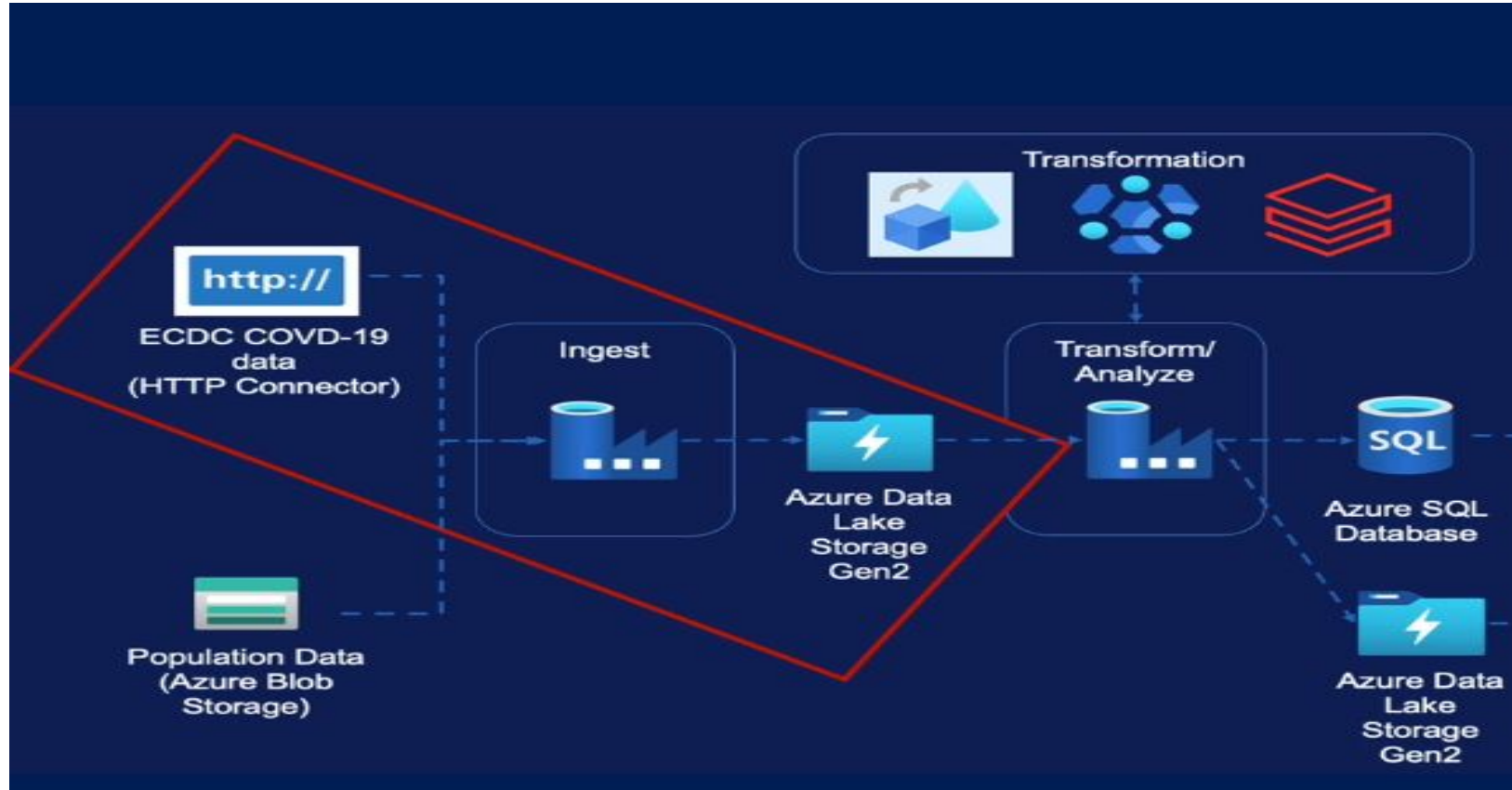


# File ingestion from HTTP to adls gen2 (multiple files)



Create HTTP Link Service and ADLS Gen2 (we already know how to create DL LS)  
Here we only focus on how to create HTTP LS)

The screenshot displays the Azure Data Factory interface. On the left, a navigation pane shows various settings categories: General, Connections, Source control, Author, and Security. The 'Connections' section is expanded, and 'Linked services' is selected. The main workspace shows the 'Linked services' page with a '+ New' button and a search bar. A modal dialog titled 'New linked service' is open on the right. It has two tabs: 'Data store' (selected) and 'Compute'. Below the tabs is a search bar containing 'Http'. A horizontal filter bar shows categories: 'All' (selected), 'Azure', 'Database', 'File', and 'Generic protocol'. In the center, a blue globe icon with 'HTTP' text is highlighted within a selection box. Below the icon, the text 'HTTP' is displayed. At the bottom of the dialog, there are 'Continue' and 'Cancel' buttons.

» Data Factory Validate all Publish all

Home Edit Monitor Tools Settings

General

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Security

Linked services

Linked service defines the connection information to a data store or compute.

+ New

Filter by name Annotations : Any

New linked service

Data store Compute

Http

All Azure Database File Generic protocol

HTTP

HTTP

Continue Cancel

## New linked service



HTTP [Learn more](#)

Name \*

Is\_http

Description

Connect via integration runtime \* ⓘ



AutoResolveIntegrationRuntime

Base URL \*

https://raw.githubusercontent.com



Information will be sent to the URL specified. Please ensure you trust the URL entered.

Server certificate validation ⓘ



Enable



Disable

Authentication type \* ⓘ

Anonymous



Connection successful



Test connection

Create

Back

Cancel

## Provide

1. Name,
2. Base URL,
3. Authentication type

# HTTP Source dataset creation:

## We can provide Relative URL Dynamically using lookup activity

Data Factory

Validate all

Publish all 1

Factory Resources

Filter resources by name

Pipelines0

Change Data Capture (preview)0

Datasets0

Data flows0

Power Query0

Use th

Set properties

Name

ds\_http\_source

Linked service \*

ls\_http

Relative URL

First row as header

☒

Import schema

☐ From connection/store

☐ From sample file

☒ None

OK

Back

Cancel

ds\_http\_source

ds\_dl\_sink



DelimitedText  
ds\_http\_source

Create parameter for Relative URL and add  
Dynamic content to it (@dataset().RelativeURL

Connection

Schema

Parameters

Linked service \*

Is\_http



Test connection



Edit



New

Base URL

https://raw.githubusercontent.com

Relative URL ⓘ

@dataset().RelativeURL



Preview data



Detect format

Compression type

No compression

Column delimiter ⓘ

Comma (,)

Row delimiter ⓘ

Default (\r,\n, or \r\n)

# Target DataLake dataset creation:

## We can provide Directory and Filename Dynamically using lookup activity

The screenshot shows the 'Set properties' dialog for a dataset named 'ds\_dl\_sink' in a Data Factory environment. The left sidebar shows the 'Factory Resources' tree with 'Datasets' selected, showing one dataset: 'ds\_http\_source'. The main area displays the 'ds\_http\_source' dataset with a 'DelimitedText' icon and the name 'ds\_http\_source'. The 'Set properties' panel on the right contains the following fields and options:

- Name:** ds\_dl\_sink
- Linked service \*:** ls\_adls
- File path:** raw / Directory / File name
- First row as header:** ☒
- Import schema:** ☐ From connection/store ☐ From sample file ☒ None

At the bottom, there are buttons for 'OK', 'Back', and 'Cancel'.

ds\_http\_source

ds\_dl\_sink



DelimitedText  
ds\_dl\_sink

Create parameter for Directory and Filename add Dynamic content to it

1. Directory (@dataset().Directory)
2. Filename (@dataset().Filename)

Connection Schema Parameters

Linked service \*

ls\_adls



Test connection



Edit



New

[Learn more](#)

File path

raw

/ @dataset().Directory

/ @dataset().Filename

Compression type

No compression

Column delimiter ⓘ

Comma (,)

Row delimiter ⓘ

Default (\r,\n, or \r\n)

Encoding ⓘ

Default(UTF-8)

Quote character ⓘ

Double quote (")


# 1. Upload JSON “http\_file\_list” file to adls gen2 and bring this into ADF dataset.

```
[
{
  "sourceBaseURL": "https://raw.githubusercontent.com",
  "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/cases_deaths.csv",
  "FolderName": "cases_deaths",
  "sinkFileName": "cases_deaths.csv"
},
{
  "sourceBaseURL": "https://raw.githubusercontent.com",
  "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/hospital_admissions.csv",
  "FolderName": "hospital_admissions",
  "sinkFileName": "hospital_admissions.csv"
},
{
  "sourceBaseURL": "https://raw.githubusercontent.com",
  "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/testing.csv",
  "FolderName": "testing",
  "sinkFileName": "testing.csv"
},
{
  "sourceBaseURL": "https://raw.githubusercontent.com",
  "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/country_response.csv",
  "FolderName": "country_response",
  "sinkFileName": "country_response.csv"
}
]
```



# Output of Lookup Activity

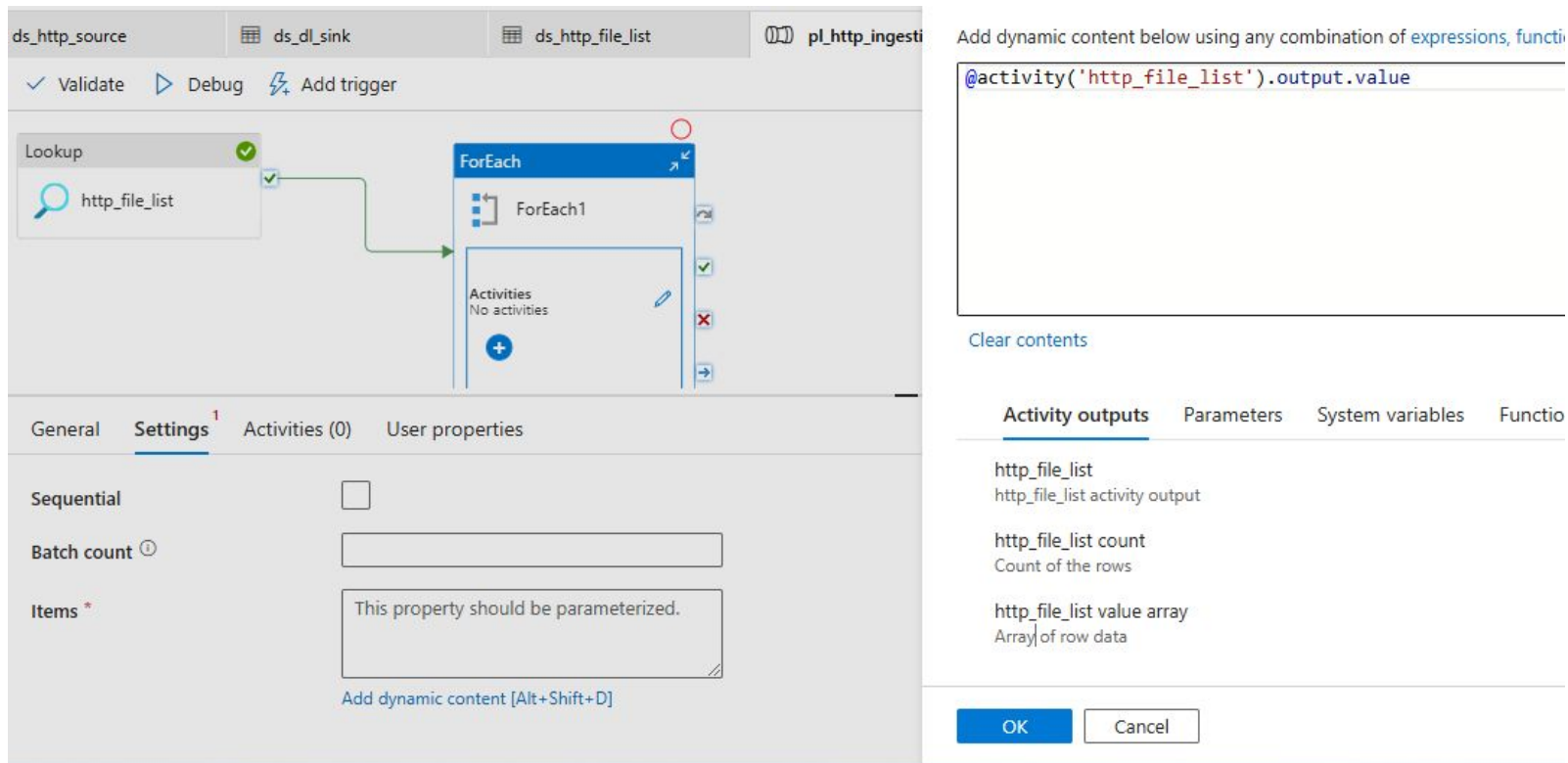
## Output

 Copy to clipboard

```
{
  "count": 4,
  "value": [
    {
      "sourceBaseURL": " https://raw.githubusercontent.com ",
      "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/cases_deaths.csv",
      "FolderName": "cases_deaths",
      "sinkFileName": "cases_deaths.csv"
    },
    {
      "sourceBaseURL": " https://raw.githubusercontent.com ",
      "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/hospital_admissions.csv",
      "FolderName": "hospital_admissions",
      "sinkFileName": "hospital_admissions.csv"
    },
    {
      "sourceBaseURL": " https://raw.githubusercontent.com ",
      "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/testing.csv",
      "FolderName": "testing",
      "sinkFileName": "testing.csv"
    },
    {
      "sourceBaseURL": " https://raw.githubusercontent.com ",
      "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/country_response.csv",
      "FolderName": "country_response",
      "sinkFileName": "country_response.csv"
    }
  ]
}
```

## Add ForEach Activity

1. Goto Setting - Items - Add dynamic content -http\_file\_list value array (i.e @activity('http\_file\_list').output.value)



The screenshot shows the Power Automate interface with the following components:

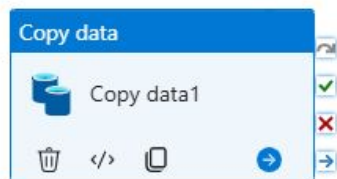
- Top Bar:** ds\_http\_source, ds\_dl\_sink, ds\_http\_file\_list, pl\_http\_ingesti.
- Buttons:** Validate, Debug, Add trigger.
- Canvas:** A 'Lookup' activity with 'http\_file\_list' is connected to a 'ForEach' activity. The 'ForEach' activity has a sub-activity 'ForEach1' and a section 'Activities' with 'No activities'.
- Settings Panel:** The 'Settings' tab is selected. It shows 'Sequential' (unchecked), 'Batch count' (empty), and 'Items' (with a message: 'This property should be parameterized.'). Below the 'Items' field is a link: 'Add dynamic content [Alt+Shift+D]'.
- Dynamic Content Panel:** A text box contains the expression: `@activity('http_file_list').output.value`. Below it is a 'Clear contents' link.
- Activity Outputs Panel:** A list of activity outputs for 'http\_file\_list':
  - http\_file\_list: http\_file\_list activity output
  - http\_file\_list count: Count of the rows
  - http\_file\_list value array: Array of row data
- Buttons:** OK, Cancel.

## Goto ForEach - Activities - add Copy Activity

dl\_sink ds\_http\_file\_list pl\_http\_ingestion\_...

✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

pl\_http\_ingestion\_adls > ForEach1



General Source<sup>1</sup> Sink<sup>1</sup> Mapping Settings User properties

Name \*  [Learn more](#)

Description

Activity state ⓘ ☒ Activated ☐ Deactivated

## Click on Copy Activity

1. Source - provide Dynamic content to Value (@item().sourceRelativeURL

✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

pl\_http\_ingestion\_adls > ForEach1

Copy data

Copy data1

⌵ </> 📄 ➡

General **Source** Sink Mapping Settings User properties

Dataset properties

Name	Value
RelativeURL	@item().sourceRelativeURL

Request method \* ⓘ GET ▼

Additional headers ⓘ

Click on Copy Activity

2. Sink - provide Dynamic content to Value  
Directory (@item().FolderName)  
Filename (@item().sinkFileName)




✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

pl\_http\_ingestion\_adls > ForEach1

**Copy data**

Copy data1

General Source **Sink** Mapping Settings User properties

Sink dataset \* ds\_dl\_sink  Open  New [Learn more](#) 

Dataset properties ⓘ

Name	Value
Directory	@item().FolderName
Filename	@item().sinkFileName

Copy behavior ⓘ

Select...

Here is the final Pipeline

✓ Validate   ▶ Debug   ⚡ Add trigger

Lookup ✓  
http\_file\_list

✓ +

ForEach ✓

ForEach1

Activities

Copy file → +

Parameters   Variables   Settings   **Output**

Showing 1 - 6 of 6 items

Activity name	↑↓	Activity status	↑↓	Activity type	↑↓	
Copy file		✓ Succeeded		Copy data		-
Copy file		✓ Succeeded		Copy data		-
Copy file		✓ Succeeded		Copy data		-
Copy file		✓ Succeeded		Copy data		-
ForEach1		✓ Succeeded		ForEach		-
http_file_list		✓ Succeeded		Lookup		-

Home > gpottabcovid19adits\_1/3338601/343 | Overview > gpottabcovid19adits | Containers >



Search

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

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: raw

Search blobs by prefix (case-sensitive)

Show deleted objects

	Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/>	 cases_deaths	12/5/2024, 8:26:46 PM					- ***
<input type="checkbox"/>	 country_response	12/5/2024, 8:26:44 PM					- ***
<input type="checkbox"/>	 hospital_admissions	12/5/2024, 8:26:45 PM					- ***
<input type="checkbox"/>	 testing	12/5/2024, 8:26:44 PM					- ***