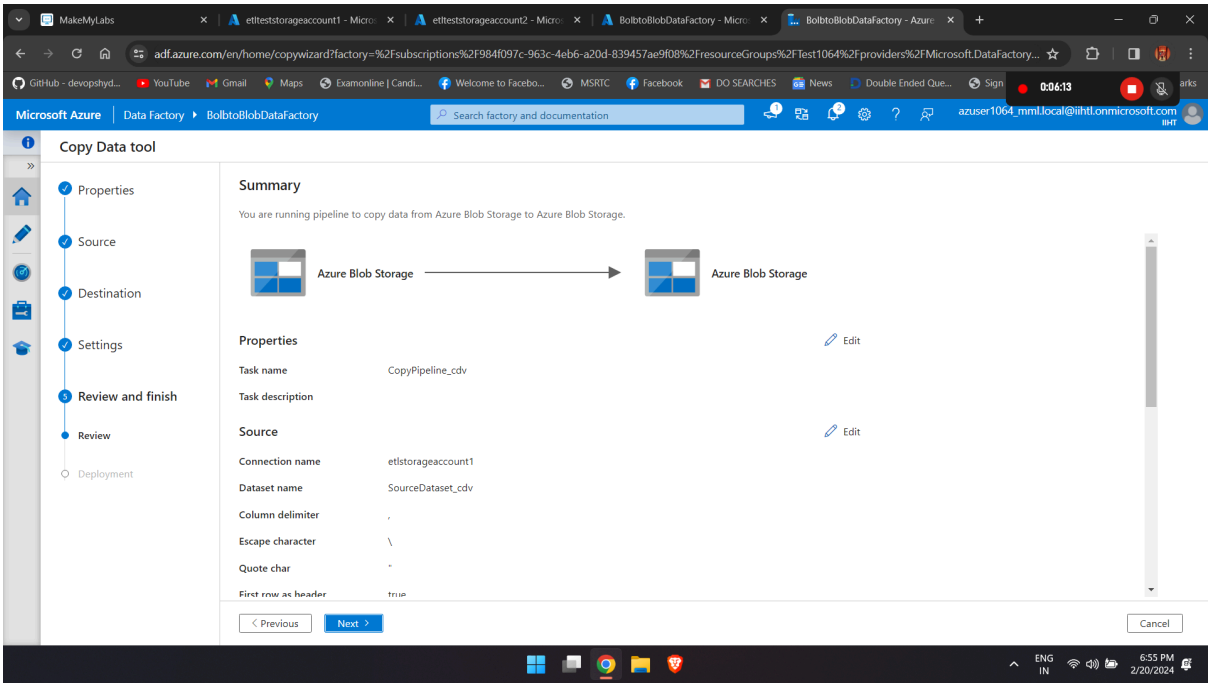
Overall, Azure Data Factory is a powerful data integration service that enables organisations to build scalable, reliable, and efficient data pipelines for ingesting, transforming, and processing data across hybrid and multi-cloud environments. It simplifies the process of managing complex data workflows and enables organisations to unlock insights from their data faster and more effectively.

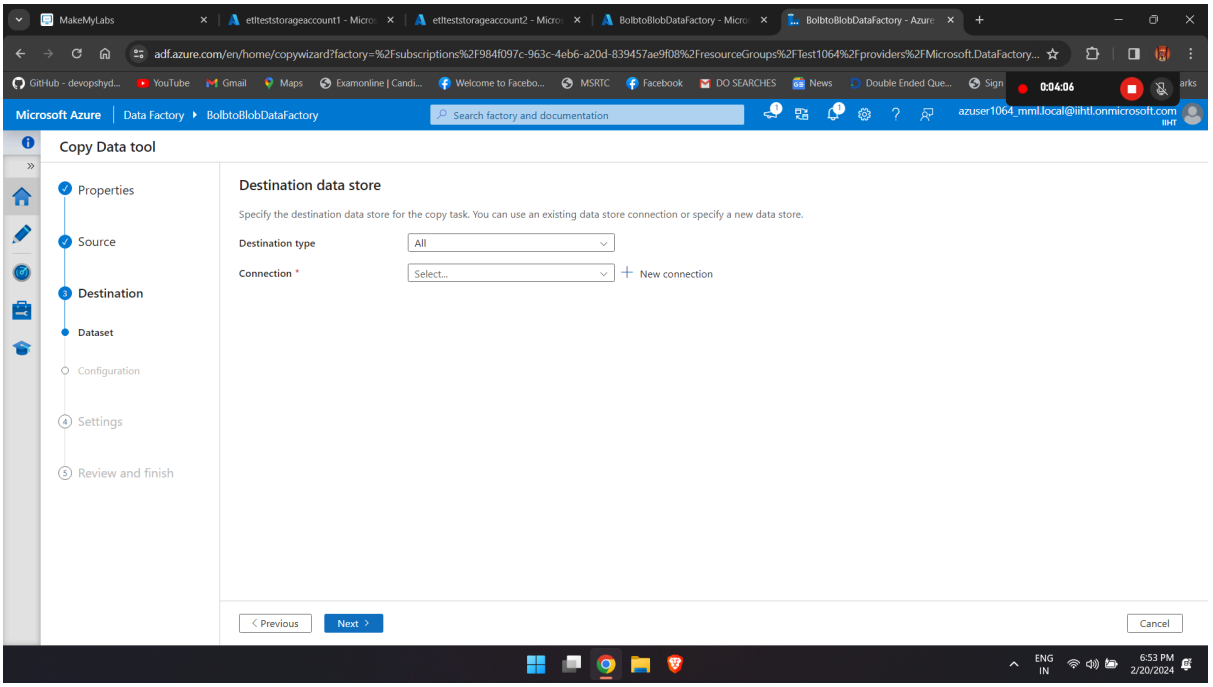
Hands On :

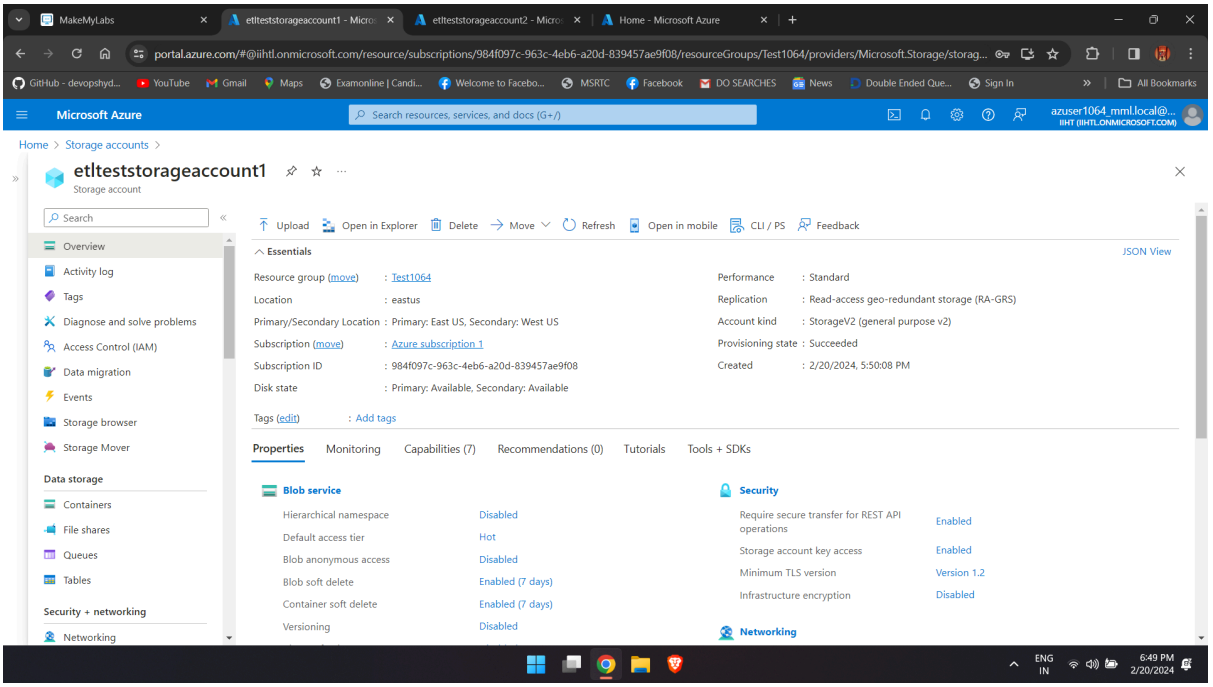

Microsoft Azure
Data Factory
BolbtoBlobDataFactory
All pipeline runs
CopyPipeline_cdv - Activity runs
Rerun
Cancel
Refresh
Update pipeline
List
Gantt
Error details
Copy_cdv
Error code
UserErrorSourceBlobNotExist
Failure type
User configuration issue
Details
How helpful or unhelpful was this error message?
Monitor in Azure Metrics
Export to CSV
Copy_cdv
Failed
Copy data
2/20/2024, 6:56:07 PM
11s
AutoResolveIntegratior
blobcontainer1//

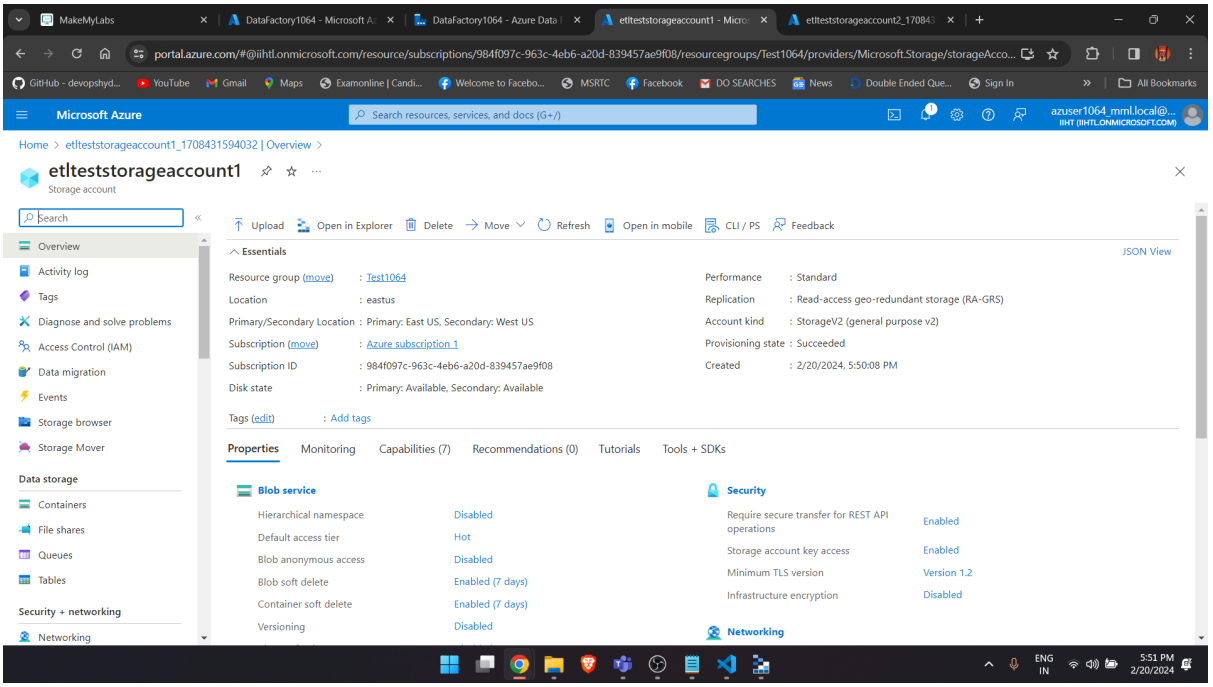

Microsoft Azure
Data Factory
BolbtoBlobDataFactory
Copy Data tool
Properties
Source
Destination
Settings
Review and finish
Review
Deployment
Azure Blob Storage
Azure Blob Storage
Validating ..
Deployment step
Status
Validating copy runtime environment
In progress
Skip validation

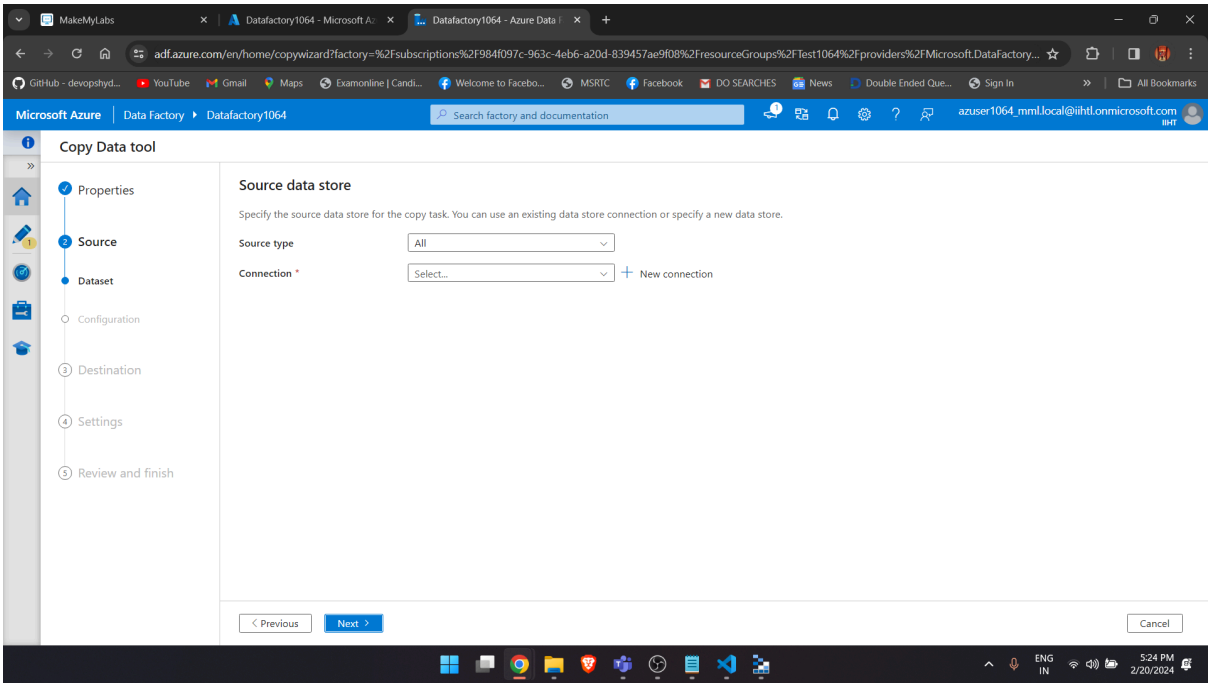

Microsoft Azure
Data Factory
BolbtoBlobDataFactory
Search factory and documentation
azuser1064_mml.local@iihtl.onmicrosoft.com
Copy Data tool
Properties
Source
Destination
Settings
Review and finish
Review
Deployment
Summary
You are running pipeline to copy data from Azure Blob Storage to Azure Blob Storage.
Destination
Edit
Connection name
etlstorageaccount1
Dataset name
DestinationDataset_cdv
Column delimiter
,
Escape character
\
Quote char
"
First row as header
true
Copy settings
Edit
Timeout
0.12:00:00
Retry
0
Retry interval (sec)
30
Secure output
false
Secure input
false
Previous
Next
Cancel


Microsoft Azure
Data Factory
BolbtoBlobDataFactory
Search factory and documentation
azuser1064_mml.local@iihtl.onmicrosoft.com
Copy Data tool
Properties
Source
Destination
Settings
Review and finish
Review
Deployment
Summary
You are running pipeline to copy data from Azure Blob Storage to Azure Blob Storage.
Azure Blob Storage
Azure Blob Storage
Properties
Edit
Task name
CopyPipeline_cdv
Task description
Source
Edit
Connection name
etlstorageaccount1
Dataset name
SourceDataset_cdv
Column delimiter
,
Escape character
\
Quote char
"
First row as header
true
Previous
Next
Cancel


MakeMyLabs
etlteststorageaccount1 - Micro
etlteststorageaccount2 - Micro
BolbtoBlobDataFactory - Micro
BolbtoBlobDataFactory - Azure
adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2FTest1064%2Fproviders%2FMicrosoft.DataFactory...
0:04:06
Microsoft Azure
Data Factory
BolbtoBlobDataFactory
Search factory and documentation
azuser1064_mml.local@iihtl.onmicrosoft.com
Copy Data tool
Properties
Source
Destination
Dataset
Configuration
Settings
Review and finish
Destination data store
Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.
Destination type
All
Connection *
Select...
New connection
Previous
Next
Cancel
ENG
IN
6:53 PM
2/20/2024


MakeMyLabs
etlteststorageaccount1 - Micro
etlteststorageaccount2 - Micro
Home - Microsoft Azure
portal.azure.com/#@iihtl.onmicrosoft.com/resource/subscriptions/984f097c-963c-4eb6-a20d-839457ae9f08/resourceGroups/Test1064/providers/Microsoft.Storage/storag...
Microsoft Azure
Search resources, services, and docs (G+/)
azuser1064_mml.local@...
IIHT (IIHTL.ONMICROSOFT.COM)
Home > Storage accounts >
etlteststorageaccount1
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
Networking
Upload
Open in Explorer
Delete
Move
Refresh
Open in mobile
CLI / PS
Feedback
Essentials
JSON View
Resource group (move) : Test1064
Location : eastus
Primary/Secondary Location : Primary: East US, Secondary: West US
Subscription (move) : Azure subscription 1
Subscription ID : 984f097c-963c-4eb6-a20d-839457ae9f08
Disk state : Primary: Available, Secondary: Available
Performance : Standard
Replication : Read-access geo-redundant storage (RA-GRS)
Account kind : StorageV2 (general purpose v2)
Provisioning state : Succeeded
Created : 2/20/2024, 5:50:08 PM
Tags (edit) : Add tags
Properties
Monitoring
Capabilities (7)
Recommendations (0)
Tutorials
Tools + SDKs
Blob service
Hierarchical namespace Disabled
Default access tier Hot
Blob anonymous access Disabled
Blob soft delete Enabled (7 days)
Container soft delete Enabled (7 days)
Versioning Disabled
Security
Require secure transfer for REST API operations Enabled
Storage account key access Enabled
Minimum TLS version Version 1.2
Infrastructure encryption Disabled
Networking
ENG
IN
6:49 PM
2/20/2024


MakeMyLabs
DataFactory1064 - Microsoft A
DataFactory1064 - Azure Data
etlteststorageaccount1 - Micro
etlteststorageaccount2_170843
portal.azure.com/#@iihtl.onmicrosoft.com/resource/subscriptions/984f097c-963c-4eb6-a20d-839457ae9f08/resourcegroups/Test1064/providers/Microsoft.Storage/storageAcco...
GitHub - devopshyd...
YouTube
Gmail
Maps
Examonline | Candi...
Welcome to Facebo...
MSRTC
Facebook
DO SEARCHES
News
Double Ended Que...
Sign In
All Bookmarks
Microsoft Azure
Search resources, services, and docs (G+/)
azuser1064_mml.local@...
IIHT (IIHTL.ONMICROSOFT.COM)
Home > etlteststorageaccount1_1708431594032 | Overview >
etlteststorageaccount1
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
Networking
Upload
Open in Explorer
Delete
Move
Refresh
Open in mobile
CLI / PS
Feedback
Essentials
JSON View
Resource group (move) : Test1064
Location : eastus
Primary/Secondary Location : Primary: East US, Secondary: West US
Subscription (move) : Azure subscription 1
Subscription ID : 984f097c-963c-4eb6-a20d-839457ae9f08
Disk state : Primary: Available, Secondary: Available
Performance : Standard
Replication : Read-access geo-redundant storage (RA-GRS)
Account kind : StorageV2 (general purpose v2)
Provisioning state : Succeeded
Created : 2/20/2024, 5:50:08 PM
Tags (edit) : Add tags
Properties
Monitoring
Capabilities (7)
Recommendations (0)
Tutorials
Tools + SDKs
Blob service
Hierarchical namespace Disabled
Default access tier Hot
Blob anonymous access Disabled
Blob soft delete Enabled (7 days)
Container soft delete Enabled (7 days)
Versioning Disabled
Security
Require secure transfer for REST API operations Enabled
Storage account key access Enabled
Minimum TLS version Version 1.2
Infrastructure encryption Disabled
Networking
ENG
IN
5:51 PM
2/20/2024


MakeMyLabs
Datafactory1064 - Microsoft A
Datafactory1064 - Azure Data
adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2FTest1064%2Fproviders%2FMicrosoft.DataFactory...
GitHub - devopshyd...
YouTube
Gmail
Maps
Examonline | Candi...
Welcome to Facebo...
MSRTC
Facebook
DO SEARCHES
News
Double Ended Que...
Sign In
All Bookmarks
Microsoft Azure
Data Factory
Datafactory1064
Search factory and documentation
azuser1064_mml.local@iihtl.onmicrosoft.com
IIHT
Copy Data tool
Properties
Source
Dataset
Configuration
Destination
Settings
Review and finish
Source data store
Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.
Source type
All
Connection *
Select...
New connection
Previous
Next
Cancel
ENG
IN
5:24 PM
2/20/2024