

# **Data Migration Pipeline from On-Prem SQL Server to AzureBlobStorage**

## Business Requirement

The Client requires a solution to migrate data from on-premises SQL Server Database to Azure Blob Storage containers using Azure Data Factory (ADF). The solution must automate the daily data copying process, delete existing files in the target containers before each run, and ensure the data is stored in the designated containers as per the defined mapping.

## Prerequisites

### SQL Server:

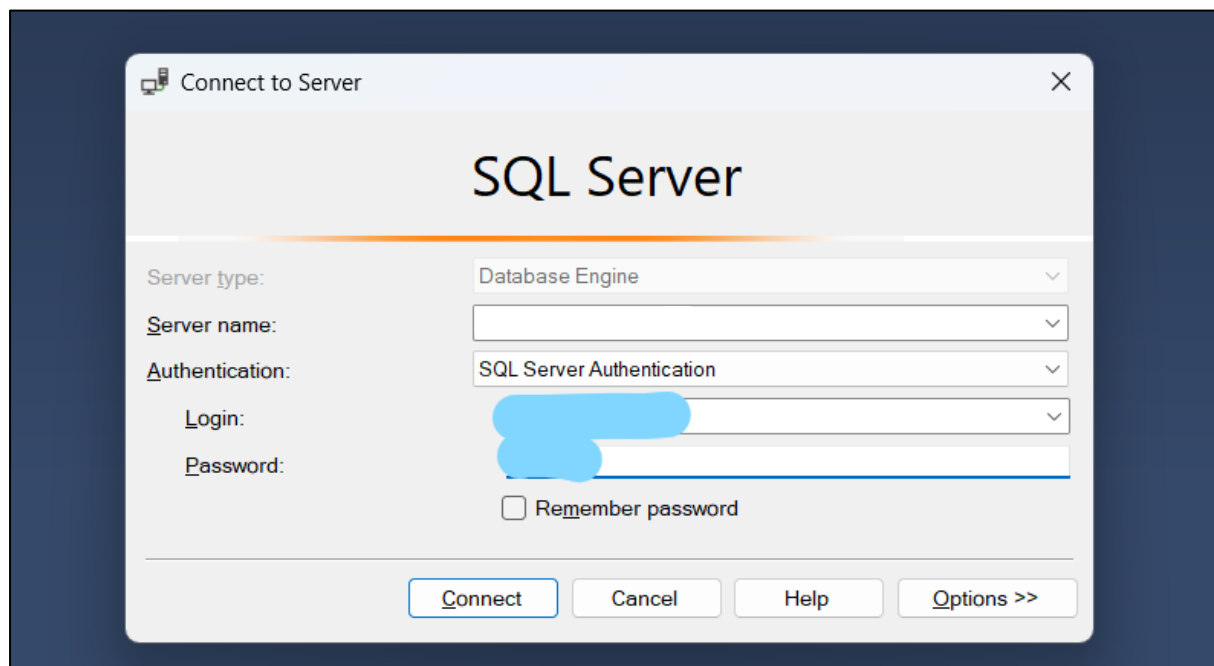
- Database: HealthcareDB
- Tables: Patients, Appointments, Hospital
- Authentication: SQL authentication
  - Username: give your desired username
  - Password: give password
- Integration Runtime: Self-Hosted Integration Runtime (IR) configured for on-premises SQL Server access.

### Azure Blob Storage:

- Containers: patientblobstorage, appointmentsblobstorage, hospitalblobstorage
- Permissions: Contributor role on the storage account for read, write, and delete operations.

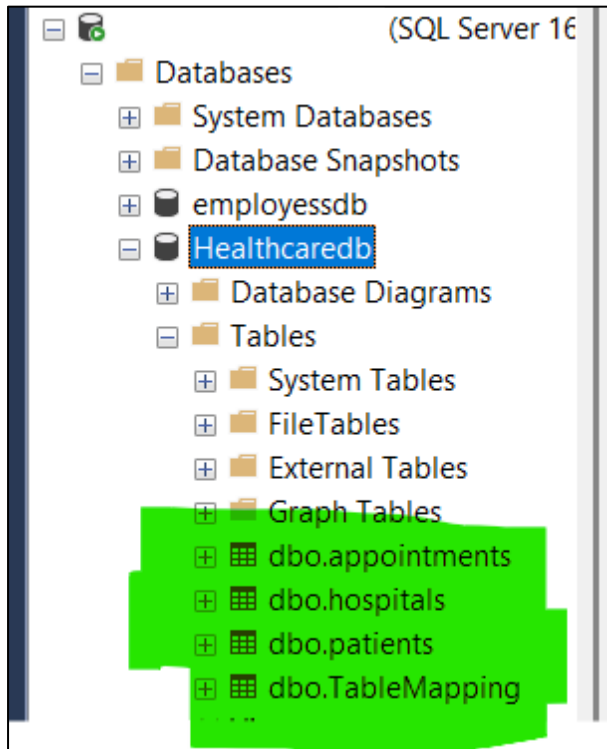
## Implementation

1. In SQL server need to create server for self-hosted because we need to migrate data to Azure Storage.



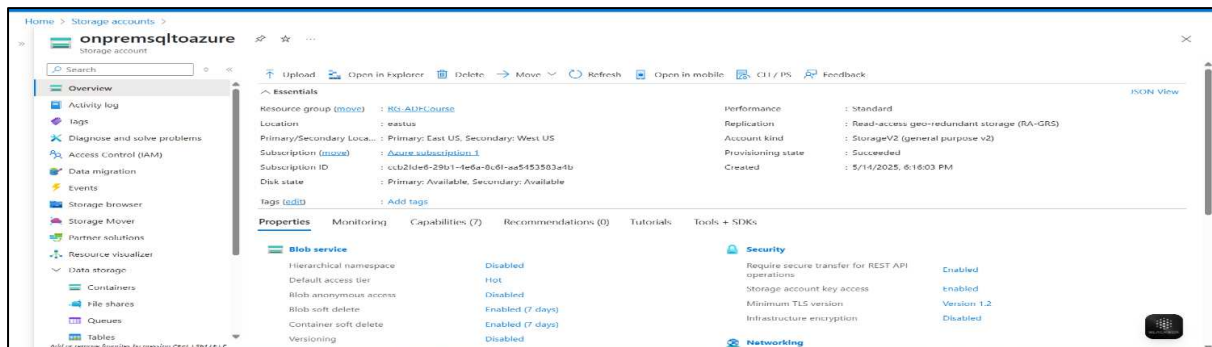
## 2. Create database and create tables in sql server (patients,appointments,hospitals,TabelMapping)

Note : TableMapping table used to map the tables with specific containers.

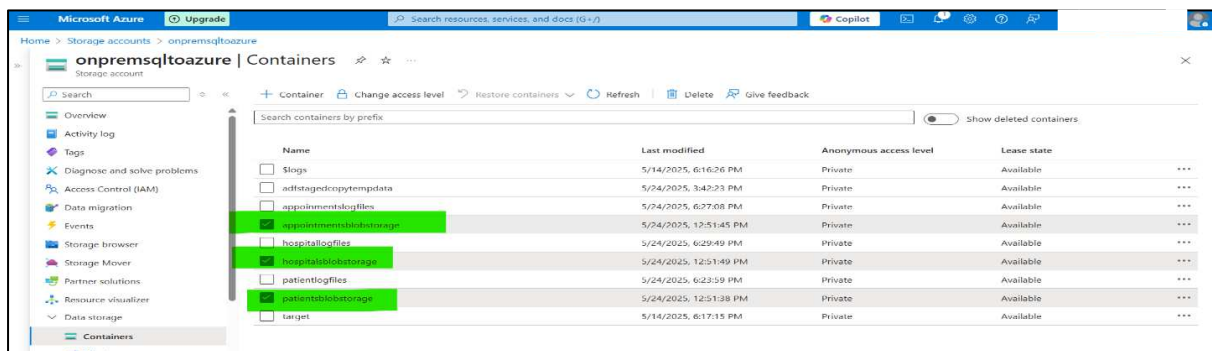


Check all the tables created successfully and fetching the data properly.

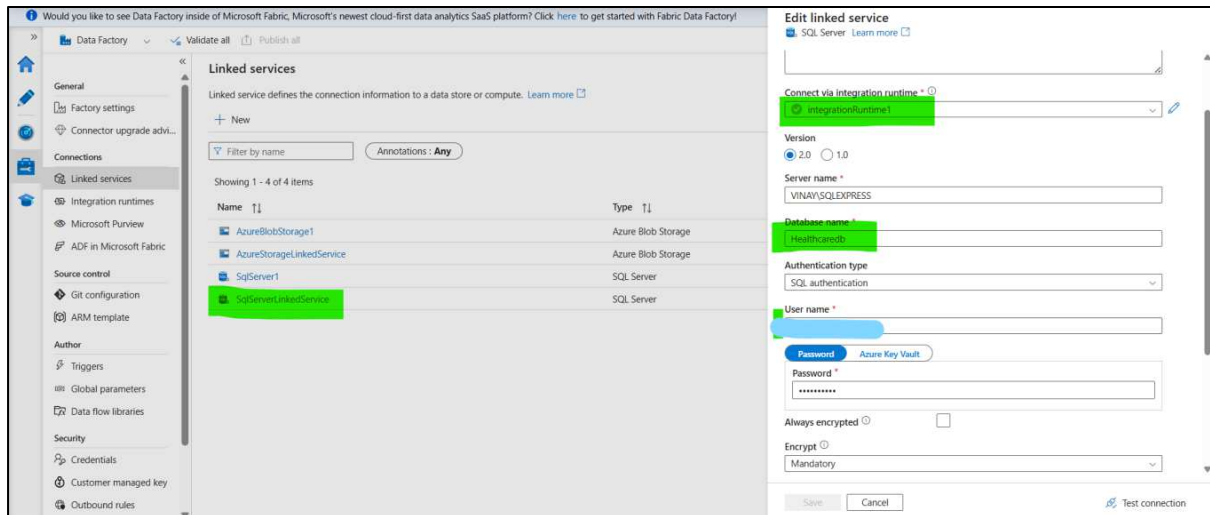
## 3. Then coming to Azure Data Factory create a Storage Account



## 4. Create three containers for each table to store data in particular container

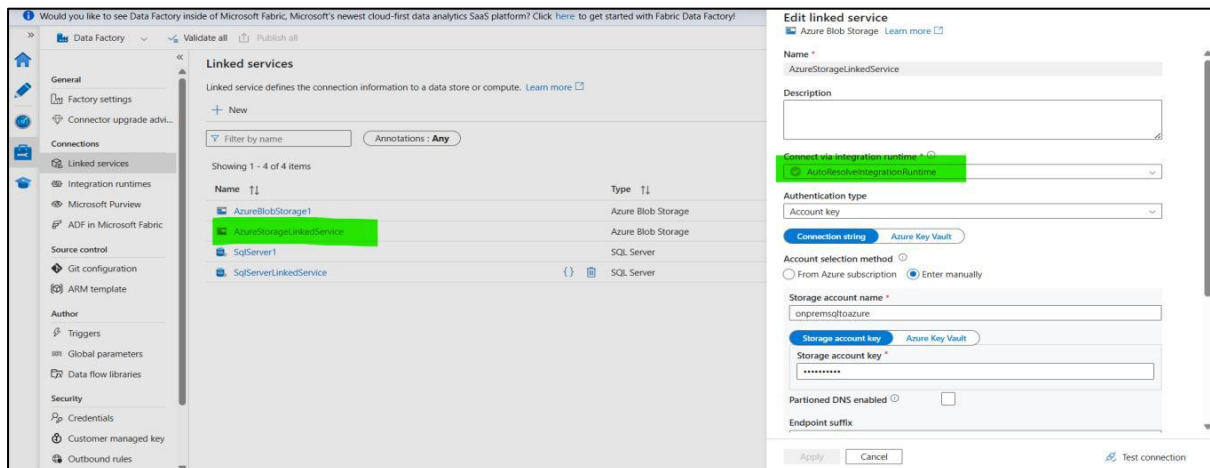


## 5. In Azure Data Factory studio, Create a Source Linked Service for On-Prem SQL Server



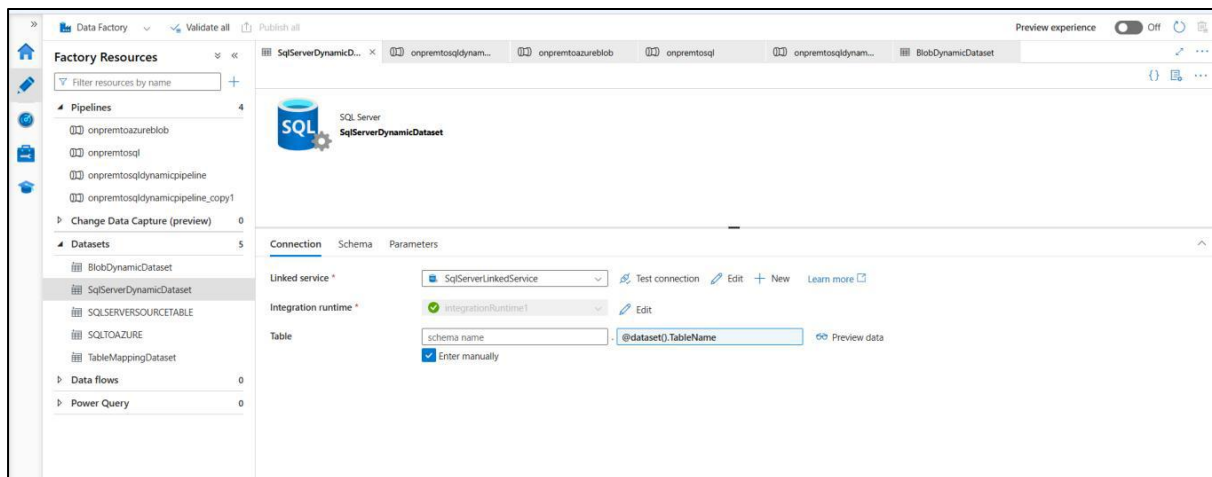
In Connect Via Integration runtime create self-hosted server because we need migrate data from on-prem to Azure Storage and give Database name, select SQL Authentication, give Username, Password which was there for on-prem SQL Server.

## 6. Create a Destination Azure Blob Storage we need to store data in Azure Blob Storage

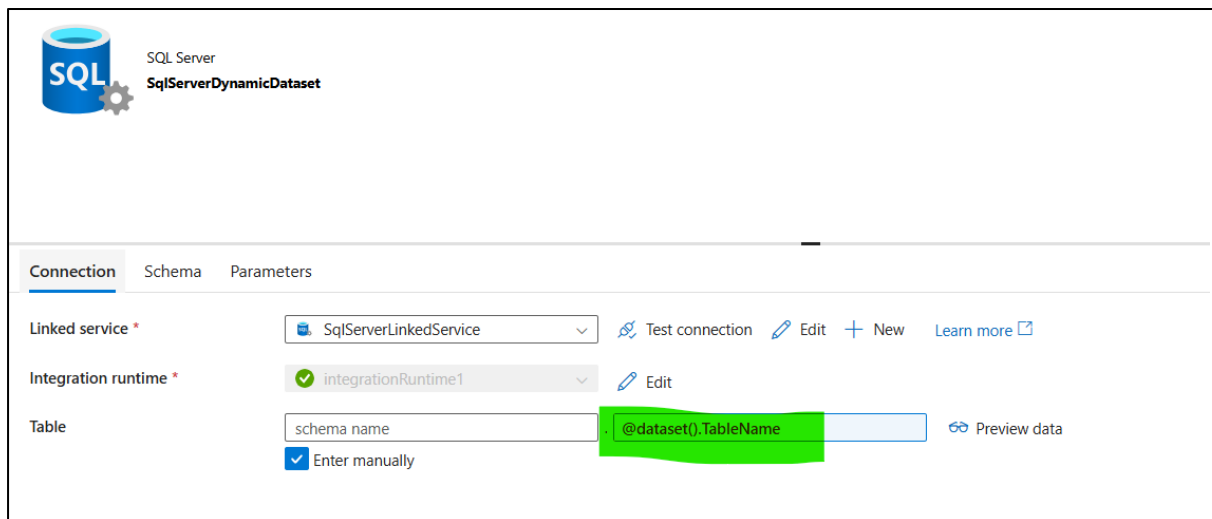


Note: After creating both Source and Destination Linked Service check the Test connection in order to check this works properly.

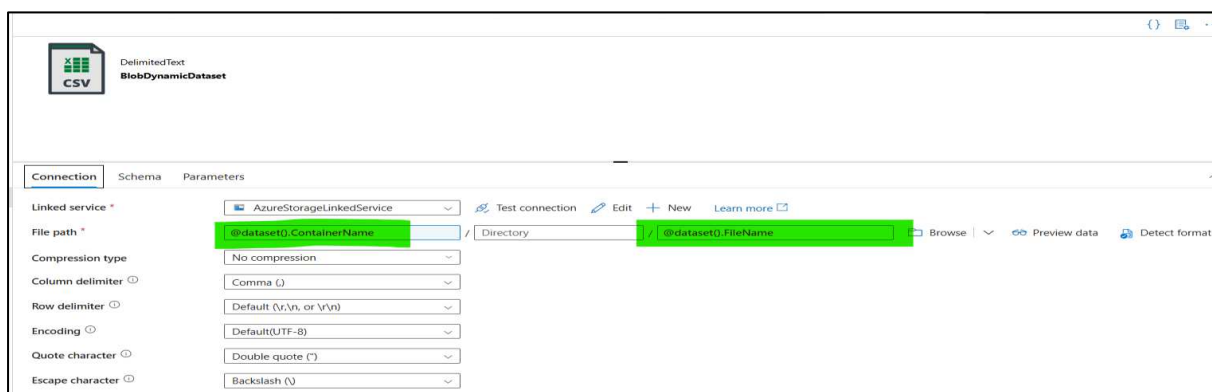
## 7. Goto Author create dataset for Source type of SQL Server



In dataset window create a parameter called table name, because we have three tables we need to set it dynamically

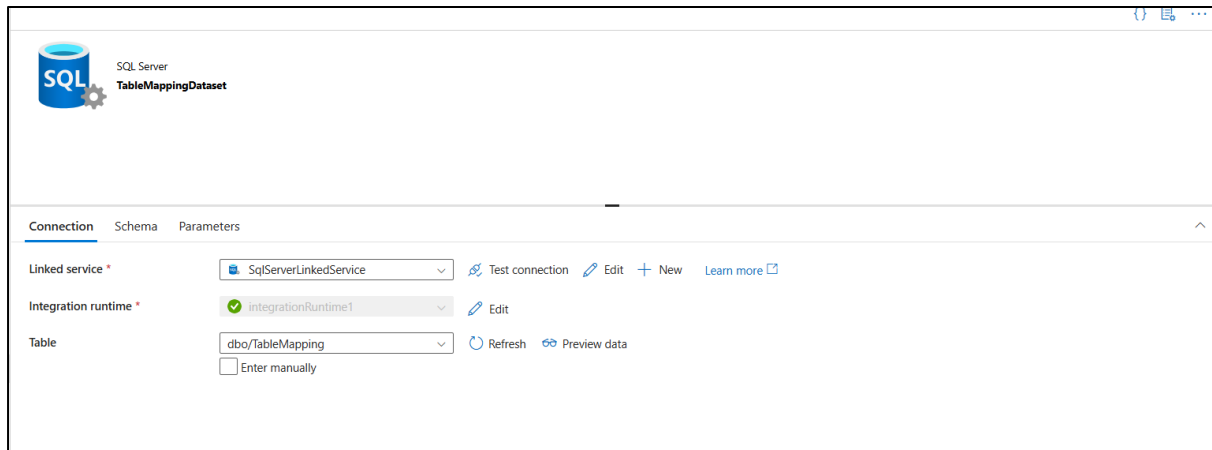


## 8. Create a Sink dataset for Destination type of Blob Storage



Create Two Parameter called ContainerName and FileName

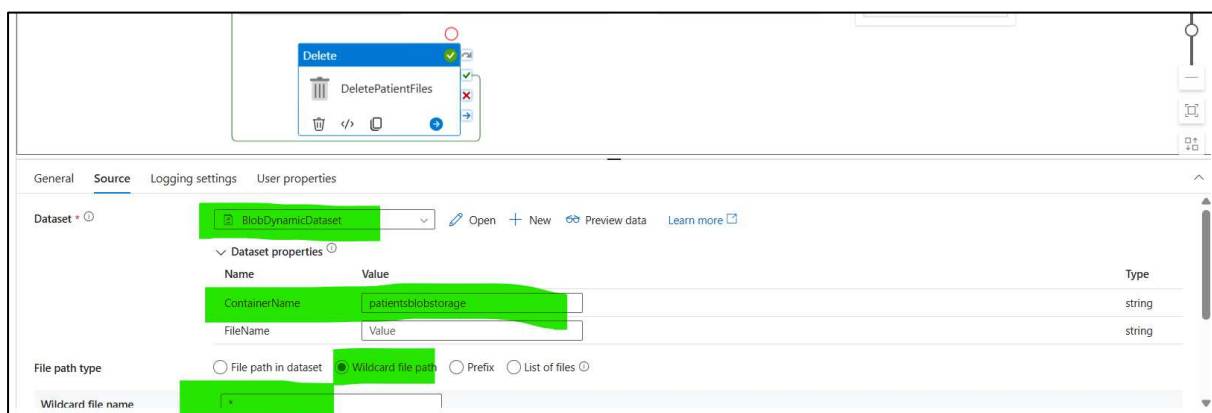
## 9. Create a dataset for TableMapping



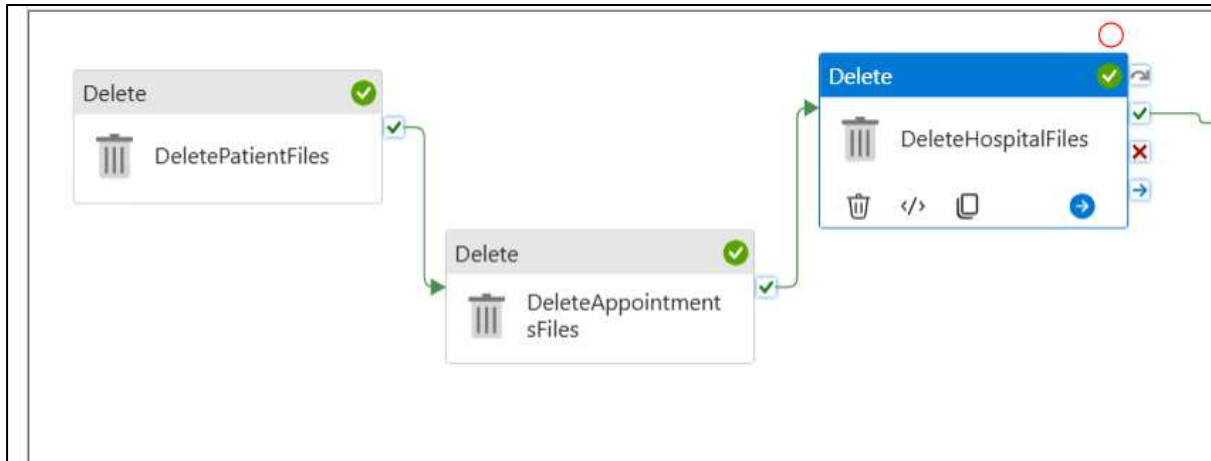
The screenshot shows the configuration page for a 'TableMappingDataset' in SQL Server. The 'Connection' tab is active, showing the 'Linked service' as 'SqlServerLinkedService', the 'Integration runtime' as 'IntegrationRuntime1', and the 'Table' as 'dbo/TableMapping'. There are buttons for 'Test connection', 'Edit', '+ New', and 'Learn more'. Below the table selection, there are 'Refresh' and 'Preview data' buttons, and an unchecked checkbox for 'Enter manually'.

After creating both Linked Services and Datasets for Source and Destination please cross check the connection by clicking the Test Connection and preview data.

- 10. Now, create a pipeline where it will dynamically migrate the data from on-prem to Azure Storage**
- 11. Create a three Delete Activities because Since the Delete activity in ADF operates on a single dataset at a time, and each dataset instance can only point to one container at a time, you need separate activities to handle each container.**
- 12. Create DeletePatientFiles, DeleteAppointmentsFiles, DeleteHospitalFiles for each storage.**
- 13. In every particular delete activity, in source select dataset. Then in container name specify every container name for example, for patient give patinetblobstorage, for appointment give appoinmentblobstorage, for hospital give hospitalblobstorage.**
- 14. In every delete activity select Wildcard file path and give wildcard name as (\*).**



The screenshot shows the configuration page for a 'Delete' activity named 'DeletePatientFiles'. The 'Source' tab is active, showing the 'Dataset' as 'BlobDynamicDataset'. Below the dataset selection, there are 'Open', '+ New', 'Preview data', and 'Learn more' buttons. The 'Dataset properties' section shows 'ContainerName' as 'patientsblobstorage' and 'FileName' as 'Value'. The 'File path type' section has 'Wildcard file path' selected. The 'Wildcard file name' field is empty.



**15. Connect every activity with each other**

**16. In every container select enable logging because we need to log files for each activity and for that create a three log files in Azure container so that it can store in particular log storage.**

Home > Storage accounts > onpremsqltoazure

onpremsqltoazure | Containers

Search containers by prefix

Show deleted containers

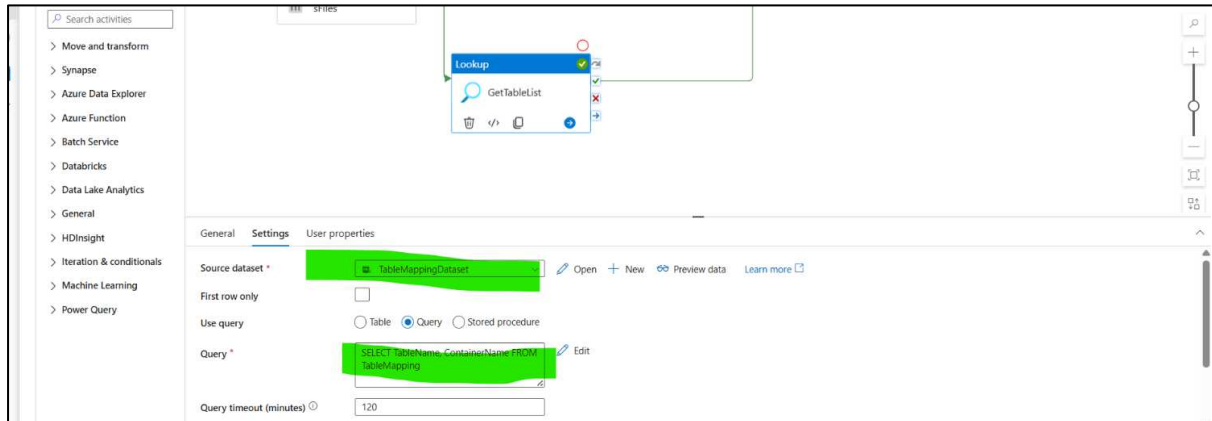
Name	Last modified	Anonymous access level	Lease state
<input type="checkbox"/> \$logs	5/14/2025, 6:16:26 PM	Private	Available
<input type="checkbox"/> adfstagedcopytempdata	5/24/2025, 3:42:23 PM	Private	Available
<input checked="" type="checkbox"/> appointmentslogfiles	5/24/2025, 6:27:08 PM	Private	Available
<input type="checkbox"/> appointmentsblobstorage	5/24/2025, 12:51:45 PM	Private	Available
<input checked="" type="checkbox"/> hospitallogfiles	5/24/2025, 6:29:49 PM	Private	Available
<input type="checkbox"/> hospitalsblobstorage	5/24/2025, 12:51:49 PM	Private	Available
<input checked="" type="checkbox"/> patientlogfiles	5/24/2025, 6:23:59 PM	Private	Available
<input type="checkbox"/> patientsblobstorage	5/24/2025, 12:51:38 PM	Private	Available
<input type="checkbox"/> target	5/14/2025, 6:17:15 PM	Private	Available

In this container whenever a delete activity will occur that deleted log files will store in this particular container.

## 17. Add Lookup Activity to canvas

Settings Tab:

- Source Dataset: Select TableMappingDataset.
- Use Query: Select Query.
- Query: SELECT TableName, ContainerName FROM TableMapping
- First row only: Uncheck (to get all rows).



## 18. Add ForEach Activity:

A. Drag a ForEach activity to the canvas and connect it to the Lookup activity (green arrow from GetTableList to ForEach).

B. Configure

- Name: IterateTables
- Settings Tab:
- Items: @activity('GetTableList').output.value (this references the Lookup output, an array of table-container pairs).
- Sequential: Check (to process tables one at a time).

## 19. Add Copy Data Activity Inside ForEach:

A. Inside the ForEach activity, drag a Copy Data activity.

B. Configure: Name: CopyTableToBlob

- Source Tab:
- Source Dataset: Select SqlServerDynamicDataset.
- Dataset Parameters:
- TableName: @item().TableName (uses the current table name from the ForEach loop).

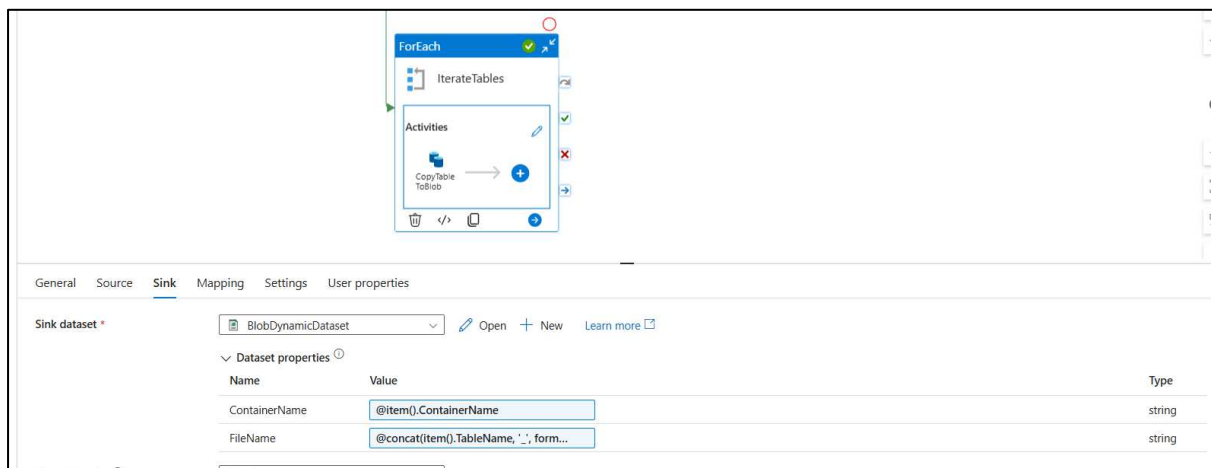


## 20. Sink Tab:

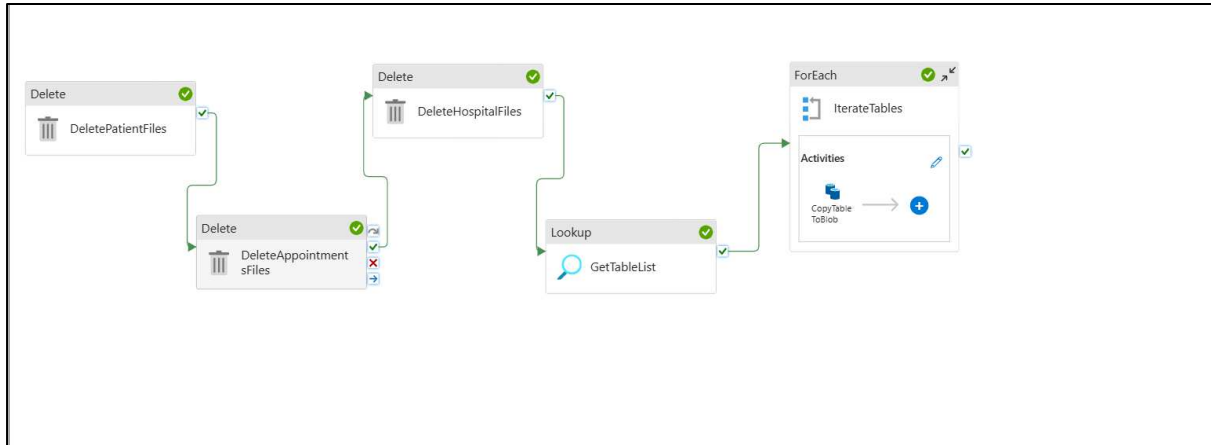
A. Sink Dataset: Select BlobDynamicDataset.

B. Dataset Parameters:

- ContainerName: @item().ContainerName (uses the current container name from the ForEach loop).
- FileName:@concat(item().TableName,'\_',formatDateTime(utcNow(),'yyyyMMdd\_HH:mm:ss'),''.csv')
- fileformat
  - a. Settings Tab
  - b. Enable staging: Check (for better performance with self-hosted IR).
  - c. Staging storage account: Select your Blob Storage account via AzureStorageLinkedService.
  - d. Staging folder: Specify a folder (e.g., staging).



## 21. Final pipeline is here



**22. Save the pipeline and click on publish all and then Debug check for container whether they are working as per requirement.**

**23. As mentioned in requirement that we need to delete the data at mentioned intervals so for that we need to add trigger**

## 24. Open CopyTableToBlobPipeline

- Go to the Trigger tab and click New/Edit.
- In the popup, click Choose trigger > New.
- Configure the trigger:
- Name: DailyTrigger
- Type: Tumbling Window Trigger
- Start Date: 2025-05-25T00:00:00Z (May 25, 2025, 00:00 UTC, which is 05:30 AM IST) just for example
- Recurrence: Every 1 Day (set Interval to 1, Frequency to Day)
- End Date: Optional (leave blank for no end date, or set a future date like 2026-05-25T00:00:00Z)
- Window Size: 1 Day (default)
- Advanced Settings:
- Time Zone: UTC (default). Since the pipeline's file names use IST (adjusted with `addMinutes(utcNow(), 330)`), the trigger time will be in UTC, but the file names will reflect IST.
- Activated: Yes
- Click OK to create the trigger. And then Click Publish All to save and activate the trigger.

## Containers

patientsblobstorage

Container

Search

Upload

Change access level

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: patientsblobstorage

Search blobs by prefix (case-...)

Show deleted blobs

Add filter

Name

Patients\_20250524\_131451.csv

Patients\_20250524\_131451.csv

Blob

Save

Discard

Download

Refresh

Delete

Overview

Versions

Snapshots

Edit

Generate SAS

1

patient\_id,first\_name,last\_name,birth\_date,gender

2

1,"John","Doe",1985-04-12 00:00:00.0000000,"M"

3

2,"Jane","Smith",1990-11-05 00:00:00.0000000,"F"

4

3,"Alice","Johnson",1975-06-30 00:00:00.0000000,"F"

5

4,"Robert","Williams",1969-01-22 00:00:00.0000000,"M"

6

5,"Michael","Brown",1982-09-14 00:00:00.0000000,"M"

7

6,"Emily","Davis",1993-07-19 00:00:00.0000000,"F"

8

7,"Daniel","Wilson",1988-03-25 00:00:00.0000000,"M"

9

8,"Laura","Moore",1995-12-10 00:00:00.0000000,"F"

10

9,"David","Taylor",1978-11-11 00:00:00.0000000,"M"

11

10,"Olivia","Anderson",1991-02-02 00:00:00.0000000,"F"

12

Csv Preview

## Patient Container

hospitalsblobstorage

Container

Search

Upload

Change access level

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: hospitalsblobstorage

Search blobs by prefix (case-...)

Show deleted blobs

Add filter

Name

hospitals\_20250524\_131615.csv

hospitals\_20250524\_131615.csv

Blob

Save

Discard

Download

Refresh

Delete

Overview

Versions

Snapshots

Edit

Generate SAS

1

hospital\_id,name,city,state

2

1,"General Hospital","New York","NY"

3

2,"City Health Center","Los Angeles","CA"

4

3,"Lakeside Medical","Chicago","IL"

5

4,"Sunrise Clinic","Houston","TX"

6

5,"Valley Hospital","Phoenix","AZ"

7

6,"Greenwood Health","San Diego","CA"

8

7,"Maple Medical","Dallas","TX"

9

8,"Riverside Care","San Jose","CA"

10

9,"Oak Hill Hospital","Austin","TX"

11

10,"Cedar Wellness Center","San Francisco","CA"

12

Csv Preview

## Hospital Container

appointmentsblobstorage

Container

Search

Upload

Change access level

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: appointmentsblobstorage

Search blobs by prefix (case-...)

Show deleted blobs

Add filter

Name

Appointments\_20250524\_131...

Appointments\_20250524\_131534.csv

Blob

Save

Discard

Download

Refresh

Delete

Overview

Versions

Snapshots

Edit

Generate SAS

1

appointment\_id,patient\_id,hospital\_id,appointment\_date,reason

2

1,1,1,2025-06-01 09:00:00.0000000,"Annual Checkup"

3

2,2,2,2025-06-02 10:30:00.0000000,"Follow-up Visit"

4

3,3,3,2025-06-03 11:15:00.0000000,"Blood Test"

5

4,4,4,2025-06-04 14:00:00.0000000,"X-Ray"

6

5,5,5,2025-06-05 08:45:00.0000000,"Physical Therapy"

7

6,6,6,2025-06-06 13:30:00.0000000,"Consultation"

8

7,7,7,2025-06-07 15:00:00.0000000,"Allergy Test"

9

8,8,8,2025-06-08 16:00:00.0000000,"Vaccination"

10

9,9,9,2025-06-09 09:30:00.0000000,"MRI Scan"

11

10,10,10,2025-06-10 11:00:00.0000000,"Surgery Consultation"

12

Csv Preview

## Appointment Container

### **Validate Output:**

1. In Azure Portal, navigate to your Blob Storage account.
2. Before the run, note the existing files in each container.
3. After the run:
  - Confirm the old files are deleted.
  - Check for new files with the current timestamp (e.g., Patients\_20250524\_181200.csv for a run at 06:12 PM IST).
  - Verify the files are still in CSV format with headers.
4. Verify that each day
  - Old files are deleted.
  - New files are created with the current date and time (e.g., Patients\_20250525\_000000.csv).

### **Meets Requirement**

- **Delete Files Daily:** The Delete activities (DeletePatientFiles, DeleteAppointmentsFiles, DeleteHospitalFiles) remove all files in the containers at the start of each pipeline run, ensuring the containers only contain the current day's data.
- **New Data Daily:** The pipeline copies fresh data from the SQL Server tables (Patients, Appointments, Hospital) to the containers each day, with file names reflecting the current date and time (e.g., Patients\_20250525\_000000.csv).
- **Scheduled Runs:** The Tumbling Window Trigger runs the pipeline daily, automating the process without manual intervention.
- **CSV Format:** The files remain in CSV format, as confirmed by the BlobDynamicDataset settings.