

Vinodh Kumar TD

Azure Data Architect

Microsoft Certified Trainer

C# Corner MVP

Blogger

21st Jan 2023



Data transformation using Databricks & ADF

Azure Data Factory

ADF Layers

Pipeline: The graphical user interface where you place widgets and draw data paths

Activity: A graphical widget that does something to your data

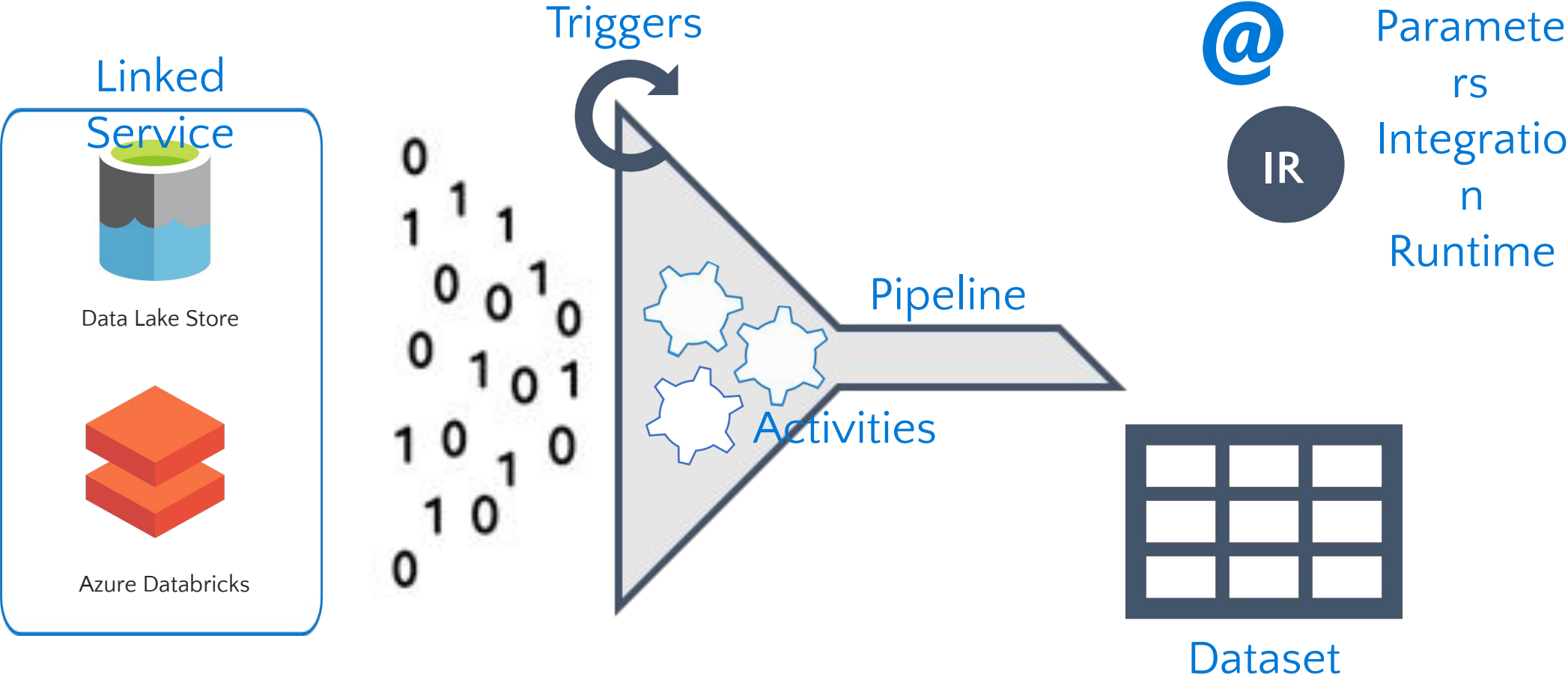
Source and Sink: The parts of an activity that specify where data is coming from and going to

Data Set: An explicitly defined set of data that ADF can operate on

Linked Service: The connection information that allows ADF to access a specific outside data resource

Integration Runtime: A glue/gateway layer that lets ADF talk to software outside of itself

Data integration with Azure Data Factory



Linked Services

Microsoft Azure | Synapse Analytics | asaworkspaceinaday84

Analytics pools

SQL pools

Apache Spark pools

External connections

Linked services

Orchestration

Triggers

Integration runtimes

Security

Access control

Managed private endpoints

Linked services

Linked services are much like connection strings, which define the connection information needed to connect to external data sources.

+ New

Showing 1 - 10 of 10 items

Name	Type
asadatalakeinaday84	Azure Data Lake Storage Gen2
asakeyvaultinaday84	Azure Key Vault
asastoreinaday84	Azure Blob Storage
asaworkspaceinaday84-WorkspaceDefaultSqlServer	Azure Synapse Analytics (for SQL)
asaworkspaceinaday84-WorkspaceDefaultStorage	Azure Data Lake Storage Gen2
sqlpool01	Azure Synapse Analytics (for SQL)
sqlpool01_highperf	Azure Synapse Analytics (for SQL)
sqlpool01_import01	Azure Synapse Analytics (for SQL)
sqlpool01_workload01	Azure Synapse Analytics (for SQL)
sqlpool01_workload02	Azure Synapse Analytics (for SQL)

New linked service

Build interactive reports

Learn more

Connect to Power BI

Search

All

Azure

Compute

Apache Impala

Azure Cosmos DB (SQL API)

Azure Data Lake Storage Gen2

Continue

New linked service (Azure Cosmos DB (SQL API))

Choose a name for your linked service. This name cannot be updated later.

Name *

asacosmosdb01

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Connection string

Azure Key Vault

Account selection method

From Azure subscription

Enter manually

Azure subscription

Select all

Azure Cosmos DB account name *

asacosmosdbinaday84

Database name *

CustomerProfile

Additional connection properties

+ New

Annotations

Create

Back

Connection successful

Test connection

Cancel

Configure and manage secrets in Azure Key Vault

Azure Key Vault protects

1. Secrets
2. Keys
3. Certificates

Data Engineers are typically concerned with accessing the data contained in Key Vault to apply to linked services

The screenshot displays the 'New linked service' configuration interface. At the top, a header reads 'New linked service'. Below it, a light blue information bar states: 'Choose a name for your linked service. This name cannot be updated later.' The form includes several fields: 'Name' (a text input with a placeholder), 'Description' (a larger text area), and 'Connect via integration runtime' (a dropdown menu currently showing 'AutoResolveIntegrationRuntime'). A red rectangular box highlights the 'Connection string' section, which contains two buttons: 'Connection string' and 'Azure Key Vault' (the latter is highlighted in blue). Below this, the 'AKV linked service' dropdown is visible, followed by an 'Edit connection' link. Further down, there are fields for 'Secret name' and 'Secret version' (with a default value of 'Use the latest version if left blank'). At the bottom, there are expandable sections for 'Annotations', 'Parameters', and 'Advanced'.

Datasets

New integration dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All

Azure

Database

File

Generic protocol

NoSQL

Services and apps

Apache Impala

Azure Cosmos DB (SQL API)

Azure Data Lake Storage Gen2

Azure Cosmos DB (MongoDB API)

Azure Data Explorer (Kusto)

Azure Database for MariaDB

Azure Data Lake Storage Gen1

Azure Database for MySQL

Azure Blob Storage

Continue

Cancel

Set properties

Choose a name for your dataset. This name can be updated at any time until it is published.

Name

asal400_customerprofile_cosmosdb

Linked service *

asacosmosdb01

Connect via integration runtime *

AutoResolveIntegrationRuntime

Collection

OnlineUserProfile01

☐ Edit

Import schema

☒ From connection/store

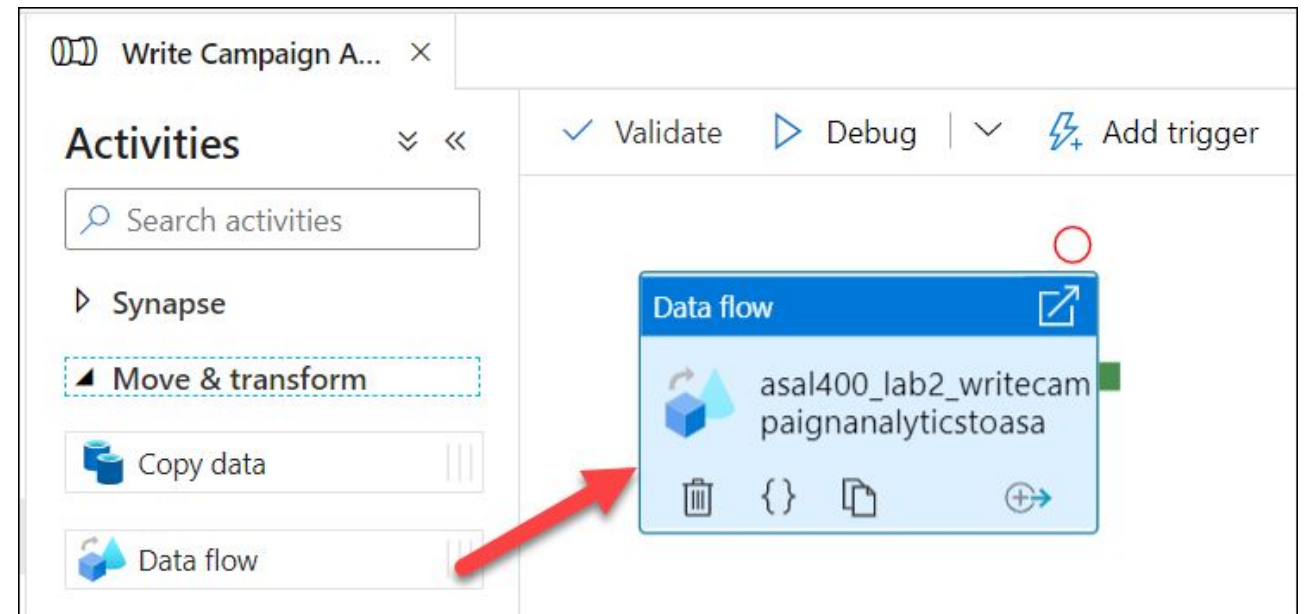
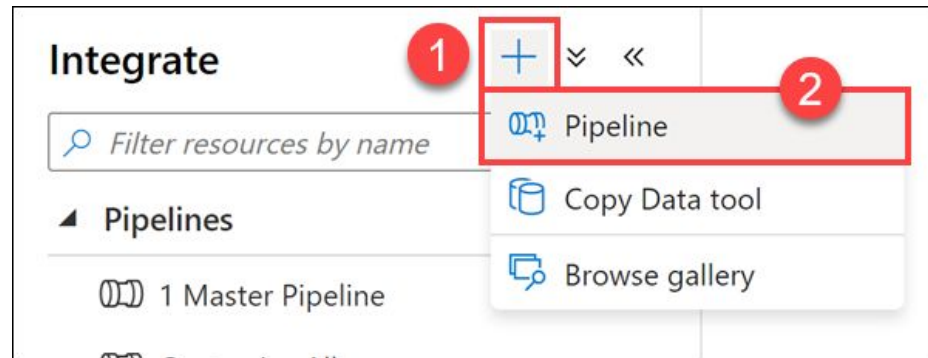
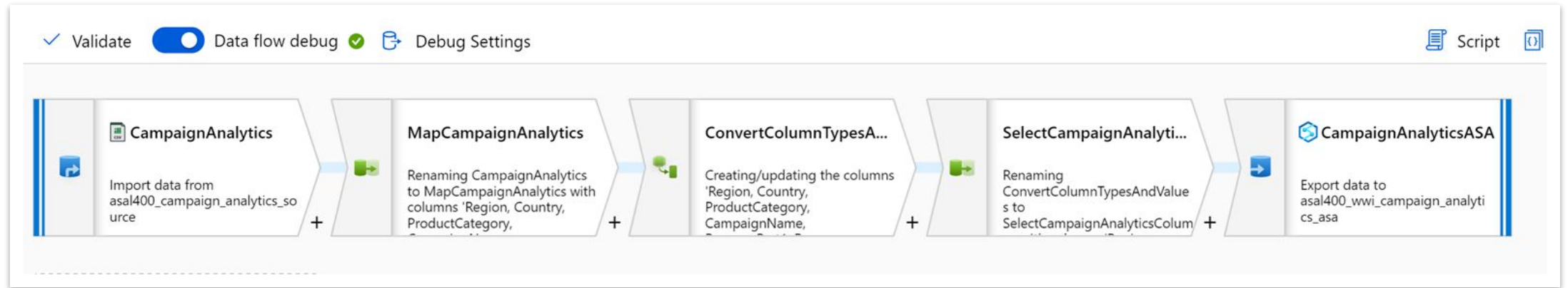
☐ None

OK

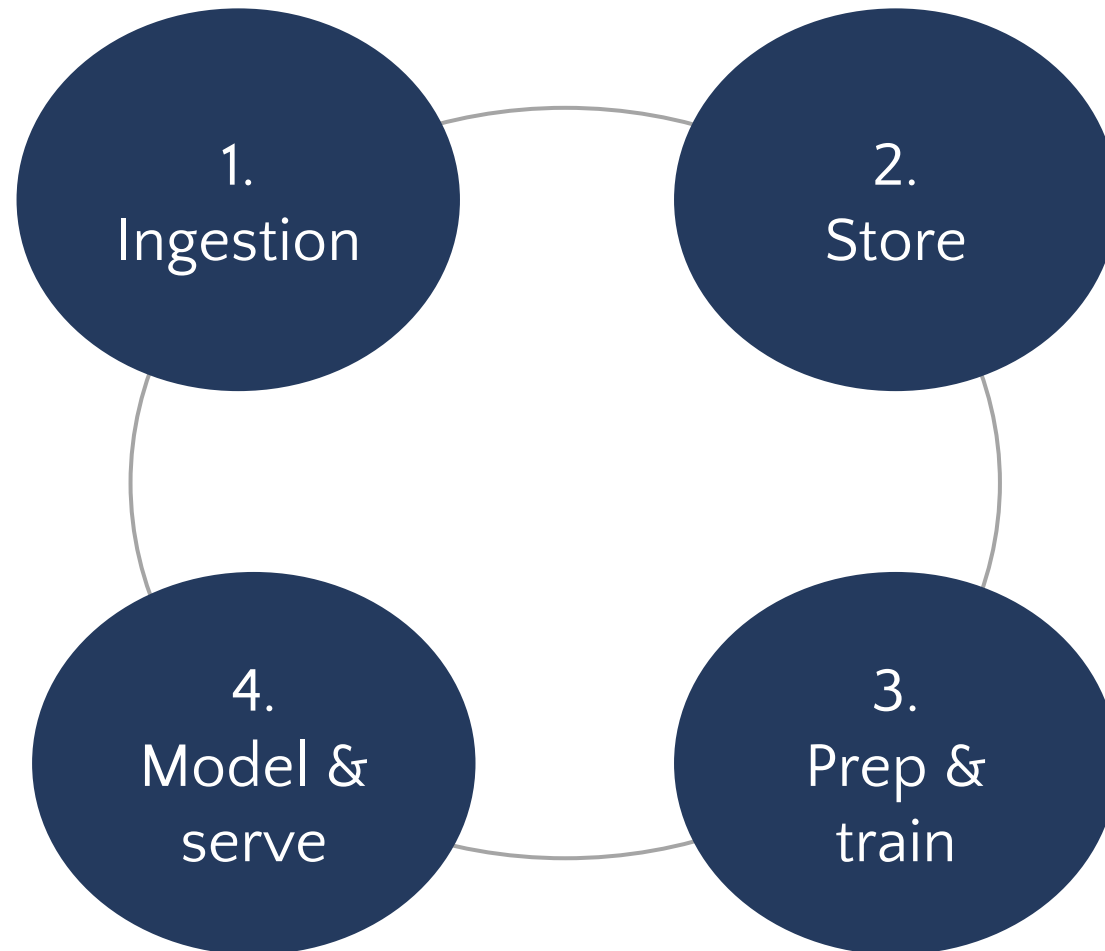
Back

Cancel

Activities and pipelines



Processing Big Data with Azure – Typical usage



Integrating data from Notebooks with Azure Data Factory

1.
Use Storage
Account

2.
Use Azure
Data Factory

3.
Create data
workflow
pipeline

4.
Add a
Notebook
Activity to the
pipeline

5.
Set
parameters
&
dependency
conditions

Programmatically creating Azure Data Factory

The screenshot shows the 'Properties' window for an Azure Data Factory resource. The 'Code' tab is selected, displaying a JSON definition for a linked service. The 'name' property is highlighted with a red box, and the 'url' property is also highlighted with a red box. The 'Apply' button is highlighted with a red box, and the 'Cancel' button is also highlighted with a red box.

Properties

General Related

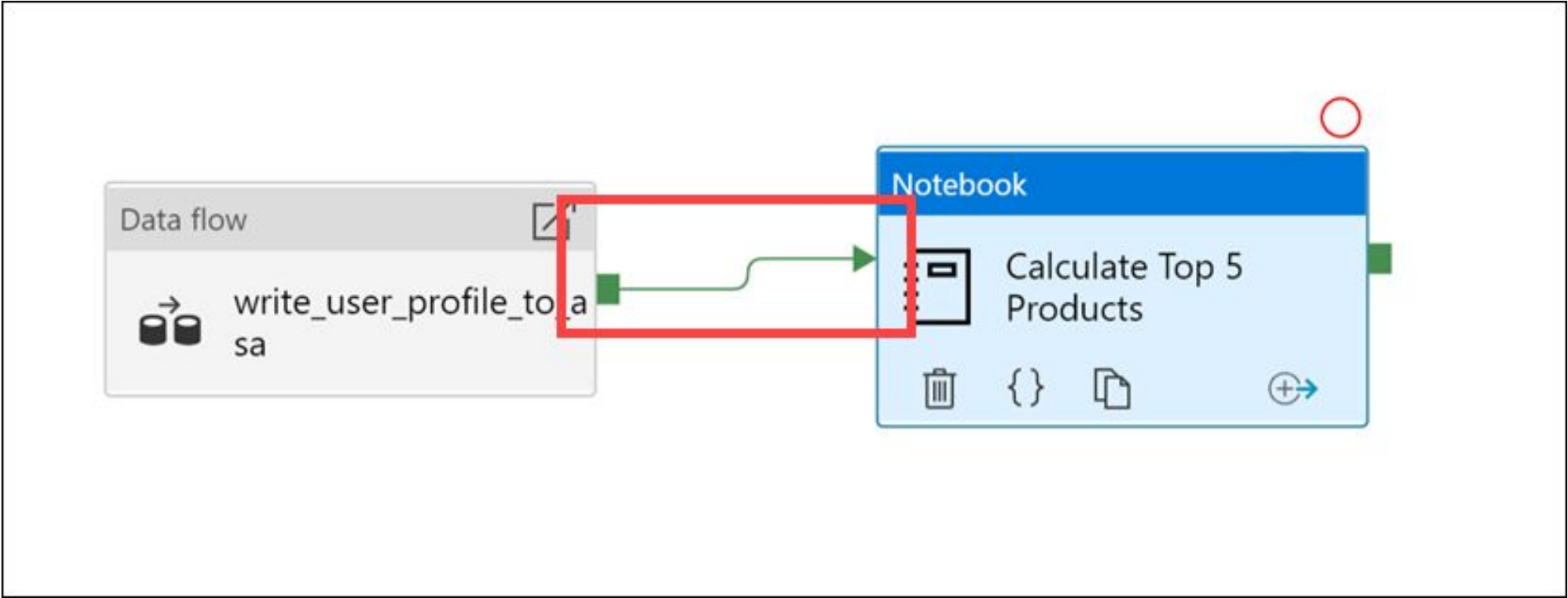
Code

Copy to clipboard

```
1 {  
2   "name": "asadatalake356357",  
3   "type": "Microsoft.Synapse/workspaces/linke  
4   "properties": {  
5     "annotations": [],  
6     "type": "AzureBlobFS",  
7     "typeProperties": {  
8       "url": "https://asadatalake356357.d  
9       "encryptedCredential": "ew0KICAiVmV  
10    }  
11  }  
12 }
```

Apply Cancel

Controlling notebook execution in a pipeline



Viewing the output or results of pipeline activities

Pipeline runs

Triggered

Debug

Rerun

Cancel options

Refresh

Edit columns

List

Gantt

Filter by run ID or name

Chennai, Kolkata, Mu...

Last 24 hours

Pipeline name :

All

Status :

All

Runs :

Latest runs

Copy fi

Triggered by :

All

Add filter



Showing 1 - 100 items






1

2

<div><input type="checkbox"/></div>	<div>Pipeline name <div>↑↓</div></div>	<div>Run start <div>↑↓</div></div>	<div>Run end <div>↑↓</div></div>	<div>Duration</div>	<div>Triggered by</div>	<div>Status <div>↑↓</div></div>	<div>Error</div>
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:10:52 PM</div>	<div>1/18/2023, 7:42:53 PM</div>	<div>00:32:01</div>	<div>8d7fdaa5-d2b2-43a...</div>	<div><div></div><div>Succeeded</div></div>	
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:10:52 PM</div>	<div>--</div>	<div>00:40:35</div>	<div>4bc1530c-2308-430...</div>	<div><div></div><div>In progress</div></div>	
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:10:52 PM</div>	<div>1/18/2023, 7:44:02 PM</div>	<div>00:33:10</div>	<div>baee8cc3-b63d-441...</div>	<div><div></div><div>Succeeded</div></div>	
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:10:52 PM</div>	<div>1/18/2023, 7:32:34 PM</div>	<div>00:21:42</div>	<div>e39cd1e4-612d-4e4...</div>	<div><div></div><div>Succeeded</div></div>	
<div><div><div></div><div>></div></div><div><div></div><div>ingestion_pipeline</div></div></div>	<div>1/18/2023, 7:10:50 PM</div>	<div>--</div>	<div>00:40:37</div>	<div>Manual trigger</div>	<div><div></div><div>In progress</div></div>		
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:03:09 PM</div>	<div>1/18/2023, 7:25:44 PM</div>	<div>00:22:34</div>	<div>142d81d3-ee1a-486...</div>	<div><div></div><div>Succeeded</div></div>	
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:03:09 PM</div>	<div>1/18/2023, 7:26:55 PM</div>	<div>00:23:45</div>	<div>38d62701-fde9-448e...</div>	<div><div></div><div>Succeeded</div></div>	
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:03:09 PM</div>	<div>1/18/2023, 7:25:02 PM</div>	<div>00:21:52</div>	<div>cca6741d-f243-452f...</div>	<div><div></div><div>Succeeded</div></div>	
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:01:36 PM</div>	<div>1/18/2023, 7:02:15 PM</div>	<div>00:00:38</div>	<div>72e90481-708b-454...</div>	<div><div></div><div>Failed</div></div>	<div><div></div></div>
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:01:21 PM</div>	<div>1/18/2023, 7:02:19 PM</div>	<div>00:00:57</div>	<div>3fef7a0c-5d2e-47a2...</div>	<div><div></div><div>Failed</div></div>	<div><div></div></div>
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 7:01:19 PM</div>	<div>1/18/2023, 7:02:22 PM</div>	<div>00:01:02</div>	<div>Manual trigger</div>	<div><div></div><div>Failed</div></div>	<div><div></div></div>
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 6:59:34 PM</div>	<div>1/18/2023, 7:22:37 PM</div>	<div>00:23:02</div>	<div>46fe0d08-59d3-4ab...</div>	<div><div></div><div>Failed</div></div>	<div><div></div></div>
<div><input type="checkbox"/></div>	<div><div></div><div>ingestion_pipeline</div></div>	<div>1/18/2023, 6:59:34 PM</div>	<div>1/18/2023, 7:05:50 PM</div>	<div>00:06:15</div>	<div>fe50b028-476d-486...</div>	<div><div></div><div>Succeeded</div></div>	

Viewing the output or results of pipeline activities

[All pipeline runs](#) >  [Pipeline Master - Activity runs](#) >  [Pipeline Wrapper - Activity runs](#) >







 Rerun  Rerun from activity  Rerun from failed activity  Refresh  Update pipeline List Gantt

Activity runs

Pipeline run ID e0950733-0629-4cbc-832d-c00eeb68e8c6

All status ▾

Showing 1 - 6 items

Activity name ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Status ↑↓
Execute Pipeline	Execute Pipeline	1/17/2023, 7:11:18 PM	00:01:23	 Succeeded
Execute Pipeline	Execute Pipeline	1/17/2023, 6:11:12 PM	01:00:03	 Succeeded
Execute Pipeline	Execute Pipeline	1/17/2023, 6:08:46 PM	00:02:26	 Succeeded
Azure Function	Azure Function	1/17/2023, 6:08:32 PM	00:00:13	 Succeeded
Switch	Switch	1/17/2023, 6:08:31 PM	01:04:26	 Succeeded
Lookup	Lookup	1/17/2023, 6:08:18 PM	00:00:12	 Succeeded