

## Custom Learnings

Day 9

## Azure Data Factory

Hands-on session:

The screenshot displays the Microsoft Azure Data Factory pipeline editor. The pipeline is named 'PL\_IDA\_01' and is currently in the 'Validate' state. The pipeline consists of the following activities:

- Lookup**: A 'Lookup' activity named 'Lookup1'.
- ForEach**: A 'ForEach' activity named 'ForEach1' that contains a 'Copy data1' activity.

The 'Output' tab at the bottom shows the execution results for the 'Copy data1' activity. The table below summarizes the data:

Activity name	Activity status	Run start	Duration	Integration runtime	User proper
Copy data1	✓ Succeeded	9/8/2023, 6:11:54 AM	12s	AutoResolveIntegration	
Copy data1	✓ Succeeded	9/8/2023, 6:11:54 AM	10s	AutoResolveIntegration	

The screenshot displays the Microsoft Azure portal showing the 'output' container in the 'idashellsa03' storage account. The container lists various CSV files, including SalesLT.Address.csv, SalesLT.Customer.csv, and SalesLT.Product.csv. The table below summarizes the data:

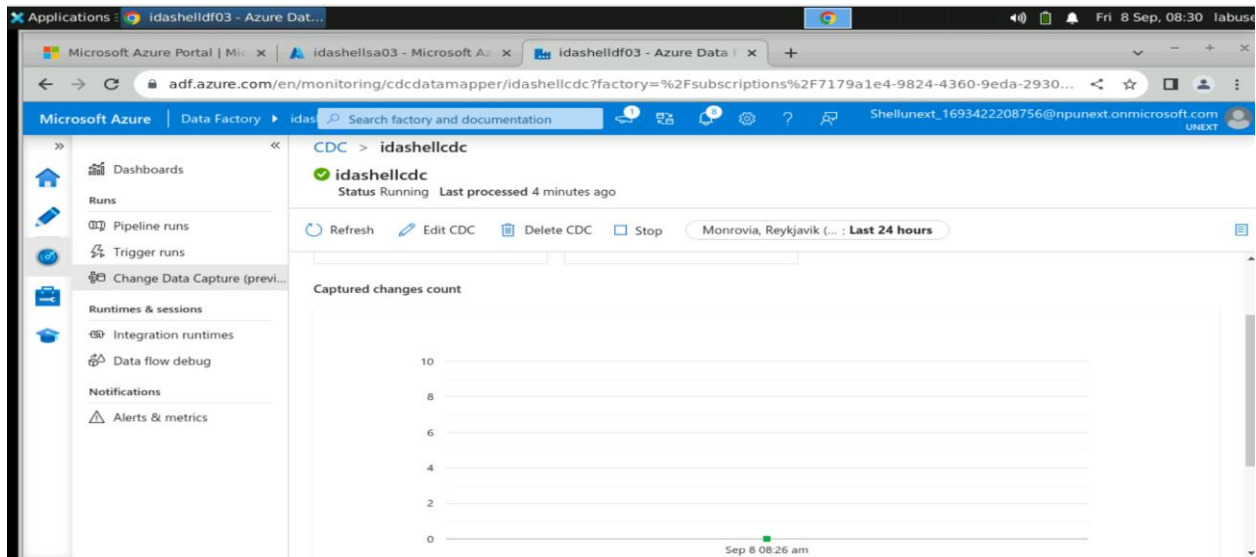
Name	Modified	Access tier	Archive status	Blob type
SalesLT.Address.csv	9/8/2023, 6:12:02 AM	Hot (Inferred)		Block blob
SalesLT.Customer.csv	9/8/2023, 6:12:05 AM	Hot (Inferred)		Block blob
SalesLT.CustomerAddress.csv	9/8/2023, 6:12:00 AM	Hot (Inferred)		Block blob
SalesLT.Product.csv	9/8/2023, 6:12:01 AM	Hot (Inferred)		Block blob
SalesLT.ProductCategory.csv	9/8/2023, 6:12:03 AM	Hot (Inferred)		Block blob
SalesLT.ProductDescription.csv	9/8/2023, 6:12:01 AM	Hot (Inferred)		Block blob
SalesLT.ProductModel.csv	9/8/2023, 6:12:01 AM	Hot (Inferred)		Block blob
SalesLT.ProductModelProductDescription.csv	9/8/2023, 6:12:03 AM	Hot (Inferred)		Block blob
SalesLT.SalesOrderDetail.csv	9/8/2023, 6:12:01 AM	Hot (Inferred)		Block blob
SalesLT.SalesOrderHeader.csv	9/8/2023, 6:12:02 AM	Hot (Inferred)		Block blob
SalesLT.vGetAllCategories.csv	9/8/2023, 6:12:03 AM	Hot (Inferred)		Block blob

Task: creating a cdc:

This screenshot shows the 'Change Data Capture' configuration page in the Microsoft Azure Data Factory portal. The left sidebar displays the 'Factory Resources' tree, with 'idashelldcdc' selected under the 'Change Data Capture (preview)' section. The main area shows the configuration for 'idashelldcdc', including a 'Source Table' dropdown set to 'SalesLT.Customer' and a 'Target Table' dropdown set to 'CustTarget'. The 'Data flow debug' toggle is turned on. A warning message at the top states: 'This preview feature is licensed to you as part of your Azure subscription. By continuing you agree to the Preview Terms and Privacy Statement'.

This screenshot shows the 'Change Data Capture' monitoring page in the Microsoft Azure Data Factory portal. The left sidebar displays the 'Factory Resources' tree, with 'idashelldcdc' selected under the 'Change Data Capture (preview)' section. The main area shows the monitoring view for 'idashelldcdc', including a table with columns: Name, Source, Target, Status, Latency, and Last processed. The table shows one item, 'idashelldcdc', with a status of 'Running' and a latency of '15 minute'.

Name	Source	Target	Status	Latency	Last processed
idashelldcdc	LS_AzureSqlDatabase	LS_AzureDataLake	Running	15 minute	3 minutes ago



Loads:

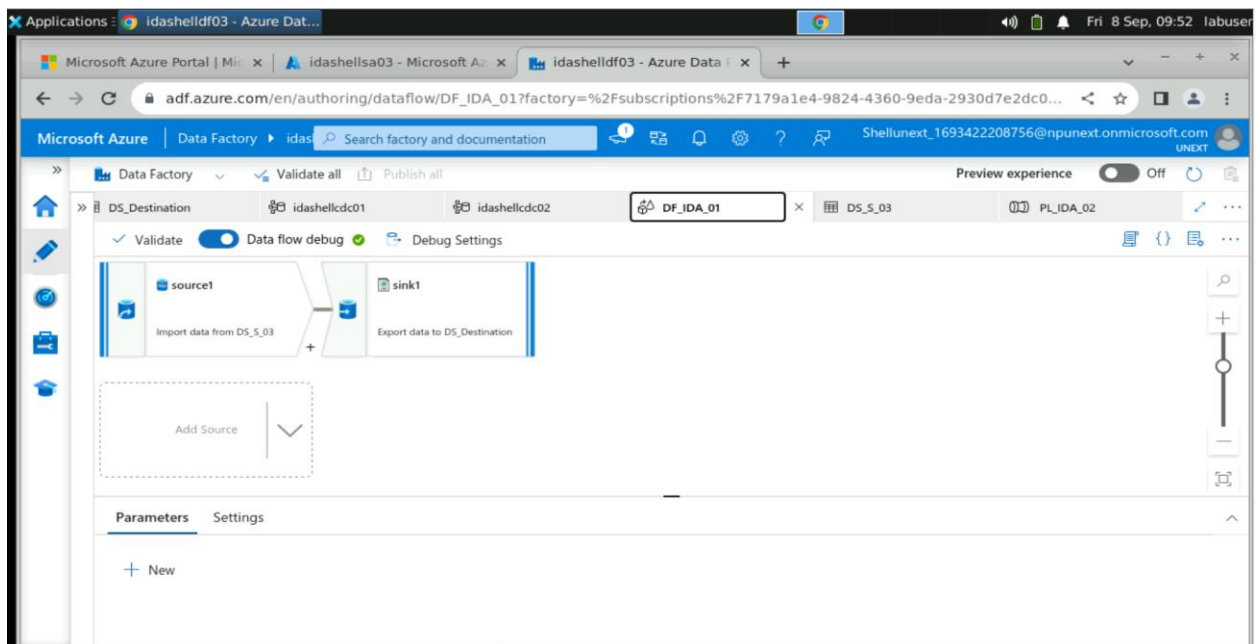
Full Load: Whole data will be loaded.

Incremental Load: Only the changes will get loaded.

Schema Drift:

1. Allows Schema Drift: keeps track of changes in the number of columns.
2. Infer Drifted Column Types: keeps track of changes in the data type of the columns.

Task: creation of cdc through data flow



The screenshot shows the Microsoft Azure Data Factory portal. The top navigation bar includes the Microsoft Azure logo, a search bar, and the user profile. The main content area displays the pipeline 'PL\_IDA\_02' in a 'Data flow' view. The pipeline status is 'Succeeded'. Below the status, a table lists the activities and their details.

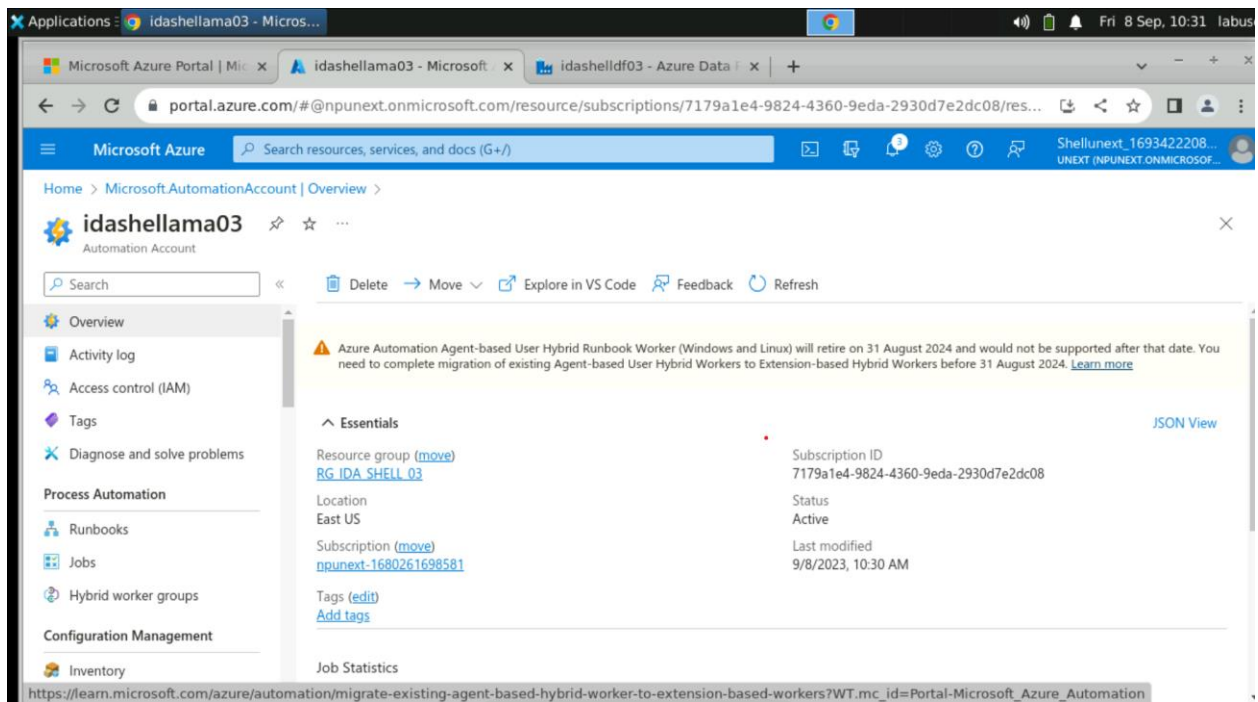
Activity name	Activity status	Run start	Duration	Integration runtime	User properties	Activity run ID
Data flow1	Succeeded	9/8/2023, 9:50:42 AM	1m 22s	debugpool-8Cores-Gei		1d08e54c-586b-4f08-l

The screenshot shows the Microsoft Azure portal displaying the 'output' container. The container is a 'Container' type. The main content area shows a list of blobs with their names, modified dates, access tiers, archive statuses, and blob types.

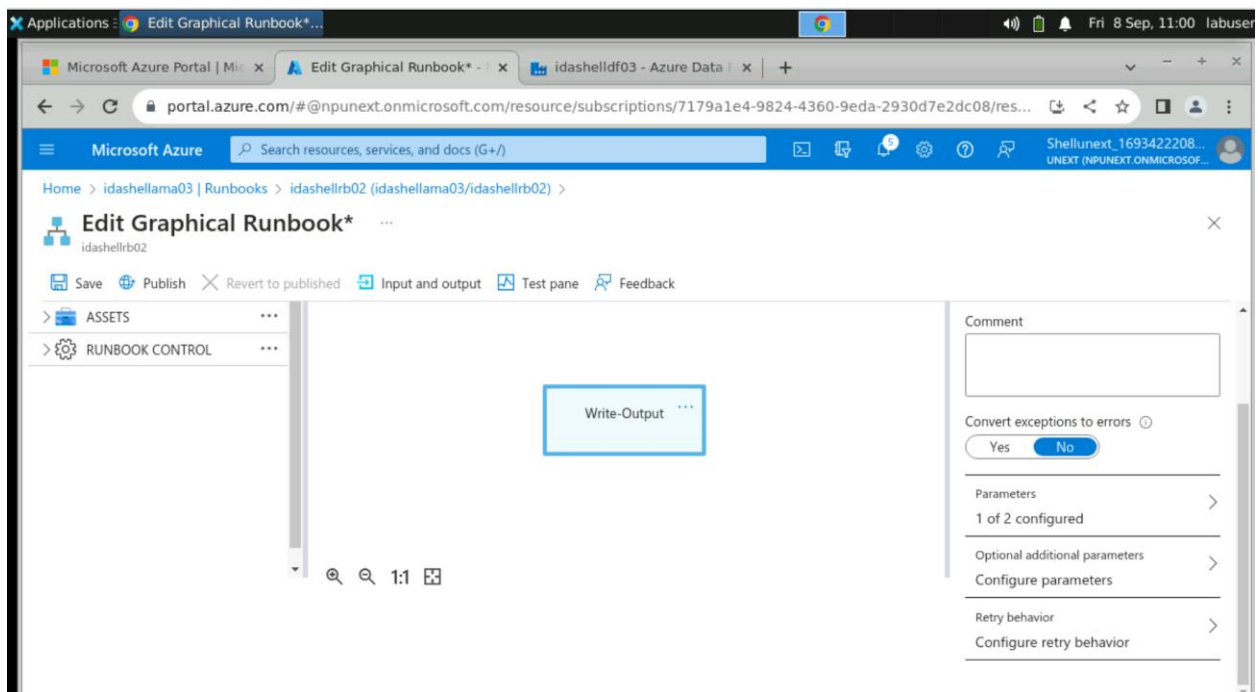
Name	Modified	Access tier	Archive status	Blob type
_SUCCESS	9/8/2023, 9:51:22 AM	Hot (Inferred)		Block blob
part-00000-74fbb7e8-7230-414c-8fe9-d22a710...	9/8/2023, 9:51:21 AM	Hot (Inferred)		Block blob
SalesLT.Address.csv	9/8/2023, 6:12:02 AM	Hot (Inferred)		Block blob
SalesLT.Customer.csv	9/8/2023, 6:12:05 AM	Hot (Inferred)		Block blob
SalesLT.CustomerAddress.csv	9/8/2023, 6:12:00 AM	Hot (Inferred)		Block blob
SalesLT.Product.csv	9/8/2023, 6:12:01 AM	Hot (Inferred)		Block blob
SalesLT.ProductCategory.csv	9/8/2023, 6:12:03 AM	Hot (Inferred)		Block blob
SalesLT.ProductDescription.csv	9/8/2023, 6:12:01 AM	Hot (Inferred)		Block blob

Azure run book:

1. Set of instructions to carry out some executions.
2. Example: Azure Automation Run Book



Creation of sample run book:



Applications

Test - Microsoft Azure - ...

Fri 8 Sep, 11:03 labuse

Microsoft Azure Portal | Mic

Test - Microsoft Azure

idashelldf03 - Azure Data

+

portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/7179a1e4-9824-4360-9eda-2930d7e2dc08/res...

Microsoft Azure

Search resources, services, and docs (G+)

Shellunext\_1693422208...  
UNETX (NPUNEXT.ONMICROSO...

Home > idashellama03 | Runbooks > idashellrb02 (idashellama03/idashellrb02) >

Test

idashellrb02

Start

Stop

Suspend

Resume

View last test

Refresh job streams

Parameters

No input parameters

Run Settings

Run on Azure

Using a hybrid runbook worker can increase test performance. [Learn more](#)

Activity-level tracing

Trace level

None Basic Detailed

No tracing enabled

Completed

test