



# Azure Data Factory Real Time Scenario

## Part 5

**List of Files in Copy Activity File Path Property**

## **Business Requirement**

- Suppose the data movement using copy activity happens with multiple files from source to sink.
- As per requirement as an Azure Data Engineer i have been assigned task by a Product Manager to move multiple files with different pattern matching like (123cities.csv, data120345.csv and username.csv) from source to sink.
- To carry out this particular task List of Files option in copy activity comes to work to move multiple files with different pattern from source to sink. It is similar to wildcard path option. Here the only difference is that wildcard works to move multiple files with same pattern and List of File works to move multiple files with different pattern.
- Further step by step you will get solid understanding to hands on implement List of Files in the simplest way.

# Steps to Demonstrate List of Files in Copy Activity

## Step 1

Open Azure Portal go to ADLS storage to create a new container.

The screenshot shows the Microsoft Azure portal interface. The browser address bar displays the URL: `portal.azure.com/#@praveenpt580gmail.onmicrosoft.com/resource/subscriptions/b46092db-d6bf-4d8c-87a9-a8aadab998d7/resourceGroups/NewR...`. The page title is "microadls | Containers". The left sidebar shows the navigation menu with "Containers" selected. The main content area shows a table of containers:

Name	Last modified	Anonymous access level	Lease state
<input type="checkbox"/> slogs	12/20/2024, 12:24:30 AM	Private	Available
<input type="checkbox"/> log	1/9/2025, 1:40:03 PM	Private	Available
<input type="checkbox"/> target	1/10/2025, 9:36:31 PM	Container	Available

## Step 2

Go to container given under data storage section click on '+' container and create a new container to store multiple files of different pattern make it name input then click on create.

The screenshot shows the Microsoft Azure portal interface. On the right, a 'New container' dialog box is open. The 'Name' field is filled with 'input'. The 'Anonymous access level' is set to 'Private (no anonymous access)'. The 'Create' button is at the bottom. In the background, the 'Containers' page is visible, showing a table of existing containers:

Name	Last modified	Anonymous access
<input type="checkbox"/> Stogs	12/20/2024, 12:24:30 AM	Private
<input type="checkbox"/> log	1/9/2025, 1:40:03 PM	Private
<input type="checkbox"/> target	1/10/2025, 9:36:31 PM	Container

### Step 3

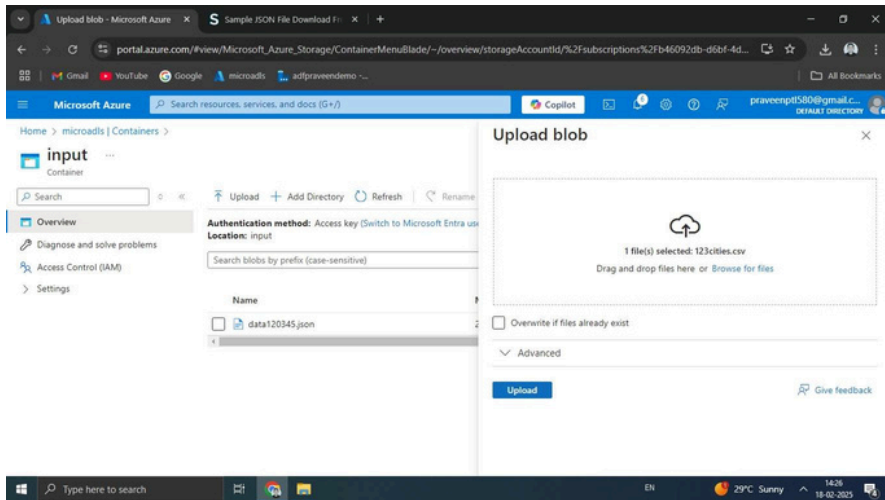
Again click on '+' container and create another new container make it name output to copy multiple files with different pattern from source (input container) then click on create.

The screenshot shows the Microsoft Azure portal interface. On the left, the 'microadls | Containers' section is active, displaying a list of containers. The main area shows the 'New container' dialog box with the name 'output' entered and the anonymous access level set to 'Private (no anonymous access)'. The 'Create' button is visible at the bottom of the dialog.

Name	Last modified	Anonymous access
<input type="checkbox"/> \$logs	12/20/2024, 12:24:30 AM	Private
<input type="checkbox"/> input	2/18/2025, 2:18:12 PM	Private
<input type="checkbox"/> log	1/9/2025, 1:40:03 PM	Private
<input type="checkbox"/> target	1/10/2025, 9:36:31 PM	Container

## Step 4

Download two .csv and one .json file from google in my computer rename it as 123cities,username for .csv file and data120345 for json file. Go to input container then click on upload and upload a new file 123cities.csv then click on upload.



## Step 5

Similarly upload data120345.json and username.csv file.

The screenshot shows the Microsoft Azure portal interface. The browser address bar displays the URL: `portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~ /overview/storageAccountId/%2Fsubscriptions%2Fb46092db-d6bf-4d...`. The page title is "input - Microsoft Azure". The left sidebar shows the "input" container selected, with options for "Overview", "Diagnose and solve problems", "Access Control (IAM)", and "Settings". The main content area shows the "input" container details, including the authentication method (Access key) and location (input). A search bar for blobs is present. Below the search bar, a table lists the uploaded files:

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> 123cities.csv	2/18/2025, 2:26:56 PM	Hot (Inferred)		Block blob	8.
<input type="checkbox"/> data120345.json	2/18/2025, 2:26:25 PM	Hot (Inferred)		Block blob	2f
<input type="checkbox"/> username.csv	2/18/2025, 2:27:34 PM	Hot (Inferred)		Block blob	1:

The Windows taskbar at the bottom shows the search bar, task view button, and system tray with the date and time: 14:27, 18-02-2025.

## Step 6

Keep output container remains empty we will copy all the files from input container here.

The screenshot shows the Microsoft Azure portal interface. The browser address bar displays the URL: `portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2Fb46092db-d6bf-4d...`. The page title is "output - Microsoft Azure". The left sidebar shows the navigation menu with "output" selected under "Containers". The main content area shows the "output" container overview. A notification box in the top right corner states: "Successfully deleted blob(s). Successfully deleted 2 blobs(s)." The "Authentication method" is set to "Access key (Switch to Microsoft Entra user account)". The "Location" is "output". There is a search bar for blobs and a table with columns: Name, Modified, Access tier, Archive status, Blob type, and Size. The table currently shows "No results". The Windows taskbar at the bottom shows the time as 14:28 on 18-02-2025.

Microsoft Azure

Home > microadls | Containers >

output  
Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

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

Location: output

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
No results					

Type here to search

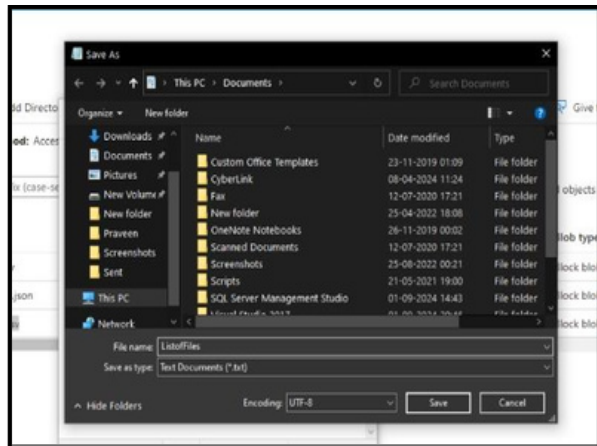
EN 29°C Sunny 14:28 18-02-2025



## Step 7

Open notepad and create a new metadata file name it as ListofFiles to store metadata 123cities.csv, data120345.json, username.csv. here.

123cities.csv  
data120345.json  
username.csv.



## Step 8

Go to input container and upload metadata text file ListofFiles.txt.

The screenshot shows the Microsoft Azure portal interface. The browser address bar displays the URL: `portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2Fb46092db-d6bf-4d...`. The page title is "input - Microsoft Azure". The left sidebar shows the navigation menu with "Overview" selected. The main content area shows the "input" container details. The "Authentication method" is "Access key (Switch to Microsoft Entra user account)". The "Location" is "input". There is a search bar with the placeholder text "Search blobs by prefix (case-sensitive)" and a toggle for "Show deleted objects". Below this is a table listing the blobs in the container.

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> 123cities.csv	2/18/2025, 2:26:56 PM	Hot (Inferred)		Block blob	8.
<input type="checkbox"/> data120345.json	2/18/2025, 2:26:25 PM	Hot (Inferred)		Block blob	21
<input type="checkbox"/> ListofFiles.txt	2/18/2025, 2:32:55 PM	Hot (Inferred)		Block blob	4.
<input type="checkbox"/> username.csv	2/18/2025, 2:27:34 PM	Hot (Inferred)		Block blob	1.

The Windows taskbar at the bottom shows the search bar with the text "Type here to search", the system tray with the date "14:33 18-02-2025" and weather "29°C Sunny".

## Step 9

Go back to Azure Portal and Open Azure Factory.

The screenshot shows the Microsoft Azure Portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and the user's profile. The main content area is titled "Azure services" and features a grid of service tiles. The "Data factories" tile is highlighted, showing a "+ Create" button and a "View" button. Below the "Azure services" section is the "Resources" section, which includes a table of recent resources.

Name	Type	Last Viewed
microadls	Storage account	21 minutes ago
adpraveendemo	Data factory (V2)	3 days ago
NewResourcePack	Resource group	3 days ago
Pay-As-You-Go	Subscription	3 days ago
prafactory	Data factory (V2)	3 days ago

The bottom of the screen shows the Windows taskbar with the search bar, task icons, and system tray information including the date and time.

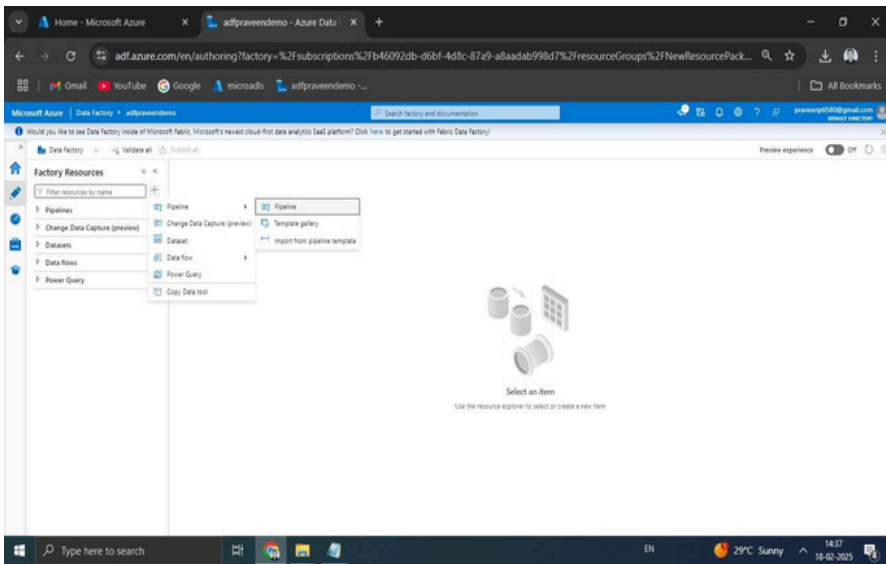
After Azure Data Factory is opened go to author tab to create a new pipeline.

The screenshot shows the Azure Data Factory 'Author' tab for a workspace named 'adfpraveendemo'. The interface includes a left-hand navigation pane with icons for Home, Author, Monitor, and Settings. The main content area features a 'New' button and four primary data processing activities: Ingest, Orchestrate, Transform data, and Configure SSIS. Below these, a 'Recent resources' section displays a table of recently used datasets and pipelines.

Name	Type	Last opened by you
DelimitedText2	Dataset	Last Sunday at 1:18 AM
DelimitedText1	Dataset	Last Sunday at 1:17 AM
pipelines	Pipeline	Last Saturday at 6:15 PM
AzureOperational	Dataset	Last Thursday at 2:29 PM
AzureOpTable1	Dataset	Last Thursday at 2:28 PM

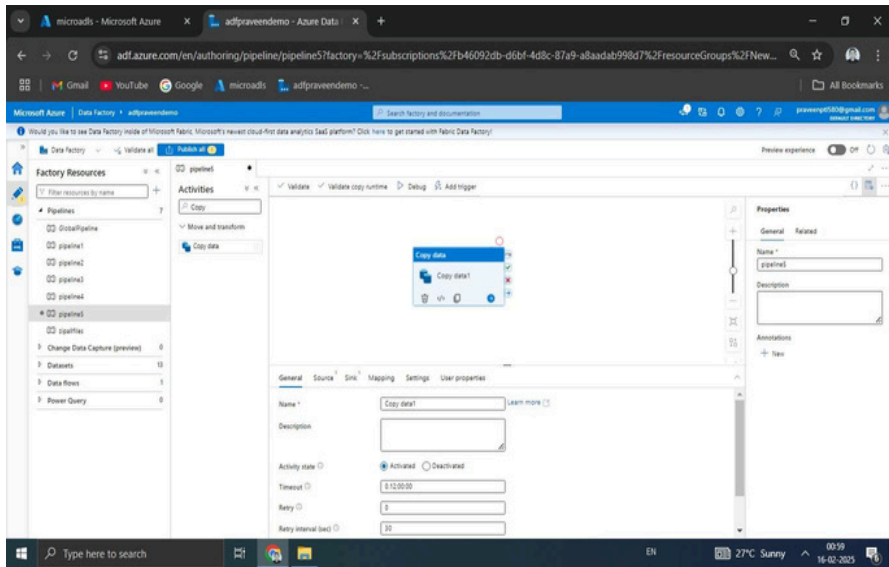
## Step 11

Click on '+' icon given near factory resources search bar and create a new pipeline.



## Step 12

**In activity search bar search copy activity and drag and drop it.**



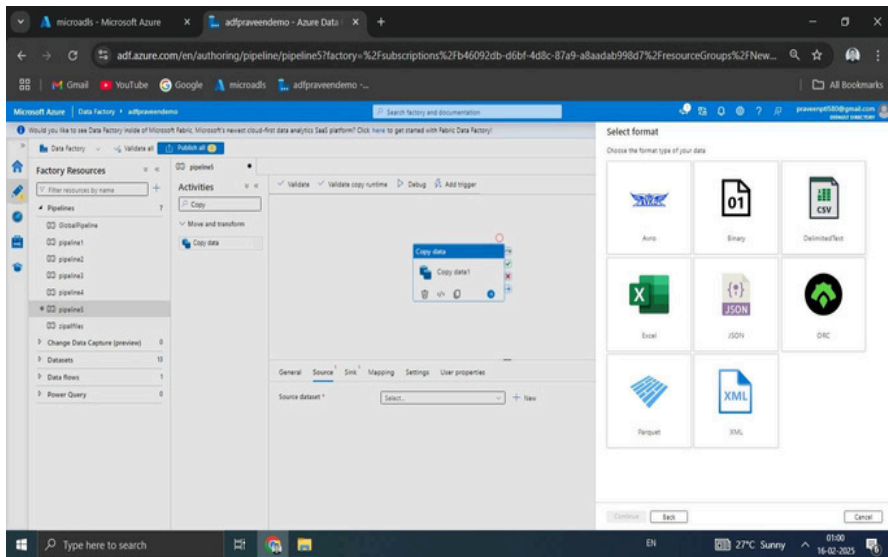
## Step 13

Go to Copy Activity source create a new source dataset for ADLS Gen2.

The screenshot displays the Microsoft Azure Data Factory portal interface. On the left, the 'Factory Resources' pane shows a list of pipelines, with 'pipeline5' selected. The main workspace shows a 'Copy' activity within a pipeline. A 'Copy data' dialog box is open, showing a 'Copy data1' activity. Below this, the 'Source dataset' dropdown is set to 'Select...'. On the right, the 'New dataset' dialog is open, showing the 'Azure Data Lake' tab. The 'Azure Data Lake Storage Gen2' option is selected. The bottom of the screen shows the Windows taskbar with the search bar, task icons, and system tray information including the date and time (01:00, 16-02-2025).

## Step 14

Select a new dataset format as Delimited Text and click on continue.





## Step 15

In set properties create a new linked service for ADLS Gen2 if you already created linked service then select it.

The screenshot displays the Microsoft Azure Data Factory portal interface. The browser address bar shows the URL: `adf.azure.com/en/authoring/pipeline/pipeline5?factory=%2Fsubscriptions%2Fb46092db-d6bf-4d8c-87a9-a8aadab998d7%2FresourceGroups%2FNew...`. The page title is "Microsoft Azure | Data Factory | adfpravedemo".

The main workspace shows a pipeline named "pipeline5" with a "Copy data" activity selected. The "Set properties" dialog is open on the right, showing the "Source" tab. The "Source dataset" dropdown is set to "Select...".

The "Set properties" dialog has the following fields:

- Name: DelimitedText9
- Linked service: Select... (Dropdown menu)
- + New: AzureDataLakeStorage1

The "Copy data" activity is shown in the center of the workspace, with a "Copy data1" button. The "Source" tab is selected, and the "Source dataset" dropdown is set to "Select...".

The Windows taskbar at the bottom shows the date and time as 01:01 on 16-02-2025, and the weather as 27°C Sunny.

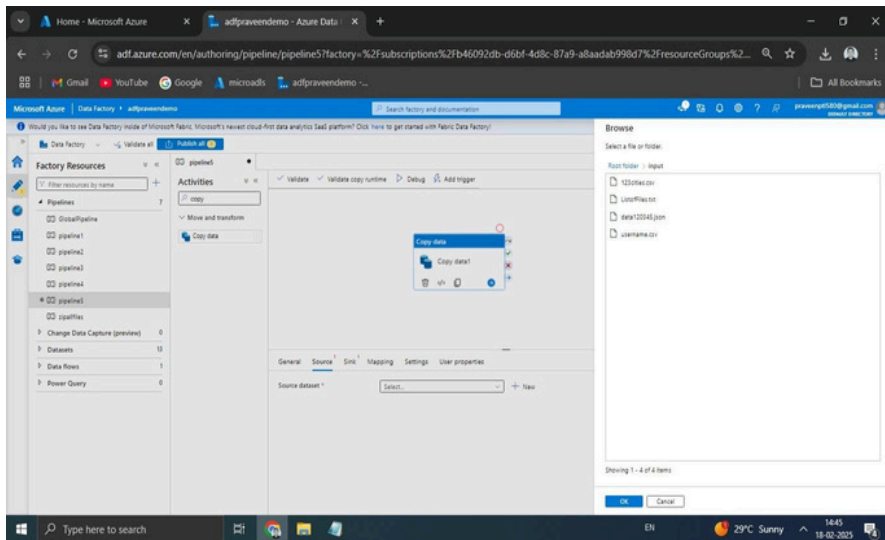
## Step 16

Now on browse button to browse input container.

The screenshot displays the Microsoft Azure Data Factory portal interface. The left sidebar shows the 'Factory Resources' tree with 'pipelines' expanded. The main workspace shows the 'Copy data' activity selected in the 'Activities' pane. A 'Copy data' dialog box is open, showing the activity name 'Copy data1'. The 'Set properties' pane on the right is active, showing the 'Source' tab. The 'Source dataset' dropdown is set to 'Select...'. The 'File path' section shows 'File system' selected for the file system, and 'Directory' for the directory. The 'First row as header' checkbox is checked. The 'Import schema' section shows 'From connection/store' selected. The 'Advanced' section is expanded. The bottom status bar shows the Windows taskbar with the search bar, taskbar icons, and system tray information including the date and time (14:43, 18-02-2025).

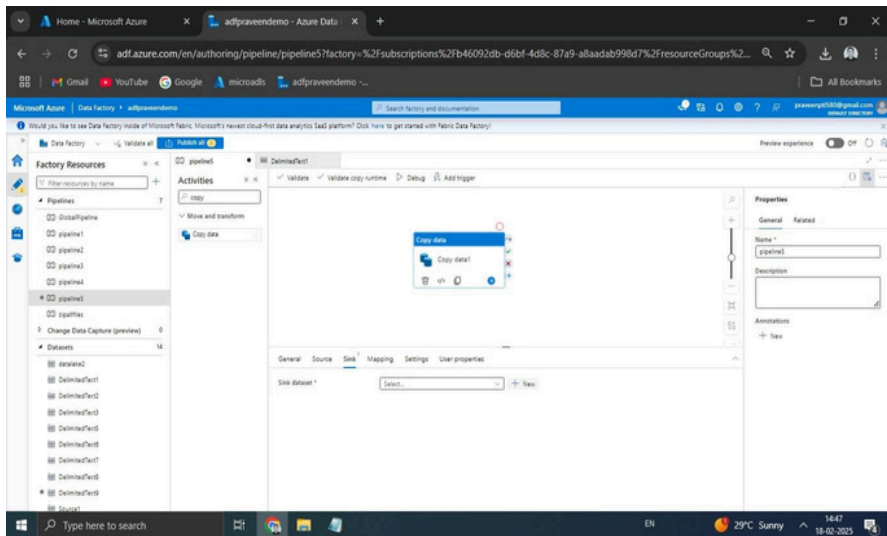
## Step 17

Go to input container don't select any file we have to copy all the files in output container then click on ok.



## Step 18

Similarly go to copy activity sink create a new sink dataset.



## Search and select data store ADLS Gen2 for new sink dataset then click on continue

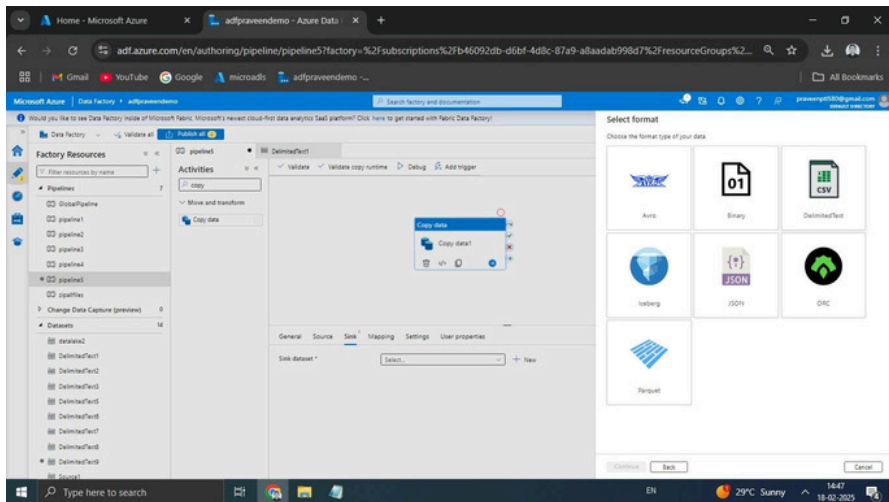
The screenshot shows the Microsoft Azure Data Factory portal interface. The browser address bar displays the URL: `adf.azure.com/en/authoring/pipeline/pipeline?factory=%2Fsubscriptions%2Fb46092db-d6bf-4d8c-87a9-a8aadab998d7%2FresourceGroups%2F...`. The page title is "Microsoft Azure | Data Factory | adfpravedemo".

The left sidebar shows the "Factory Resources" tree with "Pipelines" and "Datasets" expanded. Under "Pipelines", "GlobalPipeline" is selected. Under "Datasets", "DelimitedText1" is selected.

The main area displays the "Activities" list for "DelimitedText1", showing a "Copy" activity. The "Copy" activity is expanded, showing the "Copy data" task. The "Sink" tab is selected, and the "Sink dataset" dropdown is set to "Select...".

The right sidebar shows the "New dataset" wizard. The "Select a data store" dropdown is set to "Azure Data Lake". The "All" tab is selected, and "Azure Data Lake Storage Gen2" is chosen as the data store. The "Continue" button is visible at the bottom of the wizard.

In dataset format select any format like that i select delimited text then click on continue.



## Step 21

Click on browse button to browse output container.

The screenshot displays the Microsoft Azure Data Factory portal interface. On the left, the 'Factory Resources' pane shows a list of pipelines, with 'pipeline5' selected. The 'Activities' pane for 'pipeline5' shows a 'Copy data' activity. A context menu is open over the 'Copy data' activity, with the 'Copy data' option highlighted. The main canvas shows the 'Copy data' activity configuration. The 'Sink' tab is active, showing the 'Sink dataset' dropdown set to 'Select...'. On the right, the 'Set properties' pane is open, showing the 'Name' field set to 'DelimitedText10', the 'Linked service' set to 'AzureDataLakeStorage1', and the 'File path' field with 'File system', 'Directory', and 'File name' sub-fields. The 'First row as header' checkbox is checked. The 'Import schema' section has 'None' selected. The 'Advanced' section is collapsed. At the bottom, the Windows taskbar shows the search bar, task icons, and system tray with the date '18-02-2025' and time '14:48'.

Go to output container then click on ok.

The screenshot displays the Microsoft Azure Data Factory portal interface. The browser address bar shows the URL: `adf.azure.com/en/authoring/pipeline/pipeline5?factory=%2Fsubscriptions%2Fb46092db-d6bf-4d8c-87a9-a8aadab998d7%2FresourceGroups%2F...`. The left sidebar shows the 'Factory Resources' tree with 'pipelines' expanded, listing 'pipeline1' through 'pipeline5'. The main workspace shows the 'Copy data' activity configuration for 'DelimitedText'. The 'Sink' tab is active, showing a 'Sink dataset' dropdown menu. A 'Browse' dialog box is open on the right, prompting the user to 'Select a file or folder' and showing the 'Root folder' as 'output'. The dialog has 'OK' and 'Cancel' buttons at the bottom. The Windows taskbar at the bottom shows the search bar, task view icon, and system tray with the date '18-02-2025' and time '14:49'.



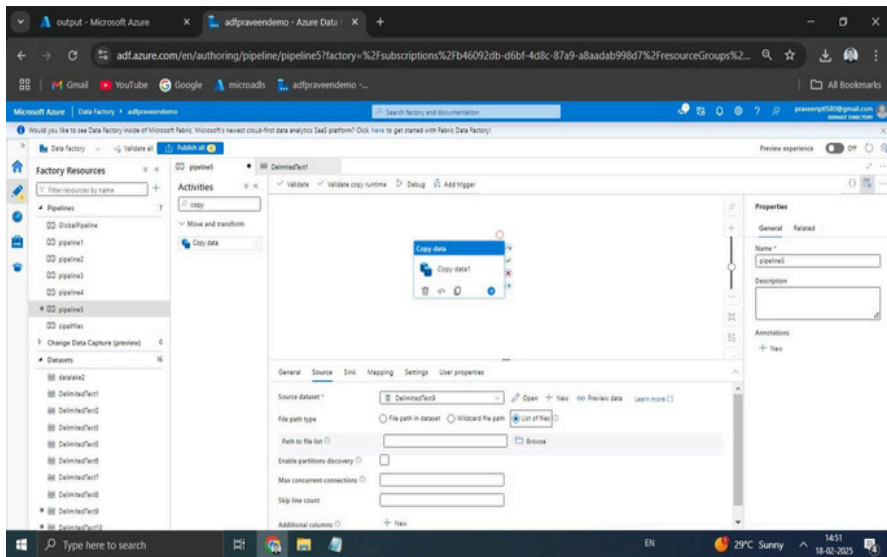
## Step 23

Go back to copy activity source select File path type as List of files.

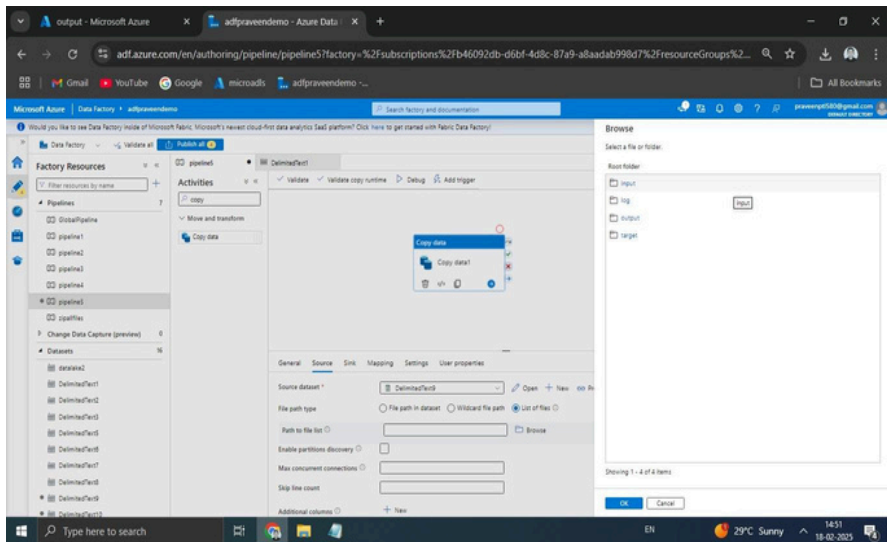
The screenshot displays the Microsoft Azure Data Factory portal interface. On the left, the 'Factory Resources' pane shows a tree view with 'Pipelines' expanded, listing 'pipeline1' through 'pipeline5'. 'pipeline5' is selected. The main workspace shows the 'Copy data' activity within 'pipeline5'. A context menu is open over the activity, with 'Copy data?' selected. Below the activity, the 'Properties' pane is visible, showing the 'General' tab. The 'Source dataset' is 'DelimitedText'. Under 'File path type', the 'List of files' option is selected. The 'Path to file list' field is empty, and a 'Browse' button is next to it. Other settings like 'Enable partitions discovery', 'Max concurrent connections', and 'Skip line count' are also visible. The bottom of the screen shows the Windows taskbar with the search bar and system tray.

## Step 24

In Path to file list click on browse button.



Go to input container to select metadata text file ListofFiles.txt to get file metadata.



## Step 26

Select metadata file ListofFiles.csv then click on ok.

The screenshot displays the Microsoft Azure Data Factory portal interface. The left sidebar shows the 'Factory Resources' tree with 'pipelines' expanded. The main canvas shows a 'Copy data' activity being configured. A 'Browse' dialog box is open on the right, showing the file selection process. The 'Source' tab is active in the configuration pane, showing the 'Source dataset' as 'DelimitedText1'. The 'File path type' is set to 'List of files'. The 'Browse' button is highlighted, and the 'List of files' option is selected. The 'Browse' dialog shows a list of files, including 'ListofFiles.csv', which is highlighted. The 'OK' button is visible at the bottom of the dialog.

Microsoft Azure | Data Factory | adfpravedemo

Would you like to see Data Factory index of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click here to get started with Fabric Data Factory!

Factory Resources

- Pipelines
- GlobalPipeline
- pipeline1
- pipeline2
- pipeline3
- pipeline4
- pipeline5
- zipfiles
- Change Data Capture (preview)
- Datasets
- dataset2
- DelimitedText1
- DelimitedText2
- DelimitedText3
- DelimitedText4
- DelimitedText5
- DelimitedText6
- DelimitedText7
- DelimitedText8
- DelimitedText9
- DelimitedText10

Activities

- copy
- Move and transform
- Copy data

Copy data

Copy data1

General Source Sink Mapping Settings User properties

Source dataset \* DelimitedText1

File path type File path in dataset Wildcard file path List of files

Path to file list

Enable partitions discovery

Max concurrent connections

Skip line count

Additional columns

Browse

Select a file or folder.

Root folder > Input

- 123files.csv
- ListofFiles.txt
- data123143.json
- username.csv

Showing 1 - 4 of 4 items

OK Cancel

## Step 27

Now execute pipeline using debug mode and wait for five to ten seconds until pipeline executes successfully.

The screenshot displays the Microsoft Azure Data Factory portal interface. The browser address bar shows the URL: `adf.azure.com/en/authoring/pipeline/pipeline5?factory=%2Fsubscriptions%2Fb46092db-d6bf-4d8c-87a9-a8aadab998d7%2FresourceGroups%2F...`. The left sidebar shows the 'Factory Resources' tree with 'Pipelines' expanded, listing 'GlobalPipeline', 'pipeline1', 'pipeline2', 'pipeline3', 'pipeline4', 'pipeline5' (selected), and 'pipeline6'. The main canvas shows the 'Copy data' activity in the 'Copy data' pipeline. The 'Properties' pane on the right shows the 'General' tab with 'Name' set to 'pipeline5'. The bottom section shows the 'Pipeline run' details for 'Copy data', indicating it is 'In progress' with a status of 'Queued'. The 'Output' tab is active, showing a table with columns: 'Activity name', 'Activity status', 'Activity', 'Run start', 'Duration', and 'Integration runtime'. The table contains one row: 'Copy data1', 'Queued', 'Copy data', '2/18/2023 2:54:23 PM', and '3s'. The Windows taskbar at the bottom shows the system clock as 14:54 on 18-02-2023.

## Step 28

Once pipeline executes successfully go back to ADLS storage go to output container click on refresh you see all the files with different pattern (123cities.csv, data120345.csv and username.csv) stored simultaneously using List of files in copy activity.

The screenshot displays the Microsoft Azure portal interface. The browser address bar shows the URL: `portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2Fb46092db-d6bf-4d...`. The page title is "output - Microsoft Azure". The left sidebar shows the navigation menu with "Home > microads | Containers > output" selected. The main content area shows the "output" container overview. The "Authentication method" is "Access key (Switch to Microsoft Entra user account)" and the "Location" is "output". The "Search blobs by prefix (case-sensitive)" field is empty. The "Show deleted objects" toggle is turned off. A table lists the files in the container:

	Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/>	123cities.csv	2/18/2025, 2:54:36 PM	Hot (Inferred)		Block blob	8...
<input type="checkbox"/>	data120345.json	2/18/2025, 2:54:36 PM	Hot (Inferred)		Block blob	21...
<input type="checkbox"/>	username.csv	2/18/2025, 2:54:36 PM	Hot (Inferred)		Block blob	11...

The Windows taskbar at the bottom shows the search bar, task view button, and system tray with the date "18-02-2025" and time "14:55".