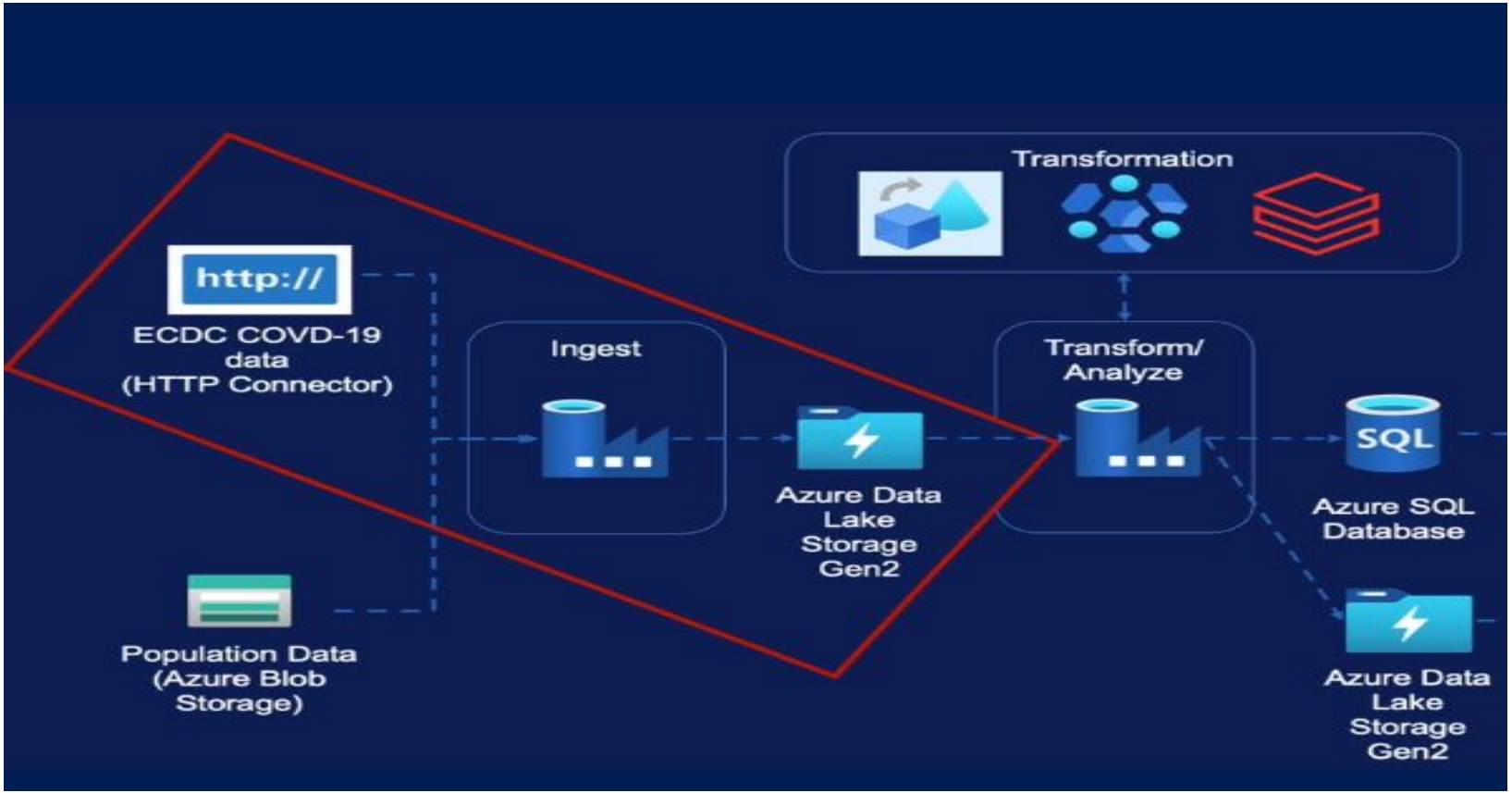
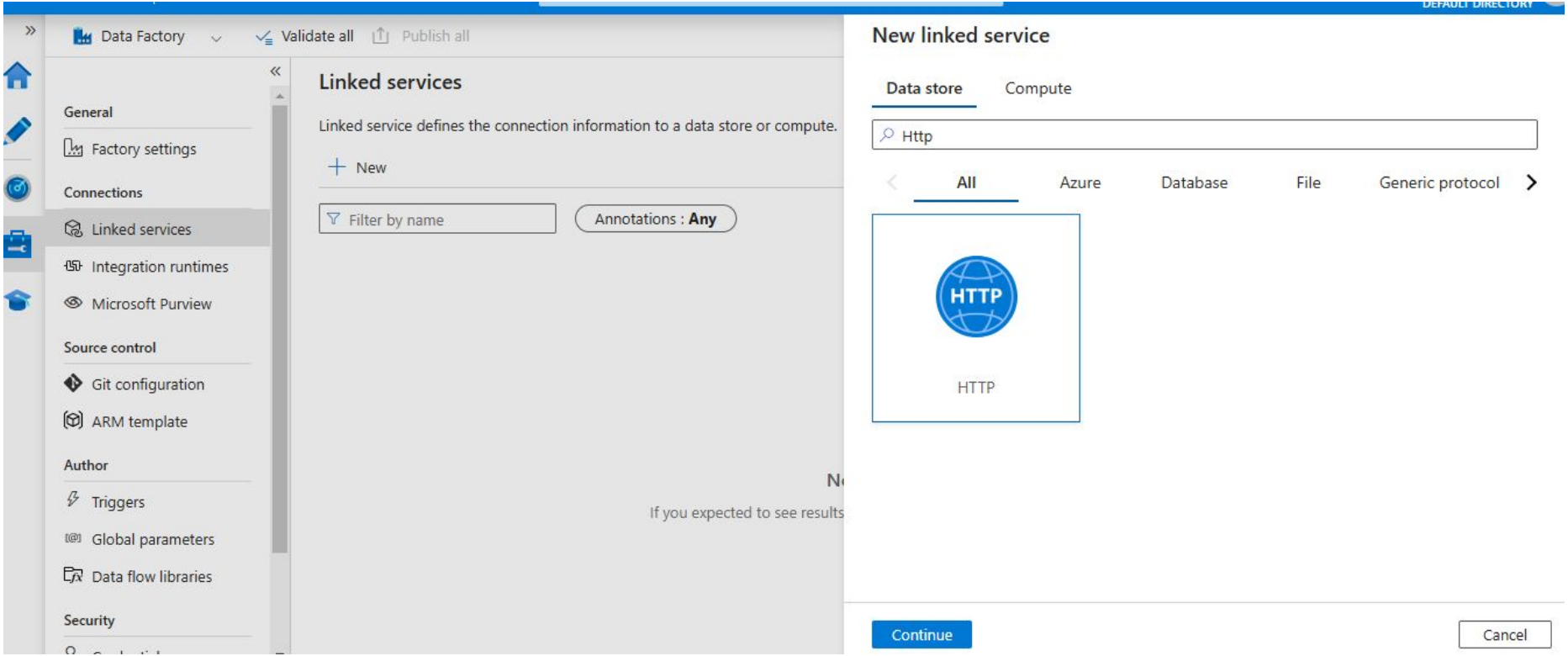
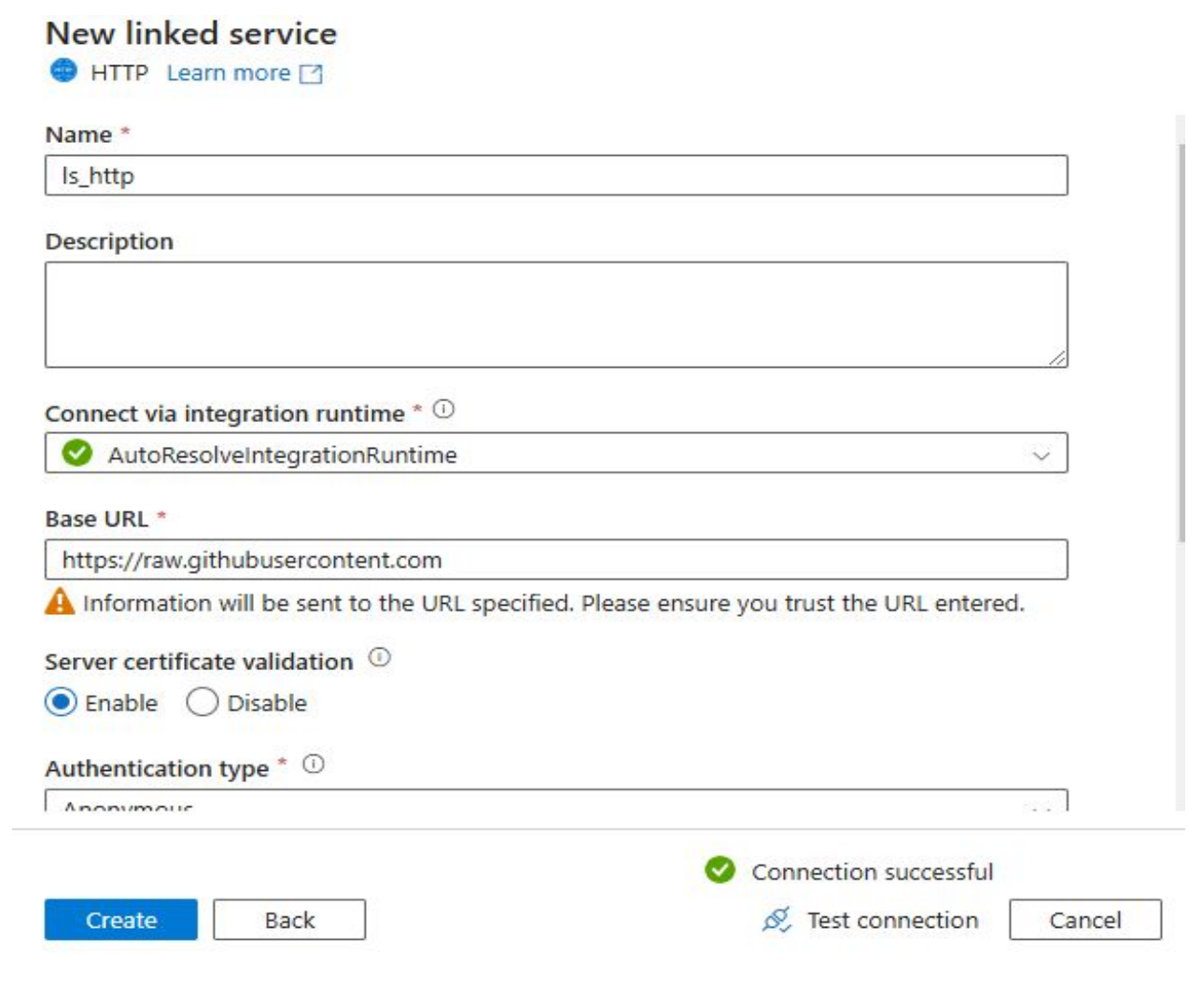
# File ingestion from HTTP to adls gen2 (multiple files)



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)

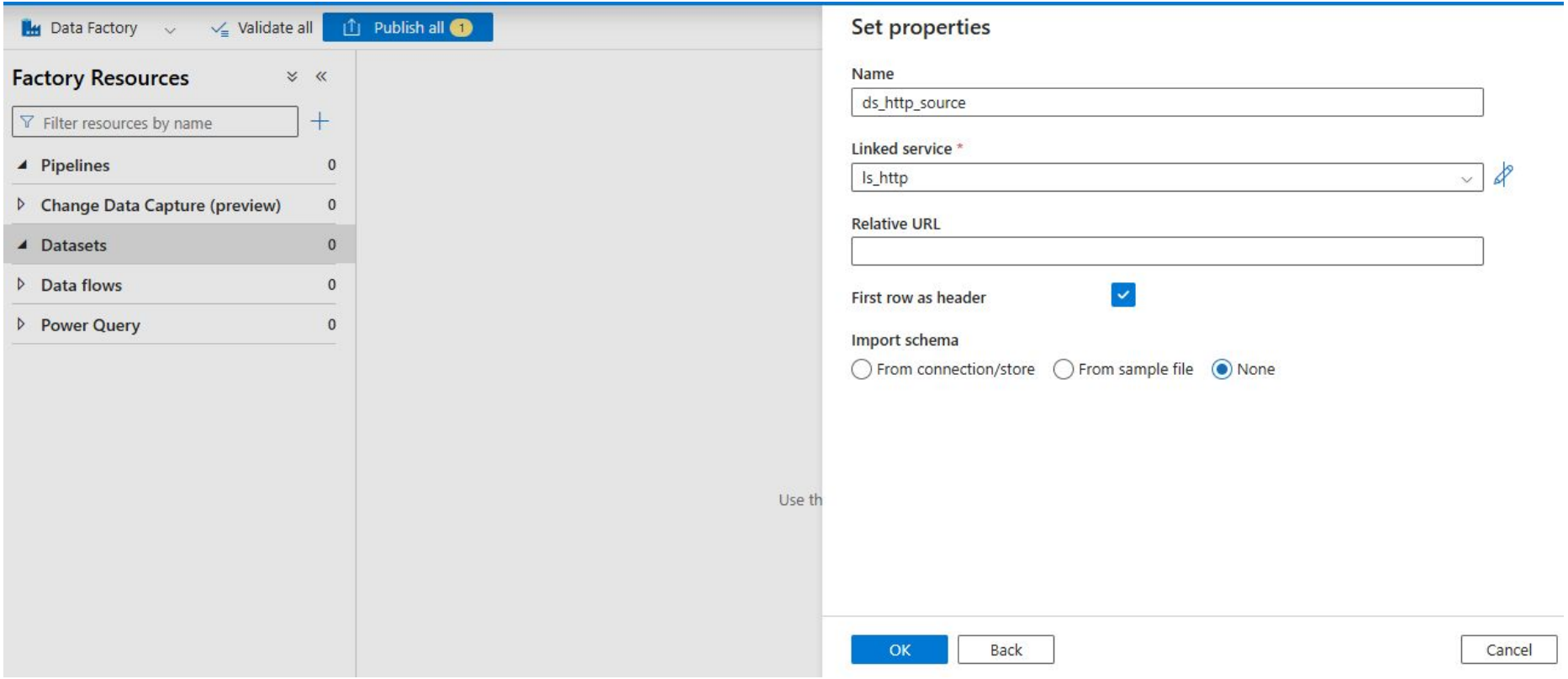





**Provide**

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

# HTTP Source dataset creation:
# We can provide Relative URL Dynamically using lookup activity



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 * ls_http | Test connection | Edit | New

Base URL https://raw.githubusercontent.com

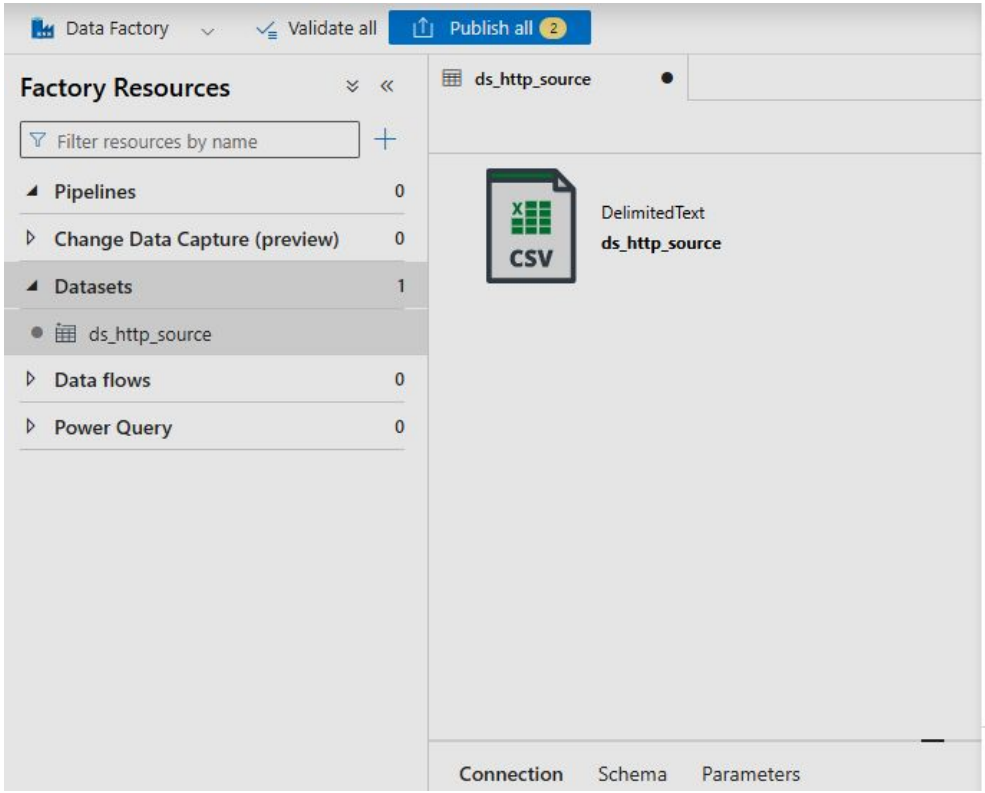
Relative URL @dataset().RelativeURL | Preview data | Detect format

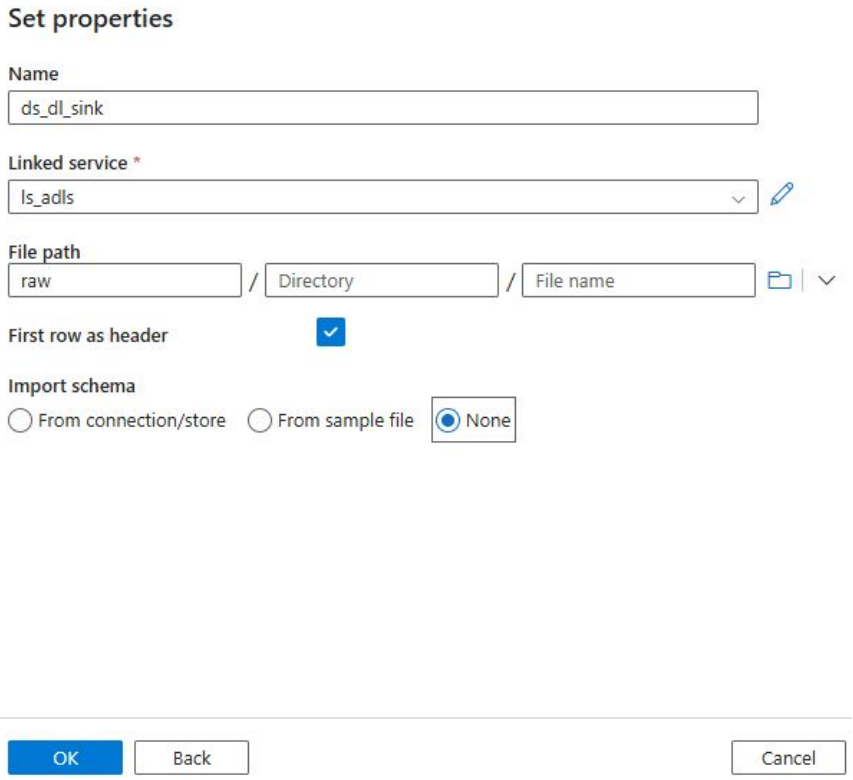
Compression type No compression

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



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

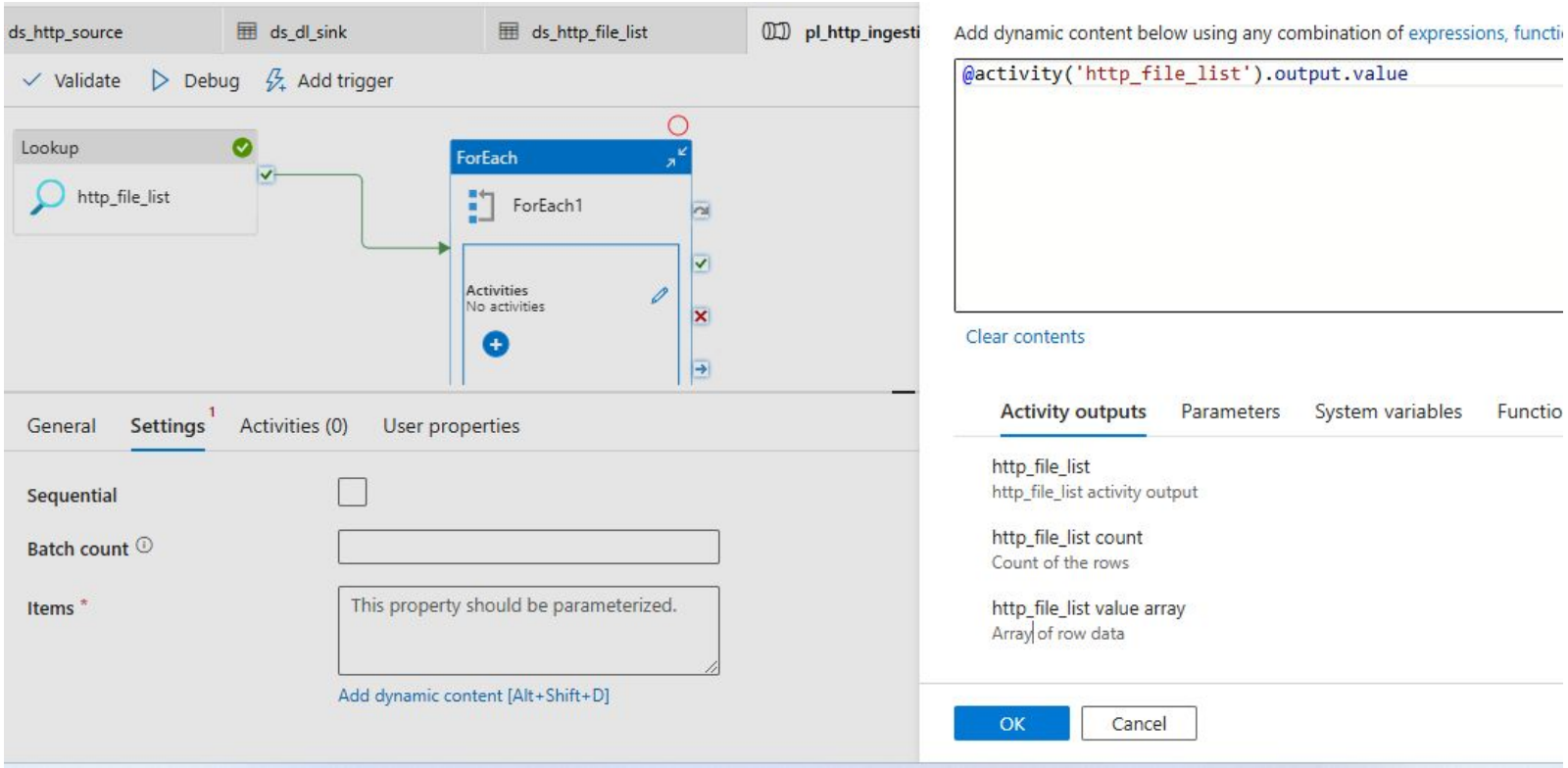
**Output**

Copy to clipboard

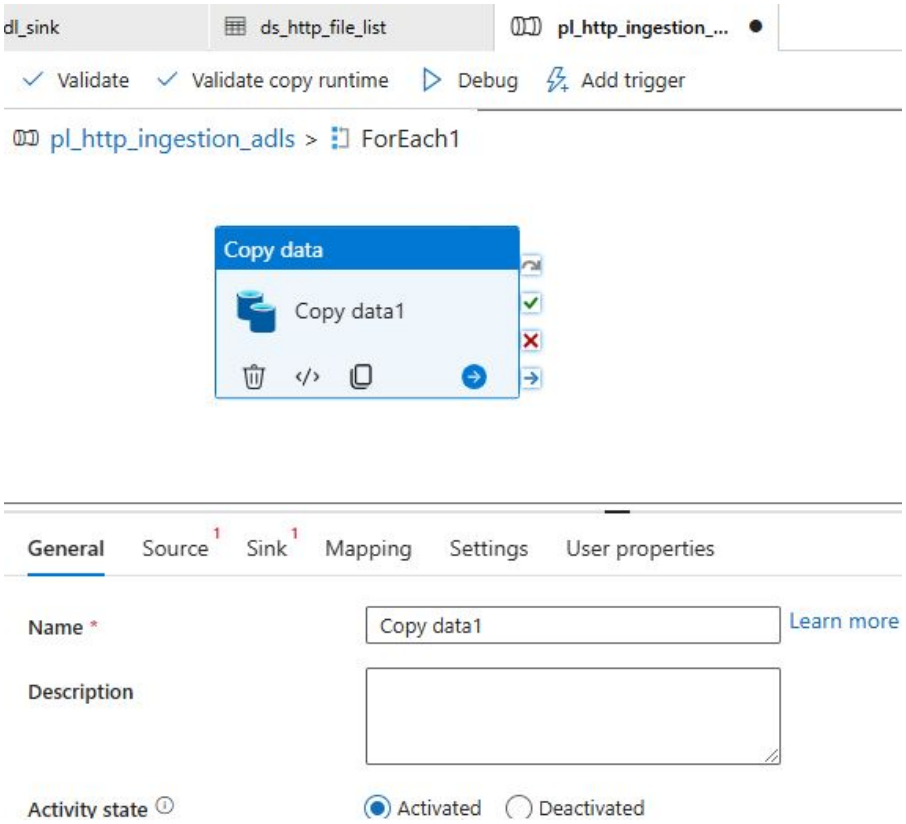

```
{
    "count": 4,
    "value": [
        {
            "sourceBaseURL": " https://raw.githubusercontent.com ",
            "sourceRelativeURL": "Gangadhar8605/Azure-Data-Engineer/refs/heads/main/covid19/cases_deaths.csv",
            "FolderName": "cases_deaths",
            "sinkFileName": "cases_deaths.csv"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

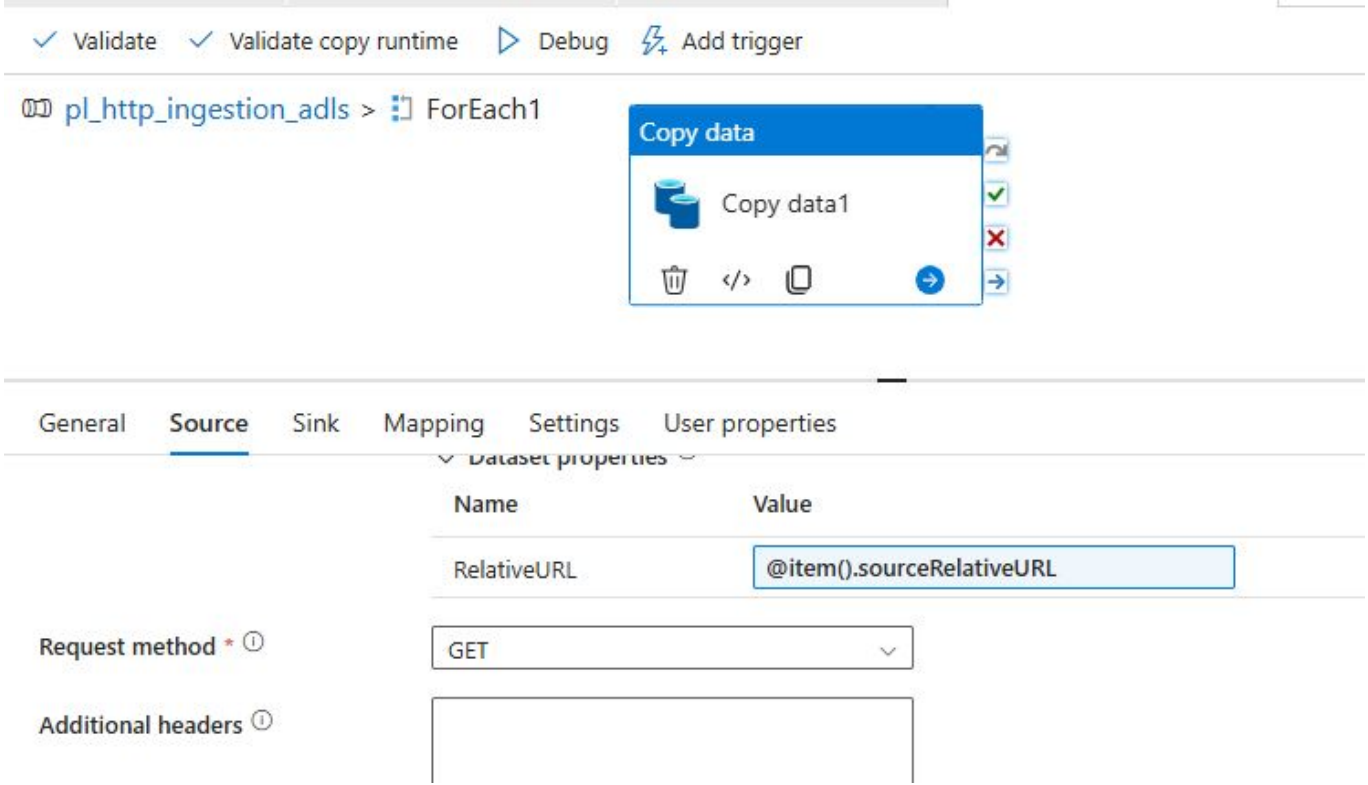


# Goto ForEach - Activities - add Copy Activity



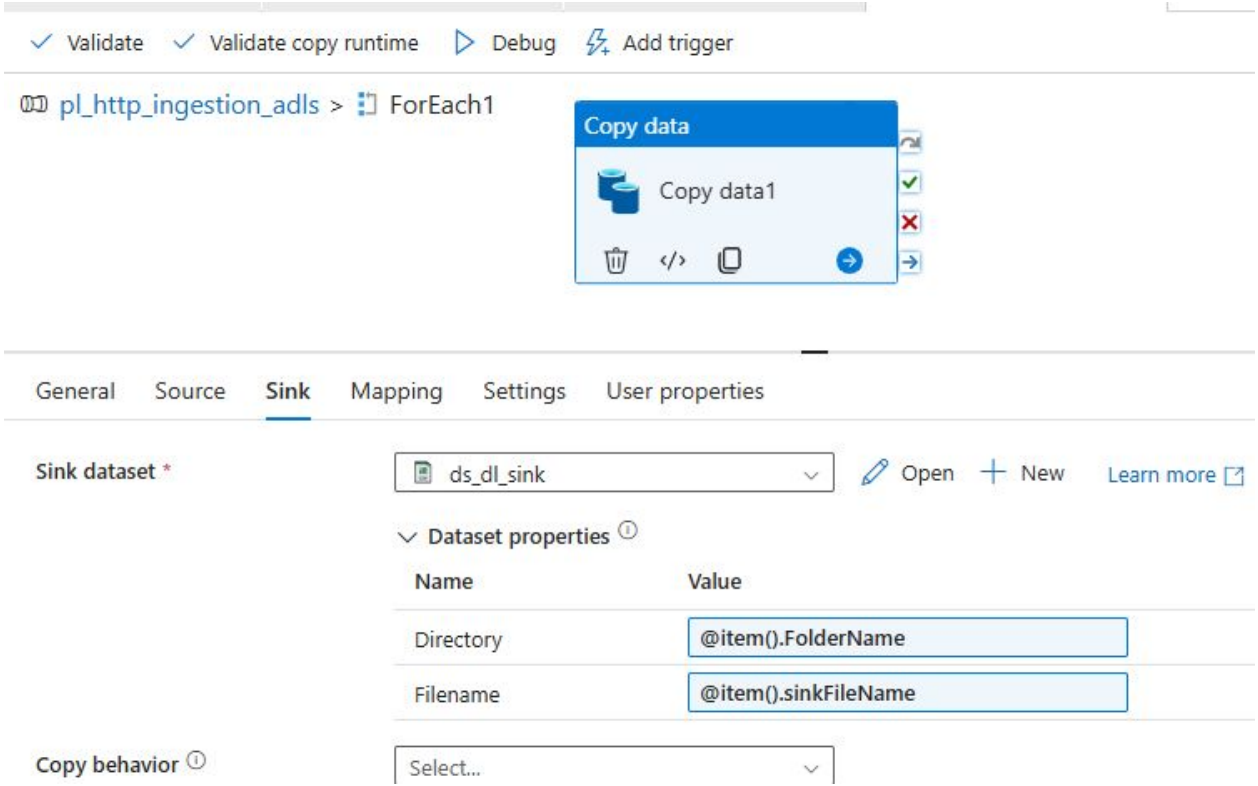
Click on Copy Activity

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

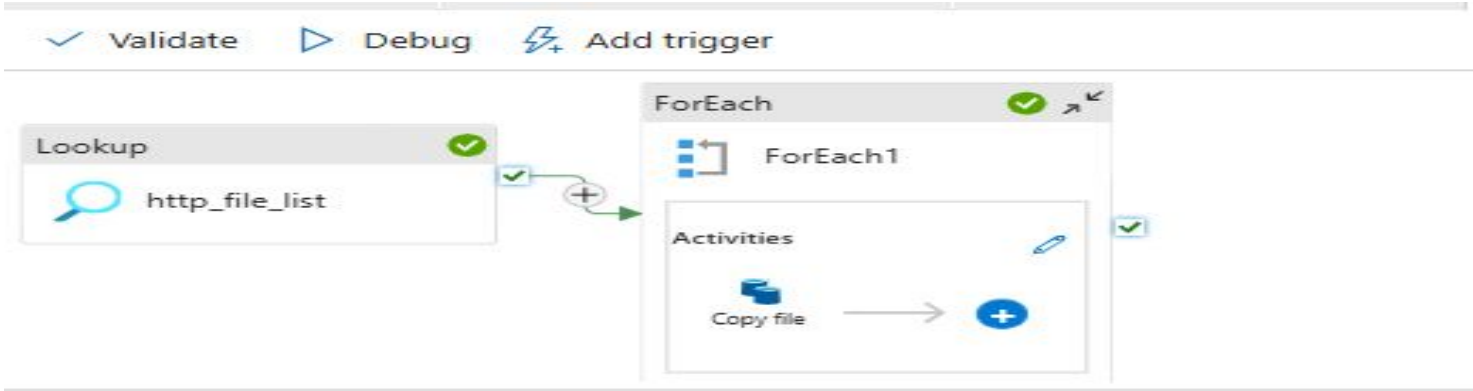


Click on Copy Activity

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



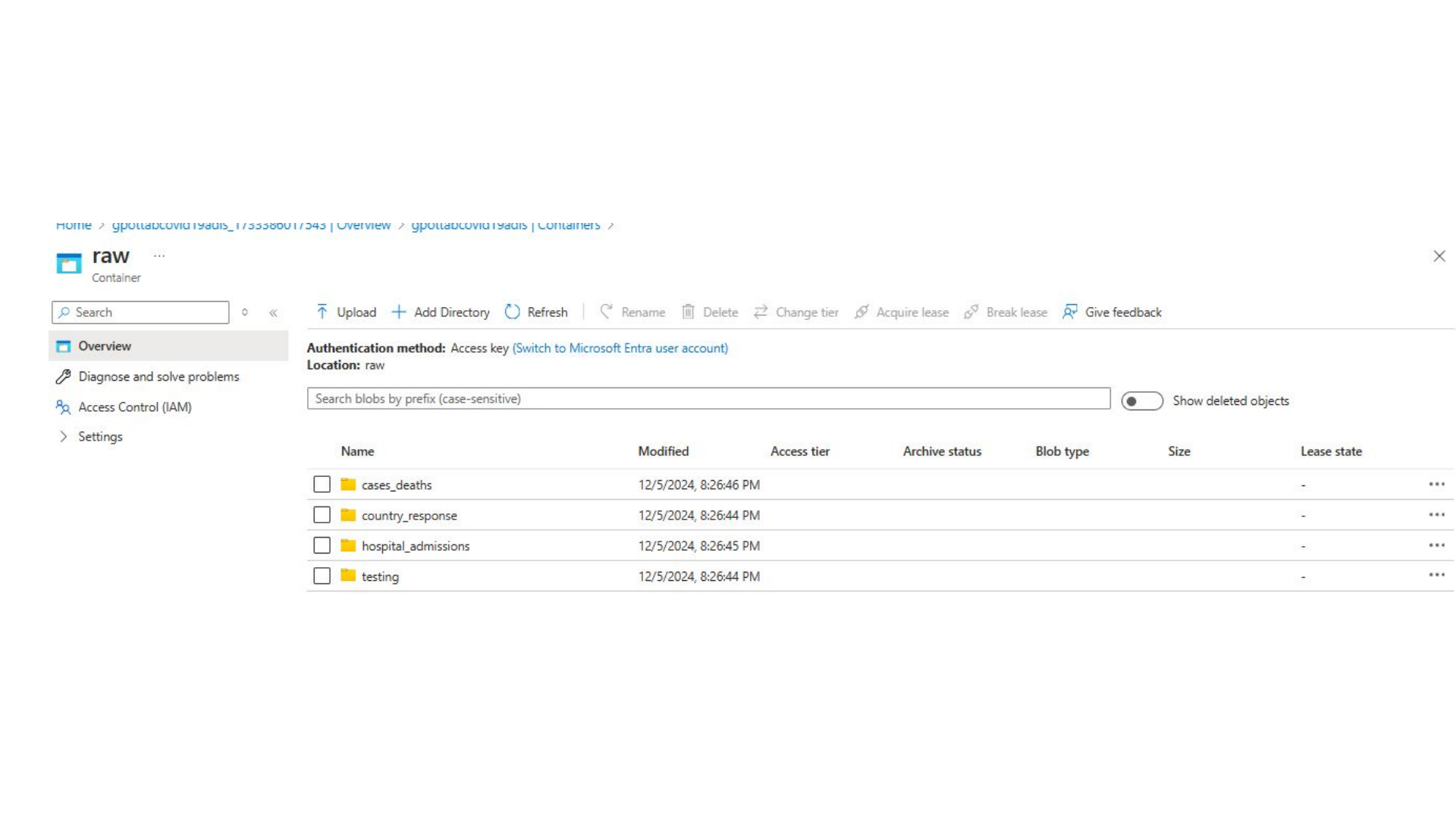
# Here is the final Pipeline




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 > gpottabcovid19adls_1733386017543 | Overview > gpottabcovid19adls | Containers >

**raw** ...
Container

Search

- Overview
- Diagnose and solve problems
- Access Control (IAM)
- Settings

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:** raw

Search blobs by prefix (case-sensitive)

Show deleted objects

| Name | Modified | Access tier | Archive status | Blob type | Size | Lease state |
|---|---|---|---|---|---|---|
| cases_deaths | 12/5/2024, 8:26:46 PM | | | | | - |
| country_response | 12/5/2024, 8:26:44 PM | | | | | - |
| hospital_admissions | 12/5/2024, 8:26:45 PM | | | | | - |
| testing | 12/5/2024, 8:26:44 PM | | | | | - |