

Assignment5 ADF Pipeline 3

Easy Questions:

Question1: Create a pipeline name 'Without_Foreach_Example' to copy the Customer & Customer address table data into ADLS without using foreach activity.

The screenshot shows the Azure Data Factory (ADF) interface for a pipeline named 'Without_Foreach_Example'. The 'Source' tab is selected in the 'Copy data' activity. The 'Source dataset' is 'SQL_generic_DS'. The 'Dataset properties' table is as follows:

Name	Value	Type
SchemaName	saleslt	string
TableName	customer	string

The 'Use query' option is set to 'Table', and the 'Query timeout (minutes)' is 120.

The screenshot shows the 'Sink' tab for the 'Copy data' activity. The 'Sink dataset' is 'ADLS_CSV_Generic_DS'. The 'Dataset properties' table is as follows:

Name	Value	Type
FilePath	customer_csv	string

The 'Copy behavior' is set to 'None'.

Data Factory | Validate all | Publish all | Preview experience | Off

Factory Resources

- ADLS_DS
- Cust_data_ADLS_DS
- Cust_data_SQL_DS
- DB_DS
- Prod_info_ADLS_DS
- Prod_info_ADLS_DS2
- Prod_info_Database_DS
- Assignment-5
 - address_info_ADLS_DS
 - address_info_sql_DS
 - ADLS_CSV_Generic_DS
 - ADLS_Json_generic_DS
 - Cust_info_ADLS_DS
 - Cust_info_SQL_DS
 - SQL_generic_DS
- Data flows

Activities

- Move & transform
 - Copy data
 - Data flow
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Copy data | Copy data1 | Copy data2

Source | Sink | Mapping | Settings | User properties

Source dataset * | SQL_generic_DS | Open | New | Preview data | Learn more

Dataset properties

Name	Value	Type
SchemaName	salesit	string
TableName	customeraddress	string

Use query | ☒ Table | ☐ Query | ☐ Stored procedure

Query timeout (minutes) | 120

Data Factory | Validate all | Publish all | Preview experience | Off

Factory Resources

- ADLS_DS
- Cust_data_ADLS_DS
- Cust_data_SQL_DS
- DB_DS
- Prod_info_ADLS_DS
- Prod_info_ADLS_DS2
- Prod_info_Database_DS
- Assignment-5
 - address_info_ADLS_DS
 - address_info_sql_DS
 - ADLS_CSV_Generic_DS
 - ADLS_Json_generic_DS
 - Cust_info_ADLS_DS
 - Cust_info_SQL_DS
 - SQL_generic_DS
- Data flows

Activities

- Move & transform
 - Copy data
 - Data flow
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Copy data | Copy data1 | Copy data2

Sink | Mapping | Settings | User properties

Sink dataset * | ADLS_CSV_Generic_DS | Open | New | Learn more

Dataset properties

Name	Value	Type
FilePath	customer_address_csv	string

Copy behavior | None

Max concurrent connections

Data Factory | Validate all | Publish all | Preview experience | Off

Factory Resources

- Copy_Customer_To_CSV_Pipe
- Copy_ProductTable_To_CSV
- Assignment-5
 - Foreach_Example
 - Ingestion_Customer_JOIN_ADLS
 - Ingestion_Customer_SQLDB_ADLS
 - Ingestion_Customer_SQLDB_ADLS_F...
 - Without_Foreach_Example
- Change Data Capture (preview)
- Datasets
 - ADLS_DS
 - Cust_data_ADLS_DS
 - Cust_data_SQL_DS
 - DB_DS
 - Prod_info_ADLS_DS
 - Prod_info_ADLS_DS2
 - Prod_info_Database_DS

Activities

- Move & transform
 - Copy data
 - Data flow
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Copy data | Copy data1 | Copy data2

Output

Pipeline run ID: a96231f0-3633-4f72-9fb8-d71c2d55e06d | View debug run consumption | Export to CSV

All status | Showing 1 - 2 of 2 items

Activity name	Status	Activity type	Run start	Duration	Log
Copy data2	Succeeded	Copy data	7/5/2023, 10:53:40 AM	7s	
Copy data1	Succeeded	Copy data	7/5/2023, 10:53:32 AM	7s	

landing Container

Search

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

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Authentication method: Access key (Switch to Azure AD User Account)

Location: landing

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

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
Address						...
CSV						...
CSV_Pipe						...
customer						...
customer_address_csv						...
customer_csv						...
customeraddress						...
CustomerFolder						...
JSON						...
customer_names_with_address.txt	7/4/2023, 10:17:23 PM	Hot (Inferred)		Block blob	12.01 KiB	Available
SalesLT.Customer.txt	7/4/2023, 11:14:48 PM	Hot (Inferred)		Block blob	212.88 KiB	Available

Creating Generic SQL Dataset and ADLS Dataset

Data Factory Validate all Publish all Preview experience Off

Factory Resources

Filter resources by name

ADLS_DS

Cust_data_ADLS_DS

Cust_data_SQL_DS

DB_DS

Prod_info_ADLS_DS

Prod_info_ADLS_DS2

Prod_info_Database_DS

Assignment-5

address_info_ADLS_DS

address_info_sql_DS

ADLS_CSV_Generic_DS

ADLS_Json_generic_DS

Cust_info_ADLS_DS

Cust_info_SQL_DS

SQL_generic_DS

Data flows

Ingestion_Custom... Foreach_Example Without_Foreach_E... ADLS_CSV_Generic_DS ADLS_Json_generic_DS SQL_generic_DS

Azure SQL Database SQL_generic_DS

Connection Schema Parameters

Linked service * sql_LS Test connection Edit New Learn more

Table @dataset().SchemaName @dataset().TableName Preview data

Edit

Expand Data Factory Validate all Publish all Preview experience Off

Factory Resources

Filter resources by name

ADLS_DS

Cust_data_ADLS_DS

Cust_data_SQL_DS

DB_DS

Prod_info_ADLS_DS

Prod_info_ADLS_DS2

Prod_info_Database_DS

Assignment-5

address_info_ADLS_DS

address_info_sql_DS

ADLS_CSV_Generic_DS

ADLS_Json_generic_DS

Cust_info_ADLS_DS

Cust_info_SQL_DS

SQL_generic_DS

Ingestion_Custom... Foreach_Example Without_Foreach_E... ADLS_CSV_Generic_DS ADLS_Json_generic_DS SQL_generic_DS

Delimited Text ADLS_CSV_Generic_DS

Connection Schema Parameters

Linked service * ADLS_LS Test connection Edit New Learn more

File path * landing @dataset().FilePath File name Browse

Compression type None

Column delimiter Comma (,)

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

Encoding Default(UTF-8)

Quote character Double quote (")

Question2: Create a pipeline name 'Foreach_Example' to copy the Customer & Customer address table data into ADLS using the one copy activity.

The screenshot shows the Azure Data Factory interface. On the left, the 'Factory Resources' pane lists various pipelines and datasets. The 'Foreach_Example' pipeline is selected. The main canvas shows the pipeline design with a 'ForEach' loop containing a 'Copy data' activity. The 'Settings' tab is active, showing the 'Batch count' and 'Items' properties. The 'Items' property is set to '@createArray('customer', 'customer...)'.

The screenshot shows the Azure Data Factory interface. On the left, the 'Factory Resources' pane lists various pipelines and datasets. The 'Foreach_Example' pipeline is selected. The main canvas shows the pipeline design with a 'Copy data' activity. The 'Source' tab is active, showing the 'Source dataset' and 'Dataset properties' section. The 'Source dataset' is set to 'SQL_generic_DS'. The 'Dataset properties' section shows 'SchemaName' as 'saleslt' and 'TableName' as '@item()'.

The screenshot shows the Azure Data Factory interface. On the left, the 'Factory Resources' pane lists various pipelines and datasets. The 'Foreach_Example' pipeline is selected. The main canvas shows the pipeline design with a 'Copy data' activity. The 'Sink' tab is active, showing the 'Sink dataset' and 'Dataset properties' section. The 'Sink dataset' is set to 'ADLS_json_generic_DS'. The 'Dataset properties' section shows 'FilePath' as '@item()'.

Factory Resources

Filter resources by name

Pipelines

- Copy_Customer_JSON_To_Folder
- Copy_Customer_to_JSON
- Copy_Customer_To_CSV_Pipe
- Copy_ProductTable_To_CSV
- Assignment-5
 - Ingestion_Customer_JOIN_ADLS
 - Ingestion_Customer_SQLDB_ADLS
 - Ingestion_Customer_SQLDB_ADLS_F...
- Foreach_Example**

Change Data Capture (preview)

Datasets

- ADLS_DS
- Cust_data_ADLS_DS
- Cust_data_SQL_DS
- DB_DS
- Prod_Info_ADLS_DS

Ingestion_Custom... Foreach_Example

Validate Debug Add trigger

ForEach

ForEach1

Activities

Copy data1

Parameters Variables Settings **Output**

Pipeline run ID: df8968d2-c3e5-4b66-864f-4b9c38bc7627 View debug run

All status List

Showing 1 - 3 of 3 items

Activity name	Status	Activity type	Run start	Duration	Log	Integration r
Copy data1	Succeeded	Copy data	7/5/2023, 10:25:28 AM	8s		AutoResolve
Copy data1	Succeeded	Copy data	7/5/2023, 10:25:28 AM	7s		AutoResolve
ForEach1	Succeeded	ForEach	7/5/2023, 10:25:27 AM	11s		

Home > rajasekhar04061998adls1 | Containers >

landing Container

Search

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

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

- Shared access tokens
- Manage ACL
- Access policy
- Properties
- Metadata

Authentication method: Access key (Switch to Azure AD User Account)

Location: landing

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
Address						-
CSV						-
CSV_Pipe						-
customer						-
customeraddress						-
CustomerFolder						-
JSON						-
customer_names_with_address.txt	7/4/2023, 10:17:23 PM	Hot (Inferred)		Block blob	12.01 KiB	Available
SalesLT.Customer.txt	7/4/2023, 11:14:48 PM	Hot (Inferred)		Block blob	212.88 KiB	Available

landing Container

Search

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

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

- Shared access tokens
- Manage ACL
- Access policy
- Properties
- Metadata

Authentication method: Access key (Switch to Azure AD User Account)

Location: landing / customer

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[.]						
saleslt.customer.json	7/5/2023, 10:25:34 AM	Hot (Inferred)		Block blob	368.62 KiB	Available

Question4: Business wants to move the data from 3 different tables (Customer, Product, CustomerAddress) to ADLS location in CSV format.

However they want the make a pipeline in such a manner that All these copy happens through an individual pipeline, which is called by one common pipeline.

Data Factory | Validate all | Publish all

Factory Resources

- Pipelines (10)
 - Copy_Customer_JSON_To_Folder
 - Copy_Customer_To_JSON
 - Copy_Customer_To_CSV_Pipe
 - Copy_ProductTable_To_CSV
 - Assignment-5
 - ForEach_Example
 - Ingestion_Customer_JOIN_ADLS
 - Ingestion_Customer_SQLDB_ADLS
 - Ingestion_Customer_SQLDB_ADLS_F...
 - pipeline1
 - Without_Foreach_Example
- Change Data Capture (preview) (0)
- Datasets (14)

Activities

- ForEach_Example
- Without_Foreach_Example
- ADLS_CSV_Generic_DS

Settings

Sequential ☒

Items: @createArray('customer', 'product', '...')

Pipeline expression builder

Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variable](#)

@createArray('customer', 'product', 'customeraddress')

Clear contents

Parameters | System variables | **Functions** | Variables

Search

Data Factory | Validate all | Publish all | Preview experience: Off

Factory Resources

- Pipelines (10)
 - Copy_Customer_JSON_To_Folder
 - Copy_Customer_To_JSON
 - Copy_Customer_To_CSV_Pipe
 - Copy_ProductTable_To_CSV
 - Assignment-5
 - ForEach_Example
 - Ingestion_Customer_JOIN_ADLS
 - Ingestion_Customer_SQLDB_ADLS
 - Ingestion_Customer_SQLDB_ADLS_F...
 - pipeline1
 - Without_Foreach_Example
- Change Data Capture (preview) (0)
- Datasets (14)
 - ADLS_DS
 - Cust_data_ADLS_DS
 - Cust_data_SQL_DS

Activities

- Move & transform
 - Copy data
 - Data flow
- Iteration & conditionals
 - ForEach

Output

Pipeline run ID: 4d71891f-c41c-46b7-a3b7-57f5f8434dec | View debug run | View debug run consumption

All status | List | Export to CSV

Showing 1 - 4 of 4 items

Activity name	Status	Activity type	Run start	Duration	Log
Copy data1	Succeeded	Copy data	7/5/2023, 11:57:38 AM	8s	
Copy data1	Succeeded	Copy data	7/5/2023, 11:57:28 AM	8s	
Copy data1	Succeeded	Copy data	7/5/2023, 11:57:19 AM	8s	
ForEach1	Succeeded	ForEach	7/5/2023, 11:57:19 AM	29s	

Home > rajasekhar04061998adls1 | Containers >

landing | Container

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

Authentication method: Access key (Switch to Azure AD User Account)

Location: landing

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

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
Address					-	...
CSV					-	...
CSV_Pipe					-	...
customer					-	...
customer_address_csv					-	...
customer_csv					-	...
customeraddress					-	...
CustomerFolder					-	...
JSON					-	...
product					-	...

Question5: In the above case problem is that every time the pipeline runs it will overwrite the data. Ensure that all the data goes into

Folder like (Customer/Year/Month/Day) to avoid any overwriting. Please rewrite the pipeline.

The screenshot shows the Azure Data Factory pipeline editor. On the left, the 'Factory Resources' pane lists various pipelines and activities. The main canvas displays a 'Copy data' activity. The 'Source' tab is selected, showing the 'Source dataset' as 'SQL_generic_DS'. Below this, the 'Dataset properties' section shows 'SchemaName' as 'saleslt' and 'TableName' as 'customer'. The 'Use query' section has 'Table' selected. The 'Query timeout (minutes)' is set to 120, and the 'Isolation level' is set to 'None'.

The screenshot shows the Azure Data Factory pipeline editor with the 'Sink' tab selected for the 'Copy data' activity. The 'Sink dataset' is 'ADLS_CSV_Generic_DS'. The 'Dataset properties' section shows 'Name' as 'ADLS_CSV_Generic_DS' and 'FilePath' as 'Customer/@{formatDateTime(utcnow(), 'yyyy')}/@{formatDateTime(utcnow(), 'MM')}/@{formatDateTime(utcnow(), 'dd')}'. The 'Copy behavior' is set to 'None'. To the right, the 'Pipeline expression builder' is open, showing the expression: `Customer/@{formatDateTime(utcnow(), 'yyyy')}/@{formatDateTime(utcnow(), 'MM')}/@{formatDateTime(utcnow(), 'dd')}`. The 'Functions' tab is selected in the expression builder.

Data Factory

Validate all

Publish all

Factory Resources

Filter resources by name

Pipelines

Copy_Customer_JSON_To_Folder

Copy_Customer_to_JSON

Copy_Customer_To_CSV_Pipe

Copy_ProductTable_To_CSV

Assignment-5

3_Tables_to_ADLS

address_greater_than_100

Foreach_Example

Ingestion_Customer_JOIN_ADLS

Ingestion_Customer_SQLDB_ADLS

Ingestion_Customer_SQLDB_ADLS...

Ingestion_Product_Addres_To_JS...

Ingestion_Product_To_JSON

Customer_data_Dynamic_folder

Without_Foreach_Example

Ingestion_Product_...

address_greater_th...

Ingestion_Product_...

Customer_data_Dy...

Validate

Debug

Add trigger

Copy data

Copy data1

Parameters

Variables

Settings

Output

Pipeline run ID: 0a79be51-8688-4cf2-b7b4-b82163d9a5da

All status

Showing 1 - 1 of 1 items

Activity name	Status	Activity type	Run start	Duration	Log
Copy data1	Succeeded	Copy data	7/5/2023, 11:04:45 PM	7s	

Upload

Add Directory

Refresh

Rename

Delete

Change tier

Acquire lease

Break lease

Give feedback

Authentication method: Access key

Location: landing / Customer / 2023 / 07 / 05

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
saleslt.customer.txt	7/5/2023, 11:04:51 PM	Hot (Inferred)		Block blob	212.88 KiB

Question 6: I am attaching a file, load this file into your ADLS location. This file contains a threshold value. Create a pipeline in such a manner that, it will first get the threshold value from this file and then check if the record count in the customer

table is more than it or not. If yes then copy the customer data from SQL db to ADLS location in JSON format.

Question 3: Create a pipeline name 'Foreach_Example_2' which solves the following business use case.

Customer data is very important for our business. Hence whenever we have more than 100 records in the customer table, we copy the customer data to another table customer_copy within the sql db.

However whenever we do this copy, we first truncate the table 'customer_copy' and then copy the data from 'customer' table

Hint : Use ADF to solve this problem.