

Azure data factory and its copy activity:-

Azure Data Factory is a cloud-based data integration service that allows you to create, schedule, and orchestrate data pipelines to move and transform data from various sources to different destinations. It provides a platform for building hybrid data integration solutions that connect data across cloud and on-premises environments.

One of the key components of Azure Data Factory is Copy Activity, which is used to move data between different data stores. Here's how you can execute and explain Copy Activity in Azure Data Factory:

1.Create a Data Factory: First, you need to create an Azure Data Factory instance in the Azure portal. This involves providing basic information like subscription, resource group, region, and name for your data factory.

2.Create Linked Services: Linked Services are connections to external data sources or destinations. You'll need to create linked services for the source and destination data stores you want to work with. For example, if you're copying data from an Azure SQL Database to Azure Blob Storage, you'll need to create linked services for both databases.

3.Create Datasets: Datasets represent the data structures within your data stores. You'll define datasets for the source and destination data, specifying the format, schema, and location of the data. This

helps Data Factory understand the structure of the data it's working with.

4.Create Pipelines: Pipelines are the workflows that define the data movement and transformation tasks. You'll create a new pipeline and add Copy Activity to it. Within the Copy Activity, you'll specify the source dataset, destination dataset, and any transformations or mappings required.

5.Configure Copy Activity: In the Copy Activity settings, you'll specify details like the source and destination linked services, the source and destination datasets, any data transformations or mappings, and options for copying data (e.g., incremental copy, parallel copy).

6.Debug and Execute Pipeline: Before deploying your pipeline to production, you can debug and test it within the Azure Data Factory interface. This allows you to verify that the data movement and transformations are working as expected. Once you're satisfied with the pipeline, you can deploy and schedule it to run automatically according to your desired frequency.

Now, let's explain some key concepts and features of Azure Data Factory Copy Activity:

- **Data Movement:** Copy Activity supports moving data between various data sources and destinations, including Azure Blob Storage, Azure SQL Database, Azure Data Lake Storage, on-premises databases, and more. It provides built-in connectors

for many common data sources and destinations, making it easy to integrate with different systems.

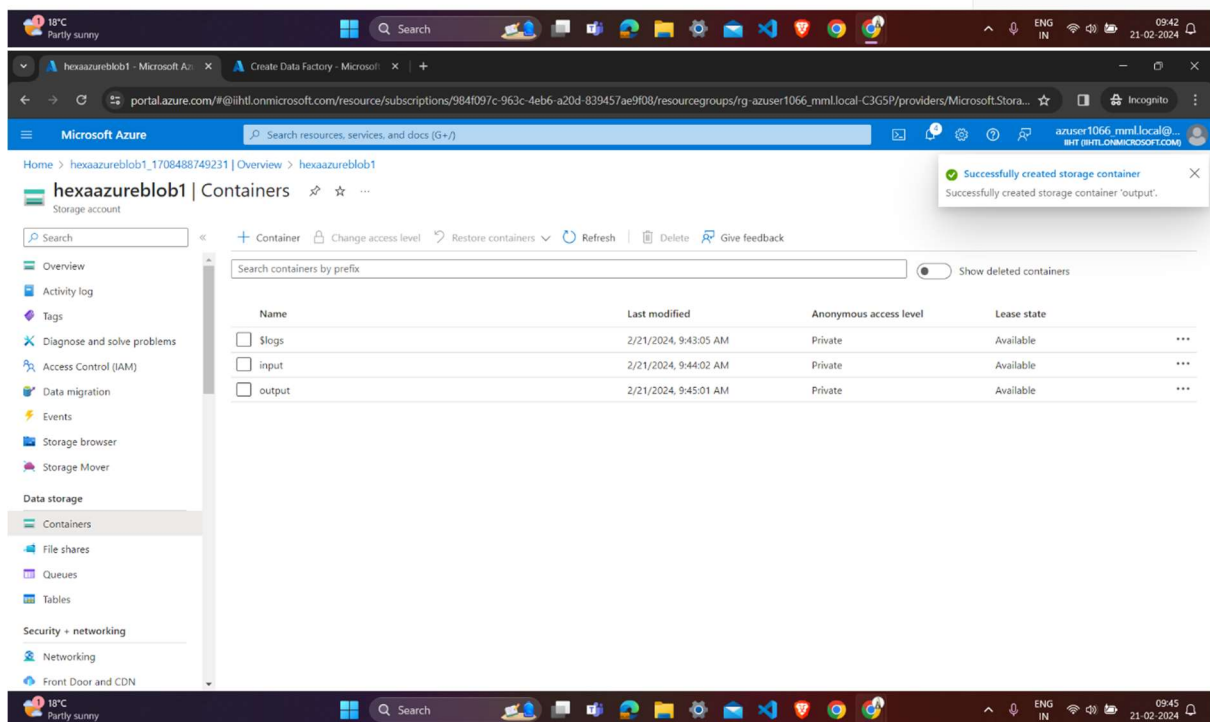
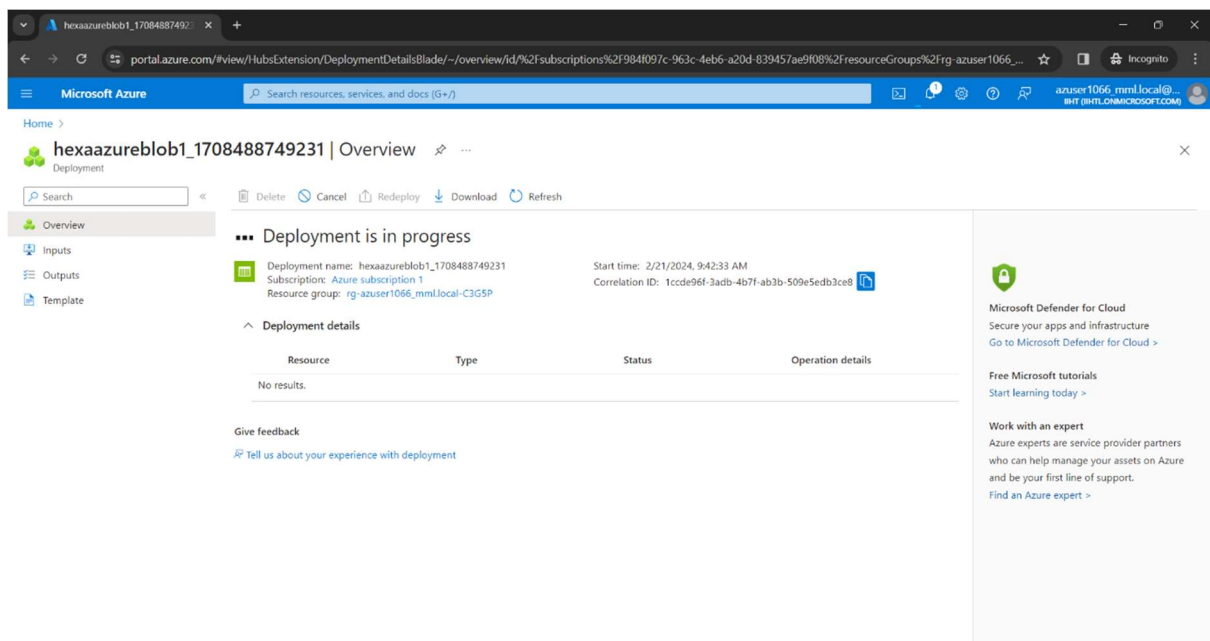
- **Data Transformation:** Copy Activity allows you to perform basic transformations on the data during the copy process, such as column mapping, data type conversion, and filtering. You can also use Azure Data Factory's data flow feature for more advanced data transformations and processing.
- **Performance and Scalability:** Copy Activity is designed for high performance and scalability, with features like parallel data transfer and automatic retry mechanisms to handle large volumes of data and ensure data integrity.
- **Monitoring and Logging:** Azure Data Factory provides monitoring and logging capabilities to track the execution of pipelines and activities, monitor data movement performance, and troubleshoot any issues that arise during the copy process.

Overall, Copy Activity in Azure Data Factory is a powerful tool for building data integration and ETL (extract, transform, load) pipelines, allowing you to move and transform data efficiently between different data sources and destinations in the cloud and on-premises environments.

Copy Activity:- Blob to Blob

Create data storage account:-

- Create a hexaazureblob1 account
- Create two container in it input and output
- Upload a csv file in the input container biba.csv.



input - Microsoft Azure

Create Data Factory - Microsoft

portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade~/~/overview/storageAccountId/%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2Fresourcegroup...

Microsoft Azure

Search resources, services, and docs (G+I)

user1066.mml.local@...
BHT (BHTLONMICROSOFT.COM)

Home > hexaazureblob1_1706488749231 | Overview > hexaazureblob1 | Containers >

input
Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: input

Search blobs by prefix (case-sensitive)

Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> Bibadata.csv	2/21/2024, 9:44:43 AM	Hot (Inferred)		Block blob	7.29 KiB	Available ***

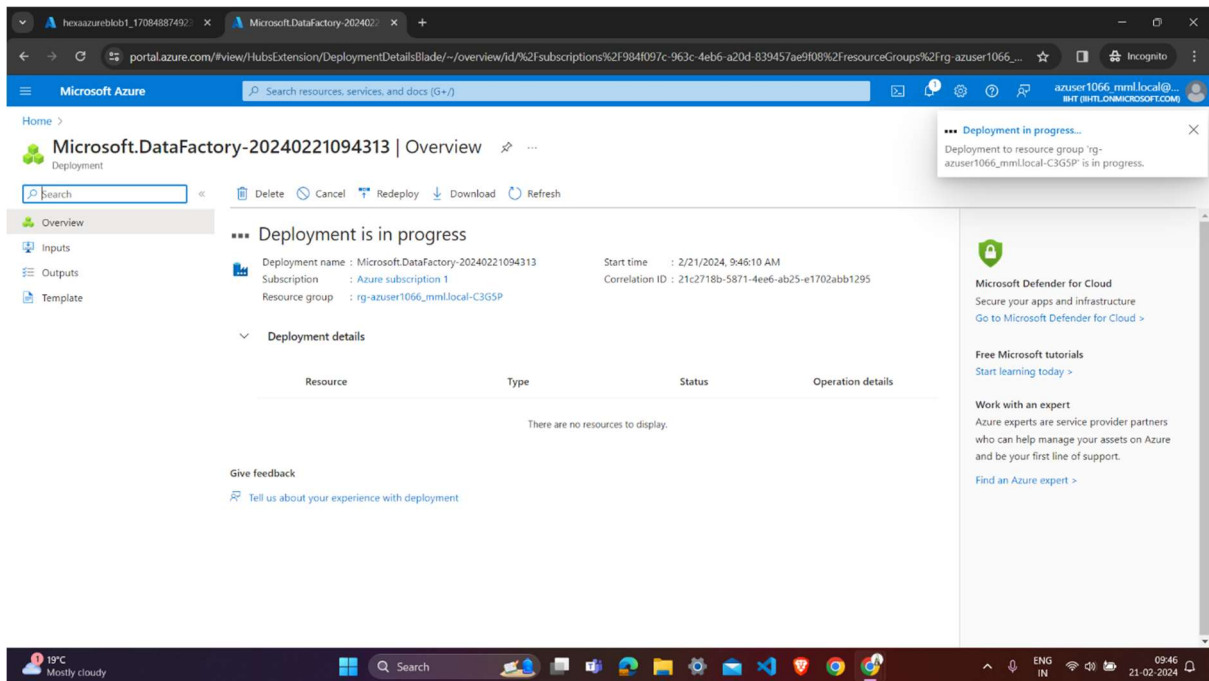
19°C
Partly sunny

Search

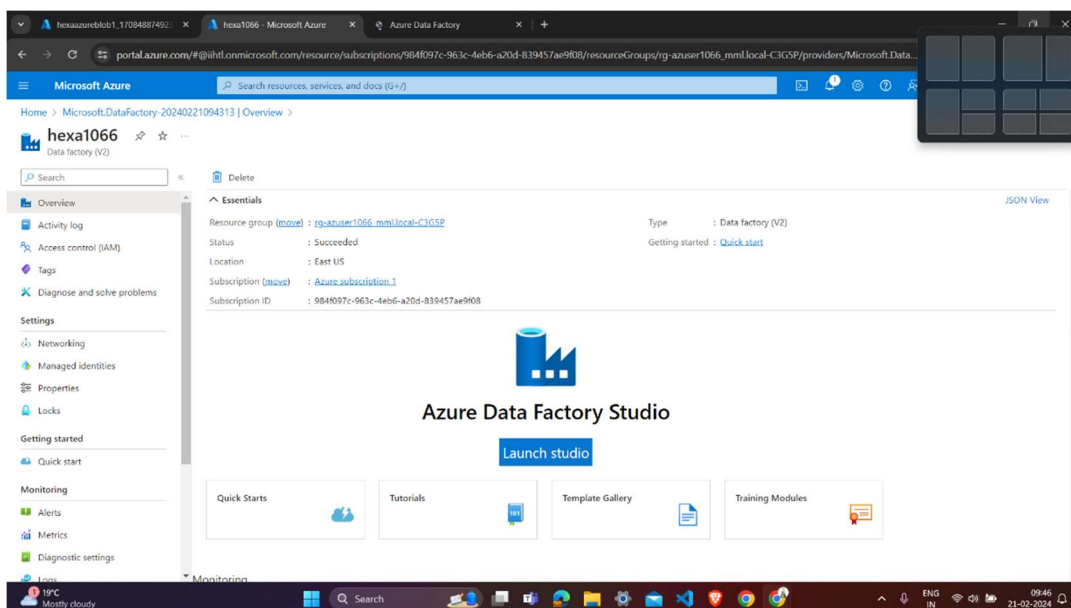
ENG
IN

09:45
21.02.2024

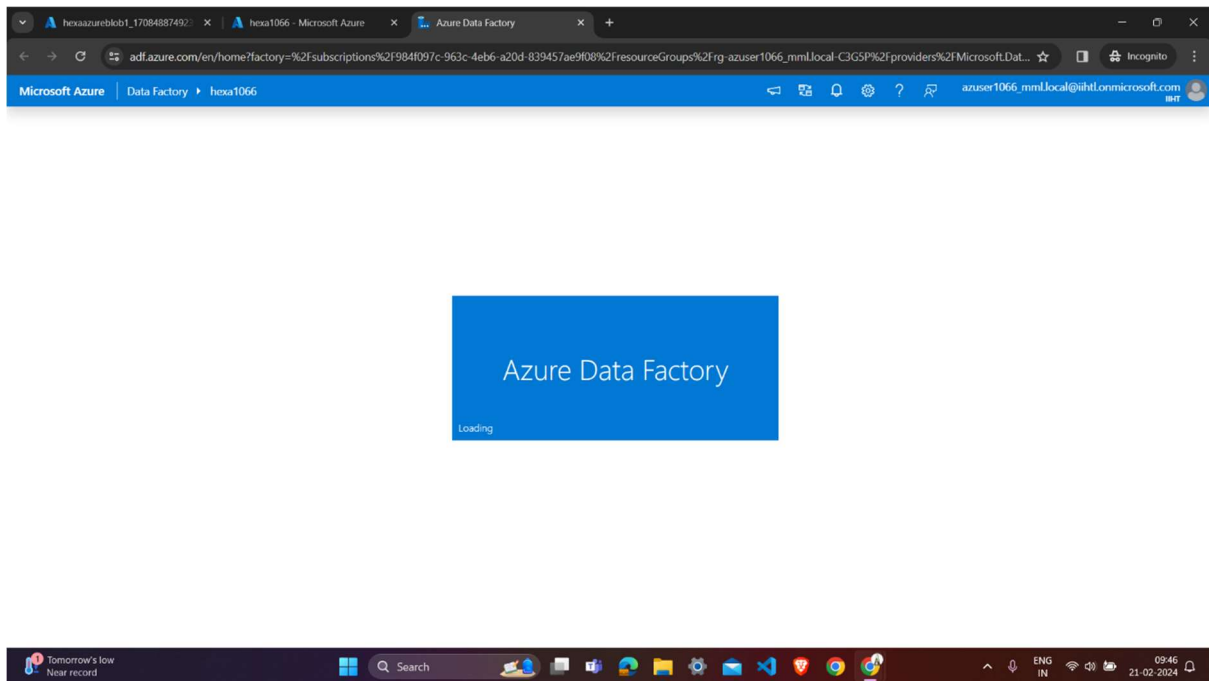
Create a data factory name hexa1066:-



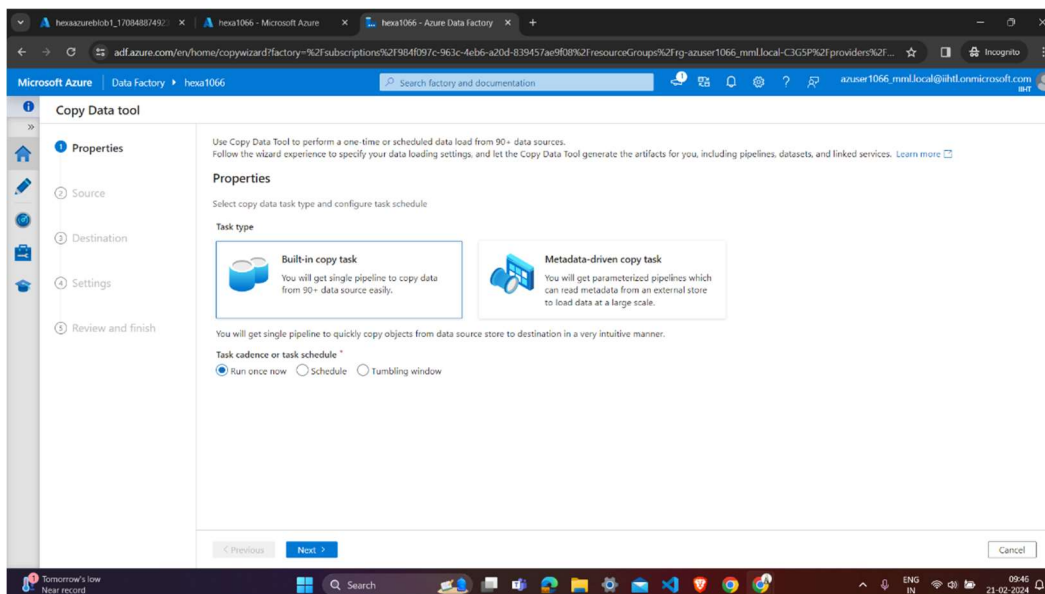
Launch Azure Data factory studio:-



Azure Data factory launching:-



Create built-in copy task:-



Enter source type (azure blob storage) and create new connection:-

Microsoft Azure | Data Factory | hexa1066

Copy Data tool

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: Azure Blob Storage

Connection: Select... + New connection

New connection

Authentication type: Account key

Connection string: Azure Key Vault

Account selection method: From Azure subscription (selected) / Enter manually

Azure subscription: Azure subscription 1 (984f097c-963c-4eb6-a20d-839457ae9f08)

Storage account name: hexaazureblob1

Additional connection properties: + New

Test connection: To linked service (selected) / To file path

Annotations: + New

Parameters: + New

Create Cancel

Connection successful

Test connection

Enter the source and destination detail :-

Microsoft Azure | Data Factory | hexa1066

Copy Data tool

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: Azure Blob Storage

Connection: hexaazureblob1 Edit + New connection

File name:

Compression type: None

Copy behavior: Select...

Max concurrent connections:

Block size (MB):

Metadata: + New

Previous Next Cancel

hexaazureblob1_17084887492 x hexa1066 - Microsoft Azure x hexa1066 - Azure Data Factory x +

adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2Frg-azuser1066_mml.local-C3G5P%2Fproviders%2F... azuser1066_mml.local@ihl.onmicrosoft.com

Copy Data tool

Properties Source Destination Settings Review and finish

Settings

Enter name and description for the copy data task, more options for data movement

Task name * CopyPipeline_h7b

Task description

Data consistency verification ☐

Enable logging ☐

Enable staging ☐

> Advanced

< Previous Next > Cancel

19°C Mostly cloudy 09:51 21.02.2024

hexaazureblob1_17084887492 x hexa1066 - Microsoft Azure x hexa1066 - Azure Data Factory x +

adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2Frg-azuser1066_mml.local-C3G5P%2Fproviders%2F... Incognito

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

Task name CopyPipeline_h7b Edit

Task description Edit

Source

Connection name hexaazureblob1 Edit

Dataset name SourceDataset_h7b Edit

File name Blobdata.csv

Container input

Destination

Connection name hexaazureblob1 Edit

< Previous Next > Cancel

19°C Mostly cloudy 09:51 21.02.2024

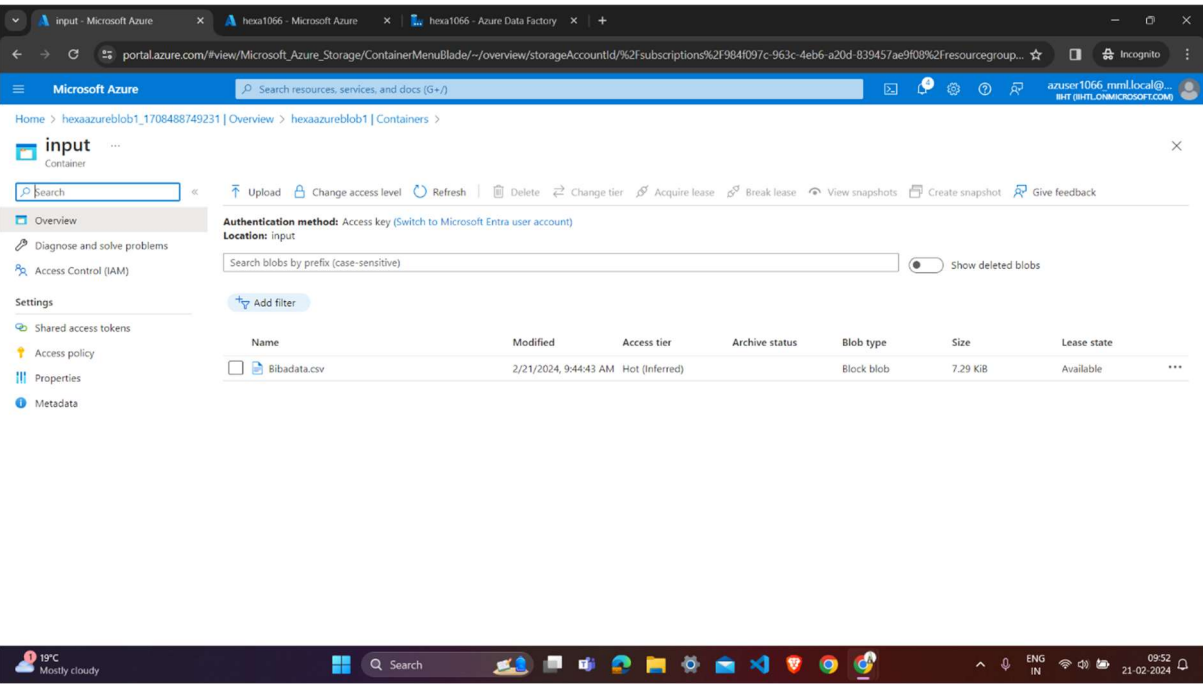
Finish Deployment:-

The screenshot displays the Microsoft Azure Data Factory 'Copy Data tool' interface. The left sidebar shows the navigation menu with steps: Properties, Source, Destination, Settings, Review and finish, Review, and Deployment. The main area shows a diagram of data flow from 'Azure Blob Storage' to 'Azure Blob Storage'. Below the diagram, the text 'Deployment complete' is displayed. A table lists the deployment steps and their status:

Deployment step	Status
Validating copy runtime environment	✓ Succeeded
> Creating datasets	✓ Succeeded
> Creating pipelines	✓ Succeeded
> Running pipelines	✓ Succeeded

Below the table, a message states: 'Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.' At the bottom of the main area, there are three buttons: 'Finish', 'Edit pipeline', and 'Monitor'. The bottom of the screen shows the Windows taskbar with the date and time as 21-02-2024 09:51.

Input file we upload biba.csv:-



Output container have copy of input biba.csv file:-

