

Custom Learnings

Day 7

Azure Data Factory

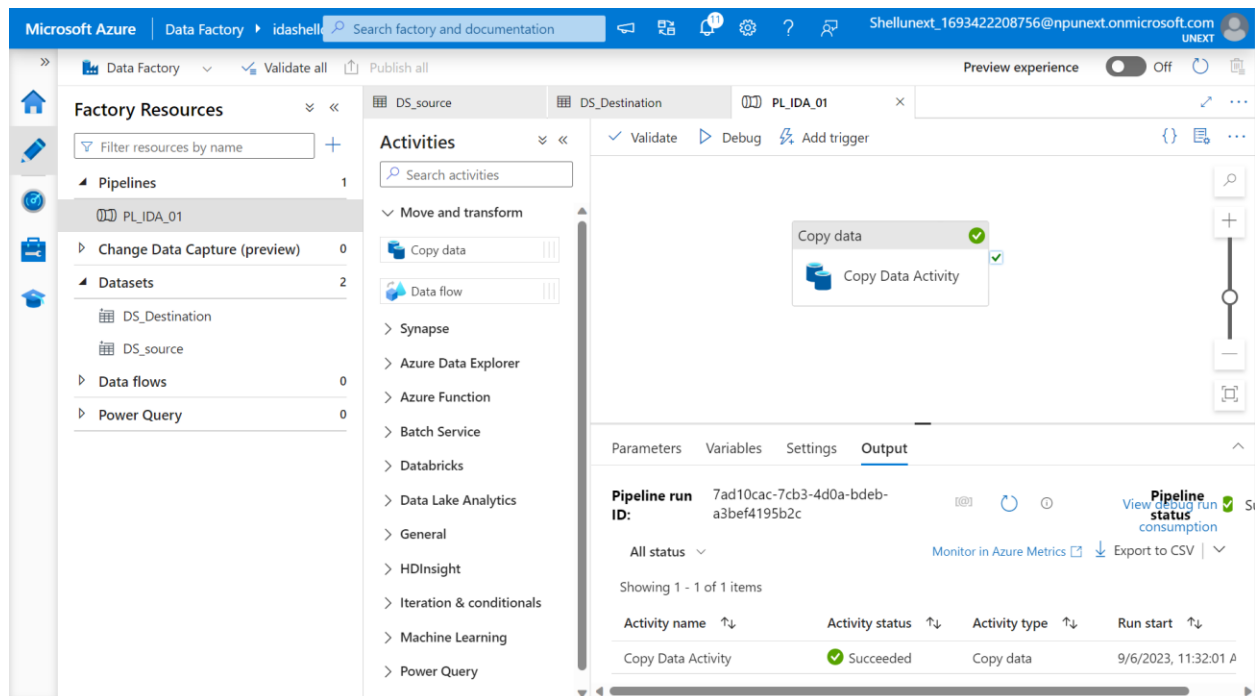
General Activities:

1. Append Variable: Supported only in array.
2. Set Variable: It is used to set/change the value of a variable.
3. Delete: Delete the values

Within a single pipeline we cannot have more than 40 activities. For this we can have execute pipeline activity. This is known as chained pipelines.

4. Lookup: To get the information about the table we can use Lookup activity.
5. System Variables: Used for audit purpose. When the pipeline runs and completes. We can create audit files and tables.
6. MetaData: To get the information about the data or files or folders like the size, name, etc.
7. Filter: To filter out the data based on some conditions.
8. Web Hook: To call an API, and perform crud operations.
9. Wait: To wait for sometime between the execution of two activities.

Pipeline creation without parameters:



The screenshot displays the Microsoft Azure Data Factory (ADF) console. On the left, the 'Factory Resources' pane shows a pipeline named 'PL_IDA_01' under the 'Pipelines' section. The main canvas shows the pipeline design with a 'Copy data' activity. The 'Copy Data Activity' is highlighted, and its 'Output' tab is active. The 'Output' tab shows the pipeline run ID '7ad10cac-7cb3-4d0a-bdeb-a3bef4195b2c' and a status of 'Succeeded'. The 'Activity name' is 'Copy Data Activity', the 'Activity status' is 'Succeeded', the 'Activity type' is 'Copy data', and the 'Run start' time is '9/6/2023, 11:32:01 A'.

Microsoft Azure Search resources, services, and docs (G+/)

Home > idashellsa03 | Containers >

input 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: input

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

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> invoice_data.csv	9/6/2023, 11:41:28 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> zipcodes.csv	9/6/2023, 11:15:28 AM	Hot (Inferred)		Block blob

Microsoft Azure Search resources, services, and docs (G+/)

Home > idashellsa03 | Containers >

output 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: output

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

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> zipcodes.csv	9/6/2023, 11:32:15 AM	Hot (Inferred)		Block blob

Pipeline with parameters:

Microsoft Azure Data Factory idashellsa03 Search factory and documentation

Validate all Publish all Preview experience Off

Factory Resources Filter resources by name

- Pipelines 1
 - PL_IDA_01
- Change Data Capture (preview) 0
- Datasets 2
 - DS_Destination
 - DS_source
- Data flows 0
- Power Query 0

Activities Search activities

- Move and 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 Data Activity

Sink dataset DS_Destination

Dataset properties

Name	Value	Type
DS_outputpar	@pipeline().paramet	string

Copy behavior Select...

Microsoft Azure

Data Factory | idashells

Search factory and documentation

Shellunext_1693422208756@npunext.onmicrosoft.com

UNETX

»

Data Factory

Validate all

Publish all

Preview experience

Off

Factory Resources

Filter resources by name

Pipelines1

PL_IDA_01

Change Data Capture (preview)0

Datasets2

DS_Destination

DS_source

Data flows0

Power Query0

Activities

Search activities

Move and 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

DS_source

DS_Destination

PL_IDA_01

Validate

Debug

Add trigger

Copy data

Copy Data Activity

Parameters

Variables

Settings

Output

Pipeline run ID: d72c2fbd-ebdd-45ad-b7cf-d72742662fd8

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start
Copy Data Activity	Succeeded	Copy data	9/6/2023, 11:55:43 A

Microsoft Azure

Search resources, services, and docs (G+/)

Shellunext_1693422208...

UNETX (NPUNEXT.ONMICROSO...

Home > idashellsa03 | Containers >

input

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: input

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
invoice_data.csv	9/6/2023, 11:41:28 AM	Hot (Inferred)		Block blob
zipcodes.csv	9/6/2023, 11:15:28 AM	Hot (Inferred)		Block blob

Microsoft Azure | Search resources, services, and docs (G+/)

Home > idashellsa03 | Containers >

output

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: output

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

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> copy_invoice_data.csv	9/6/2023, 11:55:51 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> zipcodes.csv	9/6/2023, 11:32:15 AM	Hot (Inferred)		Block blob

Settings

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

To get the meta data of input and output:

Microsoft Azure | Data Factory | idashellsa03 | Search factory and documentation

Validate all | Publish all

Preview experience: Off

Factory Resources

Filter resources by name

- Pipelines: 1
 - PL_IDA_01
- Datasets: 2
 - DS_Destination
 - DS_source
- Data flows: 0
- Power Query: 0

Activities

Search activities

- General
 - Append variable
 - Delete
 - Execute Pipeline
 - Execute SSIS package
 - Fail
 - Get Metadata
 - Lookup
 - Stored procedure
 - Script
 - Set variable
 - Validation
 - Web

Copy Data Pipeline (PL_IDA_01)

Copy data → Get Metadata

Copy Data Activity → Get Metadata1

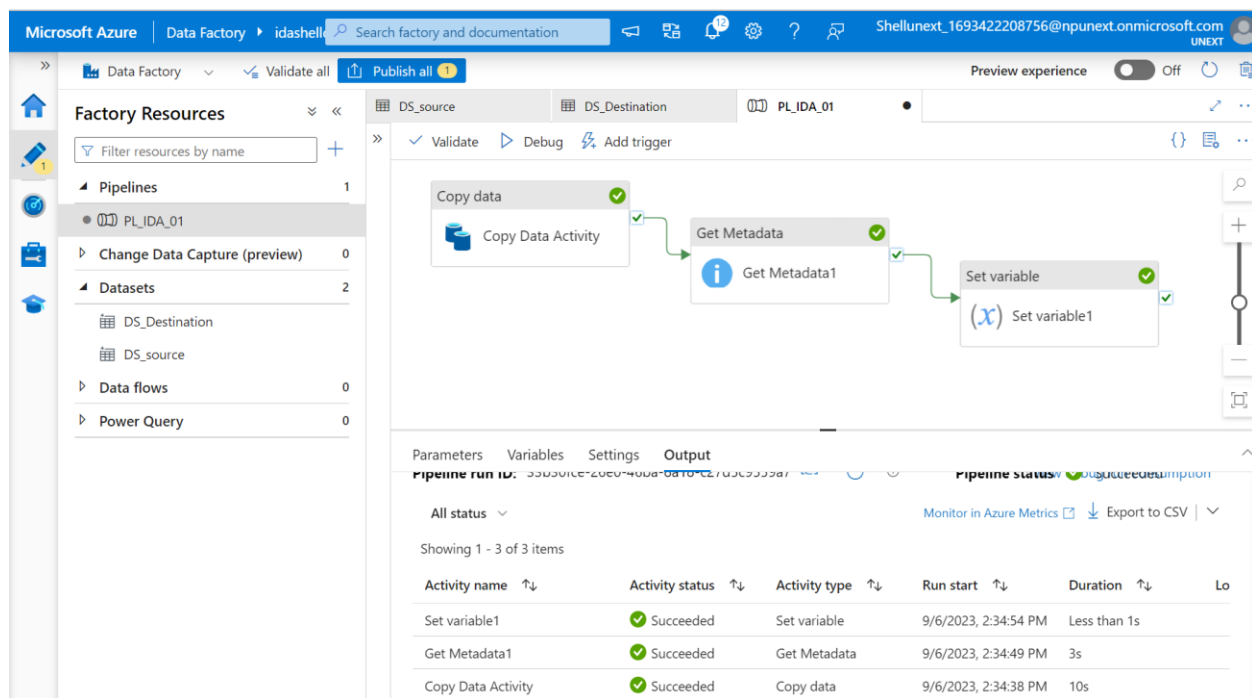
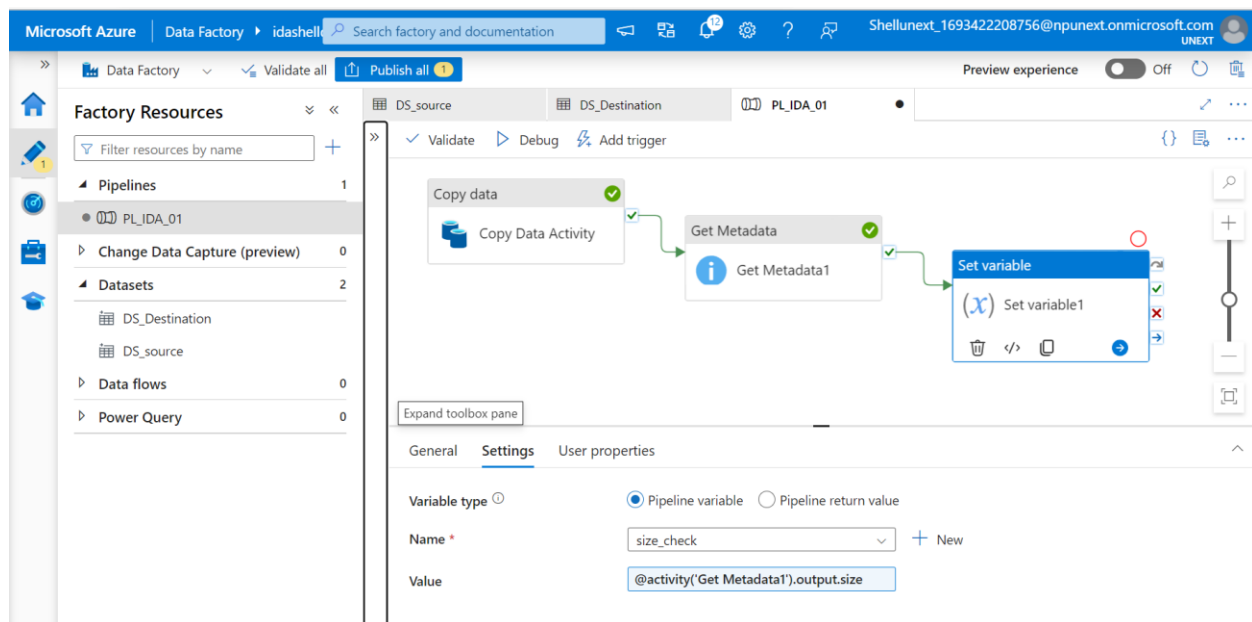
Parameters | Variables | Settings | Output

Pipeline run ID: 3d2e5f9ea106

All status

Showing 1 - 2 of 2 items

Activity name	Activity status	Activity type	Run start
Get Metadata1	Succeeded	Get Metadata	9/6/2023, 2:22:37
Copy Data Activity	Succeeded	Copy data	9/6/2023, 2:22:24



Triggers: Triggers are events or schedules that initiate the execution of data pipelines, activities, or data flows.

There are three types of triggers:

1. **Scheduled Triggers:** These triggers initiate the pipeline or activity execution based on a specified time and recurrence pattern.
2. **Tumbling Window Triggers:** These window triggers are a type of scheduled trigger used for time-based partitioning of data. Specify the window size and offset for partitioning data.
3. **Storage Event Triggers:** These triggers respond to events in azure blob storage or data lake storage, such as file creation, modification, or deletion. Define conditions based on the file path, size, or other attributes.

Monitor: To monitor the information about the trigger runs, pipeline runs, etc. We can filter the monitored logs based on success or failure and various other filters.

If condition task:

The screenshot displays the Microsoft Azure Data Factory console. The pipeline 'PL_IDA_02' is selected, showing a sequence of activities: 'Get Metadata Source', 'Set variable1', and 'If Condition1'. The 'If Condition1' activity has two paths: 'True' leading to 'Copy data' and 'False' leading to a connector. The 'Output' tab at the bottom provides a detailed log of the execution.

Activity name	Activity status	Activity type	Run start	Duration	Log
Copy data	✓ Succeeded	Copy data	9/6/2023, 3:08:23 PM	9s	
If Condition1	✓ Succeeded	If Condition	9/6/2023, 3:08:23 PM	11s	
Set variable1	✓ Succeeded	Set variable	9/6/2023, 3:08:22 PM	Less than 1s	
Get Metadata Source	✓ Succeeded	Get Metadata	9/6/2023, 3:08:19 PM	3s	

There will be more than one file in real world so we will not create if conditions for each and every file. This is a drawback of the above approach.

ForEach Activity:

We can process and check each and every file that was the disadvantage above in if-else condition.

