

Azure Data Factory

Azure data engineering: azure data engineering is collection of the following services which are inbuilt /packed in azure data engineering those are

Data integration

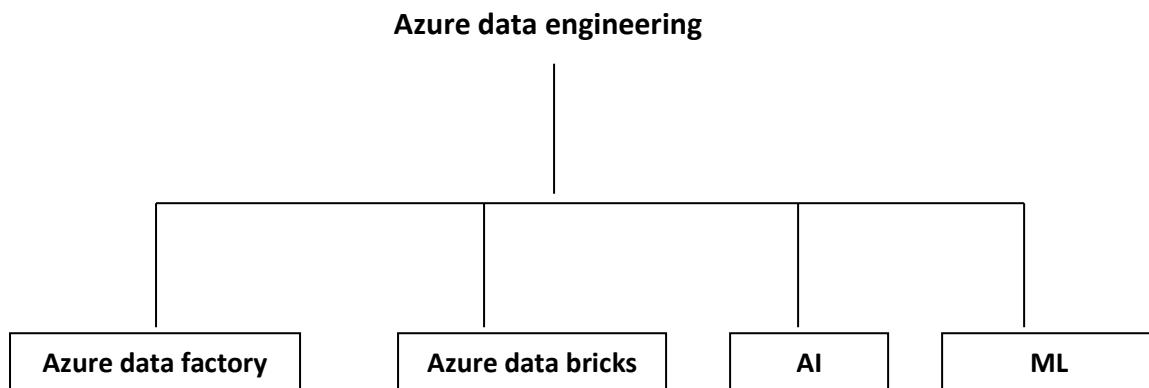
Data migration

Data summarization

Data visualization

Under this data engineering we have following

Azure data factory Azure data bricks Synapse Ai MI



Azure data factory: azure data factory is a service provided by the Microsoft to do the job (Data integration, Data migration, Data summarization, Data visualization)

Data migration: data migration is nothing but moving the data from on-premises environment to cloud environment is called data migration

Data summarization: data summarization is nothing but converting data into required form is called data summarization

```
*Untitled - Notepad
File Edit Format View Help
Azure Data Factory-->Provided by the Microsoft to the job(Data Integration,Data Migration,Data summarization)
-----
Cloud Service |--->Google Drive----->Cloud
-----
On-Premises |--->Mobiles,SD
-----
Mobiles-->16GB,32GB-(500GB)
-----
SD----->(500GB)
--
900GB-->Google Drive----->Cloud
-----
1.Accessing of data---->
2.Security----->Google
3.Unlimited----->15GB--->
-----
Data summarization
-----
Data-->Information----->Knowledge--(BI)
-----
Name----->(10(Evening Batch))-->5(Ok)5
Email---->Email_Id----->(5 (Morning Batch))
Mobile-->Mobile----->
City
Course-->Course(ADF)----->

Ln 39, Col 6 100% Windows (CRLF) UTF-8
```

```
*Untitled - Notepad
File Edit Format View Help
Data Summarization
-----
Market--(Cook some veg Curry)
-----
50--veg----->Extraction-->5Veg---->Transformation--Cooking->----->Loading-->(E T L)-->(Data Summarization)-->ADF
-----
PId Prate PQty_Sold P_MFR      P_EXP      DOS----->(Net_Sales_Amount)
123 300   10    01/01/2023  01/01/2024  20/06/2023----->Extraction Of Data
124 200   10    01/03/2023  01/03/2024  20/06/2023
125 100   20    02/02/2023  02/02/2024  20/06/2023          Prate PQty_Sold DOS

Transformation----->Loading
----- Net_Sales_Amt DOS
----- 7000        20/06/2023
3000
2000
2000
-----
7000

Source
-----
On-premises           Google Drive
-----
Files----->ADF----->Target
-----
1.IaaS-->Infrastructure----->
2.PaaS-->O/S----->Windows-->Mainframe(IBM)
3.SaaS----->
-----|
```

Cloud services providers in market

1. Azure from Microsoft
2. Gcp from Google
3. Aws form Amazon
4. ali baba

Azure Data Factory

5. sap cloud

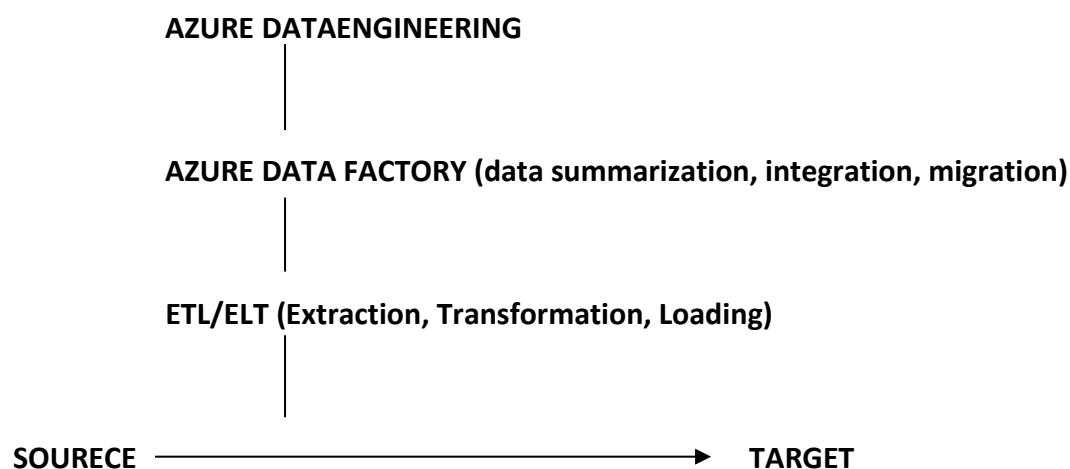
6.oracle cloud

```

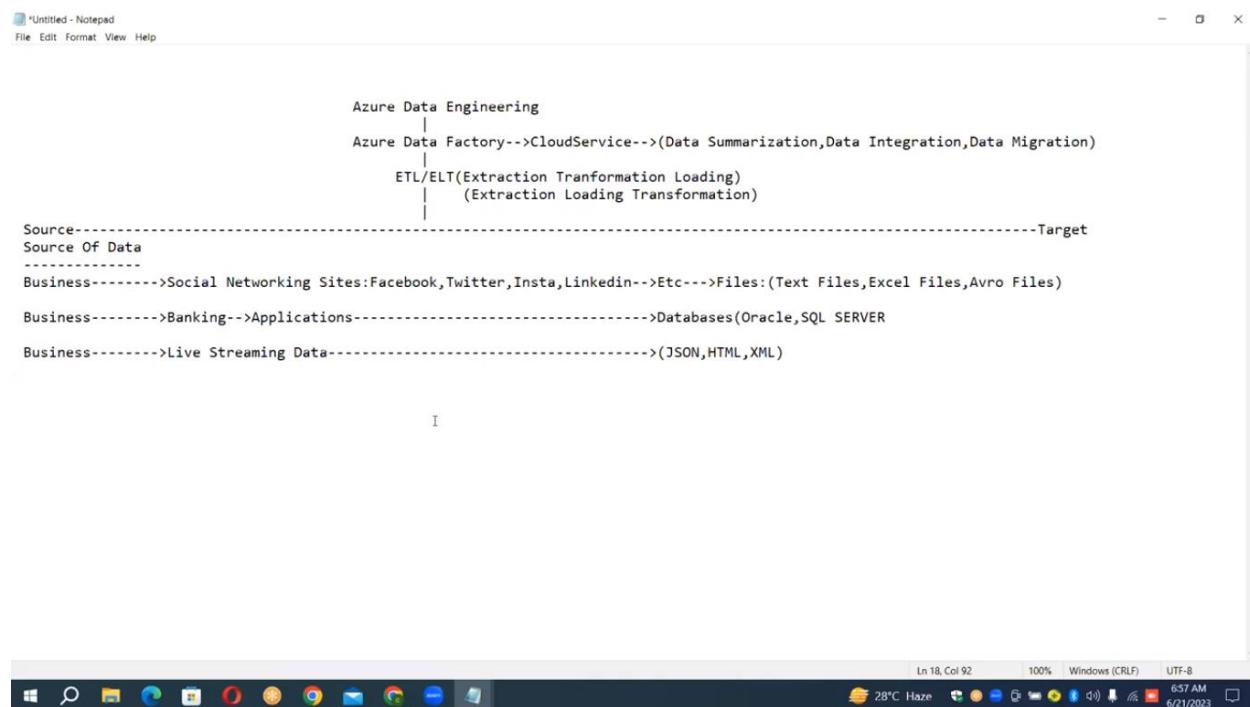
"Untitled - Notepad
File Edit Format View Help
-----
Net_Sales_Amt DOS
7000 20/06/2023
3000
2000
2000
-----
7000

Source
-----
On-premises Google Drive
-----
Files-----ADF----->Target
-----
1.Iaas-->Infrastructure-->
2.PaaS-->O/S----->Windows-->Mainframe(IBM)
3.SaaS----->
-----
1.Azure--->Microsoft--->(ADF)-->Data Summarization,Data Integration,Data migration
2.GCP----->Google
3.AWS----->Amazon
4.Ali baba
5.SAP CLOUD
6.ORACLE CLOUD
-----
```

Ln 78, Col 38 100% Windows (CRLF) UTF-8



Azure Data Factory



ETL: (EXTRACTION, TRANSFORMATION, LOADING)

ELT:(EXTRACTION ,LOADING,TRANSFORMATION)

ELT/ETL: require two things they are source ad target

BIG DATA: extremely large and diverse collections of structured, unstructured, and semi-structured data that continues to grow exponentially over time. These datasets are so huge and complex in volume.

Big data is three types they are as follows

1. Structure data

2. un-structure data

3. Semi structure data

Structure data: structure data means where the data stored in table format that data is called structure data. Table is collection of rows and column

```

-----
1.Structured Data----->Table(Rows & Columns)
Employees
-----
Name  Id   Sal   Job
-----
Allen 33  3000  Clerk
Carl  23  4000  SrClerk
  
```

Azure Data Factory

Semi structure data that is not captured or formatted in conventional ways. Semi-structured data does not follow the format of a tabular data model or relational databases because it does not have a fixed schema

```

2.Semi-Structured Data----->XML,HTML,JSON
JSON-->(Java Script Object Notation File)
(Key Value:Pair)

{
    "Name": "Allen",
    "Id" : 33,
    "Sal" : 3000,
    "Job" : "Clerk"
}
-----
XML

<Employees>
    <Name>Allen</Name>
    <Id>33</Id>
    <Sal>3000</Sal>
    <Job>Clerk</Job>
</Employees>

```

Un-structure data: Unstructured data is information that either does not have a pre-defined data model or is not organized in a pre-defined manner ex: photos, videos, voice chat

```

*Untitled - Notepad
File Edit Format View Help

Business----->Live Streaming Data----->(JSON,HTML,XML)
-----
Big-Data-(Huge Information)
-----
1.Structured Data----->Table(Rows & ----->20%
Employees
-----
Name Id Sal Job
-----
Allen 33 3000 Clerk
Carl 23 4000 SrClerk

2.Semi-Structured Data----->XML,HTML,JSON----->20%
JSON-->(Java Script Object Notation File)
(Key Value:Pair)

{
    "Name": "Allen",
    "Id" : 33,
    "Sal" : 3000,
    "Job" : "Clerk"
}
-----
XML

<Employees>
    <Name>Allen</Name>
    <Id>33</Id>
    <Sal>3000</Sal>
    <Job>Clerk</Job>
</Employees>

3.Un-Structured Data----->80%

```

80 % data we will get un-structure format only

Azure storages/Storages in azure

In azure we have 90 plus storages in that we can store structure data, semi structure data and un-structure data

Azure Data Factory

1. AZURE BLOB STORAGES
2. AZURE DATA LAKE GEN1
3. AZURE DATA LAKE GEN2
4. AZURE SQLDATABASE
5. AZURE COSMOS DB
6. AZURE ELASTIC POOL
7. AZURE SYNAPSE ANALYTICS

1. Azure Blob Storages: azure blob storage can store all the data types like structure data, semi -structure data,un-structure data

Blob stands for binary large object

2. Azure Data Lake Gen1: azure data lake will retire in 2024

3. Azure Data Lake Gen2: this is an advanced storage where the data is stored in DFS system DFS stands for distributed file system

4. Azure Sqldatabase: azure data base can store only structure data

5. Azure cosmos db is the trending data storage it can store structure data and semi-structure data only

6. Azure elastic pool: it can store only structure data

7. Azure synapse analytics: it can stores structure data

```
-----
Azure Storage/Storages in Azure-->90+
-----
1.Azure Blob Storage-->((Structured Data,Semi-Structured Data,Un-Structured Data))
2.Azure Data Lake Gen1-->2024(retire)
3.Azure Data Lake Gen2-->DFS-->((Structured Data,Semi-Structured Data,Un-Structured Data))
4.Azure SQL Database---->((Structured Data))
5.Azure COSMOS DB(NOSQL)-->((Structured Data,Semi-Structured Data))
6.Azure Elastic Pool----->((Structured Data))
7.Azure Synapse Analytics->((Structured Data)) | |
```

Azure.com → signup/sigin → portal.azure.com

Azure data factory: azure data factory has following components

- 1. Pipeline**
- 2. Linkedservices**
- 3. Datasets**
- 4. Activities**
- 5. Dataflows**
- 6. Triggers**
- 7. Integration run time**

```
*Untitled - Notepad
File Edit Format View Help
1.Azure Blob Storages--->(Structured Data,Semi-Structured Data,Un-Structured Data)
2.Azure Data Lake Gen1-->2024(retire)
3.Azure Data Lake Gen2-->DFS->(Structured Data,Semi-Structured Data,Un-Structured Data)
4.Azure SQL Database---->((Structured Data))
5.Azure COSMOS DB(NOSQL)->((Structured Data,Semi-Structured Data))
6.Azure Elastic Pool---->((Structured Data))
7.Azure Synapse Analytics->(Structured Data)
-----
Azure.com
-----
Portal.azure.com
-----
Azure Data Factory-->Components
-----
1.Pipeline
2.LinkedService
3.Datasets
4.Activities
5.Dataflows
6.Triggers
7.Integration RunTime
```

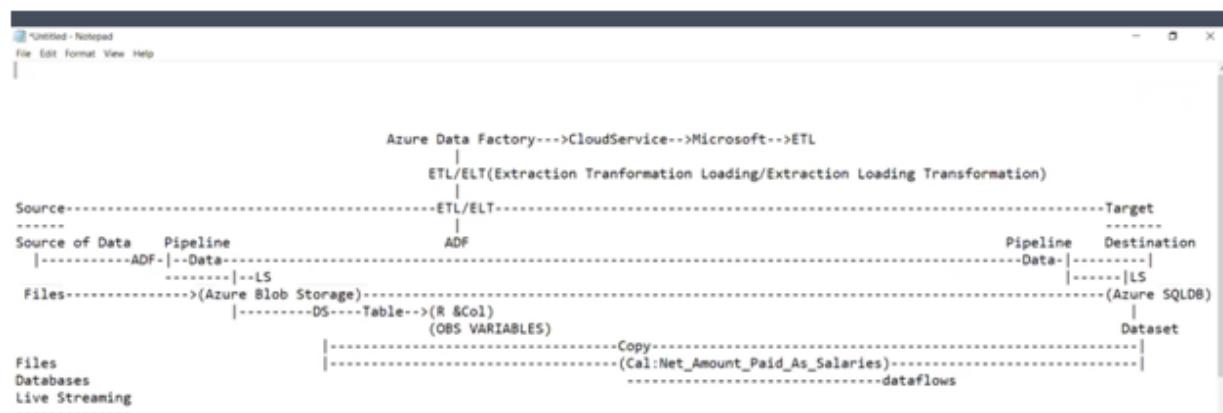
Azure Data Factory

```

*Untitled - Notepad
File Edit Format View Help
-----
Portal.azure.com
-----
Azure Data Factory-->Components
-----
1.Pipeline          |->Datapipeline
2.LinkedService
3.Datasets
4.Activities
5.Dataflows
6.Triggers
7.Integration RunTime
-----
Tank->--->---Water->--->Flats->--->Pipeline
-----
8096374412
-----
Name:
Mobile:
Email_Id:
Batch Timings:ADF 6:30AM IST

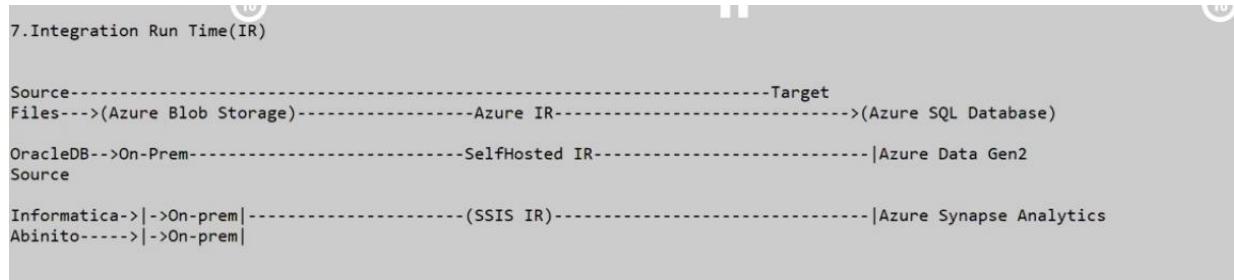
```

- 1. Pipeline:** pipeline stores the data from source to the target
- 2. Linked services:** linked services creates the relationship between two storages
- 3. Datasets:** these are called as pointers of physical location of the data
- 4. Activities:** activities are used to perform the action on the data
- 5. Dataflows:** data flows is application of transformation on the data
- 6. Triggers:** triggers are used to kick start the pipeline in automated manner (event based trigger, schedule based trigger, tumbling window trigger)



7. Integration run time:

Azure Data Factory



Case study 1

Title : copy data from azure blob storage to azure blob storage

Specification:

Source storage ----> azure blob storage

File -----> airlines.csv

Sink data storage -----> azure blob storage

Steps

1. Create resource group
2. Create source storage account
3. Create sink storage account
4. Create ADF session
5. Select mange tile: create link service
6. Create data sets
7. Create pipeline
8. Assign datasets to pipeline
9. Trigger/debug

Create resource group: goto portal.azure.com --> click on create resource group----> click on create ---->assign the resource group--->click on create review---->click on create-goto resource

Azure Data Factory

The screenshot shows the Microsoft Azure Home page. In the top navigation bar, there are tabs for 'ADF (Azure Data Factory) Mr. S' and 'Home - Microsoft Azure'. Below the tabs, the URL is 'portal.azure.com/#home'. The main content area is titled 'Azure services' and features a 'Resource groups' card with a 'Create' button and a 'View' button. Other cards include 'Storage accounts', 'Azure Synapse Analytics', 'SQL databases', 'Elastic Job agents', and a 'More services' link. Below the cards, there's a 'Free training from Microsoft' section with a link to 'Control and organize Azure resources with Azure...'. A 'Recent' tab is selected under 'Resources', showing a list of recent items including 'Name' and 'een viewed recently'. At the bottom of the page, there are links for 'View all resources', 'Navigate' (with Subscriptions, Resource groups, All resources, and Dashboard), and 'Tools' (Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, Cost Management). The status bar at the bottom shows the URL 'https://portal.azure.com/#blade/HubsExtension/BrowseResourceGroupBlade/resourceType/Microsoft.Resources%2fsubscriptions%2fresourceGroups', the date '6/26/2023', and the time '6:51 AM'.

The screenshot shows the 'Resource groups' blade in Microsoft Azure. The title bar says 'Resource groups - Microsoft Azure'. The main content area displays a search bar with 'Filter for my needs...' and a dropdown menu with 'Subscription equals CloudsEngineerSameer'. Below the search bar, it says 'Showing 0 to 0 of 0 records.' There are buttons for 'Subscription ↑↓' and 'Location ↑↓'. At the bottom, there are grouping and view options ('No grouping' and 'List view'). The center of the page displays a large 'No resource groups to display' message with a 'Create resource group' button. The status bar at the bottom shows the URL 'https://portal.azure.com/#view/HubsExtension/BrowseResourceGroups', the date '6/26/2023', and the time '6:52 AM'.

Azure Data Factory

Assign resource

Click on review&create

Validation passed.

Click on create

Azure Data Factory

ADF (Azure Data Factory) Mr. San... | Resource groups - Microsoft Azure | New Tab | [portal.azure.com/#view/HubsExtension/BrowseResourceGroups](#)

Microsoft Azure | Search resources, services, and docs (G+/-)

Home > Resource groups

Default Directory (cloudsengineersameer@gmail.onmicrosoft.com)

+ Create Manage view Refresh Export to CSV Open query Assign tags

Filter for any field... Subscription equals CloudsEngineerSameer Location equals all Add filter

Showing 1 to 1 of 1 records.

Name	Subscription	Location
Azure6_30AMGroup	CloudsEngineerSameer	East US

No grouping List view

Click on goto resource

Resource group created
Creating resource group 'Azure6_30AMGroup' in subscription 'CloudsEngineerSameer' succeeded.
[Go to resource group](#) [Pin to dashboard](#)

< Previous Page of 0 Next > Give feedback

Windows taskbar: 654 AM 6/26/2023

ADF (Azure Data Factory) Mr. San... | Azure6_30AMGroup - Microsoft Azure | New Tab | [portal.azure.com/#@cloudsengineersameer@gmail.onmicrosoft.com/resource/subscriptions/8ef4b258-ea7b-4877-a5e1-e49adf8e4567/resourcegroups/Azure6_30AMGroup/over...](#)

Microsoft Azure | Search resources, services, and docs (G+/-)

Home > Azure6_30AMGroup

Resource group

+ Create Manage view Delete resource group Refresh Export to CSV Open query Assign tags Move Delete Export template Open in mobile

Search

Overview

Activity log, Access control (IAM), Tags, Resource visualizer, Events

Settings

Deployments, Security, Policies, Properties, Locks

Cost Management

Cost analysis, Cost alerts (preview), Budgets, Advisor recommendations

Monitoring

Insights (preview), Alerts

Essentials

Subscription (move) : CloudsEngineerSameer, Deployment : No deployments, Location : East US

Subscription ID : 8ef4b258-ea7b-4877-a5e1-e49adf8e4567, Tags (edit) : Click here to add tags

Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

Show 0 to 0 of 0 records. Show hidden types

Name Type Location

No resources match your filters
Try changing or clearing your filters.
[Create resources](#) [Clear filters](#)

Learn more Give feedback

Windows taskbar: 654 AM 6/26/2023

Open new tab on browser

2.Create source storage account: goto portal.azure.com -->click on storage accounts---->click on create----> mention storage account name----->click on review ----->click on create---->goto resource

Click on storage account

Storage accounts

Click on create

Create storage account

No storage accounts to display

Create a storage account to store up to 500TB of data in the cloud. Use a general-purpose storage account to store object data, use a NoSQL data store, define and use queues for message processing, and set up file shares in the cloud. Use the Blob storage account and the hot or cool access tiers to optimize your costs based on how frequently your object data is accessed.

Create storage account

Azure Data Factory

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription: CloudsEngineerSameer
Resource group: Azure6_30AMGroup

Instance details

Storage account name: sourceblobdurgo99
Region: (US) East US

Review

Basics

Subscription	CloudsEngineerSameer
Resource Group	Azure6_30AMGroup
Location	eastus
Storage account name	sourceblobdurgo99
Deployment model	Resource manager
Performance	Standard
Replication	Read-access geo-redundant storage (RA-GRS)

Advanced

Enable hierarchical namespace	Disabled
Enable network file system v3	Disabled
Allow cross-tenant replication	Enabled
Access tier	Hot
Enable SFTP	Disabled
Large file shares	Disabled

Networking

Network connectivity	Public
----------------------	--------

Create

Azure Data Factory

Your deployment is complete

Deployment name: sourceblobdurgasoft99_1687742926279
Subscription: CloudEngineerSameer
Resource group: Azure6_30AMGroup

Start time: 6/26/2023, 6:58:55 AM Correlation ID: 93f914f7-1fd1-4a38-a498-214bd957ffec

Deployment succeeded
Deployment 'sourceblobdurgasoft99_1687742926279' to resource group 'Azure6_30AMGroup' was successful.

Next steps

Go to resource (highlighted with a red arrow)

Click on goto resource (highlighted with a black box)

Give feedback
Tell us about your experience with deployment

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 and be your first line of support.
Find an Azure expert >

Properties	
Blob service	Hierarchical namespace: Disabled Default access tier: Hot Blob public access: Enabled Blob soft delete: Enabled (7 days) Container soft delete: Enabled (7 days) Versioning: Disabled Change feed: Disabled NFS v3: Disabled Allow cross-tenant replication: Enabled
File service	Log file sharing: 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	Allow access from: All networks Number of private endpoint connections: 0 Network routing: Microsoft network routing Access for trusted Microsoft services: Yes Endpoint type: Standard

Open new tab

3. Create sink storage account: goto portal.azure.com---->click on storage accounts---->click on creates---->mention storage account name---->click on review---->click on create---->goto resource

Azure Data Factory

Click on storage account

Click on create

The screenshots show the Azure portal interface for creating a new storage account. The first screenshot shows the main Azure services dashboard with a callout for the 'Storage accounts' button. The second screenshot shows the 'Storage accounts' blade with the 'Create' button highlighted.

Azure Data Factory

Provide sink storage account name

Click on review

Click on create

Azure Data Factory

Your deployment is complete

Deployment name: sinkdurgasoft99_1687743076352
Subscription: CloudEngineerSameer
Resource group: Azure6_30AMGroup

Start time: 6/26/2023, 7:01:25 AM
Correlation ID: 0402582c-feeb-4a6f-a349-5e3a652ba3bb

Deployment succeeded
Deployment 'sinkdurgasoft99_1687743076352' to resource group 'Azure6_30AMGroup' 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 and be your first line of support.
Find an Azure expert >

sourceblobdurgasoft99 | Overview >

Storage account

Search

Upload Open in Explorer Delete Move Refresh Open in mobile CU/PS Feedback

Essentials

Resource group (move) : Azure6_30AMGroup
Location : East US
Primary/Secondary Location : Primary: East US, Secondary: West US
Subscription (move) : CloudEngineerSameer
Subscription ID : 8ef4b258-ea7b-4877-a5e1-e49adfe4567
Disk state : Primary: Available, Secondary: Available
Tags (edit) : Click here to add tags

Properties Monitoring Capabilities (7) Recommendations (0) Tutorials Tools + SDKs

Blob service

Hierarchical namespace	Disabled	Security	Require secure transfer for REST API operations	Enabled
Default access tier	Hot	Storage account key access	Enabled	
Blob public access	Enabled	Minimum TLS version	Version 1.2	
Blob soft delete	Enabled (7 days)	Infrastructure encryption	Disabled	
Container soft delete	Enabled (7 days)			
Versionsing	Disabled			
Change feed	Disabled			
NFS v3	Disabled			
Allow cross-tenant replication	Enabled			

File service

Allow access from	All networks
Number of private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes
Endpoint type	Standard

25°C Mostly cloudy 7:02 AM 6/26/2023

Open new tab

4.Create ADF session:goto portal.azure.com----->click on data factories ----->click on create---->give the instance name---->review&create----->create

Azure Data Factory

Click on data factories

Click on create

Microsoft Azure | portal.azure.com/#create/Microsoft.DataFactory | Create Data Factory - Microsoft | Paused | ...

Search resources, services, and docs (G+)

Home > Data factories > Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

One-click to create data factory with sample pipeline and datasets. Try it

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ○ CloudsEngineerSameer

Resource group * ○ **Select the resource**

Create new

Instance details

Name * ○ **Enter instance name**

Region * ○ East US

Version * ○ V2

Click on review&create

Review + create < Previous Next : Git configuration >

Give feedback

Microsoft Azure | portal.azure.com/#create/Microsoft.DataFactory | Create Data Factory - Microsoft | Paused | ...

Search resources, services, and docs (G+)

Home > Data factories > Create Data Factory

Validation Passed

Basics Git configuration Networking Advanced Tags Review + create

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offerings for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the Azure Marketplace Terms for additional details.

Basics

Subscription	CloudsEngineerSameer
Resource group	Azure6_30AMGroup
Name	InstanceBlobToBlobCopy
Region	East US
Version	V2

Networking

Connect via Public endpoint

Click on create

Give feedback Download a template for automation

Deployment succeeded
Deployment 'Microsoft.Template-20230626070612' to resource group 'Azure6_30AMGroup' was successful.

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 and be your first line of support.
Find an Azure expert >

Essentials

- Resource group (move) : Azure6_30AMGroup
- Status : Succeeded
- Location : East US
- Subscription (move) : CloudEngineerSameer
- Subscription ID : 8ef4b258-ea7b-4877-a5e1-e49adfe4567

Azure Data Factory Studio

Launch studio

Monitoring

- PipelineRuns
- ActivityRuns
- TriggerRuns

5. Select mange tile: create link service

After creating source storage account click on goto source

Azure Data Factory

Click on goto source

Your deployment is complete

Deployment name: sourceblobfirstds_1687916207280
Subscription: CloudEngineerSameer
Resource group: SampleRGFirst

Start time: 6/28/2023, 7:06:58 AM
Correlation ID: 00f18fd-4c09-44bc-8ecf-879c12fa5058

Deployment details

Next steps

Go to resource (highlighted with a red arrow)

Give feedback

Tell us about your experience with deployment

Deployment succeeded
Deployment 'Sourceblobfirstds_1687916207280' to resource group 'SampleRGFirst' was successful.

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 and be your first line of support. Find an Azure expert >

In the left blade in data storages click on containers ---->click on container----> give the container name----->click on create----->click on container----->click on upload----->click on browser file----->select the file----->click on upload

Click on container

sourceblobfirstds | Containers

Storage account

Containers (highlighted with a red arrow)

File shares
Queues
Tables

Search containers by prefix:

Name	Last modified
\$logs	6/28/2023, 7:07:30 AM

30°C Mostly cloudy 7:07 AM 6/28/2023

Azure Data Factory

New container

Name * Show delete

Public access level

Enter container name

Click on create

Create

Name	Last modified	Public access level	Lease state
Slogs	6/28/2023, 7:07:30 AM	Private	Available

Successfully created storage container
Successfully created storage container 'input'.

Click on container

Name	Last modified	Public access level	Lease state
Slogs	6/28/2023, 7:07:30 AM	Private	Available
input	6/28/2023, 7:08:07 AM	Private	Available

Click on upload

Click on browser

Azure Data Factory

The screenshot shows the Azure Storage Container Overview page for the 'input' container. On the right, a 'Upload blob' dialog box is open, showing a file named 'airlines.csv' selected for upload. A red arrow points to the 'Upload' button, which is highlighted with a box and labeled 'Click on upload'.

Same as create sink data storage account and container

The screenshot shows the Azure Data Factory Studio interface. It displays the 'InstanceBlobToBlobCopyDSFirst' data factory details, including its resource group, status, location, and subscription information. Below this, there are sections for 'Azure Data Factory Studio' (with a 'Launch studio' button), 'Monitoring' (PipelineRuns, ActivityRuns, TriggerRuns), and 'Automation' (Tasks (preview)). The status bar at the bottom indicates it's 7:14 AM on 6/28/2023.

Azure Data Factory

After creating azure data factory click on launch studio---->select mange---->select link service---->click on new----> give link service name---->select storage account---->click on test connection ---->click create

The screenshot shows the Azure Data Factory interface for the 'InstanceBlobToBlobCopyDSFirst' factory. It displays four main features:

- Ingest:** Copy data at scale once or on a schedule.
- Orchestrate:** Code-free data pipelines.
- Transform data:** Transform your data using data flows.
- Configure SSIS:** Manage & run your SSIS packages in the cloud.

Below these features, there's a section for 'Recent resources' which currently shows 'No items to show'.

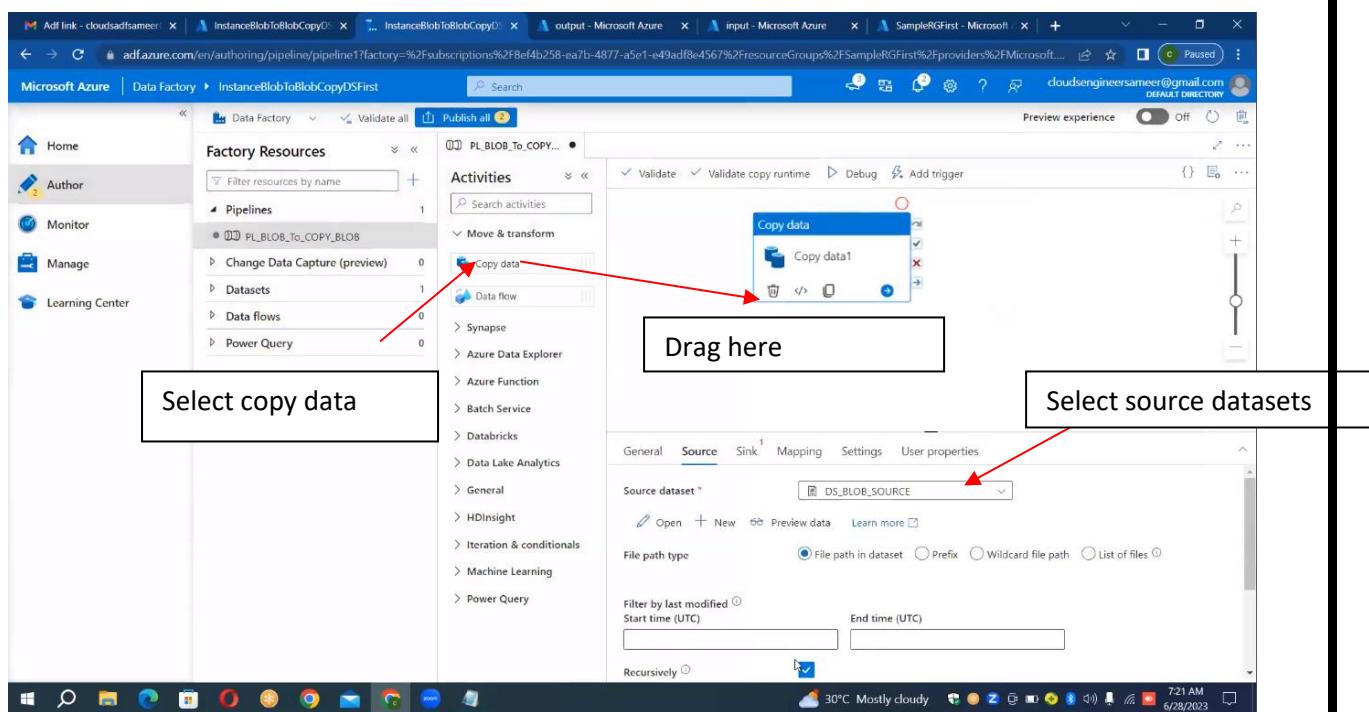
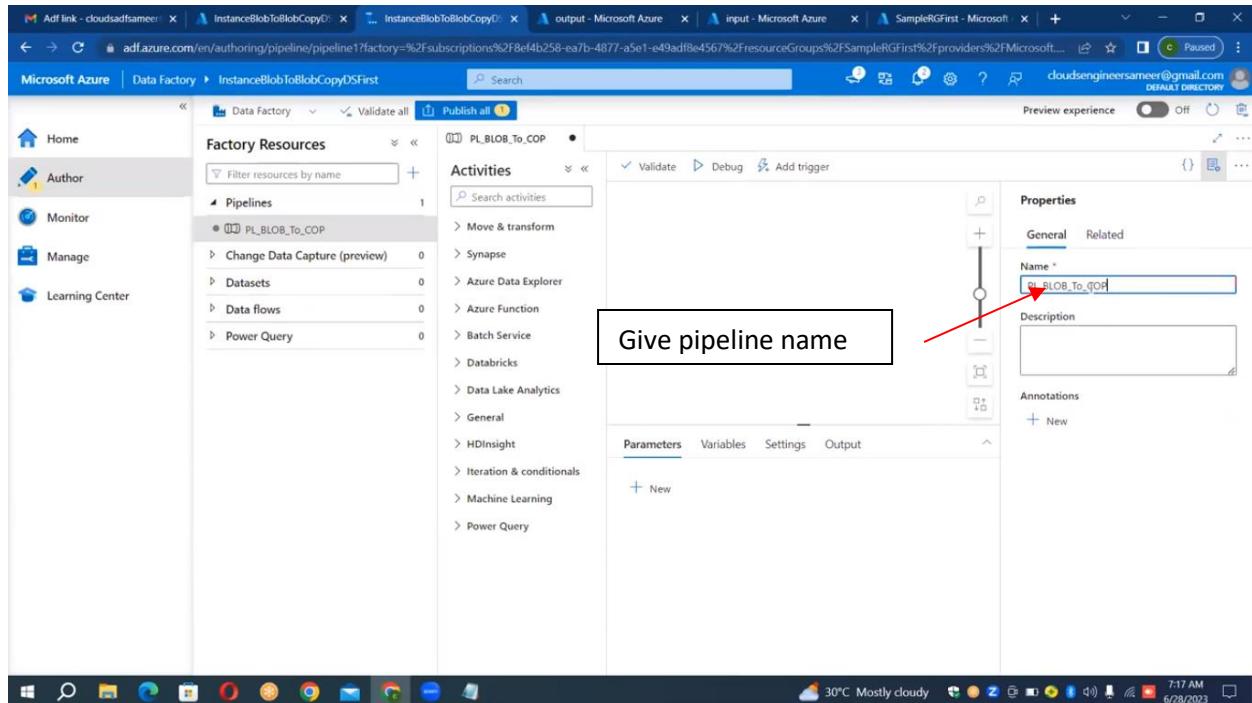
The screenshot shows the 'New linked service' dialog in the 'Author' blade of the Azure Data Factory interface. The 'Source link service name' field is highlighted with a red box. The 'Select storage account' field is also highlighted with a red box. A large red arrow points from the 'Source link service name' field towards the 'Create' button at the bottom right of the dialog. The 'Click on create' text is enclosed in a red box at the bottom right of the dialog area.

Same as create linked service for sink data storage

After creating linked service click on author

Azure Data Factory

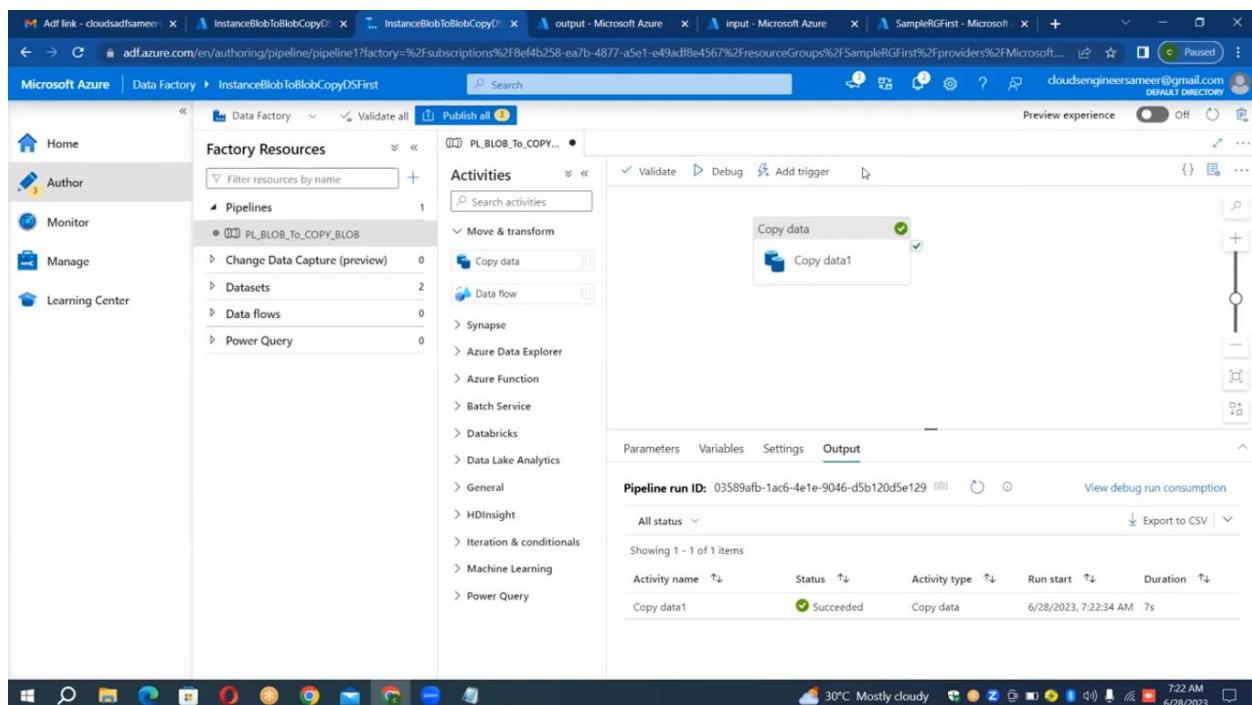
Author---->select pipeline----->new pipeline----->give name----->select move & transform from activities----->click on source ----->select source data sets----->click on sink----->select sink datasets ----->click on debug



Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Author' section is selected. In the center, a pipeline named 'PL_BLOB_To_COPY...' is displayed. Under the 'Activities' section, a 'Copy data' activity is selected. Below it, the 'Sink' tab is highlighted. A red box with the text 'Click on sink' points to the 'Sink dataset' dropdown, which contains the value 'DS_BLOB_SINK'. Another red box with the text 'Enter sink datasets' points to the same dropdown.

This screenshot is similar to the one above, showing the Microsoft Azure Data Factory pipeline configuration. The 'Sink' tab is still selected. A red box with the text 'Click on debug' points to the 'Debug' button located in the top right corner of the pipeline editor.



Case Study:

Title: Copy Data from Azure Blob Storage to Azure Data Lake Gen2

```

Case Study : 2
Title      : Copy Data From Azure Blob Storage to Azure Data Lake Gen2
-----
Specifications:
-----
Source Data Storage : Azure Blob Storage
File      : Airlines.csv
Sink Data Storage   : Azure Data Lake Gen2
-----
```

1. Create Resource Group

2. Create Storage Account (Source)

3. Create Storage Account (Sink)

4. Create Adf Session

5. Create Linked Services

6. Create Datasets

7. Create Pipeline**8. Select Activity And Assign The Data Sets To The Activity****9. Debug/Trigger****Components in ADF**

Pipeline: carries the data from source to the target

Linked services: these are used to connect the source to target with ADF

Datasets: these are the pointers to the physical location of the data

Activities: the action need to be performed on the data

Dataflows: these are the transformations in ADF

Triggers/debug: these are used to run a pipeline

Integration run time(IR):

```
(Amazon)Cloud-----AzureIR/AutoresolveIR----->Clouds(Azure)
Integration Run Time(IR): (OracleDB)On-Premises---SelfHostedIR----->Clouds(Azure)

*Untitled - Notepad
File Edit Format View Help
-----
1.Create resource group
2.Create Storage Account--(Source)
3.Create Storage Account--(Sink)
4.Create ADF Session
5.Create/Define Linked Services
6.Create Datasets--(Source & Target)
7.Create Pipeline
8.Select Activity and Assign the datasets to the Activity
9.Debug/Trigger
-----
Components in ADF
-----
Pipeline:Carries the data from to the target
Linked Services:These are used to connect the source & Target with ADF
Datasets:These are the pointers to the physical location of the data .
Activities: The Action need to be performed on the data.
Dataflows : These are the transformations in ADF
Triggers/Debug: Are used to run a Pipeline.
Integration Run Time(IR):
-----
```

Open the browser portal.azure.com----->click on resource group----->click on create----->enter the resource group name----->click on review create----->click on create----->click on resource

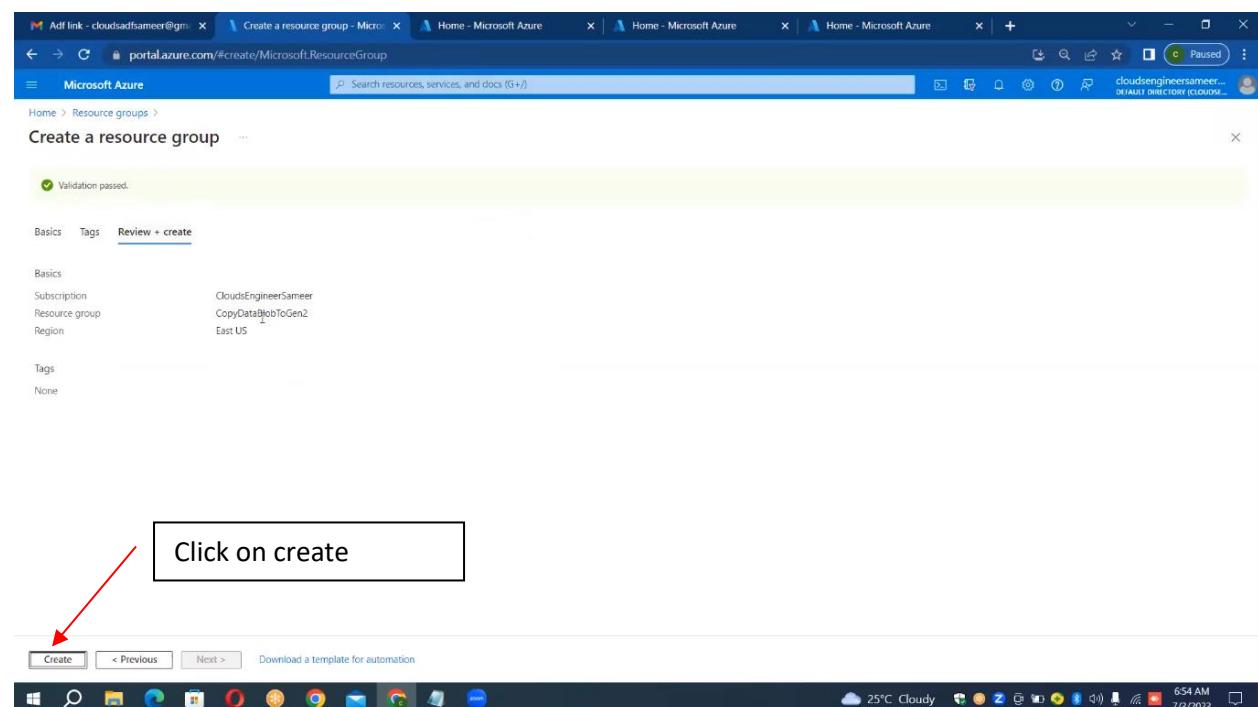
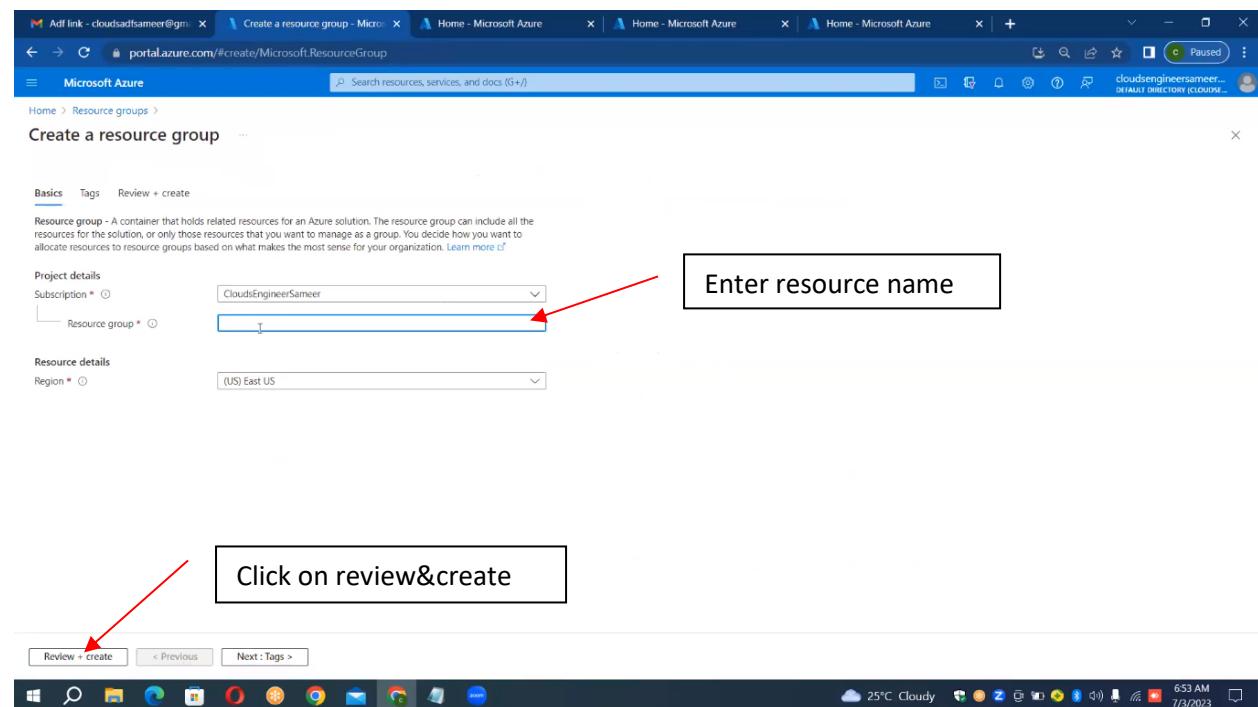
Azure Data Factory

click on resource group

The screenshot shows the Microsoft Azure portal homepage. At the top, there are several tabs for different services like Home - Microsoft Azure. Below the tabs, there's a search bar and a navigation bar with icons for Azure services such as Resource groups, Subscriptions, Storage accounts, Data factories, Azure Synapse Analytics, Azure Arc, SQL databases, and More services. A red arrow points to the 'Create a resource' button, which is located next to the Resource groups icon. To the left of the main content area, there's a box with the text 'click on resource group'. Below the navigation bar, there's a section for 'Resources' with a table showing one item: 'CloudsEngineerSameer' (Subscription type) last viewed 6 hours ago. Further down are sections for 'Navigate' (Subscriptions, Resource groups, All resources, Dashboard), 'Tools' (Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, Cost Management), and 'Useful links'.

Click on create

The screenshot shows the 'Resource groups' page in the Microsoft Azure portal. At the top, there are tabs for Home, Resource groups, and more. The main content area shows a table with no records found. The 'Create' button is highlighted with a red arrow. The table has columns for Name, Subscription, and Location. There are also buttons for Refresh, Export to CSV, Open query, and Assign tags. The status bar at the bottom shows it's 6:53 AM on 7/3/2023.



Azure Data Factory

Resource groups

Subscription equals CloudsEngineerSameer

Location equals all

Subscription: CloudsEngineerSameer

Location: East US

Click on resource group

Overview

Subscription (move) : CloudsEngineerSameer

Subscription ID : 8ef4b258-ea7b-4877-a5e1-e49ad8e4567

Tags (edit) : Click here to add tags

Resources Recommendations

No resources match your filters

Create resources Clear filters

Open new tab portal.azure.com----->click on storage account----->click on create----->enter storage account name----->click on review----->click on create----->click on goto resource

Azure Data Factory

Click on storage account

Create storage account

Azure Data Factory

Storage account name:

Region: (US) East US

Basics Advanced Networking Data protection Encryption Tags Review

Review < Previous Next : Advanced >

Storage account name: sourceblobs

Region: (US) East US

Basics Advanced Networking Data protection Encryption Tags Review

Review < Previous Next : Advanced >

Azure Data Factory

The screenshot shows the 'Create a storage account' wizard in Microsoft Azure. The 'Review' tab is selected. A red arrow points to the 'Create' button at the bottom left of the page.

Click on create

Create

The screenshot shows the 'sourceblobs_1688347688878 | Overview' page in Microsoft Azure. The 'Deployment' section indicates 'Your deployment is complete'. A red arrow points to the 'Go to resource' button under the 'Next steps' section.

Click on resource

Go to resource

Azure Data Factory

Open new tab portal.azure.com---->click on storage account---->enter sink storage account name---->---->select the advance tab---->click on enable hierarchical namespace---->click on review---->click on create---->click on goto resource

The screenshot shows the Microsoft Azure portal homepage. In the top navigation bar, there are several tabs: 'Adf link - cloudsadfsameer@gmail.com', 'CopyDataBlobToGen2 - Microsoft', 'sourceblobs - Microsoft Azure', 'Home - Microsoft Azure', 'Home - Microsoft Azure', and a paused status indicator. Below the navigation bar, the 'Microsoft Azure' logo and a search bar are visible.

In the main content area, there is a section titled 'Azure services' with various icons: 'Create a resource', 'Resource groups', 'Subscriptions', 'Storage accounts' (which is highlighted with a red arrow and a callout box), 'Azure DataFactory', 'Data factories', 'Azure Synapse Analytics', 'Azure Arc', 'SQL databases', and 'More services'. Below this, there is a 'Resources' section with a 'Recent' tab, a list of resources including 'CloudsEngineerSameer', and a 'Last Viewed' section showing '6 hours ago'. There is also a 'Navigate' section with links to 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. The bottom of the screen shows the browser's address bar with the URL 'https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.Storage/storageAccounts/sourceblobs', the taskbar with various pinned icons, and the system tray showing the date and time as '7/3/2023 7:59 AM'.

Azure Data Factory

Click on create

Enter sink storage account name

The screenshot shows the 'Create a storage account' wizard in the Azure portal. The 'Advanced' tab is highlighted with a red arrow. A callout box labeled 'Select the advance tab' points to the 'Advanced' tab. The 'Subscription' dropdown is set to 'CloudsEngineerSameer' and the 'Resource group' dropdown is set to 'CopyDataBlobToGen2'. The 'Storage account name' is 'sinkgen2ds' and the 'Region' is '(US) East US'. The 'Review' button is visible at the bottom.

The screenshot shows the 'Create a storage account' wizard in the Azure portal. The 'Advanced' tab is selected. In the 'Hierarchical Namespace' section, the 'Enable hierarchical namespace' checkbox is checked, indicated by a red arrow and a callout box labeled 'Select enable name space'.

Azure Data Factory

Access protocols

Click on review

Review

< Previous | Next : Networking >

Networking

Click on create

Create

< Previous | Next > | Download a template for automation

Deployment succeeded
Deployment 'sinkgen2ds_1688348129000' to resource group 'CopyDataBlobToGen2' was successful.

Deployment details

- Deployment name: sinkgen2ds_1688348129000
- Subscription: CloudsEngineerSameer
- Resource group: CopyDataBlobToGen2
- Start time: 7/3/2023, 7:05:39 AM
- Correlation ID: 0ba0d349-41b2-4daf-8d2d-a2b429dcfb21

Next steps

Click on goto resource

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 and be your first line of support. Find an Azure expert >

Overview

Essentials

Resource group (move)	: CopyDataBlobToGen2	Performance	: Standard
Location	: East US	Replication	: Read-access geo-redundant storage (RA-GRS)
Primary/Secondary Location	: Primary: East US, Secondary: West US	Account kind	: StorageV2 (general purpose v2)
Subscription (move)	: CloudsEngineerSameer	Provisioning state	: Succeeded
Subscription ID	: 8ef4b258-ea7b-4877-a5e1-e49adfe4567	Created	: 7/3/2023, 7:05:47 AM
Disk state	: Primary: Available, Secondary: Available		
Tags (edit)	: Click here to add tags		

Properties

Data Lake Storage

Hierarchical namespace	Enabled	Require secure transfer for REST API operations	Enabled
Default access tier	Hot	Storage account key access	Enabled
Blob public access	Enabled	Minimum TLS version	Version 1.2
Blob soft delete	Enabled (7 days)	Infrastructure encryption	Disabled
Container soft delete	Enabled (7 days)		
Versioning	Disabled		
Change feed	Disabled		
NFS v3	Disabled		
SFTP	Disabled		

Security

Allow access from	All networks
Number of private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes
Endpoint type	Standard

Networking

Allow access from	All networks
Number of private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes
Endpoint type	Standard

File service

Data can be stored in containers, fileshares, queues, tables

Container is nothing but folder

Open source storage account tab----->select container----->click on create container----->give container name----->click on create----->double click on container----->click on upload----->click on browser file----->select the file---->click on upload

Azure Data Factory

sourceblobs | Containers

Click on container

+ Container Change access level Restore containers Refresh Delete Give feedback

Name	Last modified	Public access level
Slogs	7/3/2023, 6:58:53 AM	Private

New container

Name * Show delete

Public access level

Create Give feedback

Azure Data Factory

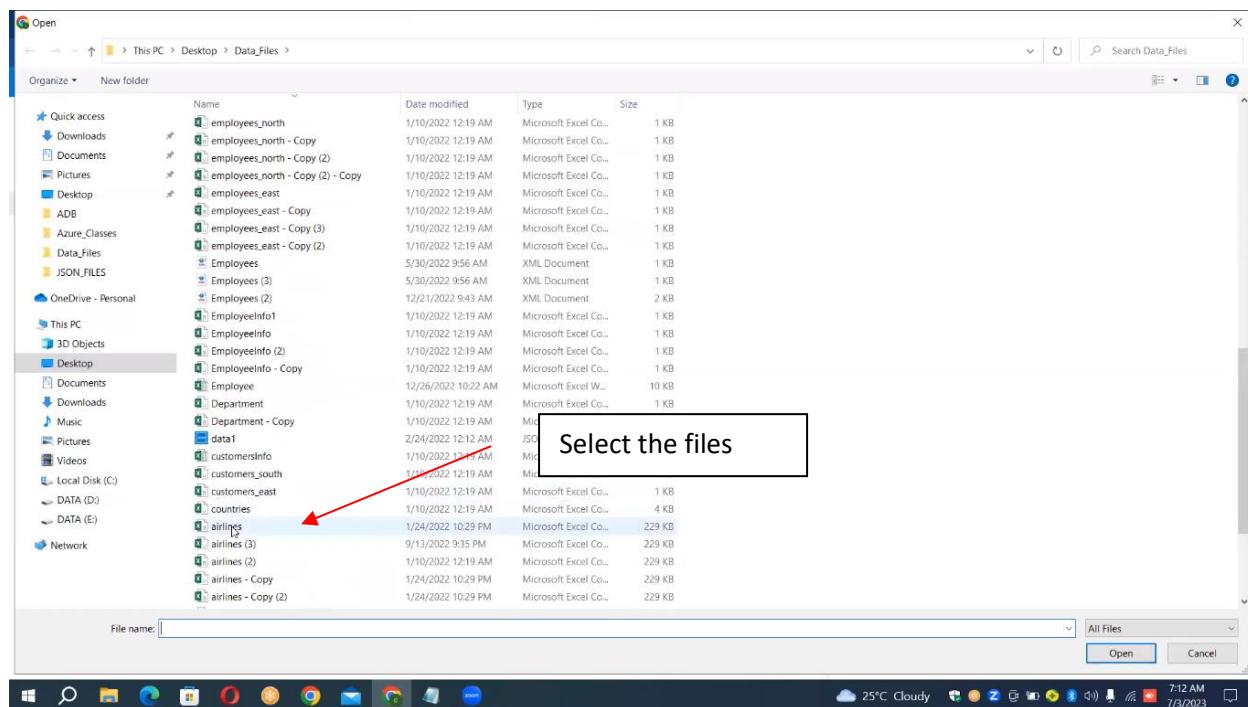
The screenshot shows the Microsoft Azure Storage account interface. On the left, a sidebar lists various storage services: Access Control (IAM), Data migration, Events, Storage browser, Storage Mover, Data storage (Containers, File shares, Queues, Tables), Security + networking (Networking, Azure CDN, Access keys, Shared access signature, Encryption, Microsoft Defender for Cloud), and Data management (Redundancy, Data protection). The main area displays a table of containers. The table has columns for Name, Last modified, Public access level, and Lease state. It contains two entries: 'Logs' (Last modified 7/3/2023, 6:58:53 AM, Private, Available) and 'input' (Last modified 7/3/2023, 7:11:19 AM, Private, Available). A red arrow points from a box labeled 'Click on create' to the 'Create' button in the top right corner of the container list. The status bar at the bottom shows '25°C Cloudy' and the date '7/3/2023'.

This screenshot shows the same Microsoft Azure Storage account interface after a new container has been created. The 'input' container is now listed in the table. A red arrow points from a box labeled 'Double click on folder' to the 'input' container entry. A success message in the top right corner states 'Successfully created storage container' and 'Successfully created storage container 'input''. The status bar at the bottom shows '25°C Cloudy' and the date '7/3/2023'.

Azure Data Factory

Screenshot of the Microsoft Azure Storage Container Overview page for the 'input' container. A red arrow points to the 'Upload' button in the top navigation bar. A callout box with the text 'Click on upload' is positioned below the 'Upload' button.

Screenshot of the 'Upload blob' dialog box. A red arrow points to the 'Browse for files' button. A callout box with the text 'Click on browse file' is positioned below the 'Browse for files' button.



Upload blob

1 file(s) selected: airlines.csv
Drag and drop files here or [Browse for files](#)

Overwrite if files already exist

[Upload](#) [Give feedback](#)

Click on upload

Azure Data Factory

If you want see the data double clik on file

If we want to see the data double click on file

Open sink storage account tab ---->click on container---->click on create container---->give container name---->click on create---->double click on container

Click on create container

Azure Data Factory

New container

Name *

Public access level Private (no anonymous access) Private (no anonymous)

Create

New container

Name *

Public access level Private (no anonymous access) Private (no anonymous access)

Create

Azure Data Factory

Home > sinkgen2ds_1688348129000 | Overview > sinkgen2ds

sinkgen2ds | Containers

Storage account

Search Container Change access level Restore containers Refresh Delete Give feedback

Name Last modified Public access level Lease state

Slogs	7/3/2023, 7:06:11 AM	Private	Available
output	7/3/2023, 7:14:12 AM	Private	Available

Show deleted containers

Container Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Data storage Security + networking Networking Access keys Shared access signature Encryption Microsoft Defender for Cloud Data management

Windows taskbar: 25°C Cloudy 7:14 AM 7/3/2023

Open new browser portal.azure.com---->click on data factory ---->click on create---->select the resource group--->enter instance name--->click on review create---->click on create---->click on goto resource----->click on launch studio----->click on manage----->click on linked services--->click on new---->enter linked services name---->select storage account---->click on create

Ad link - cloudsadfsameer@gmail.com | CopyDataBlobToGen2 - Microsoft | input - Microsoft Azure | output - Microsoft Azure | Home - Microsoft Azure

Microsoft Azure Search resources, services, and docs (G+)

Azure services

- Create a resource
- Storage accounts
- Resource groups
- Subscriptions
- Azure Databricks
- Data factories

Data factories

+ Create View More services

Click on data factory

Resources

Recent Favorite

Name	Type	Last Viewed
sinkgen2ds	Storage account	9 minutes ago
CopyDataBlobToGen2	Resource group	9 minutes ago
CloudsEngineerSameer	Subscription	7 hours ago

See all

Navigate

- Subscriptions
- Resource groups
- All resources
- Dashboard

Tools

- Microsoft Learn
- Azure Monitor
- Microsoft Defender for Cloud
- Cost Management

https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft...

Windows taskbar: 25°C Cloudy 7:15 AM 7/3/2023

Azure Data Factory

Click on create

No data factories to display

Create data factory



Select resource

Enter instance name



Azure Data Factory

Click on review create

The screenshot shows the 'Create Data Factory' wizard in the Azure portal. The 'Basics' tab is selected. The 'Subscription' dropdown is set to 'CloudsEngineerSameer'. The 'Resource group' dropdown is set to 'CopyDataBlobToGen2'. The 'Name' field contains 'InstanceCopy'. The 'Region' dropdown is set to 'East US'. The 'Version' dropdown is set to 'V2'. At the bottom, there is a 'Review + create' button highlighted by a red arrow.

Click on create

The screenshot shows the 'Create Data Factory' wizard in the Azure portal. A green checkmark indicates 'Validation Passed'. The 'Review + create' tab is selected. The 'TERMS' section contains a detailed legal agreement. Below it, the 'Basics' section shows the same configuration as the previous step. In the 'Networking' section, the 'Connect via' dropdown is set to 'Public endpoint'. At the bottom, there is a 'Create' button highlighted by a red arrow.

Azure Data Factory

[ADF Link - cloudsadfsameer@gmail.com](#) | [CopyDataBlobToGen2 - Microsoft](#) | [input - Microsoft Azure](#) | [output - Microsoft Azure](#) | [Microsoft.DataFactory-20230703071542](#) | Paused

Microsoft Azure | Search resources, services, and docs (G+)

Microsoft.DataFactory-20230703071542 | Overview

Your deployment is complete

Deployment name : Microsoft.DataFactory-20230703071542
Subscription : CloudsEngineerSameer
Resource group : CopyDataBlobToGen2

Start time : 7/3/2023, 7:16:24 AM
Correlation ID : ffd0243b-f055-4514-b4c3-edab5005cd70

Deployment details
Next steps
[Go to resource](#)

Give feedback
Tell us about your experience with deployment

Deployment succeeded
Deployment 'Microsoft.DataFactory-20230703071542' to resource group 'CopyDataBlobToGen2' was successful.

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 and be your first line of support.
Find an Azure expert >

[ADF Link - cloudsadfsameer@gmail.com](#) | [CopyDataBlobToGen2 - Microsoft](#) | [input - Microsoft Azure](#) | [output - Microsoft Azure](#) | [InstanceCopyDataFromBlobToGen2](#) | Paused

Microsoft Azure | Search resources, services, and docs (G+)

Microsoft.DataFactory-20230703071542 | Overview

InstanceCopyDataFromBlobToGen2

Essentials

Resource group (move) : CopyDataBlobToGen2
Status : Succeeded
Location : East US
Subscription (move) : CloudsEngineerSameer
Subscription ID : 8ef4b258-ea7b-4877-a5e1-e49adf8e4567

Type : Data factory (V2)
Getting started : [Quick start](#)

Azure Data Factory Studio

Launch studio

Quick Starts
Tutorials
Template Gallery
Training Modules

Monitoring

PipelineRuns
ActivityRuns
TriggerRuns

Azure Data Factory

The screenshot shows the Azure Data Factory Studio interface. On the left, there's a navigation sidebar with sections like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Getting started, Monitoring, Automation, and Tasks (preview). The main area features a large blue factory icon and the text "Azure Data Factory Studio". Below it is a "Launch studio" button, which is highlighted with a red arrow. To the right of the studio are four cards: Quick Starts, Tutorials, Template Gallery, and Training Modules.

The screenshot shows the Azure Data Factory Manage interface. On the left, there's a navigation sidebar with Home, Author, Monitor, Manage (highlighted with a red arrow), and Learning Center. The main area displays the title "InstanceCopyDataFromBlobToGen2" and four cards: Ingest, Orchestrate, Transform data, and Configure SSIS. Below these cards is a section titled "Recent resources" with a message "No items to show". At the bottom, there's a Windows taskbar showing the date and time as 7/3/2023 7:17 AM.

Microsoft Azure | Data Factory > InstanceCopyDataFromBlobToGen2

General

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Security

Credentials

Customer managed key

Outbound rules

Managed private endpoints

Workflow orchestration manager

No linked service to show

Create linked service

7:18 AM 7/3/2023

Microsoft Azure | Data Factory > InstanceCopyDataFromBlobToGen2

General

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Security

Credentials

Customer managed key

Outbound rules

Managed private endpoints

Workflow orchestration manager

No linked service to show

Create linked service

7:18 AM 7/3/2023

Azure Data Factory

Microsoft Azure | Data Factory > InstanceCopyDataFromBlobToGen2

New linked service

Azure Blob Storage Learn more

Name *

Description

Connect via integration runtime *

Authentication type

Connection string Azure Key Vault

Account selection method From Azure subscription Enter manually

Azure subscription

Storage account name *

Additional connection properties

Create Back Test connection Cancel

Microsoft Azure | Data Factory > InstanceCopyDataFromBlobToGen2

New linked service

Azure Blob Storage Learn more

Name *

Description

Connect via integration runtime *

Authentication type

Connection string Azure Key Vault

Account selection method From Azure subscription Enter manually

Azure subscription

Storage account name *

Additional connection properties

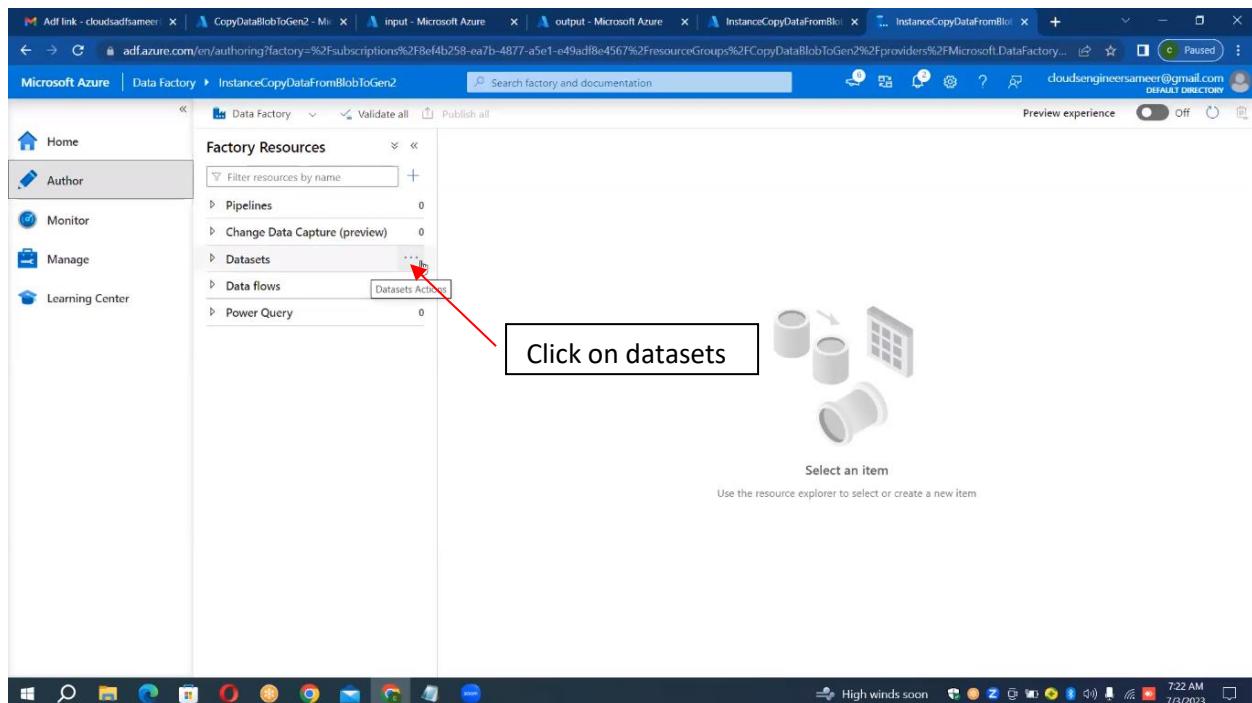
Create Back Test connection Cancel

Same as create azure data lake gen2

Here we have to create two data sets

After that goto author ----->select datasets---->newdatasets---->enter dataset name----->select linked services---->click on folder browser----click on folder---->select file----->click on ok

Azure Data Factory



Microsoft Azure | Data Factory > InstanceCopyDataFromBlobToGen2

Preview experience Off

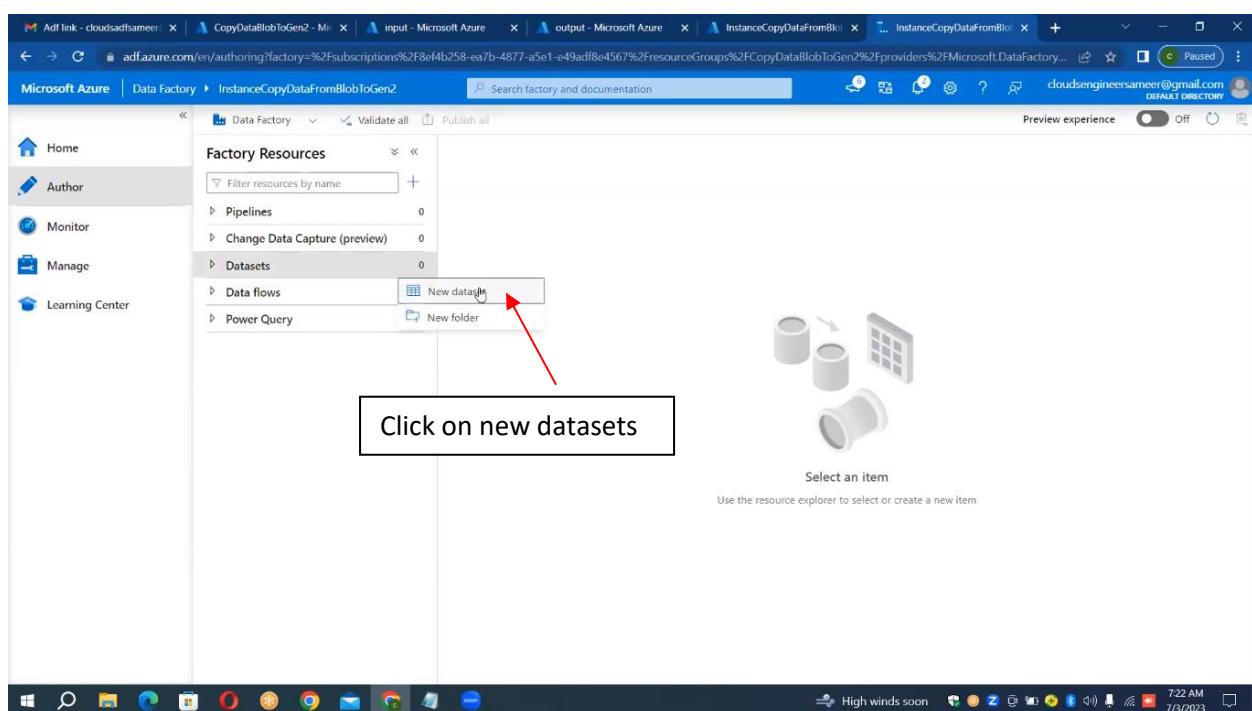
Data Factory <> Validate all Publish all

Factory Resources <>

Filter resources by name +

- > Pipelines 0
- > Change Data Capture (preview) 0
- > Datasets 0 **Click on datasets**
- > Data flows 0 Datasets Actions
- > Power Query 0

Select an item
Use the resource explorer to select or create a new item



Microsoft Azure | Data Factory > InstanceCopyDataFromBlobToGen2

Preview experience Off

Data Factory <> Validate all Publish all

Factory Resources <>

Filter resources by name +

- > Pipelines 0
- > Change Data Capture (preview) 0
- > Datasets 0 **Click on new datasets**
- > Data flows **New dataset**
- > Power Query **New folder**

Select an item
Use the resource explorer to select or create a new item

Azure Data Factory

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All Azure Database File Generic protocol NoSQL Services and apps

Amazon Marketplace Web Service	Amazon RDS for Oracle	Amazon RDS for SQL Server
Amazon Redshift	Amazon S3	Amazon S3 Compatible
HDFS	Parquet	ORC

Continue Cancel

Select format

Choose the format type of your data

Avro	Binary	CSV
Excel	JSON	ORC
Parquet	XML	

Continue Back Cancel

Click on csv type

Azure Data Factory

The screenshot shows the 'Set properties' dialog for a dataset named 'DelimitedText1'. The 'Name' field contains 'DelimitedText1'. The 'Linked service' dropdown is set to 'Select...'. A red arrow points from the text 'Enter the name' to the 'Name' input field. Another red arrow points from the text 'Select source link' to the 'Linked service' dropdown.

The screenshot shows the 'Set properties' dialog for a dataset named 'DS_BLOB_SOURCE'. The 'Name' field contains 'DS_BLOB_SOURCE'. The 'Linked service' dropdown is set to 'LS_BLOB_SOURCE'. The 'File path' section shows 'Container' / 'Directory' / 'File name'. A red arrow points from the text 'Click on browser' to the 'File name' input field.

Azure Data Factory

Select the file

Click on ok

DS_BLOB_SOURCE

Connection Schema Parameters

Linked service * LS_BLOB_SOURCE Test connection Edit + New Learn more

File path * input / Directory / airlines.csv

Compression type None

Column delimiter Comma (,)

Row delimiter Default (\r\n or \n)

Encoding Default(UTF-8)

Quote character Double quote (")

Escape character Backslash (\)

First row as header

Create another datasets

Click on new datasets

New dataset

DS_BLOB_SOURCE

Connection Schema Parameters

Linked service: LS_BLOB_SOURCE

File path: input / Directory / airlines.csv

Compression type: None

Column delimiter: Comma (,)

Row delimiter: Default (\r\n, or \n\r)

Encoding: Default(UTF-8)

Quote character: Double quote (")

Escape character: Backslash (\)

First row as header:

Dataset gen 2

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more ↗

Select a data store

Search

All	Azure	Database	File	Generic protocol	NoSQL	Services and apps
Amazon Marketplace Web Service	Amazon RDS for Oracle	Amazon RDS for SQL Server				
Amazon Redshift	Amazon S3	Amazon S3 Compatible				

Continue Cancel

Azure Data Factory

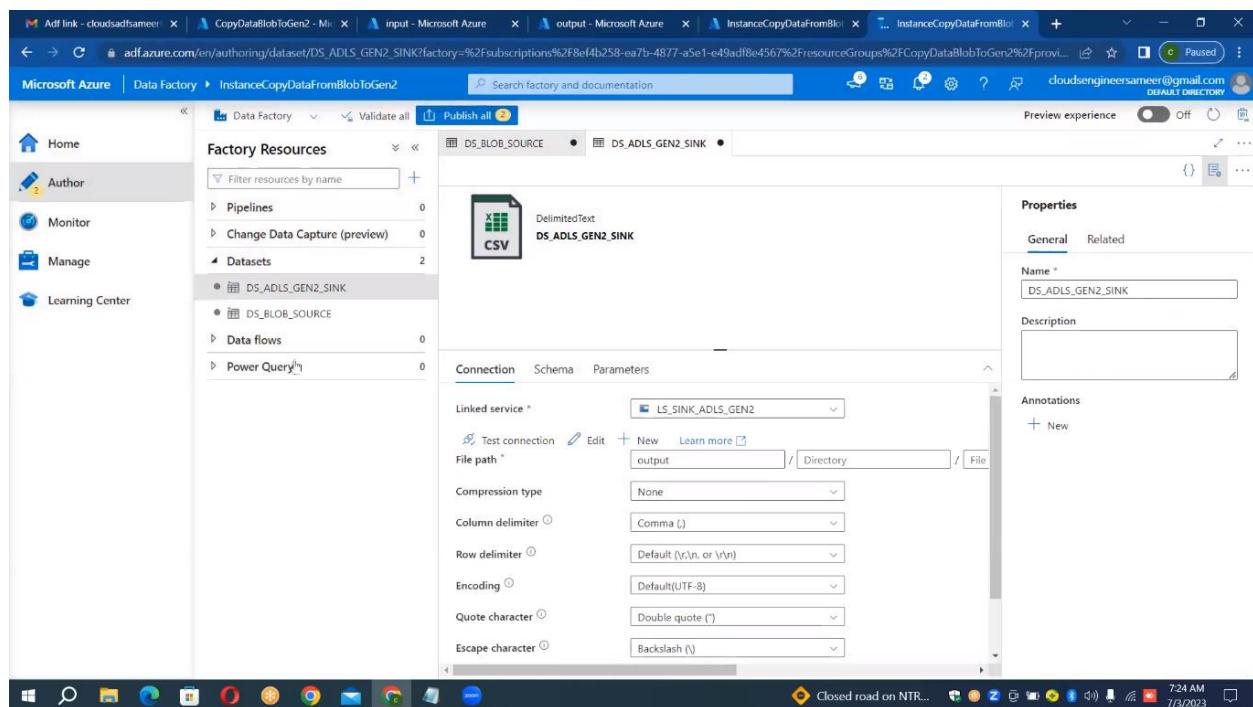
The screenshot shows the Azure Data Factory interface. On the left, the navigation bar includes Home, Author, Monitor, Manage, and Learning Center. The main area displays 'Factory Resources' with a list of Pipelines (0), Change Data Capture (preview) (0), Datasets (1), Data flows (0), and Power Query (0). A dataset named 'DS_BLOB_SOURCE' is selected. The 'Connection' tab is active, showing 'Linked service': LS_BLOB_SOURCE, 'File path': input, and 'Compression type': None. The 'Format' section shows a preview of a CSV file and a grid of options: Avro, Binary, DelimitedText (selected), Excel, JSON, ORC, Parquet, and XML. A red arrow points from a box labeled 'Select csv' to the 'DelimitedText' icon.

The screenshot shows the Azure Data Factory interface. The 'Dataset Properties' dialog is open, titled 'Set properties'. It contains fields for 'Name' (set to 'DelimitedText'), 'Linked service' (set to 'Select...'), and 'First row as header' (checkbox checked). A red arrow points from a box labeled 'Select link service' to the 'Linked service' dropdown menu. The background shows the same dataset configuration screen as the previous image.

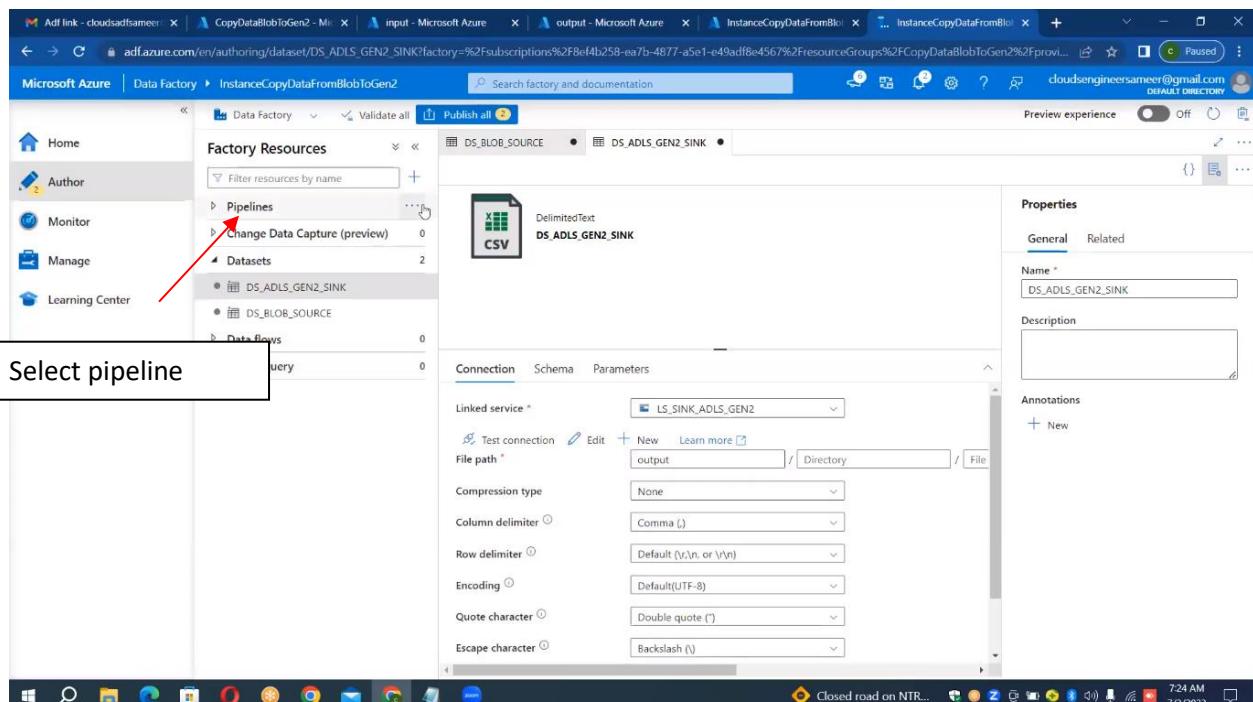
Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a navigation sidebar with options like Home, Author, Monitor, Manage, and Learning Center. The main area shows 'Factory Resources' with a list of Pipelines (0), Change Data Capture (preview) (0), Datasets (1), Data flows (0), and Power Query (0). A dataset named 'DS_BLOB_SOURCE' is selected. To its right is a detailed configuration pane titled 'Set properties'. In this pane, under the 'Connection' tab, the 'Linked service' is set to 'LS_BLOB_SOURCE' and the 'File path' is set to 'input'. A red arrow points from the text 'Click on browser' to the 'File path' input field. Other settings include 'Compression type' (None), 'Column delimiter' (Comma (,), checked), 'Row delimiter' (Default (\r\n, \n, or \r\n)), 'Encoding' (Default(UTF-8)), 'Quote character' (Double quote (')), 'Escape character' (Backslash (\)), and 'First row as header' (checked). At the bottom of the pane are 'OK', 'Back', and 'Cancel' buttons.

This screenshot continues from the previous one, showing the 'Browse' dialog that appeared after clicking the 'File path' input field. The dialog title is 'Browse' and it says 'Select a file or folder.' Under 'Root folder', the 'output' folder is selected, indicated by a red arrow pointing to a callout box labeled 'Select the output'. At the bottom of the dialog are 'OK' and 'Cancel' buttons. A red arrow also points from the text 'Click on ok' to the 'OK' button.



Select pipeline--->new pipeline--->give the pipeline--->select activity ---->select move and transform---->select copy data----->drag it to right side----->goto source----->select source dataset---->select the sink---->select sinkdataset---->change file extention----->click on debug



Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory interface. In the center, there is a 'New Pipeline' button highlighted with a red arrow and a callout box containing the text 'Click on new pipeline'. To the right, there is a 'Properties' panel with tabs for 'General' and 'Related'. Under the 'General' tab, the 'Name' field is populated with 'DS_ADLS_GEN2_SINK'. Below the name field, there is a 'Description' field and an 'Annotations' section.

The screenshot shows the Microsoft Azure Data Factory interface. In the center, there is a 'Properties' panel with tabs for 'General' and 'Related'. Under the 'General' tab, the 'Name' field is empty and has a red asterisk indicating it is required. A red arrow points to the 'Name' field with a callout box containing the text 'Give new pipeline name'. Below the name field, there is a 'Description' field and an 'Annotations' section.

Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (2), and 'Data flows' (0). The main area displays the pipeline structure: DS_BLOB_SOURCE → DS_ADLGEN2_SINK → PL_BLOB_TO_ADL... . The 'Activities' pane on the right shows a list of available activities, with 'Move & transform' highlighted by a red arrow. A callout box with the text 'Click on move transform' is positioned over the 'Move & transform' item.

This screenshot shows the same Azure Data Factory pipeline editor interface. The 'Move & transform' activity has been expanded, and the 'Copy data' option is highlighted with a red arrow. A callout box with the text 'Select copy data' is positioned over the 'Copy data' icon.

Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (2), and 'Data flows'. The main area displays a 'Copy data' activity named 'Copy data1'. The 'Source' tab is selected in the activity settings panel. A red arrow points to the 'Source' tab, and a callout box contains the text 'Click on source'.

The screenshot shows the Microsoft Azure Data Factory pipeline editor with the 'Source' tab selected in the 'Copy data' activity settings. A red arrow points to the 'Select...' button in the 'Source dataset' dropdown menu, which is open to show two options: 'DS_ADLS_GEN2_SINK' and 'DS_BLOB_SOURCE'. A callout box contains the text 'Select source datasets'.

Azure Data Factory

Click on sink

Select the sink data sets

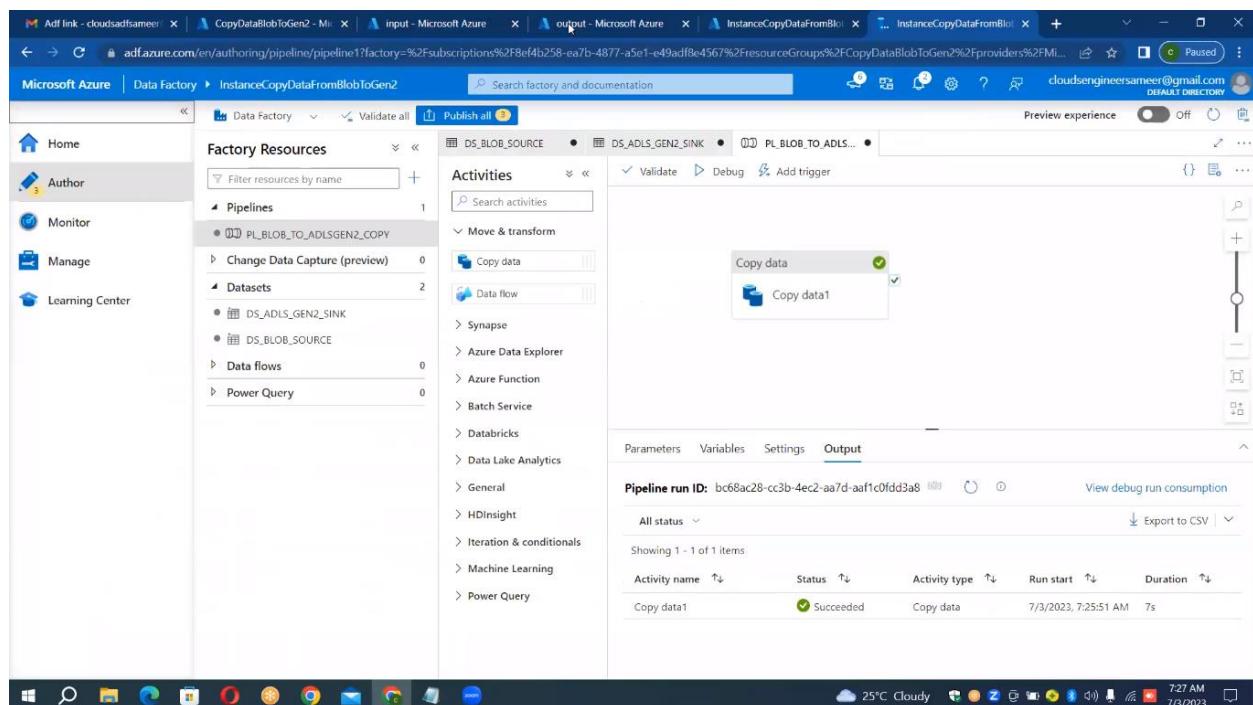
Change the file extension

Azure Data Factory

Click on debug

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the navigation pane includes Home, Author (selected), Monitor, Manage, and Learning Center. The main area displays Factory Resources, Pipelines (1), Datasets (2), and Data flows (0). A specific pipeline named 'PL_BLOB_TO_ADLSGEN2_COPY' is selected. The pipeline details pane shows a 'Copy data' activity named 'Copy data1'. The toolbar at the top has buttons for Publish all, Validate all, Debug (highlighted with a red arrow), and Add trigger. A callout box with the text 'Click on debug' is positioned over the Debug button.

The screenshot shows the Microsoft Azure Data Factory pipeline run details page. It displays a summary of the pipeline run, including the Pipeline run ID: bc68ac28-cc3b-4ec2-aa7d-aaf1c0fd3a8, status All status (Succeeded), and duration 7s. Below this, a table lists the activities: Activity name (Copy data1), Status (Succeeded), Activity type (Copy data), Run start (7/3/2023, 7:25:51 AM), Duration (7s), Log (AutoResolveIntegration), Integration runtime (f3f1b06d-e0e4-4f4b-bb31-0b96e), User properties (empty), and Activity run ID (empty). The table has columns for Activity name, Status, Activity type, Run start, Duration, Log, Integration runtime, User properties, and Activity run ID. A callout box with the text 'View debug run consumption' is located near the top right of the table.



Case study: 3

Title: copy data from azure blob storage to azure blob storage by using single account

Specification:

Source data storage: azure blob storage --->account1

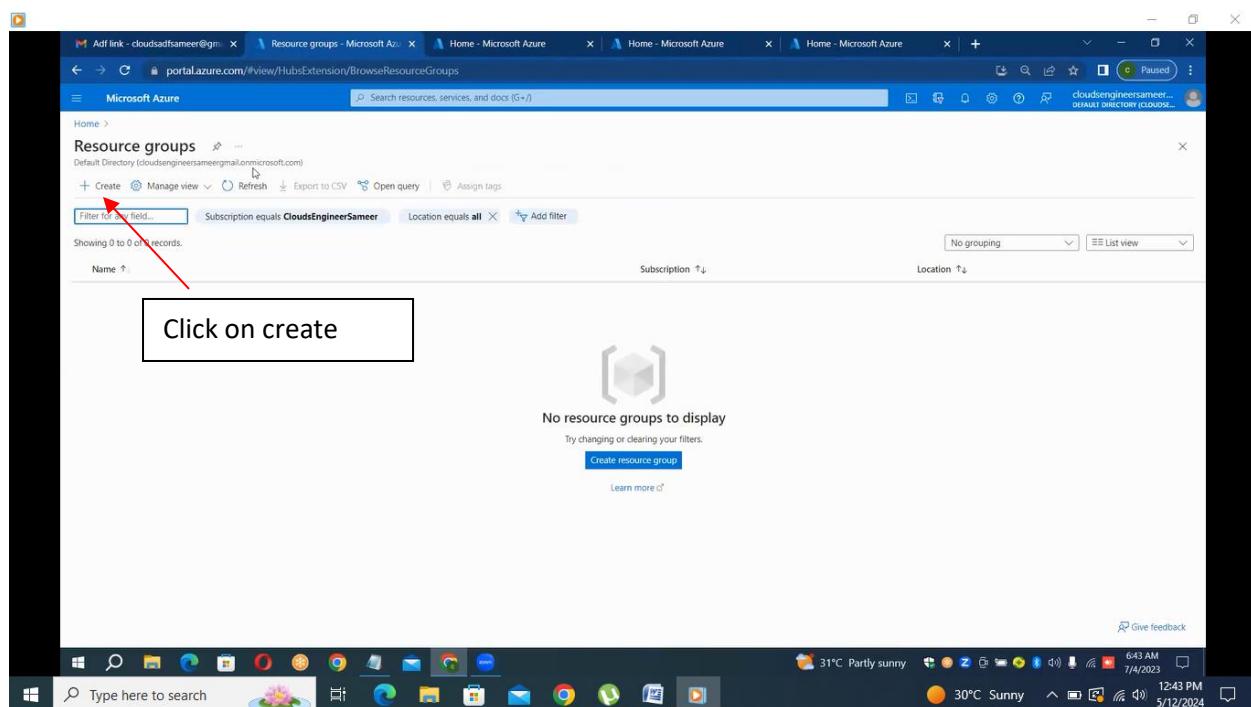
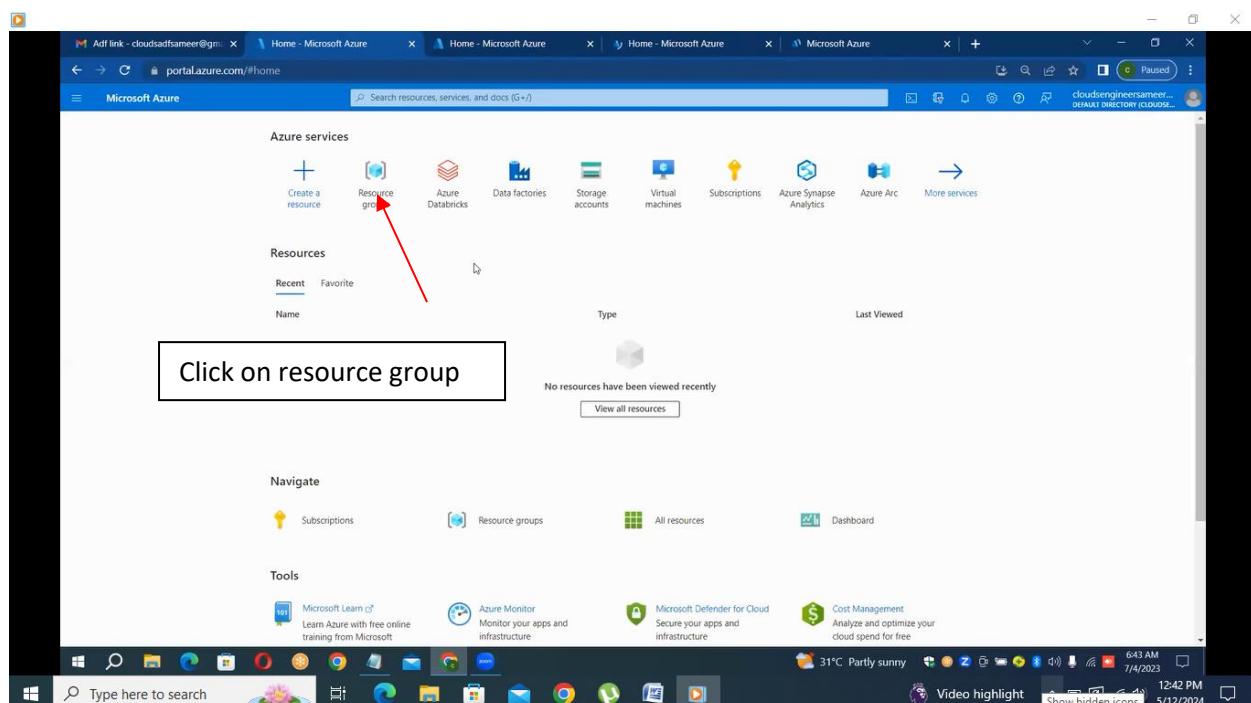
Containers : inputs

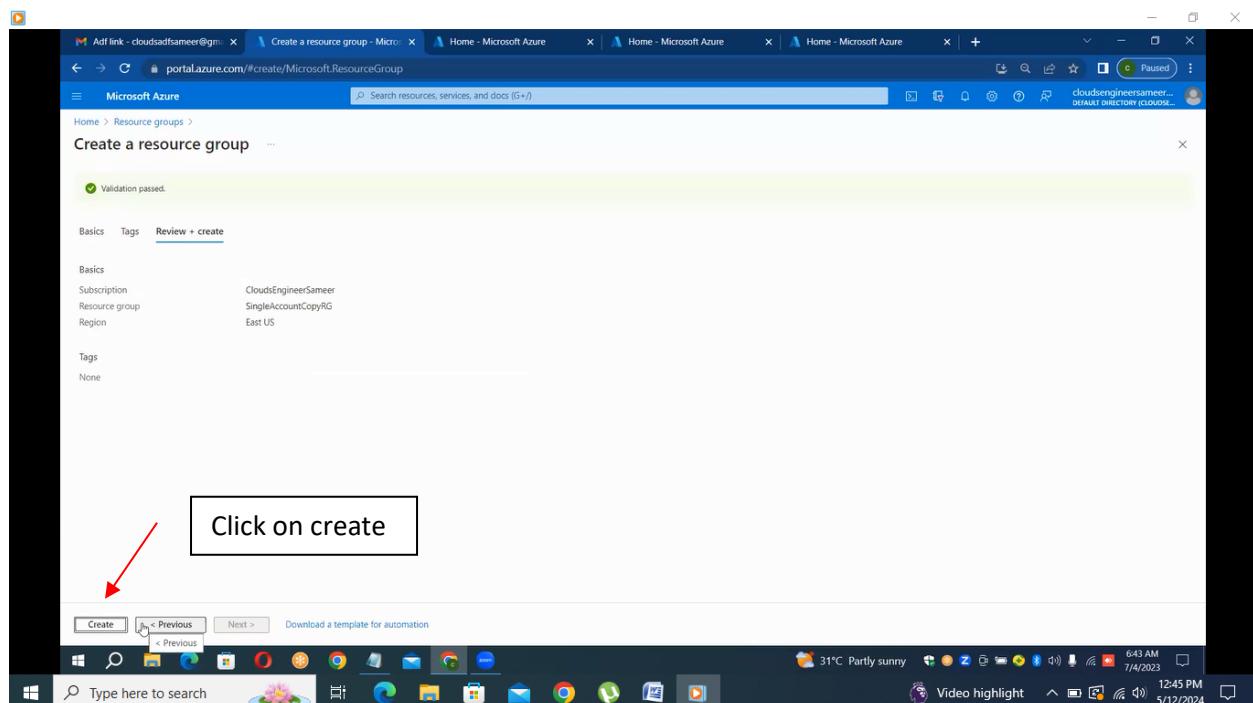
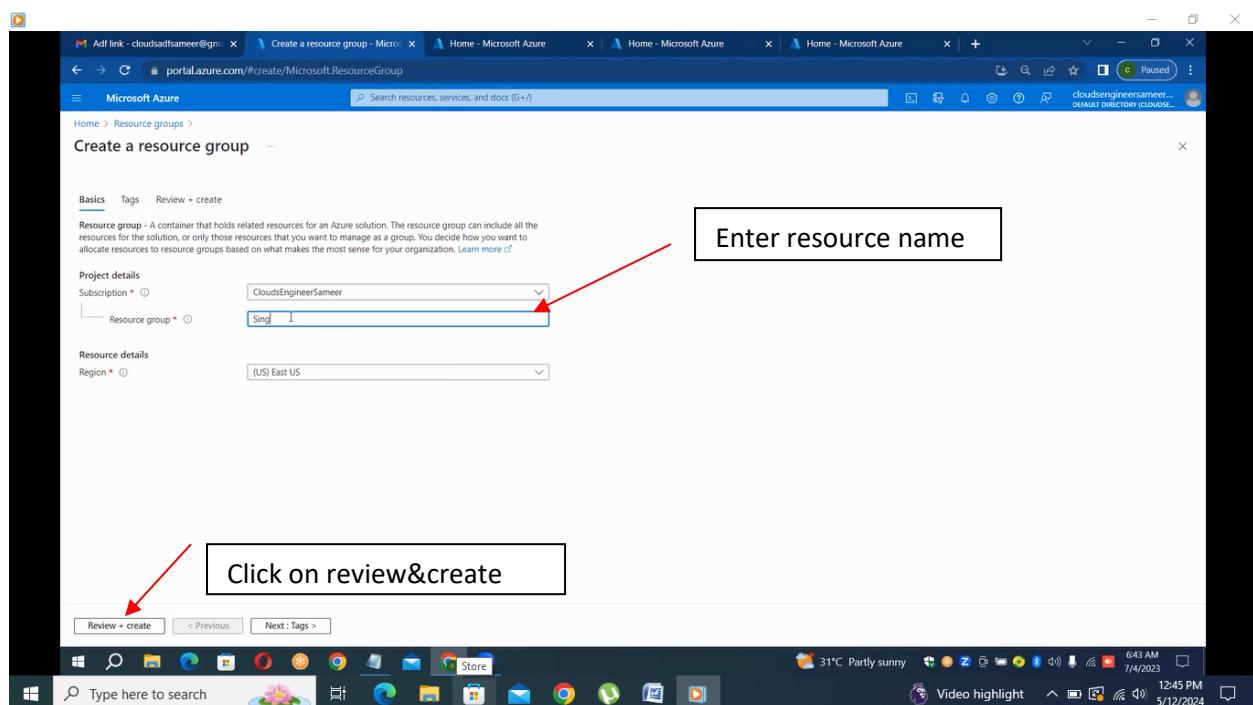
File : airlines.csv

Sink data storage : azure blob storage ---->account1

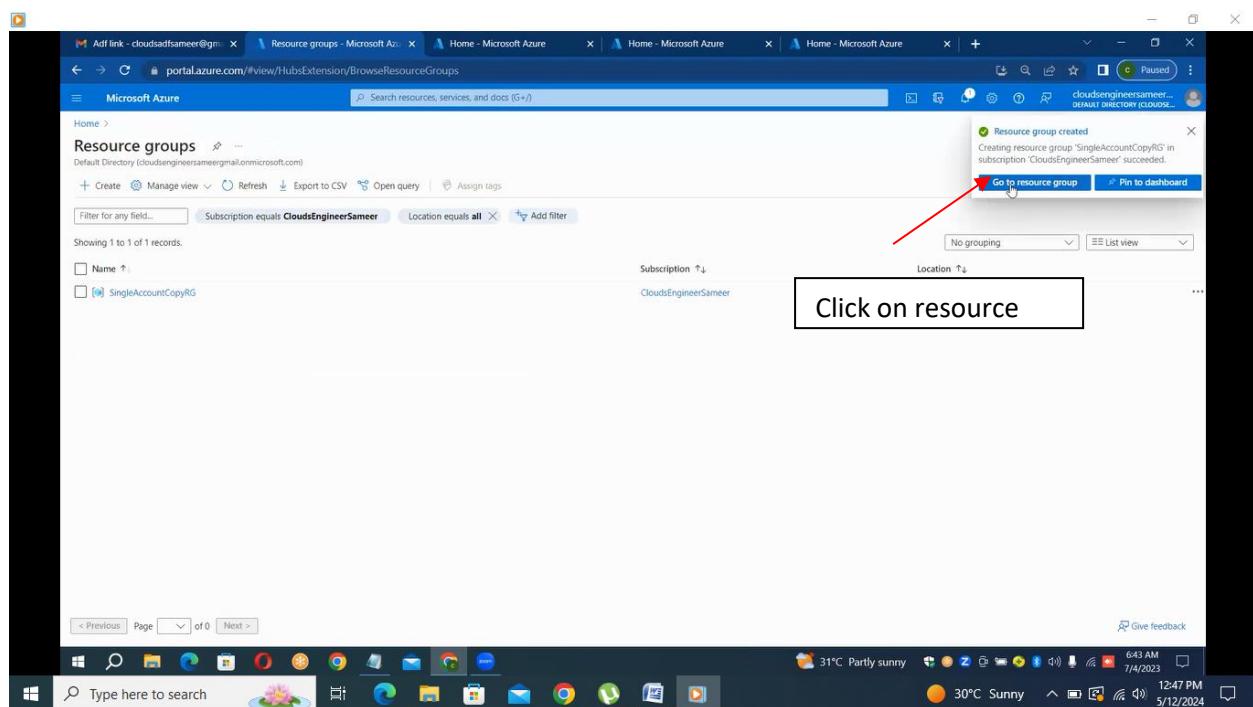
Containers : output

Open new tab portal.azure .com---->click on resource group---->click on create---->give the resource group---->click on review and create---->click on create----> goto resource

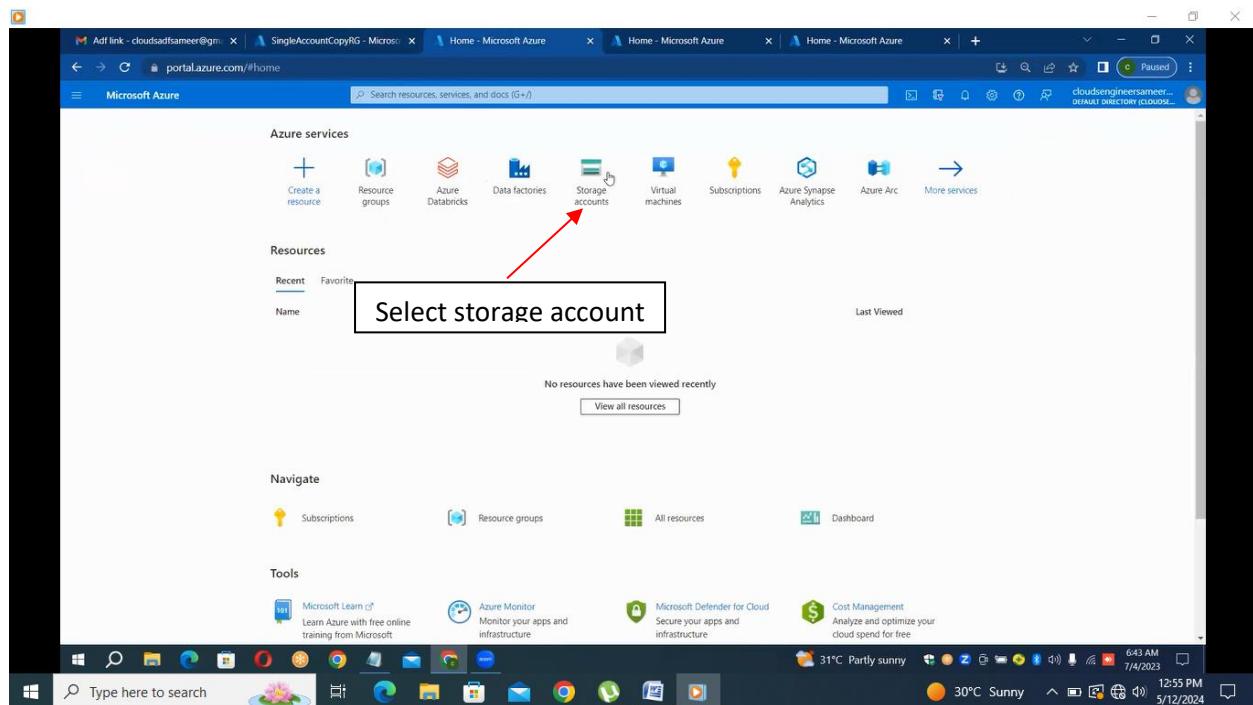




Azure Data Factory

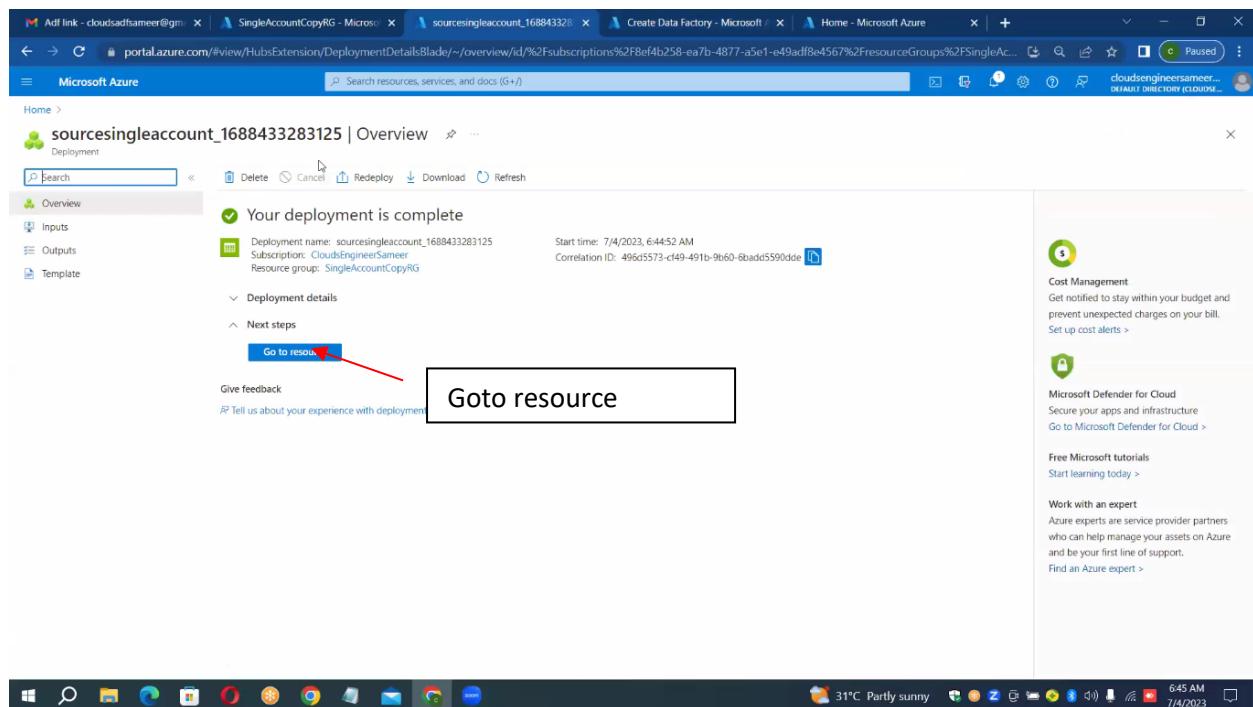
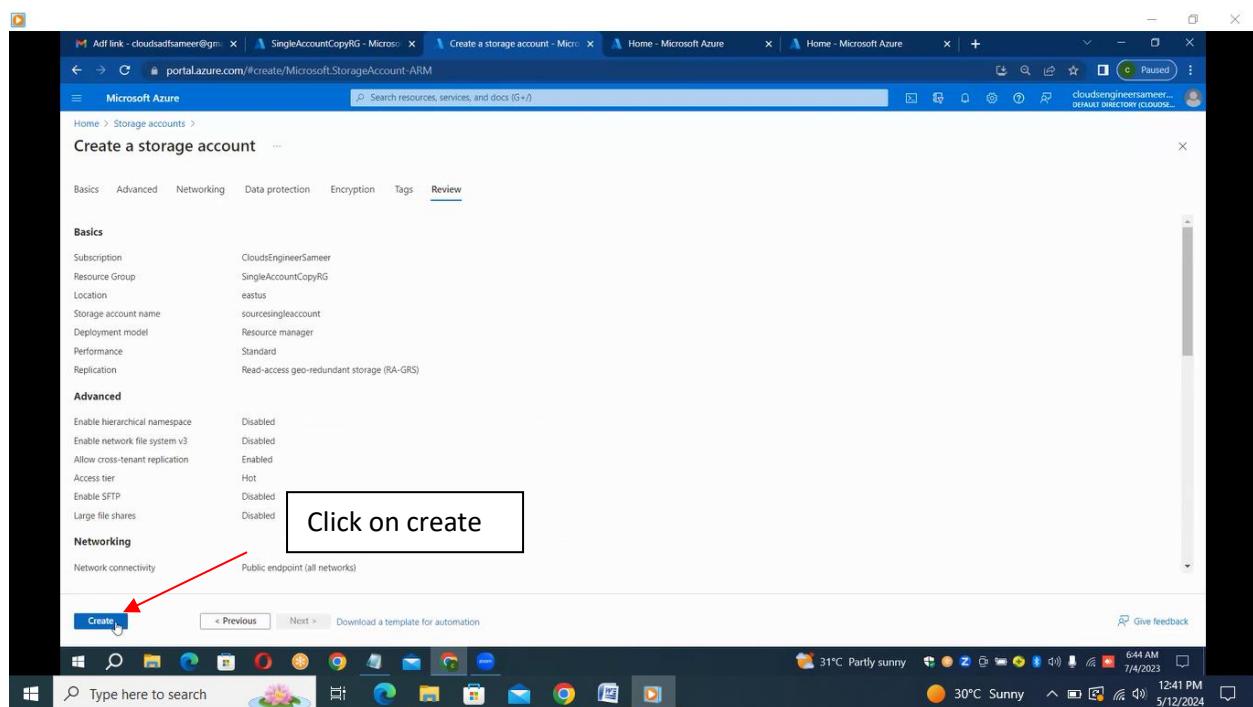


Open new tab portal.azure.com---->select storage account---->click on create---->give storage account name----->click on review----->click on create---->click on resource



The screenshot shows the Microsoft Azure Storage accounts page. At the top, there is a search bar and filter options. Below that, a table header includes columns for Name, Type, Resource group, Location, and Subscription. A red arrow points to the '+ Create' button in the top-left corner of the main content area. A callout box with the text 'Click on create' is positioned over the '+ Create' button. The central message says 'No storage accounts to display'.

The screenshot shows the 'Create a storage account' wizard in Microsoft Azure. It has tabs for Basics, Advanced, Networking, Data protection, Encryption, Tags, and Review. The Basics tab is selected. The 'Project details' section shows a subscription of 'CloudEngineerSameer' and a resource group of 'SingleAccountCopyRG'. The 'Instance details' section shows a storage account name of '(US) East US'. A red arrow points to the 'Review' button at the bottom left of the form. A callout box with the text 'Enter storage account name' is positioned over the storage account name field. The Windows taskbar at the bottom shows the date and time as 7/4/2023 12:37 PM.



Source account is ready from the left blade ---->select container----->click on new container---->gine the container name---->create

Click on container

Azure Data Factory

The screenshot shows the 'Containers' section of the Azure Storage account 'sourcesingleaccount'. A red arrow points to the '+ Create' button at the top left of the list area. A callout box contains the text 'Click on new container'.

The screenshot shows the 'New container' dialog box overlaid on the Storage account interface. A red arrow points to the 'Name' input field where 'ind' is typed. A callout box contains the text 'Enter container name'.

Successfully created storage container
Successfully created storage container 'input'.

Name	Last modified	Public access level	Lease state
Slogs	7/4/2023, 6:45:26 AM	Private	Available
Input	7/4/2023, 6:47:11 AM	Private	Available

Same as create output container

In single account we create input and output then goto input

Successfully created storage container
Successfully created storage container 'output'.

Name	Last modified	Public access level	Lease state
Slogs	7/4/2023, 6:45:26 AM	Private	Available
Input	7/4/2023, 6:47:11 AM	Private	Available
Output	7/4/2023, 6:47:27 AM	Private	Available

Azure Data Factory

Screenshot of the Azure Storage Container 'input' showing the upload interface. A red arrow points to the 'Upload' button.

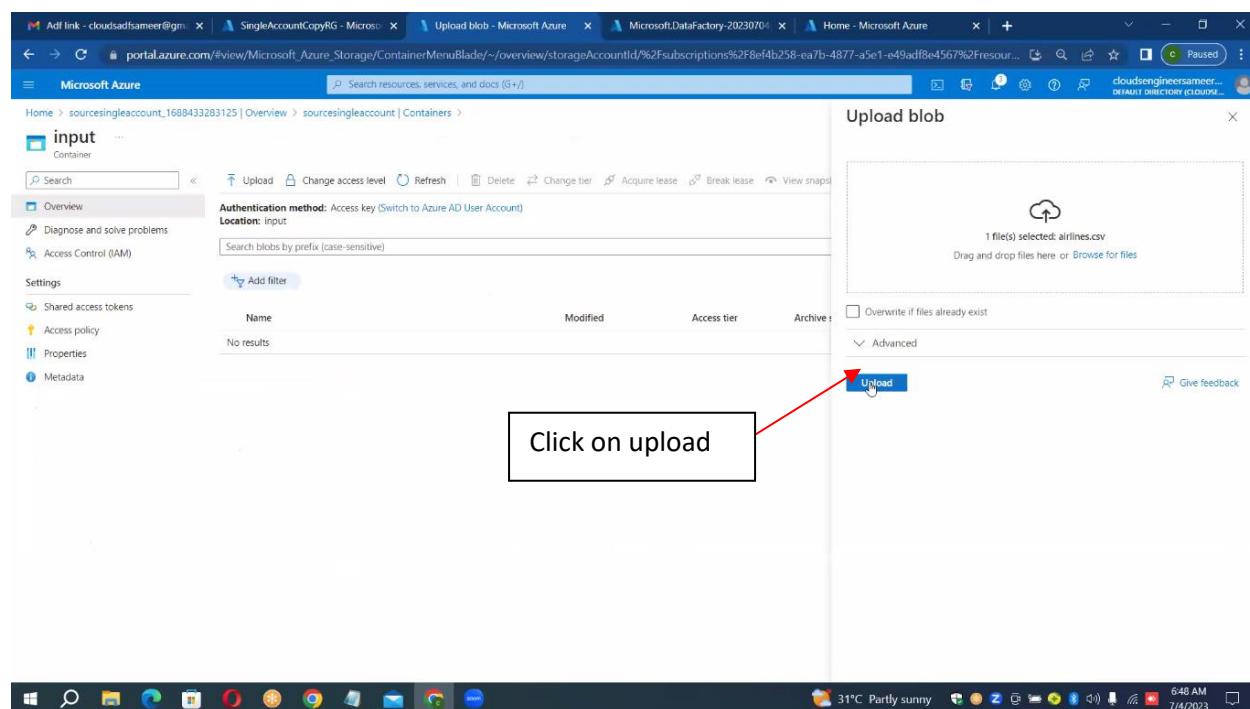
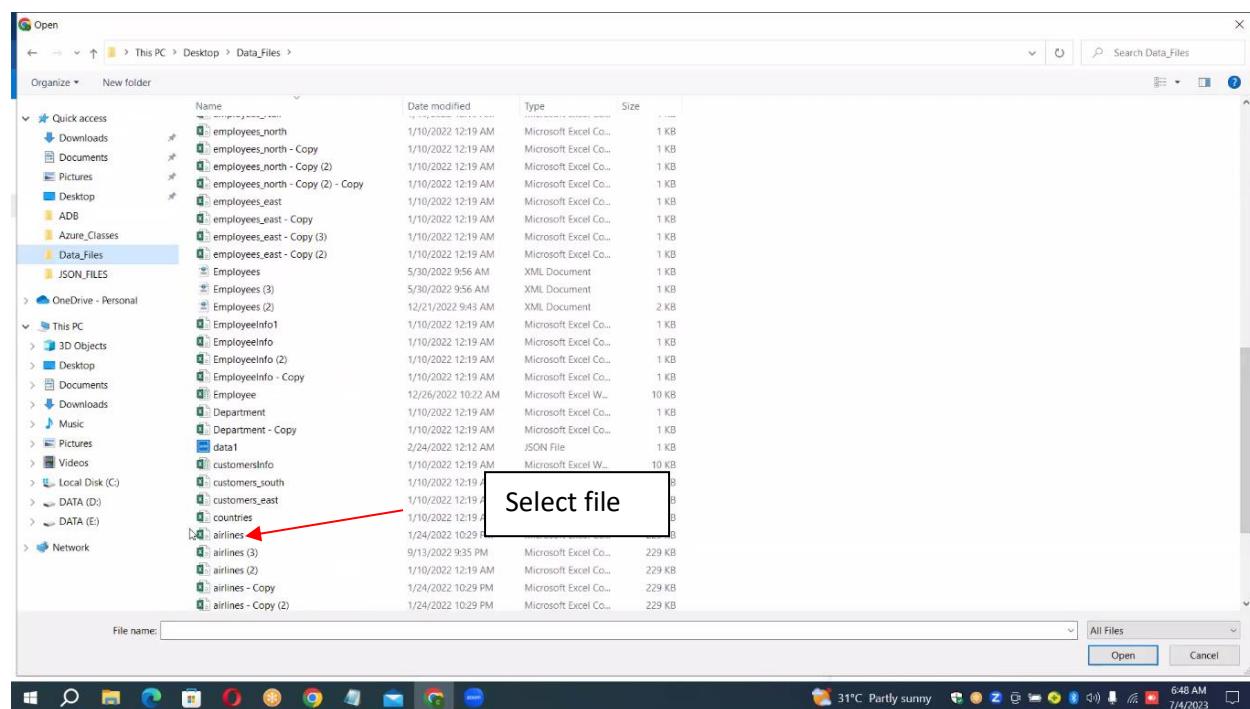
Click on upload

Sharing progress bar: 14:13 - Sharing Started to 14:29 - 30:28

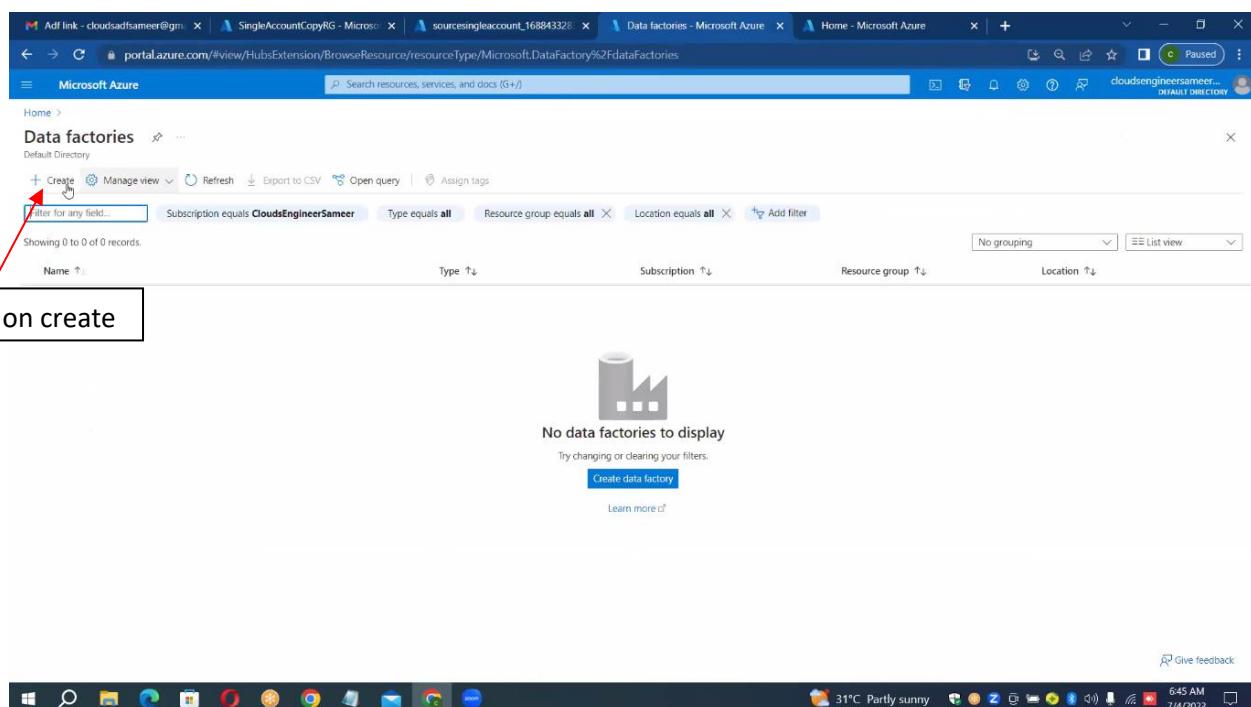
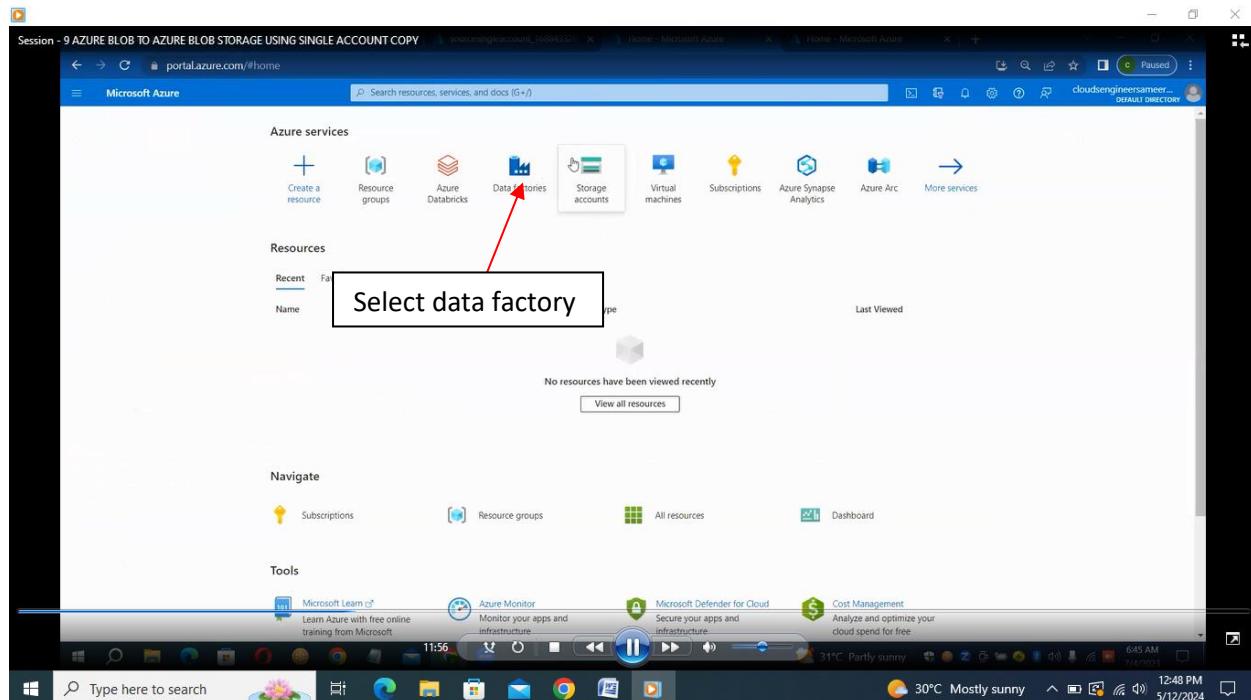
Screenshot of the 'Upload blob' dialog box. A red arrow points to the 'Process files' button.

Click on browser file

Azure Data Factory



Open new tab portal.azure.com---->select data factory---->click on create---->select resource group--->enter instance name---->click on review create---->click on create---->goto resource



Azure Data Factory

Select resource group
SingleAccountCopyRG

Name * InstanceSI

Click on review create

Select resource group

Enter instance name

Review + create

Validation Passed

Click on create

Create

Azure Data Factory

Your deployment is complete

Deployment name : Microsoft.DataFactory-20230704064514
Subscription : CloudsEngineerSameer
Resource group : SingleAccountCopyRG

Start time : 7/4/2023, 6:46:00 AM
Correlation ID : 54ac6589-b8ab-4072-990d-44b542ebaf79

Deployment details

Next steps

Go to resource

Click on resource

Give feedback
Tell us about your experience with deployment

Essentials

Resource group (move) : SingleAccountCopyRG
Status : Succeeded
Location : East US
Subscription (move) : CloudsEngineerSameer
Subscription ID : 8ef4b258-ea7b-4877-a5e1-e49adfe4567

Type : Data factory (V2)
Getting started : Quick-start

Azure Data Factory Studio

Launch studio

Click on launch studio

Quick Starts
Tutorials
Template Gallery
Training Modules

PipelineRuns
ActivityRuns
TriggerRuns

Click on manage

The screenshot shows the Microsoft Azure Data Factory interface. In the top left, there's a navigation bar with tabs like 'Adf link - cloudsadfsameer', 'SingleAccountCopyRG - M...', 'sourcesingleaccount - Micr...', 'InstanceSngleAccountCopy', 'InstanceSngleAccountCopy', 'Home - Microsoft Azure'. Below the navigation is a search bar 'Search factory and documentation' and a user profile 'cloudsengineersameer@gmail.com DEFAULT DIRECTORY'. The main content area has a title 'Data factory InstanceSngleAccountCopy'. On the left, a sidebar menu includes 'Home', 'Author', 'Monitor', **Manage** (which is highlighted with a red arrow), and 'Learning Center'. The main content area features four cards: 'Ingest' (Copy data at scale once or on a schedule.), 'Orchestrate' (Code-free data pipelines.), 'Transform data' (Transform your data using data flows.), and 'Configure SSIS' (Manage & run your SSIS packages in the cloud.). To the right is a 3D diagram illustrating data flow between various components.

Click on linkservice

This screenshot shows the 'Management' section of the Azure Data Factory interface. The left sidebar lists options like 'Home', 'Author', 'Monitor', **Manage** (highlighted with a red arrow), and 'Learning Center'. Under 'Manage', the 'Linked services' option is also highlighted with a red arrow. The main content area is titled 'Linked services' and contains a message 'No linked service to show' with a 'Create linked service' button. The bottom of the screen shows a Windows taskbar with icons for various applications and the date/time '7/4/2023 6:49 AM'.

Azure Data Factory

Click on new link service

No linked service to show

If you expected to see results, try changing your filters or create a new linked services.

Create linked service

New linked service

Data store Compute

Select blob storage

All Azure Database File Generic protocol NoSQL Services and apps

Amazon S3 Compatible	Azure Blob Storage	Azure Table Storage
SAP TABLE	Web Table	

Continue Cancel

Azure Data Factory

New linked service

Azure Blob Storage Learn more

Name * LS

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Authentication type Account key

Connection string Azure Key Vault

Account selection method From Azure subscription Enter manually

Azure subscription Select all

Storage account name * sourcesingleaccount

Additional connection properties

Create Back Test connection Cancel

New linked service

Azure Blob Storage Learn more

Connection string Azure Key Vault

Account selection method From Azure subscription Enter manually

Azure subscription CloudsEngineerSameer (8ef4b258-ea7b-4877-a5e1-e49adf8e4567)

Storage account name * sourcesingleaccount

Additional connection properties

+ New

Test connection To linked service To file path

Annotations

+ New

> Parameters > Advanced

Create Back Test connection Cancel

Azure Data Factory

Microsoft Azure | Data Factory > InstanceSingleAccountCopy

Search factory and documentation

Data Factory > Validate all > Publish all

Author

- Home
- Author**
- Monitor
- Manage
- Learning Center

General

- Factory settings
- Connections
- Linked services**
- Integration runtimes
- Microsoft Purview
- Source control
- Git configuration
- ARM template

Showing 1 - 1 of 1 items

Name	Type	Related	Annotations
LS_BLOB_SOURCE_SINK	Azure Blob Storage	0	

Triggers

Global parameters

Data flow libraries

Security

Credentials

Customer managed key

Outbound rules

Preview experience: Off

Successfully created: Successfully created LS_BLOB_SOURCE_SINK (Linked service).

Windows taskbar: 31°C Partly sunny, 6:50 AM, 7/4/2023

Microsoft Azure | Data Factory > InstanceSingleAccountCopy

Search factory and documentation

Data Factory > Validate all > Publish all

Author

- Home
- Author**
- Monitor
- Manage
- Learning Center

Factory Resources

- Pipelines
- Change Data Capture (preview)
- Datasets**
- Data flows
- Power Query

Set properties

Name: DS_BLOB

Linked service: Select...

First row as header:

OK Back Cancel

1 author

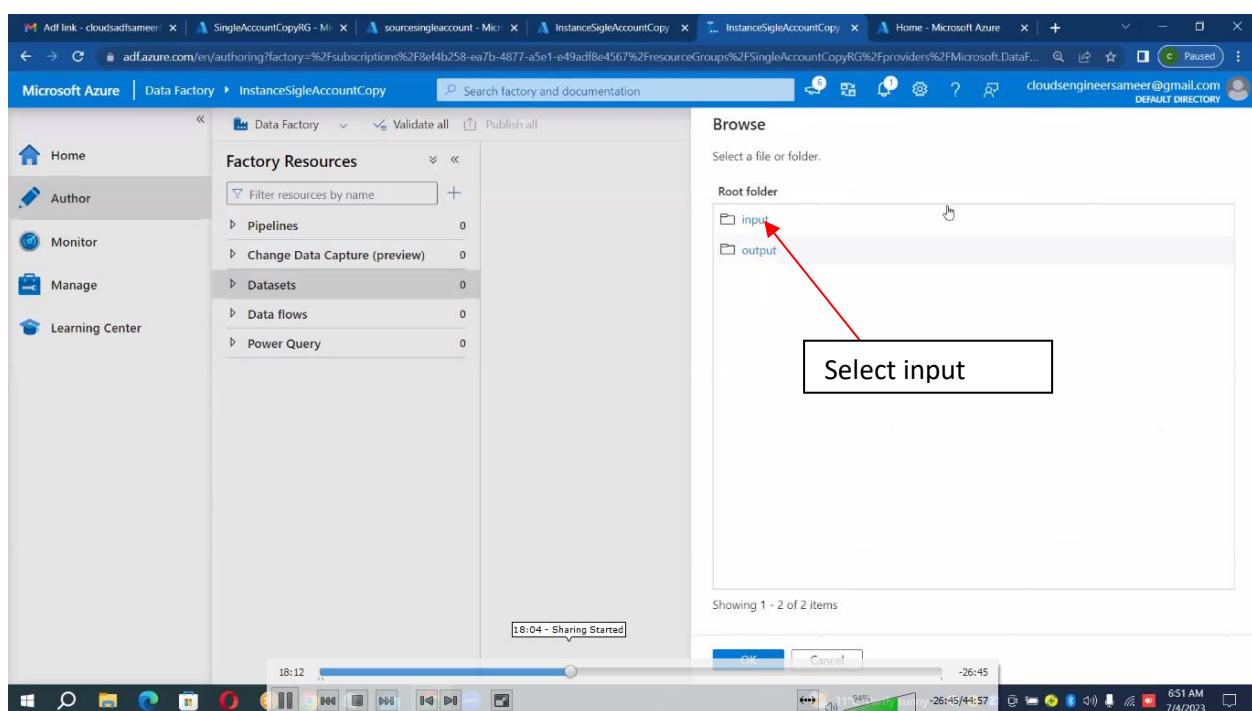
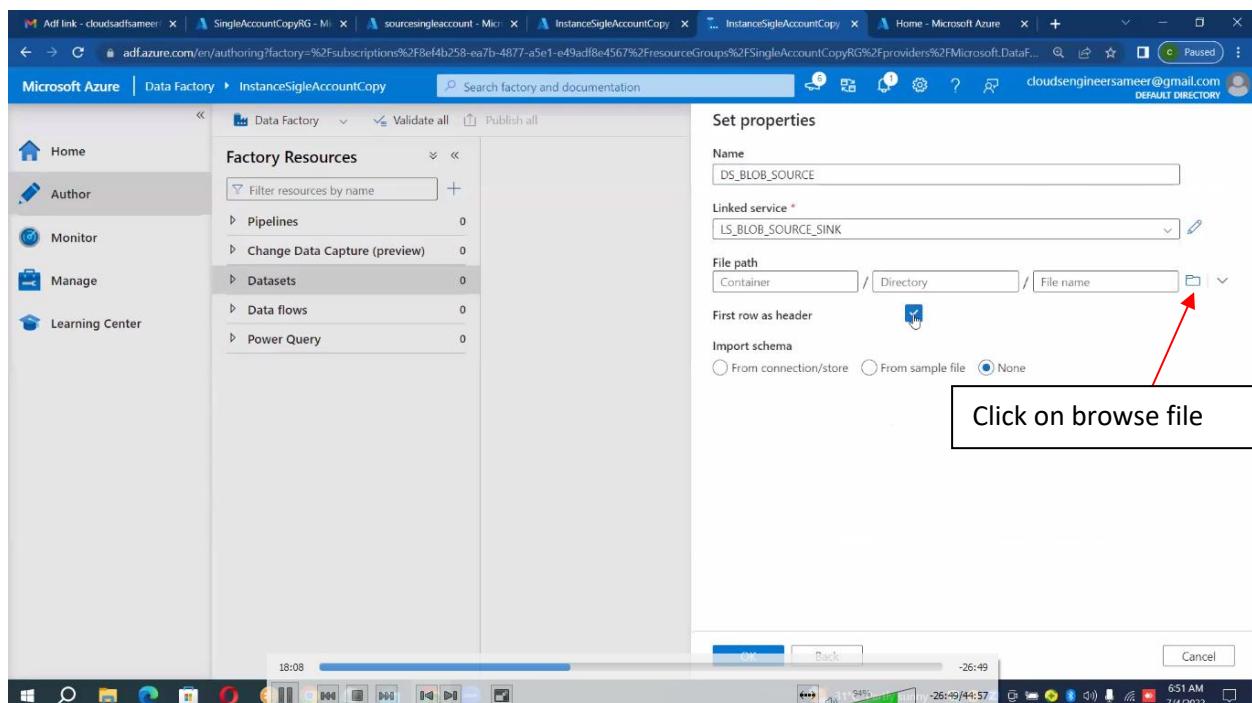
2. Datasets

3. Enter datasets

4. Select link service

Windows taskbar: 31°C Partly sunny, 6:51 AM, 7/4/2023

Azure Data Factory



Create another container output---->click on new datasets---->enter blob--->select csv file---->give the data set name--->select linkservice----->select browser ----->select output--->click on ok

Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory Author interface. On the left sidebar, the 'Author' tab is selected. In the main area, under 'Factory Resources', there is a 'Datasets' section containing one item named 'DS_BLOB_SOURCE'. This dataset is of type 'DelimitedText' and is connected to a 'Linked service' named 'LS_BLOB_SOURCE_SINK'. Below the dataset, there are buttons for 'New dataset' and 'New folder'. On the right side, there is a 'Properties' panel for 'DS_BLOB_SOURCE' with tabs for 'General' and 'Related'. The 'General' tab shows the name 'DS_BLOB_SOURCE' and a description field. There are also sections for 'Annotations' and a preview experience toggle.

This screenshot shows the same Microsoft Azure Data Factory interface as the previous one, but with a different focus. A red box highlights the 'New dataset' button at the bottom of the 'Datasets' section. Another red box highlights the search bar in the 'Select a data store' grid, which contains various data store options like Amazon Marketplace Web Service, Amazon S3 Compatible, Azure Blob Storage, etc. The 'Enter blob' button is also highlighted in a white box.

Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the navigation bar includes Home, Author (selected), Monitor, Manage, and Learning Center. In the center, under 'Factory Resources', there is a list of datasets, with 'DS_BLOB_SOURCE' currently selected. To the right, a 'Select format' dialog is open, displaying various file formats: Avro, Binary, DelimitedText (selected), Excel, JSON, XML, and others. A red arrow points to the 'CSV' icon in the 'DelimitedText' row. A callout box with the text 'Click on csv file' is positioned over the 'CSV' icon.

The screenshot shows the Microsoft Azure Data Factory interface. The 'Set properties' dialog is open for the dataset 'DS_BLOB_SOURCE'. The 'Name' field is empty and highlighted with a red border, with an error message 'Dataset name is required.' displayed above it. Other fields in the dialog include 'Linked service *' (with a dropdown menu 'Select...'), 'First row as header' (with a checked checkbox), and buttons for 'OK', 'Back', and 'Cancel'. A red arrow points to the 'Name' input field. A callout box with the text 'Enter data set name' is positioned over the 'Name' field.

Azure Data Factory

Set properties

Name: DS_BLOB_SOURCE

Linked service: LS_BLOB_SOURCE_SINK

File path: Container / Directory / File name

First row as header:

Import schema: From connection/store From sample file None

Click on browse

Browse

Select a file or folder.

Root folder

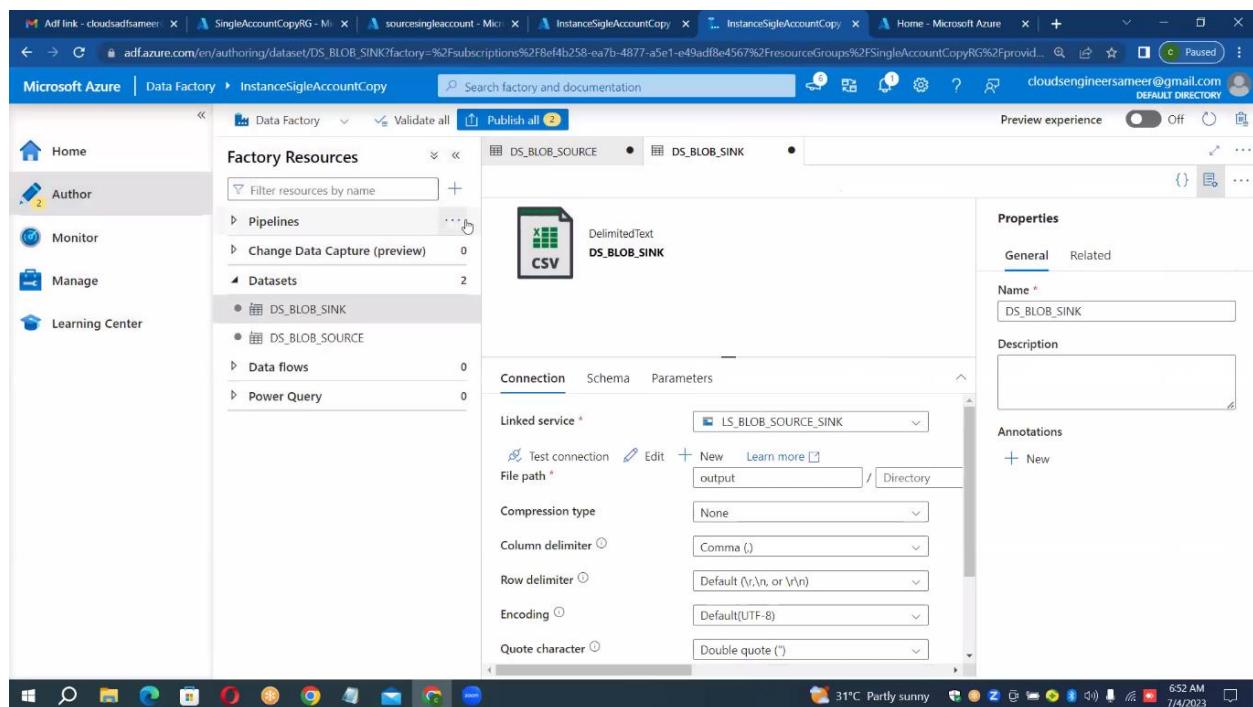
- input
- output**

Select the output

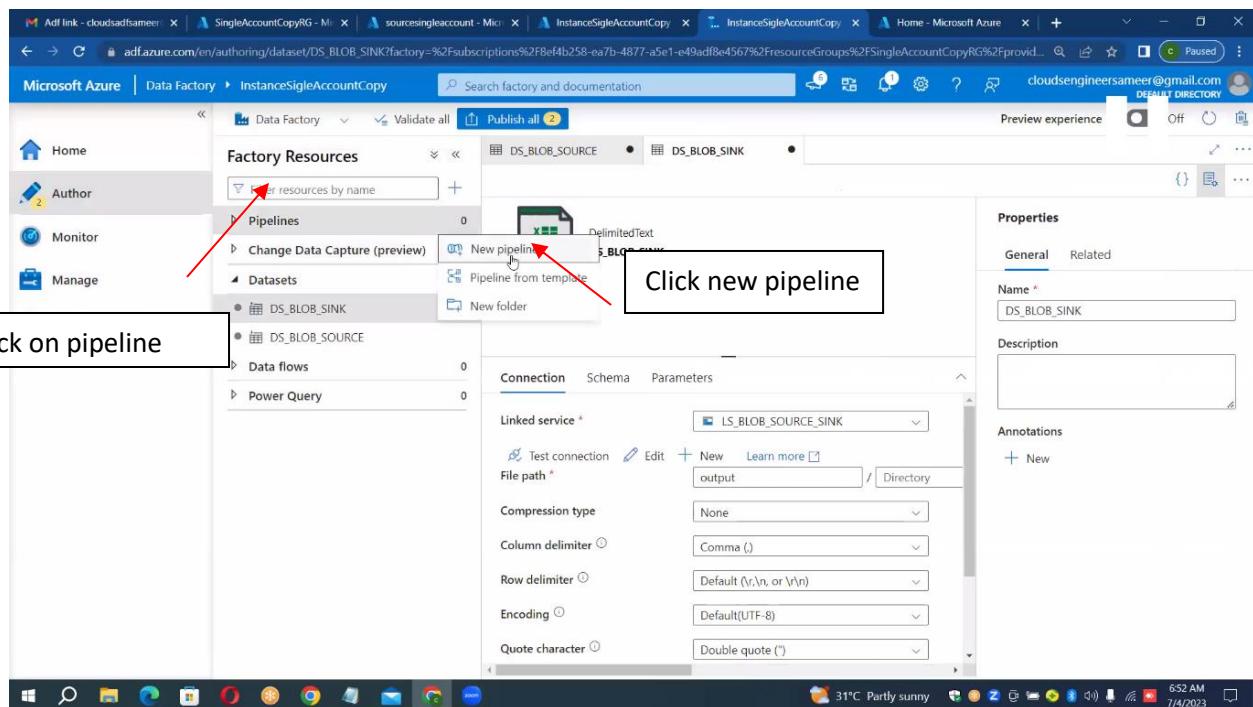
Click on ok

Showing 1 - 2 of 2 items

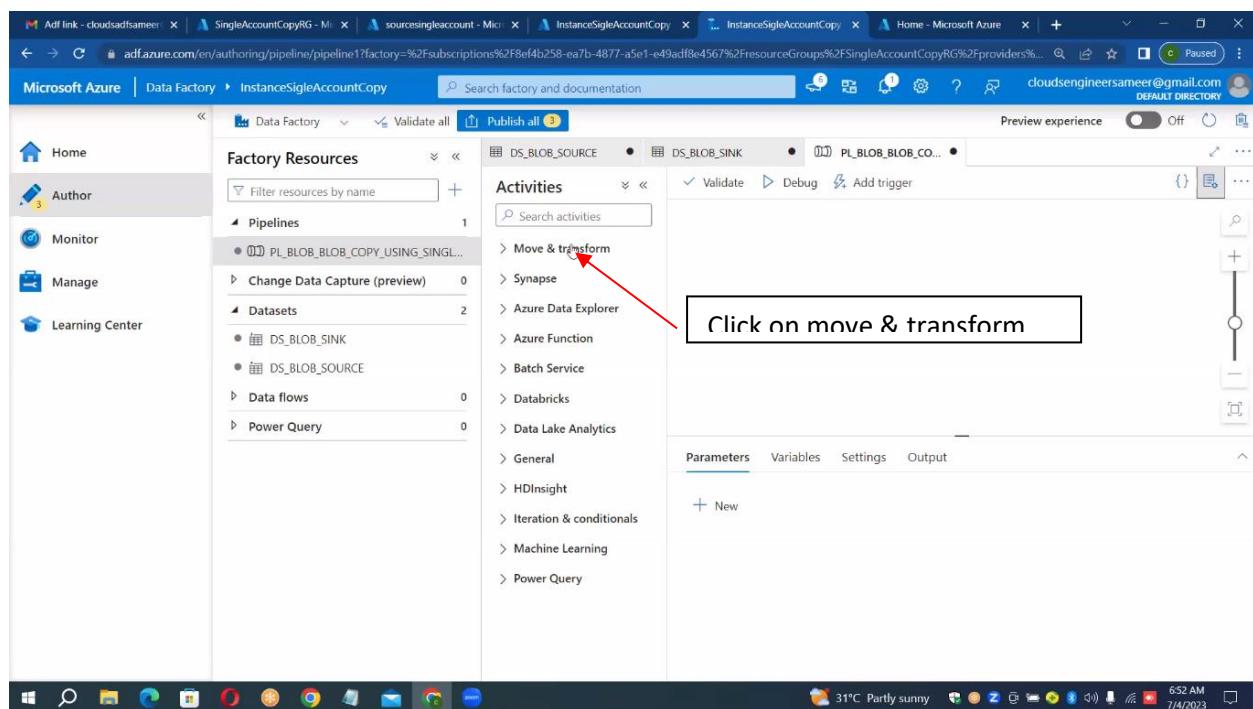
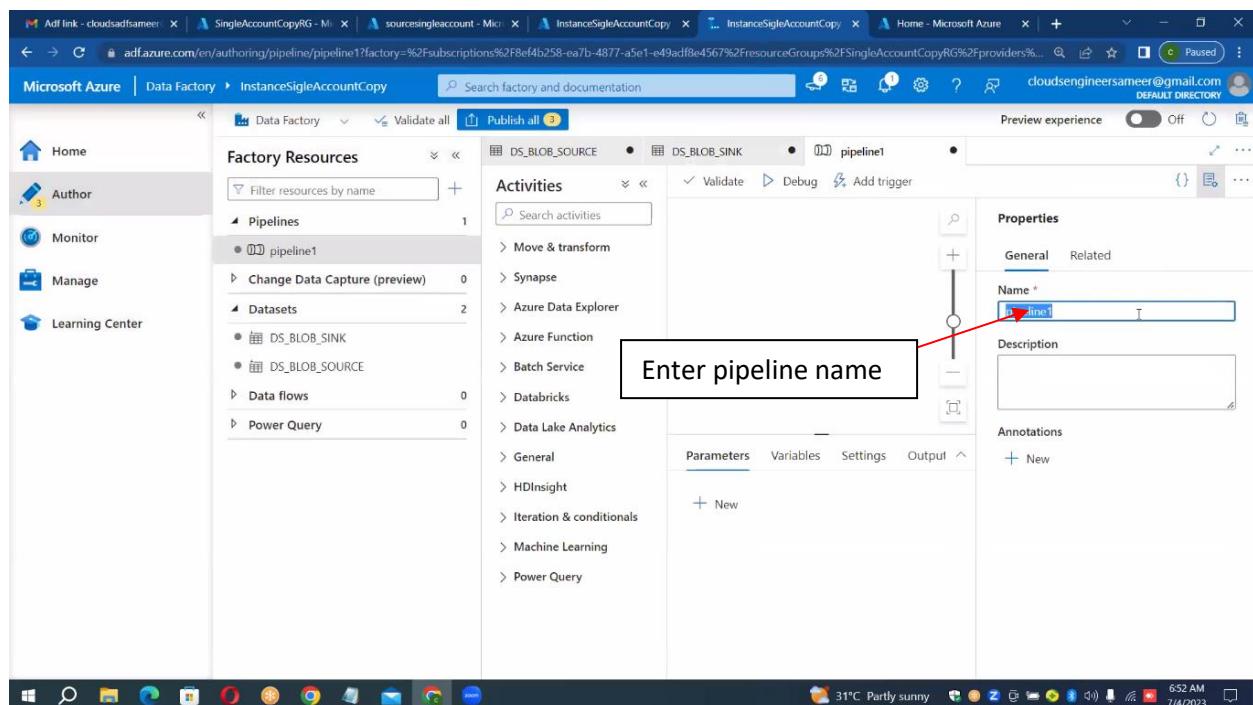
OK Cancel

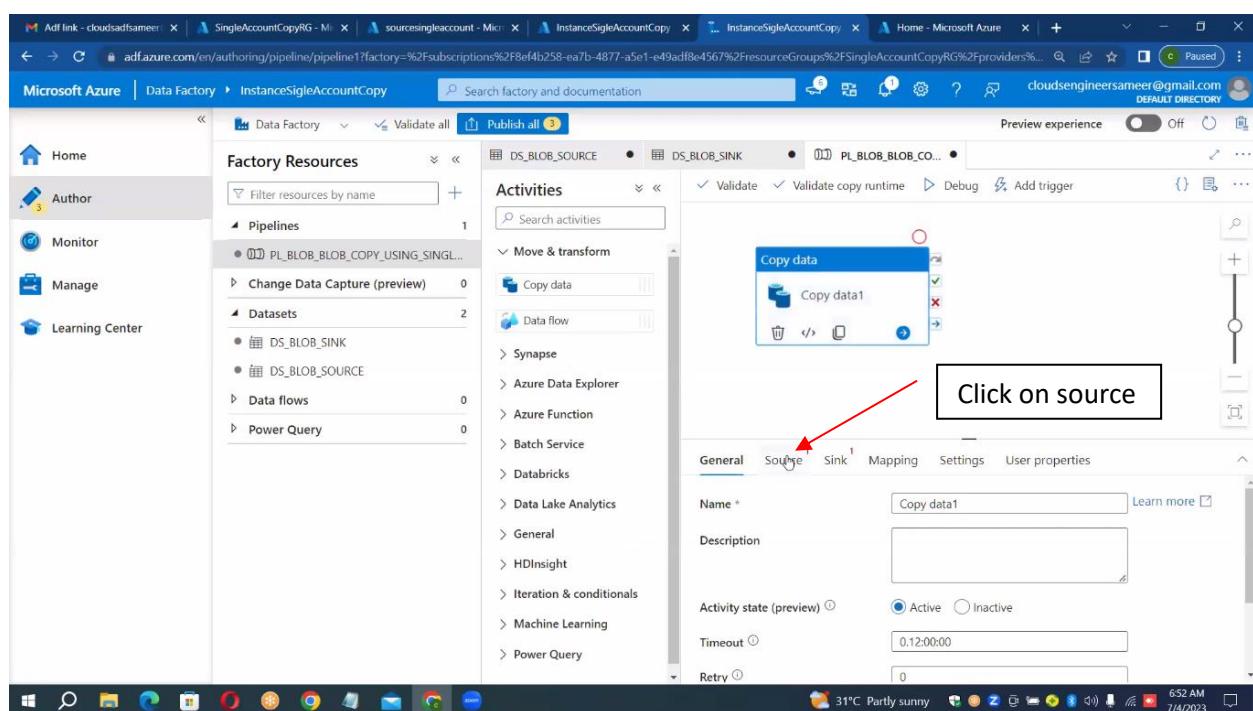
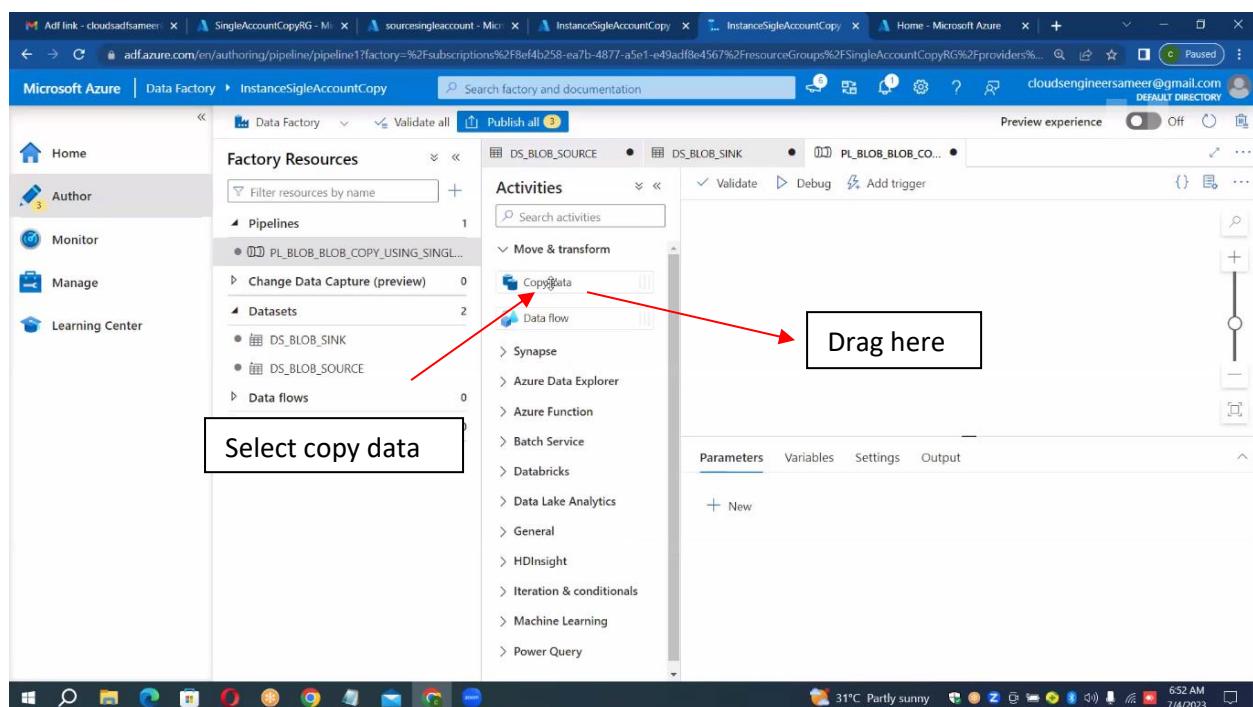


Goto pipeline---->click on new pipe line ---->give pipeline name---->select move and transform---->select copy data---->drag the copy data to right side---->click on source---->select data set---->click on sink----->select sink data set---->change extent name--->click on debug



Azure Data Factory

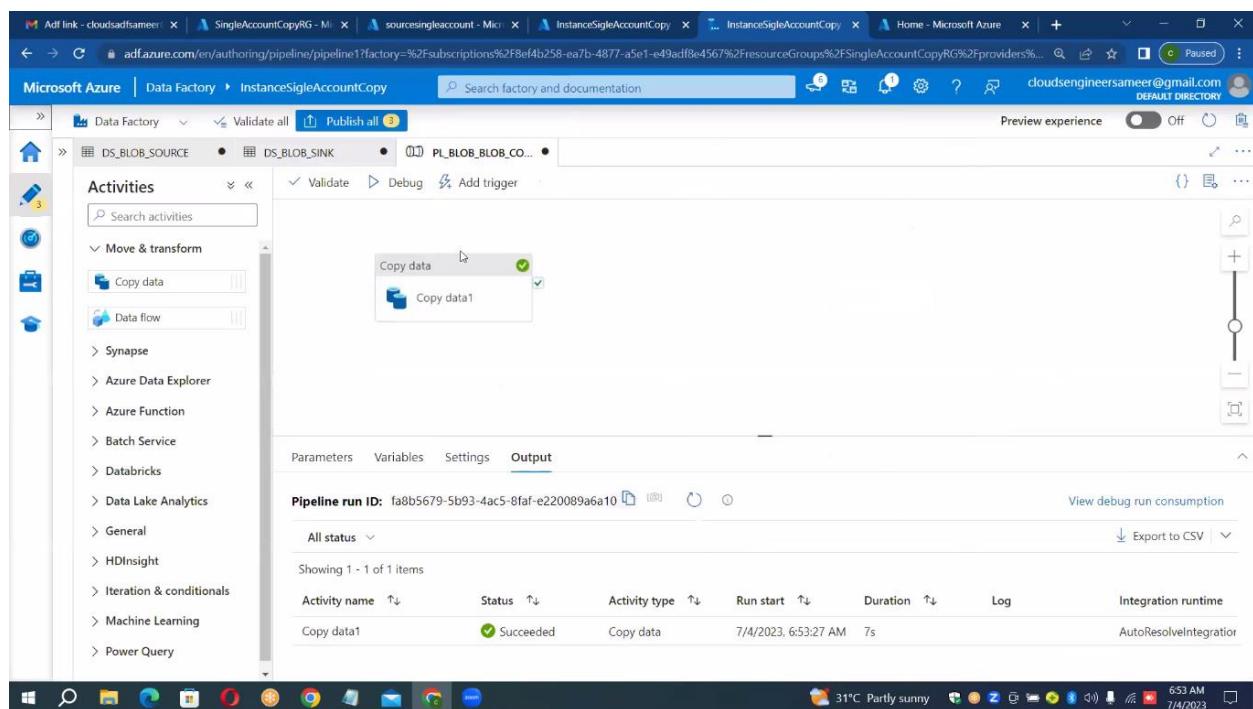
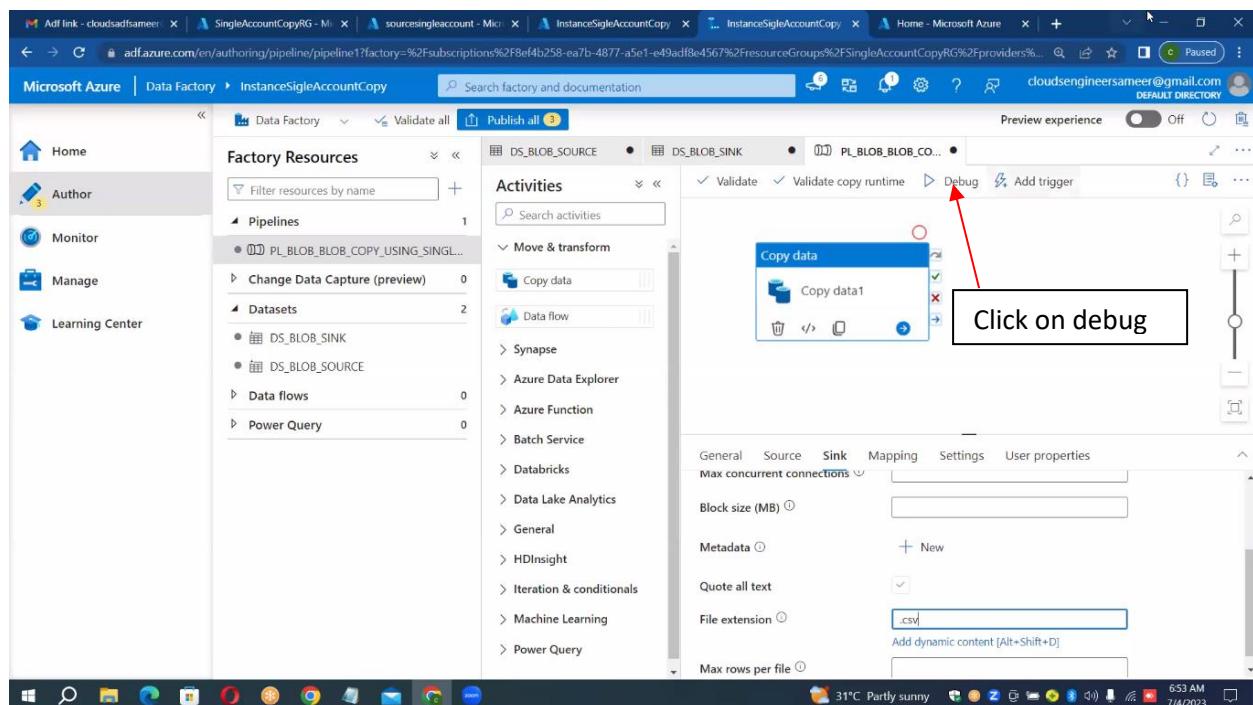




Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (2), and 'Data flows'. The main area displays a 'Copy data' activity named 'Copy data1'. The 'Sink' tab is selected, and a red arrow points to the 'Sink dataset' dropdown menu, which contains the text 'Select...'. A box labeled 'Select sink data set' is overlaid on the interface.

This screenshot continues from the previous one, showing the 'Sink' tab of the 'Copy data' activity. The 'File extension' input field is highlighted with a red arrow and contains the value '.txt'. A box labeled 'Change the extension name' is overlaid on the interface. Other visible settings include 'Max concurrent connections', 'Block size (MB)', 'Metadata', 'Quote all text', and 'Max rows per file'.



The screenshot shows the Azure Storage Explorer interface. A container named 'output' is selected. Inside, there is a single blob named 'airlines.csv'. The blob details are as follows:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
airlines.csv	7/4/2023, 6:53:33 AM	Hot (Inferred)		Block blob	228.37 KB	Available

Case study -4

Title: copy data from azure blob storage to azure data Lake Gen2 using multiple files and options

Specifications:

Source data storage : azure blob storage account

Files :

Sink data storage : azure data Lake Gen2

Files :

Without disrupting the existing container in source storage account we can create another container

Click on container ---->give new container name---->click on create---->select input 1---->select upload---->click on browser---->add some four files----->click on upload

Azure Data Factory

The screenshot shows the 'Containers' section of the Azure Storage account 'sourcesingleaccount'. A red arrow points to the '+ Container' button. A black box highlights the 'Input' container, with the text 'Click container' overlaid.

Name	Last modified	Public access level	Lease state
Slogs	7/4/2023, 6:45:26 AM	Private	Available
Input	7/4/2023, 6:47:11 AM	Private	Available
Output	7/4/2023, 6:47:27 AM	Private	Available

The screenshot shows the 'Containers' section of the Azure Storage account 'sourcesingleaccount'. A red arrow points to the 'Name' input field in the 'New container' dialog, which contains the letter 'I'. A black box highlights the 'Create' button, with the text 'Enter container name' overlaid.

New container

Name *

Public access level Private (no anonymous access) Advanced

Click on create

Azure Data Factory

The screenshot shows the Azure Storage account containers page. On the left, there's a sidebar with options like Overview, Activity log, Tags, and Data storage (Containers). The main area lists containers: Slogs, input, and input1. A red arrow points from a callout box labeled "Click on input1" to the input1 row. In the top right corner, a success message says "Successfully created storage container 'input1'".

The screenshot shows the Azure Storage account input1 container blobs page. On the left, there's a sidebar with options like Overview, Settings, and Properties. The main area shows a table with columns: Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. A red arrow points from a callout box labeled "Click on upload" to the "Upload" button in the top left of the interface.

Azure Data Factory

Upload blob

Drag and drop files here
or
Browse for files

Click on browser file

Select the file

Name	Date modified	Type	Size
Students	1/27/2023 7:58 PM	JSON File	1 KB
StudentsInfo	11/22/2022 9:01 PM	JSON File	1 KB
Students	2/24/2022 12:32 AM	JSON File	1 KB
SrcEmployees1 (2)	1/23/2023 8:32 AM	Microsoft Excel Co...	1 KB
SrcEmployees1	1/23/2023 8:32 AM	Microsoft Excel Co...	1 KB
SrcEmployees	5/18/2022 7:11 PM	Microsoft Excel W...	7 KB
SrcEmployees	1/23/2023 8:11 AM	Microsoft Excel Co...	1 KB
SrcEmployees (2)	5/18/2022 7:11 PM	Microsoft Excel W...	7 KB
SrcEmployees (2)	3/24/2023 8:09 PM	Microsoft Excel Co...	1 KB
SingleLineJSON99	12/1/2022 10:49 AM	JSON File	1 KB
SingleLineJSON1	12/1/2022 10:48 AM	Text Document	1 KB
SingleLineJSON	12/1/2022 10:34 AM	JSON File	1 KB
SCD	2/13/2023 10:31 PM	Text Document	6 KB
Sales	1/10/2022 12:19 AM	Microsoft Excel Co...	11 KB
routes	1/10/2022 12:19 AM	Microsoft Excel Co...	2,322 KB
planes	1/10/2022 12:19 AM	Microsoft Excel Co...	8 KB
MultipleDelimiter	12/8/2022 9:50 AM	Text Document	1 KB
Movies	10/25/2022 8:11 PM	Microsoft Excel Co...	2,230 KB
Items_Details.csv	11/15/2022 8:33 PM	Microsoft Excel W...	6 KB
Items_Details.csv	1/25/2023 8:37 PM	Microsoft Excel Co...	1 KB
Items_Details.csv (2)	1/25/2023 8:38 PM	Microsoft Excel Co...	1 KB
Items_Details	11/15/2022 8:41 PM	Microsoft Excel Co...	1 KB
Items	3/15/2023 8:23 PM	Text Document	1 KB
ItemCount	11/15/2022 9:30 AM	Text Document	3 KB
export	4/7/2023 11:35 AM	Microsoft Excel Co...	256 KB
EmployeesNested1	12/21/2022 9:48 AM	XML Document	2 KB
EmployeesNested	12/21/2022 9:43 AM	XML Document	2 KB
EmployeesNested (2)	12/21/2022 9:43 AM	XML Document	2 KB

Azure Data Factory

Click on upload

Successfully uploaded blob(s)
Successfully uploaded 7 blob(s).

Then create sink storage account

Click on storage account---->click on create --->enter storage account---->select advance -
----->click on review ---->click on create---goto resource --->click on container ----->click
new container---->give container name---->click on create

Azure Data Factory

Screenshot of the Microsoft Azure portal showing the storage account creation process.

The top navigation bar shows several tabs: Adf link - cloudsadfsameer, SingleAccountCopyRG - M..., input1 - Microsoft Azure, InstanceSingleAccountCopy, InstanceSingleAccountCopy, Home - Microsoft Azure, and a paused tab. The URL is portal.azure.com/#home.

The main menu includes Azure services (Create a resource, Resource groups, Azure Databricks, Data factories, Storage accounts, Virtual machines, Subscriptions, Azure Synapse Analytics, Azure Arc, More services) and Resources (Recent, Favorite).

A red box highlights the "Storage accounts" icon in the Azure services menu, with a callout "Click on storage account".

The resources section shows a message: "No resources have been viewed recently" and a "View all resources" button.

The bottom navigation bar includes links for Subscriptions, Resource groups, All resources, and Dashboard.

The taskbar at the bottom shows various pinned icons and the system clock indicating 7:02 AM on 7/4/2023.

Screenshot of the Microsoft Azure portal showing the creation of a new storage account.

The top navigation bar shows several tabs: Adf link - cloudsadfsameer, SingleAccountCopyRG - M..., input1 - Microsoft Azure, InstanceSingleAccountCopy, InstanceSingleAccountCopy, Storage accounts - Microsoft, and a paused tab. The URL is portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.Storage%2FStorageAccounts.

The main menu includes Storage accounts (Default Directory), Subscriptions, Resource groups, All resources, and Dashboard.

A red box highlights the "+ Create" button in the Storage accounts menu, with a callout "Click on create".

The storage account list table shows one record:

Name	Type	Kind	Resource group	Location	Subscription
sourcestorageaccount	Storage account	StorageV2	SingleAccountCopyRG	East US	CloudsEngineerSameer

The bottom navigation bar includes links for Subscriptions, Resource groups, All resources, and Dashboard.

The taskbar at the bottom shows various pinned icons and the system clock indicating 7:02 AM on 7/4/2023.

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *: CloudEngineerSameer

Resource group *: SingleAccountCopyRG

Storage account name *:

Region *: (US) East US

Review < Previous Next : Advanced > Give feedback

Advanced **Networking** Data protection Encryption Tags Review

Hierarchical Namespace

Access protocols

Blob and Data Lake Gen2 endpoints are provisioned by default [Learn more](#)

Enable SFTP

Enable network file system v3

Blob storage

Allow cross-tenant replication

Review < Previous Next : Networking > Give feedback

Ad link - cloudsadfsameer | SingleAccountCopyRG - Microsoft Azure | input1 - Microsoft Azure | InstanceSingleAccountCopy | InstanceSingleAccountCopy | Create a storage account -

portal.azure.com/#create/Microsoft.StorageAccount-ARM

Microsoft Azure

Search resources, services, and docs (G+)

cloudengineersameer... DEFAULT DIRECTORY

Home > Storage accounts >

Create a storage account ...

Basics Advanced Networking Data protection Encryption Tags Review

Subscription: CloudsEngineerSameer
Resource Group: SingleAccountCopyRG
Location: eastus
Storage account name: sinkstoragemultifiles
Deployment model: Resource manager
Performance: Standard
Replication: Read-access geo-redundant storage (RA-GRS)

Advanced:

- Enable hierarchical namespace: Enabled
- Enable network file system v3: Disabled
- Allow cross-tenant replication: Disabled
- Access tier: Hot
- Enable SFTP: Disabled
- Large file shares: Disabled

Networking:

Network connectivity

Create Click on create

30:00 - Sharing Started

Give feedback

30:15 -14:42 7:03 AM 7/4/2023

Ad link - cloudsadfsameer | SingleAccountCopyRG - Microsoft Azure | input1 - Microsoft Azure | InstanceSingleAccountCopy | InstanceSingleAccountCopy | sinkstoragemultifiles_1688434404603 -

portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2fsubscriptions%2f8ef4b258-ea7b-4877-a5e1-e49ad8e4567%2fresourceGroups%2fSingleAcc...

Microsoft Azure

Search resources, services, and docs (G+)

cloudengineersameer... DEFAULT DIRECTORY

Home >

sinkstoragemultifiles_1688434404603 | Overview

Deployment

Deployment succeeded Deployment 'sinkstoragemultifiles_1688434404603' to resource group 'SingleAccountCopyRG' was successful.

Go to resource Pin to dashboard

Search

Overview Inputs Outputs Template

Your deployment is complete

Deployment name: sinkstoragemultifiles_1688434404603
Subscription: CloudsEngineerSameer
Resource group: SingleAccountCopyRG

Start time: 7/4/2023, 7:03:33 AM Correlation ID: f1791b9a-37d9-45e3-aa86-c523f525a00e

Deployment details Next steps Go to resource Click on goto resource

Give feedback Tell us about your experience with deployment

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 and be your first line of support. Find an Azure expert >

31°C Partly sunny 7:04 AM 7/4/2023

Data storage

- Containers (selected)
- File shares
- Queues
- Tables

Properties

Data Lake Storage

Hierarchical namespace	Enabled
Default access tier	Hot
Blob public access	Enabled
Blob soft delete	Enabled (7 days)
Container soft delete	Enabled (7 days)
Versioning	Disabled
Change feed	Disabled
NFS v3	Disabled
SFTP	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

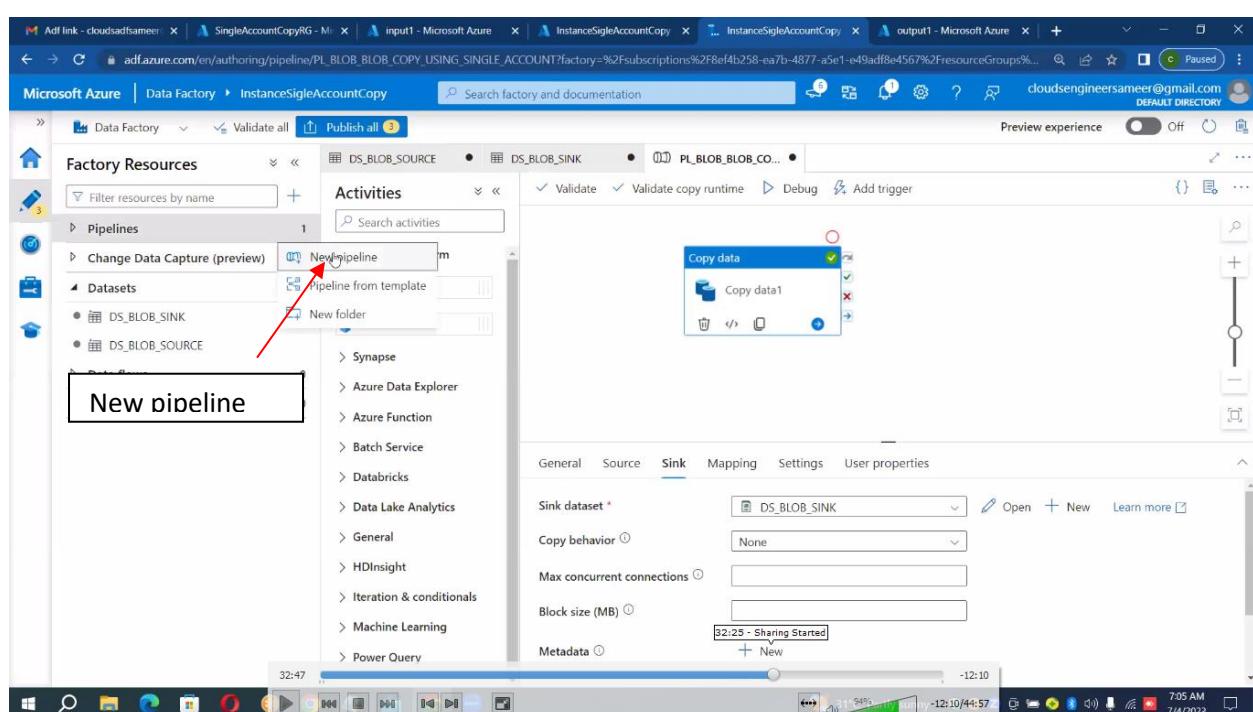
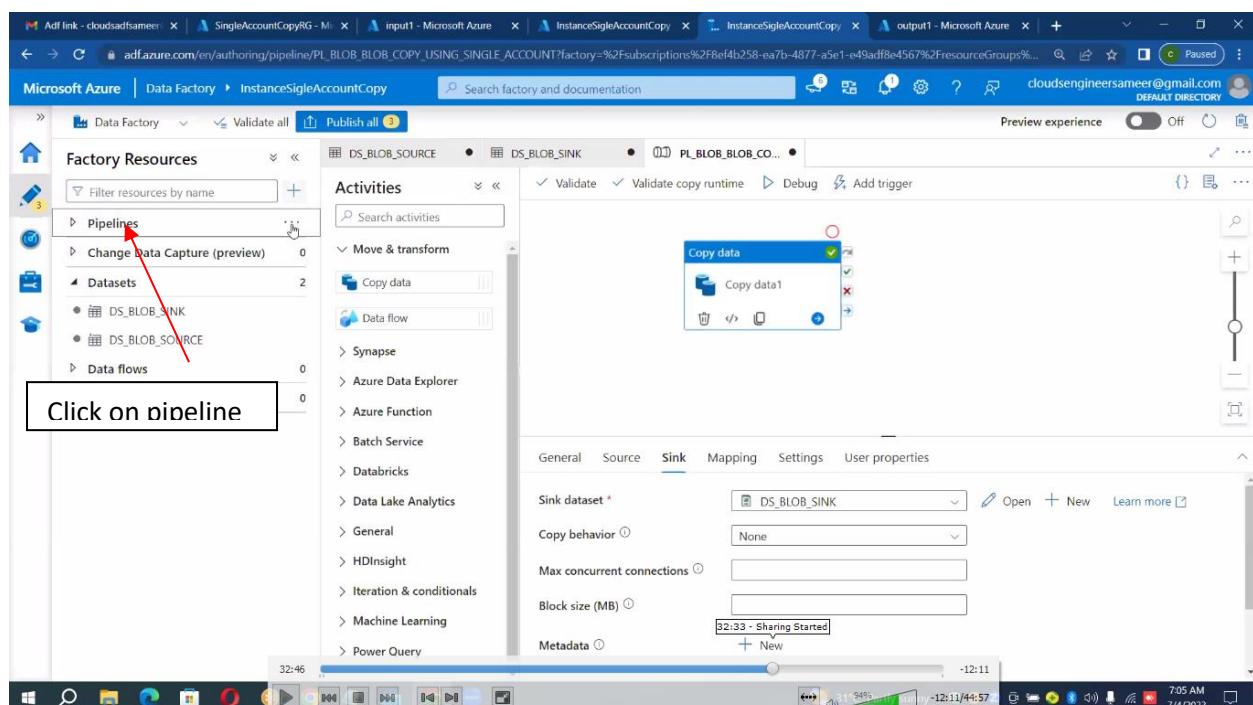
Allow access from	All networks
Number of private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes

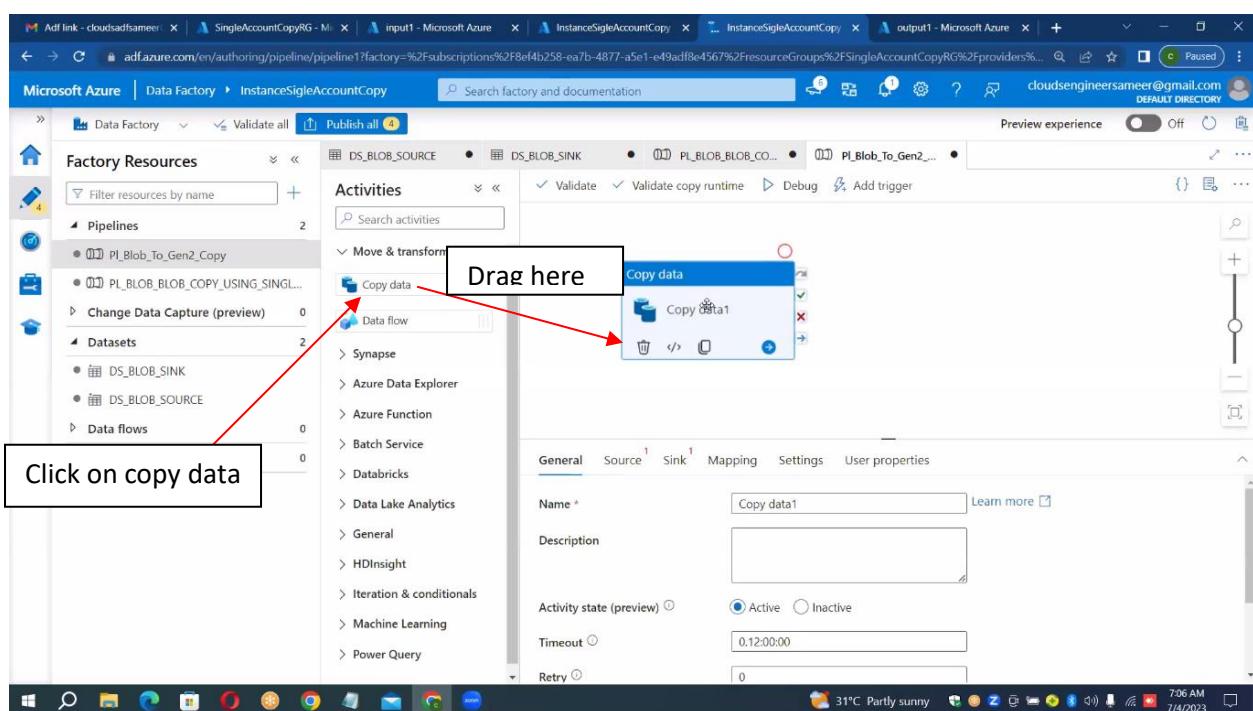
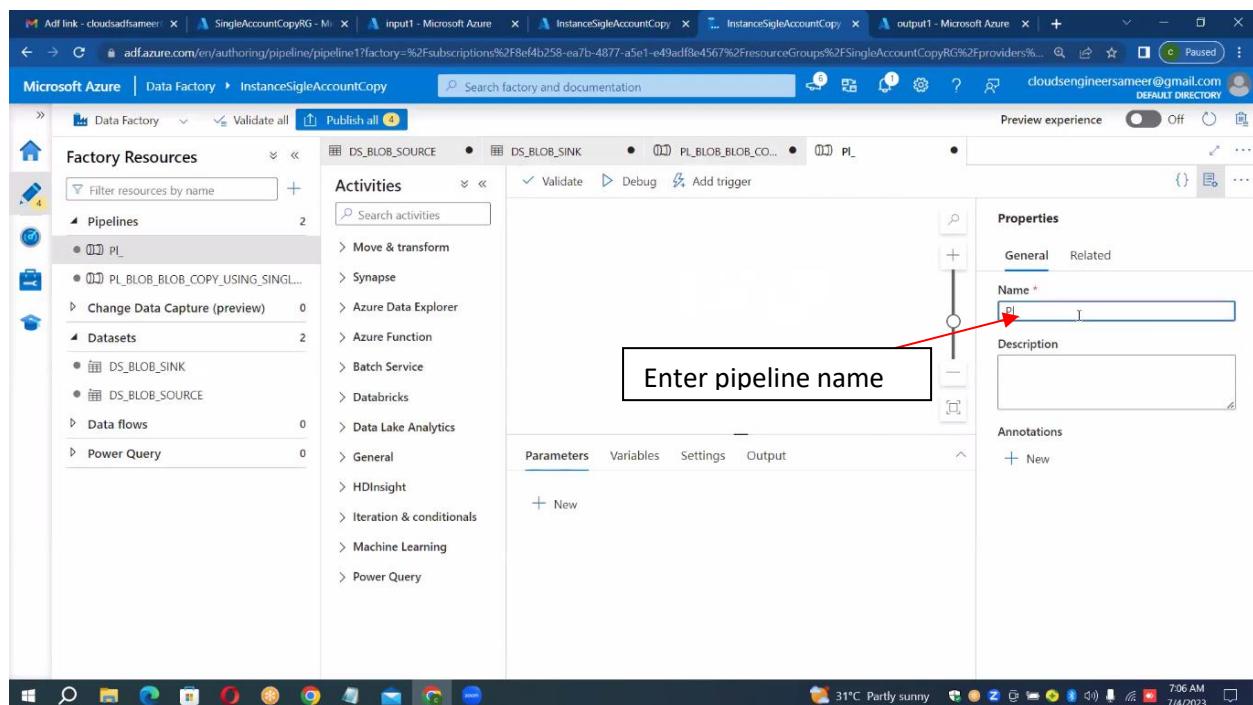
So source and sink storage is ready then we can create new pipeline in existing data factory

Open data factory click on pipeline---->new pipeline--->give the pipeline name---->select activity ---->select move transform----->select copy data ---->select source---->select source data---->select sink ----->select sink data and link service

Note: create sink data set here

Note:From file path option select file path type wildcard file path





Azure Data Factory

Click on debug

Select wildcard file path

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
airlines.csv	7/4/2023, 7:12:49 AM	Hot (Inferred)		Block blob	228.37 KB	Available
countries.csv	7/4/2023, 7:12:49 AM	Hot (Inferred)		Block blob	3.79 kB	Available
export.csv	7/4/2023, 7:12:49 AM	Hot (Inferred)		Block blob	255.6 kB	Available
Movies.csv	7/4/2023, 7:12:49 AM	Hot (Inferred)		Block blob	2.18 MiB	Available
punes.csv	7/4/2023, 7:12:49 AM	Hot (Inferred)		Block blob	7.01 kB	Available
routes.csv	7/4/2023, 7:12:49 AM	Hot (Inferred)		Block blob	2.27 MiB	Available
Sales.csv	7/4/2023, 7:12:49 AM	Hot (Inferred)		Block blob	10.21 kB	Available