

AZURE DOCUMENTATION

NAME : P. SANJANA REDDY

EMP ID:MT4020

Table of Content

SNO	USER STORY	PAGE NO
1.	Create an ADF Pipeline to read and write files in Azure Data Lake Gen 2 using Azure Data Factory	9
2.	Copy multiple files with same type of data(schema) from ADLS Gen 2 to Azure SQL	18
3.	Create a pipeline to copy 4 different tables from Azure SQL Database to ADLS Gen1/Gen2	27

User Story 1:

Create an ADF Pipeline to read and write files in Azure Data Lake Gen 2 using Azure Data Factory

Description: Azure Data Lake Gen 2(ADLS) is a storage service provided by Microsoft Azure which allows you to save the data in multiple format.

- Create an input and output container with the naming convention input_modakid and output_modakid(example input_vkg92599)
- Upload the files attached in this story to the input container of the datalake
- Create a pipeline in Azure Data Factory(ADF) to migrate the files from input to output container.

Acceptance Criteria:

- Validate the migration of files using UI
- open the csv file to verify if the data was migrated in the same format
- Record (Screenshots) and document the steps
- verify the story with your team lead

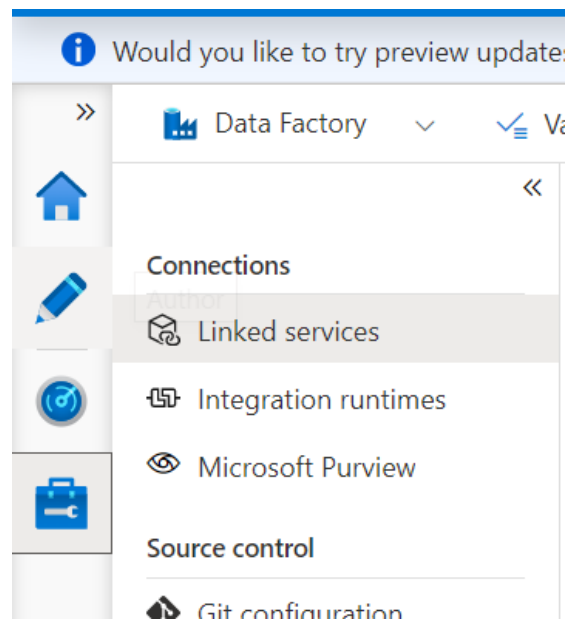
Story Points: 1

Linked Service for Source :

- Created a linked service to link your data store to the Data Factory or Synapse Workspace.
- Linked services are much like connection strings, which define the connection information needed for the service to connect to external resources.
- To create a new linked service in Azure Data Factory Studio.
- Select the Manage tab and then linked services, where you can see any existing linked services you defined. Select New to create a new linked service.
- After selecting New to create a new linked service you will be able to choose any of the supported connectors and configure its details accordingly.
- Selected the Azure Storage gen 2 to create linked Service for it.

Steps Screenshots :

- Select the Manage tab and then linked services, where you can see any existing linked services you defined.



- Select New to create a new linked service.

Linked services

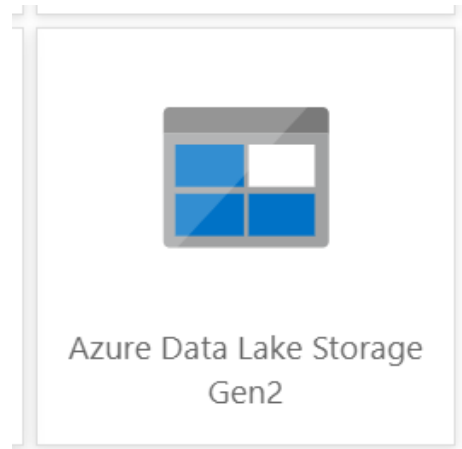
Linked service defines the connection information to a data store or compute. [Learn more](#)

[+ New](#)

Filter by name

Annotations : **Any**

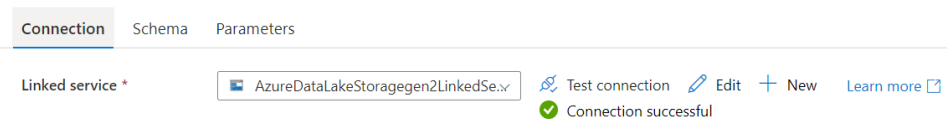
- Choose any of the supported connectors and configure its details accordingly.
- Here selected the Azure data lake Storage 2 was selected.



- Created a Linked Service for Azure data lake Storage 2 .

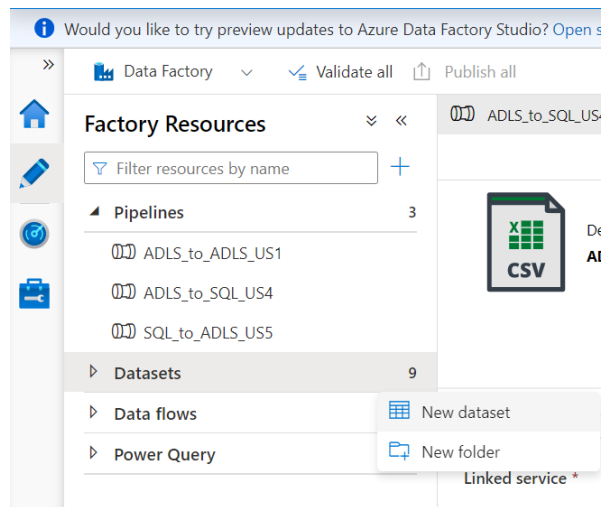
Testing Connection :

- After creation of Connection it was tested.
- The test was successful.

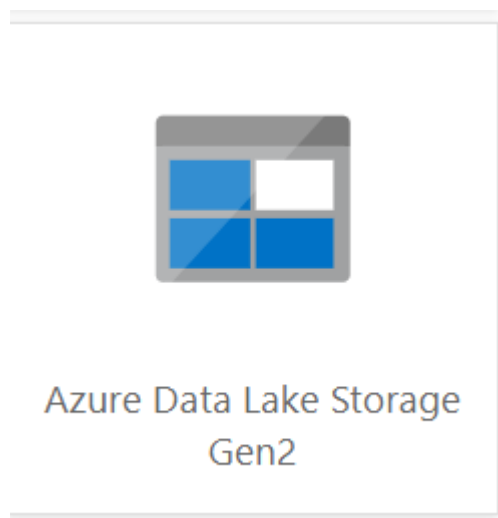


Creating a Dataset for Source :

- An activity can take zero or more input datasets, and produce one or more output datasets.
- An input dataset represents the input for an activity in the pipeline, and an output dataset represents the output for the activity.
- Datasets identify data within different data stores, such as tables, files, folders, and documents.
- Select New dataset to create a new dataset.



- Select the data store for creating a data store.



- Select the Linked Service which was created for that data store.

Name

AmazonMWSObject1

Linked service *

Select...

Preview the dataset :

- Previewed data to see whether dataset was created or not successfully.

Preview data


Linked service: AzureDataLakeStorageegen2LinkedService1

Object:

Prop_0	Prop_1	Prop_2
id	ename	state
5001	Sanjana	Telangana
5002	Venkata	Telanaga
5005	Manal	Tamil Nadu

Dataset for Destination :

- In the same way created a data set for destination .



DelimitedText
 ADLS_US1_Destination

Connection
 Schema
 Parameters

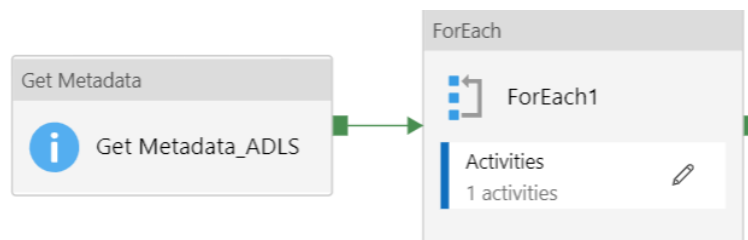
Linked service *

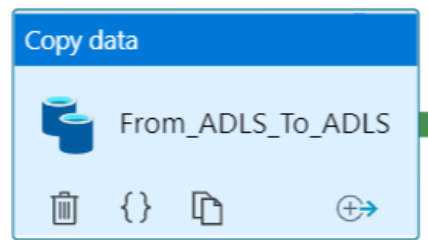
AzureDataLakeStorageegen2LinkedSe.v
 Test connection
 Edit
 + New
 Learn more

File path *

blobstorage-adlsgen2 / fromADLS / @dataset().filename
 Browse
 Preview data

Pipeline to read and write files in Azure Data Lake Gen 2 :





Activities :

Get Metadata :

- Activity can be used to retrieve the metadata of any type of file, folder or relational database table in Azure Data Factory.
- The Output of the “ Get Metadata ” Activity can be used in the “ Conditional Expressions ” to perform Validation, or, Consume the Metadata in the Subsequent Activities.

ForEach :

- The ForEach Activity defines a repeating control flow in an Azure Data Factory or Synapse pipeline.
- This activity is used to iterate over a collection and executes specified activities in a loop.

Copy data :

- In Azure Data Factory and Synapse pipelines, you can use the Copy activity to copy data among data stores located on-premises and in the cloud.
- After you copy the data, you can use other activities to further transform and analyze it.
- You can also use the Copy activity to publish transformation and analysis results for business intelligence (BI) and application consumption.

Get Metadata :

General **Settings** User properties

Dataset * ADLSgen2_US1_Source1 Open New [Learn more](#)

Field list * New Delete

☐ Argument

☐ Child items

- In Mata data dataset was given with was created before.
- In Field list argument was given as Child Items.

For Each :

General **Settings** Activities (1) User properties

Sequential ☐

Batch count ⓘ

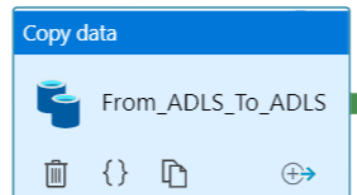
Items @activity('Get Metadata_ADLS').output...

- The For Each will take items from the Get Meta data output.
- @activity('Get Metadata_ADLS').output.childItems was given in items.
- In this we are having one activity that is copy activity.

Copy Activity :

- This is the Copy Activity.

ADLS_to_ADLS_US1 > ForEach1



- All the required fields were provided in source.

General **Source** Sink Mapping Settings User properties

Source dataset * ADLSgen2_US4_Source

[Open](#) [+ New](#) [Preview data](#) [Learn more](#)

▼ Dataset properties ⓘ

Name	Value	Type
filename	<input type="text" value="@item().name"/>	string

- All the required fields were provided in destination.

General Source **Sink** Mapping Settings User properties

Sink dataset * ADLS_US1_Destination [Open](#) [+ New](#) [Learn more](#)

▼ Dataset properties ⓘ

Name	Value	Type
filename	<input type="text" value="@item().name"/>	string

Reading and Writing files in Azure Data Lake Gen 2 :

- Files were sent from source to destination.
- Files were read and written into the Azure Data Lake Gen 2.

Authentication method: Access key (Switch to Azure AD User Account)
 Location: blobstorage-adlsgen2 / fromADLS

☐ Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state	
<input type="checkbox"/> [-]							...
<input type="checkbox"/> Employee1.csv	7/4/2022, 10:23:59 PM	Hot (Inferred)		Block blob	156 B	Available	...
<input type="checkbox"/> Employee2.csv	7/4/2022, 10:23:58 PM	Hot (Inferred)		Block blob	97 B	Available	...
<input type="checkbox"/> Employee3.csv	7/4/2022, 10:23:59 PM	Hot (Inferred)		Block blob	102 B	Available	...
<input type="checkbox"/> Employee4.csv	7/4/2022, 10:23:58 PM	Hot (Inferred)		Block blob	94 B	Available	...

User Story 4:

Copy multiple files with same type of data(schema) from ADLS Gen 2 to Azure SQL

Description:

- Upload 4 files into a folder in ADLS Gen2 storage (All the files must contain same type of data i.e. same schema)
- Create a table in Azure SQL database which maps with the schema of files uploaded in datalake
- Create a pipeline to copy data from all the files in the datalake to a single table in Azure SQL Database

Acceptance Criteria:

- Calculate the total rows in Source files and validate the count with the rows in Azure SQL Database
- Record (Screenshots) and document the steps
- Verify the story with your team lead

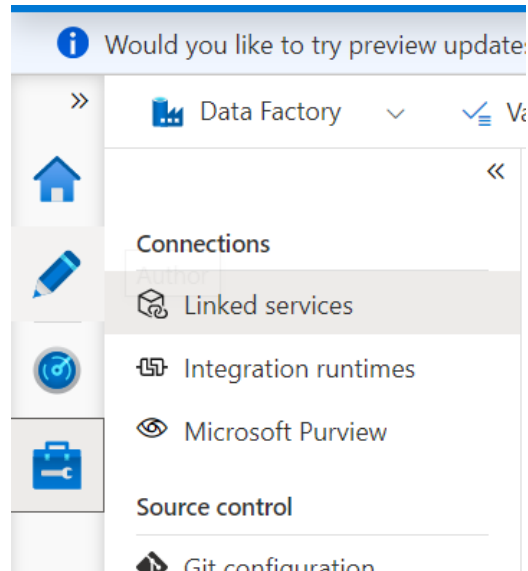
Story Points: 2

Explanation :

Linked Service for Source :

- Created a linked service to link your data store to the Data Factory or Synapse Workspace.
- Linked services are much like connection strings, which define the connection information needed for the service to connect to external resources.
- To create a new linked service in Azure Data Factory Studio.
- Select the Manage tab and then linked services, where you can see any existing linked services you defined. Select New to create a new linked service.
- After selecting New to create a new linked service you will be able to choose any of the supported connectors and configure its details accordingly.
- Selected the Azure Storage gen 2 to create linked Service for it.

Steps Screenshots:



- Select the Manage tab and then linked services, where you can see any existing linked services you defined.
- Select New to create a new linked service.

Linked services

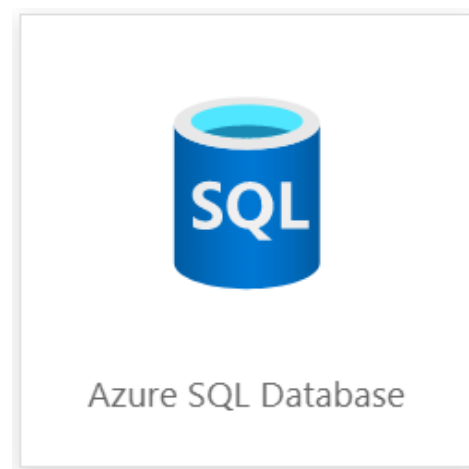
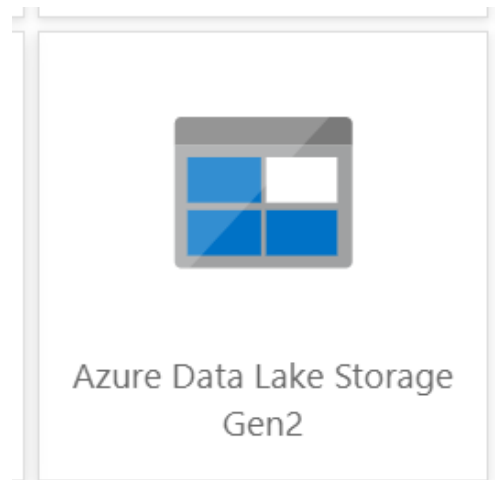
Linked service defines the connection information to a data store or compute. [Learn more](#)

New

Filter by name

Annotations : **Any**

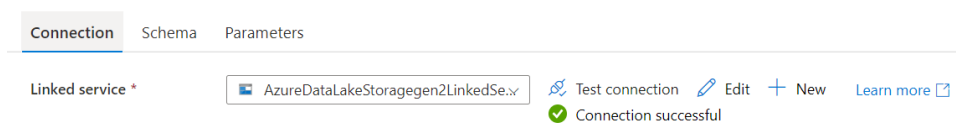
- Choose any of the supported connectors and configure its details accordingly.
- Here selected the Azure data lake Storage 2 was selected.

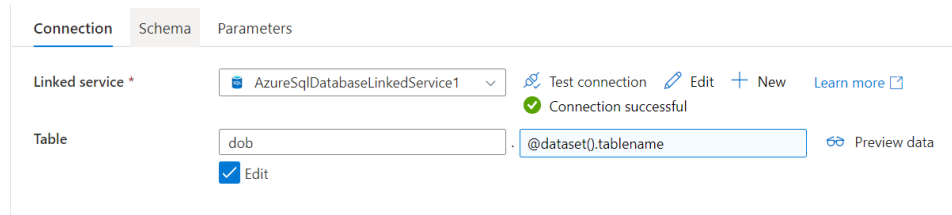


- Created a Linked Service for Azure data lake Storage 2 .

Testing Connection :

- After creation of Connection it was tested.
- The test was successful.

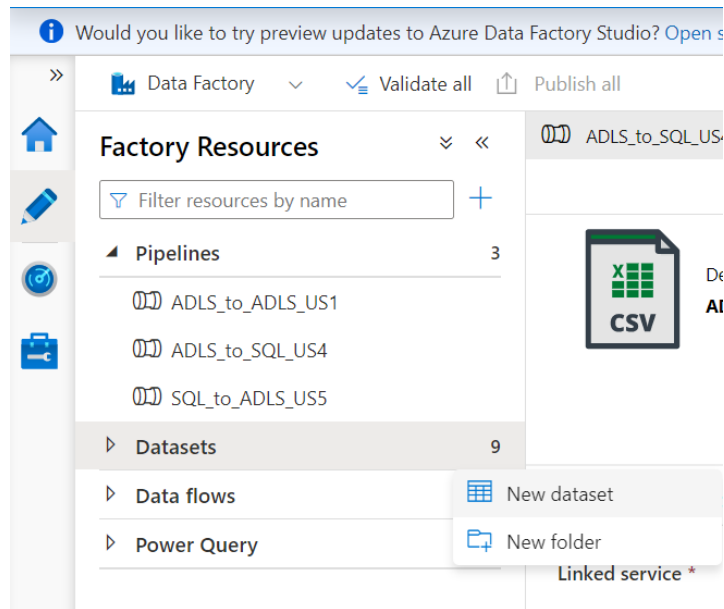




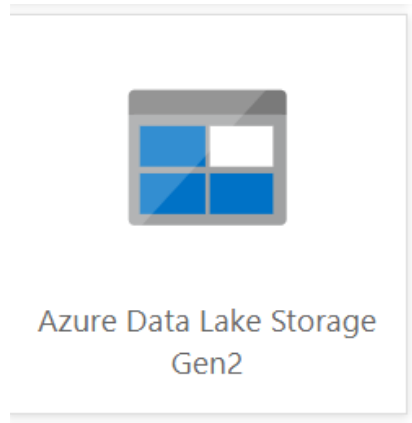
- Hence Connection was established.

Creating a Dataset for Source :

- An activity can take zero or more input datasets, and produce one or more output datasets.
- An input dataset represents the input for an activity in the pipeline, and an output dataset represents the output for the activity.
- Datasets identify data within different data stores, such as tables, files, folders, and documents.
- Select New dataset to create a new dataset.



- Select the data store for creating a data store.



- Select the Linked Service which was created for that data store.

Name

AmazonMWSObject1

Linked service *

Select...

Preview the dataset :

- Previewed data to see whether dataset was created or not successfully.

Preview data


Linked service: AzureDataLakeStorageGen2LinkedService1

Object:

id	ename	state
5001	Sanjana	Telangana
5002	Venkata	Telanaga
5005	Manal	Tamil Nadu
5003	Anaida	Kerala
5004	Nikita	Telangana

Dataset for Destination :

- In the same way created a data set for destination .



Azure SQL Database
AzureSqlTable_US4_Destination

Connection
 Schema
 Parameters

Linked service *

AzureSqlDatabaseLinkedService1
 Test connection
 Edit
 + New
 Learn more

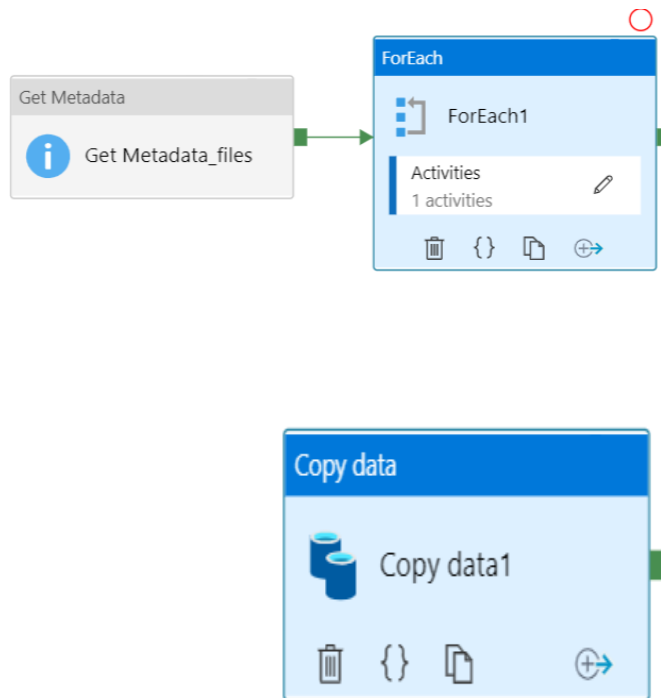
Connection successful

Table

dob
 @dataset().tablename
 Preview data

Edit

Pipeline to read and write files in Azure Data Lake Gen 2 :



Activities :

Get Metadata :

- Activity can be used to retrieve the metadata of any type of file, folder or relational database table in Azure Data Factory.
- The Output of the “ Get Metadata ” Activity can be used in the “ Conditional Expressions ” to perform Validation, or, Consume the Metadata in the Subsequent Activities.

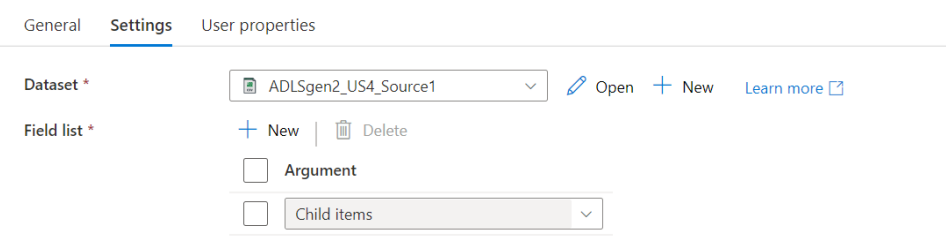
ForEach :

- The ForEach Activity defines a repeating control flow in an Azure Data Factory or Synapse pipeline.
- This activity is used to iterate over a collection and executes specified activities in a loop.

Copy data :

- In Azure Data Factory and Synapse pipelines, you can use the Copy activity to copy data among data stores located on-premises and in the cloud.
- After you copy the data, you can use other activities to further transform and analyze it.
- You can also use the Copy activity to publish transformation and analysis results for business intelligence (BI) and application consumption.

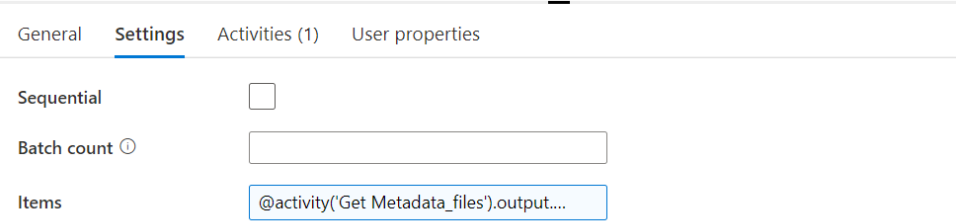
Get Metadata:



The screenshot shows the 'Settings' tab for the 'Get Metadata' activity. The 'Dataset' dropdown is set to 'ADLSgen2_US4_Source1'. The 'Field list' dropdown is set to 'Child items'. There are also links for 'Open', 'New', and 'Learn more'.

- In Metadata dataset was given with was created before.
- In Field list argument was given as Child Items.

For Each :



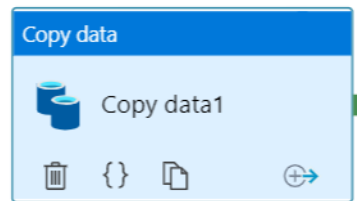
The screenshot shows the 'Settings' tab for the 'ForEach' activity. The 'Sequential' checkbox is checked. The 'Batch count' is set to 1. The 'Items' field is set to '@activity('Get Metadata_files').output...'.

- The For Each will take items from the Get Meta data output.
- @activity('Get Metadata_ADLS').output.childItems was given in items.
- In this we are having one activity that is copy activity.

Copy Activity :

- This is the Copy Activity.

ADLS_to_SQL_US4 > ForEach1



- All the required fields were provided in source.

General **Source** Sink Mapping Settings User properties

Source dataset *

[Open](#) [New](#) [Preview data](#) [Learn more](#)

Dataset properties

Name	Value	Type
filename	<input type="text" value="@item().name"/>	string

- All the required fields were provided in destination.

GeneralSourceSinkMappingSettingsUser properties

Sink dataset *

AzureSqlTable_US4_Destination

Open

New

Learn more

Dataset properties

Name	Value	Type
tablename	Employeeedetails	string

Write behavior

Insert

Upsert

Stored procedure

- Data from multiple files were copied into single table.

```
1 SELECT TOP (1000) * FROM [dbo].[Employee_Details_Us4]
```

ResultsMessages

Search to filter items...

id	ename	state
3001	Sathwi	MAnipur
5004	Saniya	Telangana
5001	Sanjana	Telangana

User Story 5:

Create a pipeline to copy 4 different tables from Azure SQL Database to ADLS Gen1/Gen2.

Description:

- Create 4 different tables in Azure SQL Database
- Create a single pipeline to copy all the 4 tables from source to destination
- Create an alert to send message to the team when the pipeline run is successful

Acceptance Criteria:

- Validate the migration of data
- Verify the alert email with your team lead
- Record (Screenshots) and document the steps
- Verify the story with your team lead

Story Points: 3

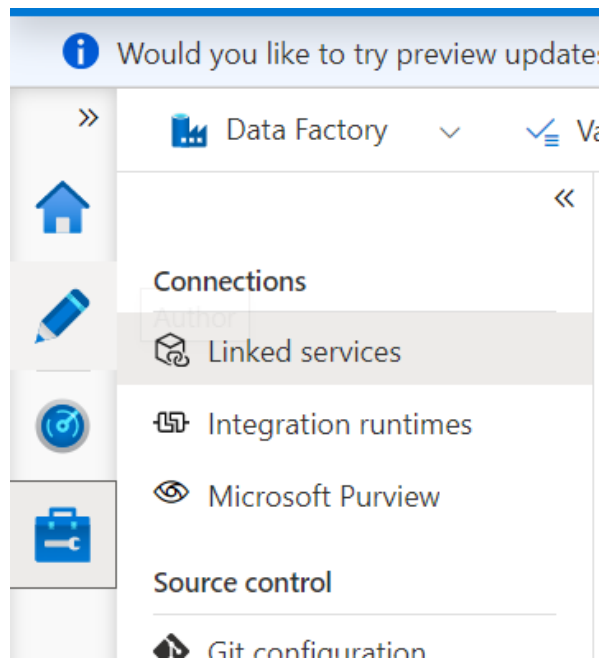
Explanation :

Linked Service for Source :

- Created a linked service to link your data store to the Data Factory or Synapse Workspace.
- Linked services are much like connection strings, which define the connection information needed for the service to connect to external resources.
- To create a new linked service in Azure Data Factory Studio.
- Select the Manage tab and then linked services, where you can see any existing linked services you defined. Select New to create a new linked service.
- After selecting New to create a new linked service you will be able to choose any of the supported connectors and configure its details accordingly.
- Selected the Azure Storage gen 2 to create linked Service for it.

Steps Screenshots :

- Select the Manage tab and then linked services, where you can see any existing linked services you defined.



- Select New to create a new linked service.

Linked services

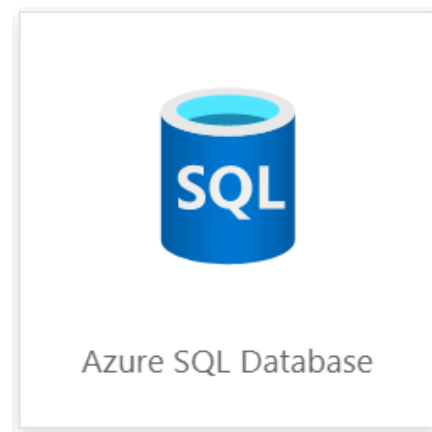
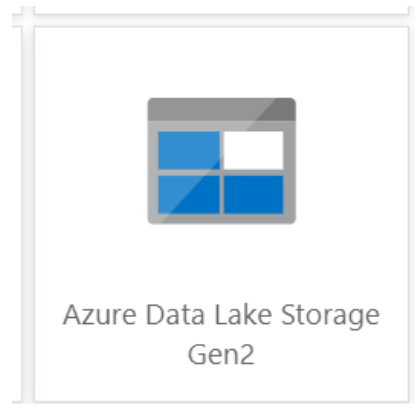
Linked service defines the connection information to a data store or compute. [Learn more](#)

New

Filter by name

Annotations : **Any**

- Choose any of the supported connectors and configure its details accordingly.
- Here selected the Azure data lake Storage 2 was selected.



- Created a Linked Service for Azure data lake Storage 2 .

Testing Connection :

- After creation of Connection it was tested.
- The test was successful.

Connection

Schema

Parameters

Linked service *

AzureDataLakeStorageegen2LinkedSe.v

Test connection

Edit

New

Learn more

Connection successful

Connection

Schema

Parameters

Linked service *

AzureSqlDatabaseLinkedService1

Test connection

Edit

New

Learn more

Connection successful

Table

dob

@dataset().tablename

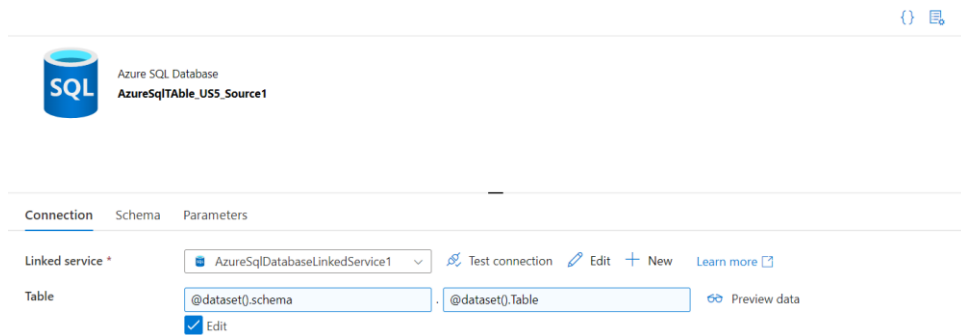
Preview data

Edit

- Hence Connection was established.

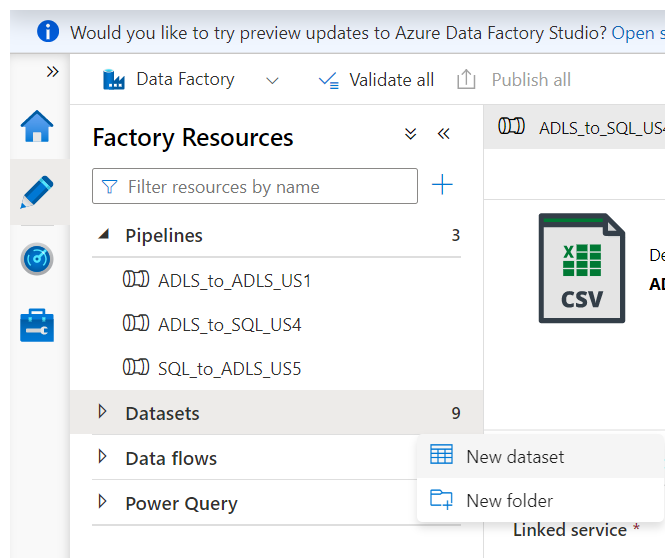
Dataset for source :

- Created the dataset for source.

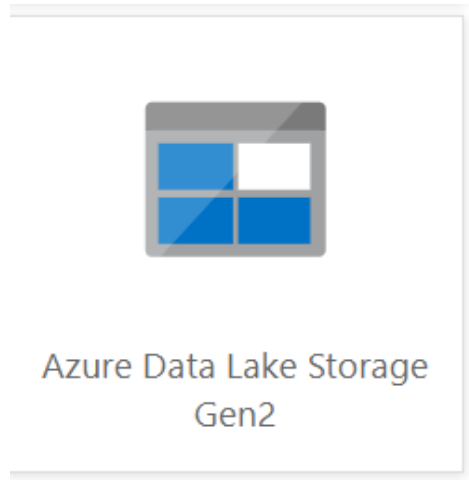


Dataset for Destination :

- An activity can take zero or more input datasets, and produce one or more output datasets.
- An input dataset represents the input for an activity in the pipeline, and an output dataset represents the output for the activity.
- Datasets identify data within different data stores, such as tables, files, folders, and documents.
- Select New dataset to create a new dataset.



- Select the data store for creating a data store.



- Select the Linked Service which was created for that data store.

Name

AmazonMWSObject1

Linked service *

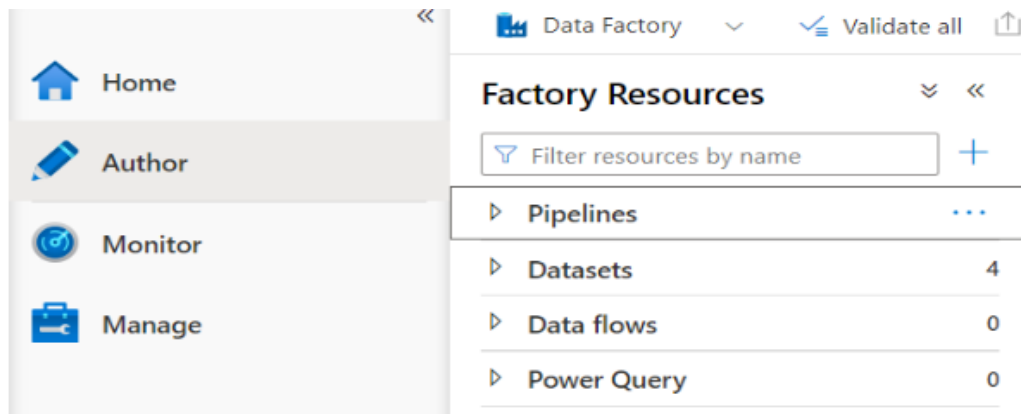
Select...



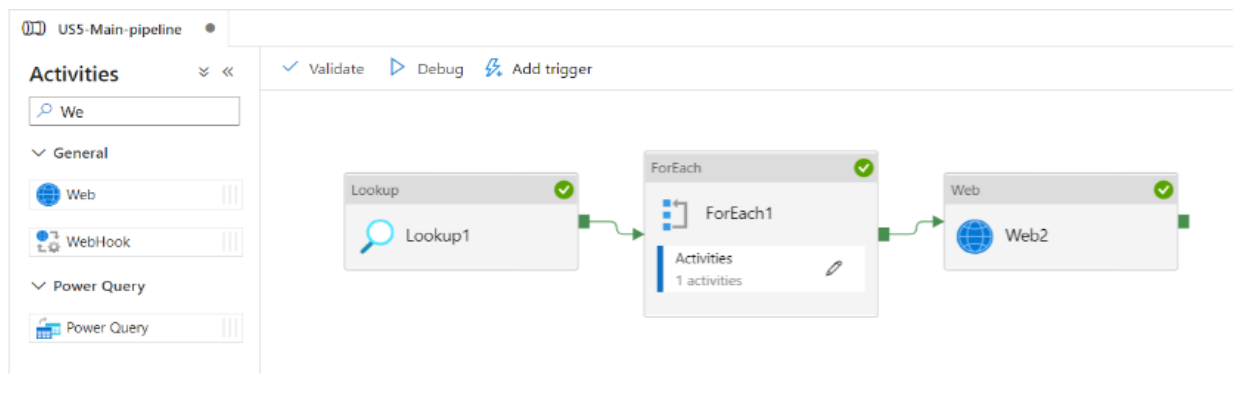
Step-1: Pipeline was created using below steps

- To create a pipeline, first opened azure data factory and clicked on **Author tab** and in the pipelined tab, clicked on create new pipeline
- From the available list of **Activities**, selected **Lookup** activity to read multiple files from the source container
- Next selected **ForEach** activity to get each file from the previous configuration
- In the ForEach activity, selected **copy data** activity to copy the data from input container to output container
- After the ForEach activity, **Web activity** was taken to get the email notification

Author tab :



Pipeline created :







Step-2: Configuration for the pipeline

Lookup Configuration

- A Lookup activity is used to read data from any datastore in azure
- **Azure_Sql_Sourc1_us5_DS** was taken as dataset for lookup activity and in the body of the query, a query was written to read all the tables


General **Settings** User properties

Source dataset * AzureSqlTable_US5_Source

 Open  New  Preview data [Learn more](#) 

First row only ☐

Use query ☐ Table ☒ Query ☐ Stored procedure

Query 

```
select TOP 4 * from
INFORMATION_SCHEMA.TABLES where
table_type='BASE TABLE';
```

ForEach Configuration :

- ForEach Activity takes input from the output of GetMetadata activity. So in the items field, output of childitems was taken from the dynamic content parameters

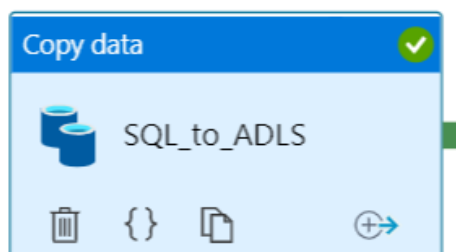
General **Settings** Activities (1) User properties

Sequential ☐

Batch count ⓘ

Items @activity("Lookup_for_sql").output.value

Copy Activity Configuration :



- Source configuration

General **Source** Sink Mapping Settings User properties

Source dataset * AzureSqlTable_US5_Source1

[Open](#) [New](#) [Preview data](#) [Learn more](#)

Dataset properties

Name	Value	Type
schema	@item().TABLE_SCHEMA	String
Table	@item().TABLE_NAME	String

• Sink Configuration

General Source **Sink** Mapping Settings User properties

Sink dataset * ADLS_US5_destination [Open](#) [New](#) [Learn more](#)

Dataset properties

Name	Value	Type
fileName	@concat(item().TABLE_SCHEMA,item()...	String

Copy behavior None

Web app Configuration

- In the web app, URL for connecting to logic app was given by copying it from the logic app designer
- POST method was given in the method field and in the body field, dynamic content parameters were given

Body field of Web Activity

Add dynamic content

```
{
  "pipelinename": "@{pipeline().Pipeline}",
  "datafactoryname": "@{pipeline().DataFactory}",
  "runid": "@{pipeline().RunId}",
  "Numberoffiles": "@{activity('Lookup1').output.count}"
}
```

Web Activity configuration

General **Settings** User properties

URL *

Method *

Headers [+ New](#)

Body

Datasets [+ Add dataset reference](#)

Linked services [+ Add linked service reference](#)

Integration runtime *


Output :

- The output was verified in the output container.

<input type="checkbox"/>		[.]					...
<input type="checkbox"/>		dboemployee11.csv	7/8/2022, 11:09:16 PM	Hot (Inferred)	Block blob	252 B	Available
<input type="checkbox"/>		dboEmployeeDetails.csv	7/8/2022, 11:09:17 PM	Hot (Inferred)	Block blob	224 B	Available
<input type="checkbox"/>		dobempp.csv	7/8/2022, 11:09:17 PM	Hot (Inferred)	Block blob	355 B	Available
<input type="checkbox"/>		dobsalesman1.csv	7/8/2022, 11:09:17 PM	Hot (Inferred)	Block blob	212 B	Available

- Email notification:**

ADF email notification



Sanjana Peddi

To: Sanjana Peddi

Fri 7/8/2022 11:11 PM

Hi Sanjana,

This is an email notification of US5 of Azure Sprint.

pipelineName:SQL_to_ADLS_US5

datafactoryname:sanjanamt4020adf

runid :98c0f33f-924d-4240-b61a-549438a72c33

nooffiles : 4

Reply

Forward