

Azure Data Factory

- Tips & Tricks

About me




- ✓ Passionate about Data 😊
- ✓ Founder at Softentity
- ✓ 8+ years of experience in the industry
- ✓ Certified Microsoft Data Engineer

Connect with me:

 LinkedIn: <https://www.linkedin.com/in/iulianatuhasu/>

 Email: tuhasu.luliana@gmail.com

 Blog: iulianatuhasu.com

Tip: Always check for pipeline dependencies

- Modifying a pipeline without considering its dependencies may lead to unexpected issues.
- Changes in a connected pipeline can have a cascading effect on other processes.



How do you check if a pipeline is called by another pipeline?

Properties -> Related tab

The screenshot displays the Azure Data Studio interface. On the left, the 'Activities' pane lists various components: Move & transform, Synapse, Azure Data Explorer, Azure Function, Batch Service, Databricks, and Data Lake Analytics. The main workspace shows a 'ForEach' activity containing a 'Stored procedure1' activity. On the right, the 'Properties' panel is open, with the 'Related (1)' tab selected. This tab shows a list of pipelines, with 'NestedForeachLoop' highlighted. Below the list, a red message states: 'The current pipeline is called by the NestedForeachLoop pipeline.'

Tip: Invoke ADF pipeline from another pipeline

- Modular and Reusable Design
- Workflow Orchestration (master pipeline that executes multiple pipelines in a sequential or parallel order)
- Ability to add Nested Foreach Logic
- Team collaboration



How to invoke a pipeline from another pipeline?

The screenshot displays the SSDT interface for a pipeline named 'Copy_Invoke_Pipel...'. The design view shows a 'Copy data' task connected to a 'Products' data store, which then flows into an 'Execute Pipeline' task. The 'Execute Pipeline' task is configured to execute 'ExecStoredProced...' (twice). The 'Settings' tab is active, showing the 'Invoked pipeline' dropdown set to 'ExecStoredProcedureWithParams' and the 'Wait on completion' checkbox checked. The 'Open' and 'New' buttons are visible next to the dropdown.

Copy_Invoke_Pipel... X

>> Validate Debug Add trigger

Copy data

Products

Execute Pipeline

ExecStoredProced...
ExecStoredProcedure...

General Settings User properties

Invoked pipeline * ExecStoredProcedureWithParams Open + New

Wait on completion ☒

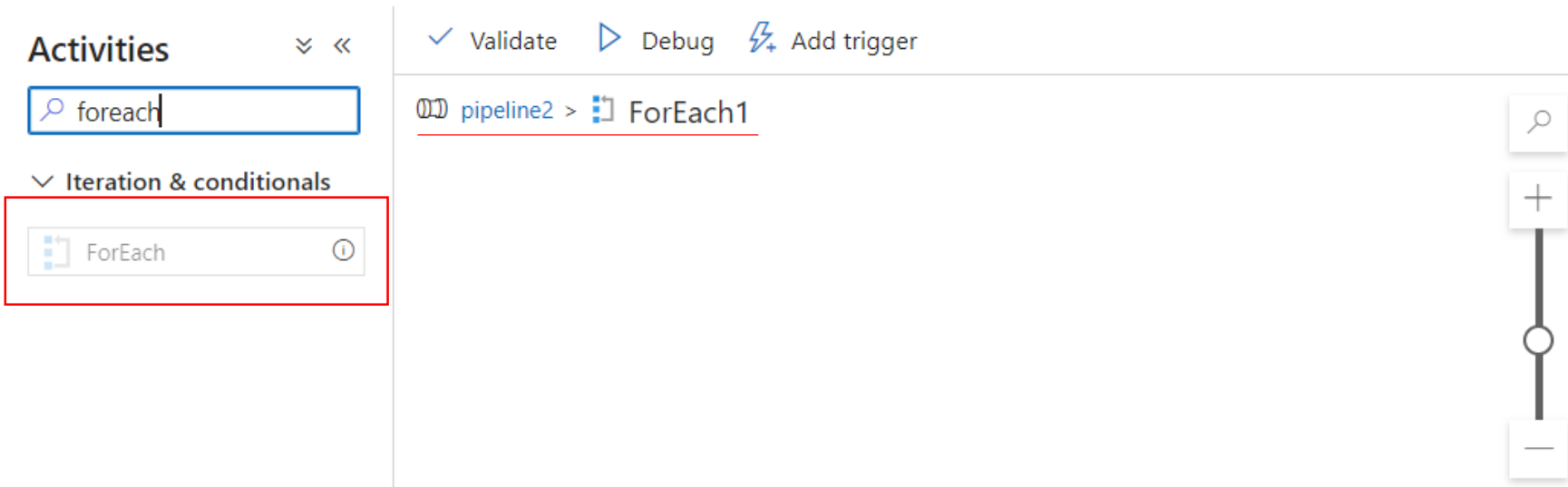
How to use Nested Foreach?

```
- demo2 (Root Folder)
|- 2023 (Year)
  |- 07 (Month)
    |- 23 (Day)
      |- 04 (Hour)
        |- file1.avro (Avro File)
        |- 05 (Hour)
          |- file2.avro (Avro File)
          |- ...
        |- 24 (Day)
          |- 01 (Hour)
            |- file3.avro (Avro File)
            |- 02 (Hour)
              |- file4.avro (Avro File)
              |- ...
          |- 08 (Month)
            |- ...
          |- ...
        |- 2024 (Year)
```

Task: Daily data extraction of .avro files in a .csv file

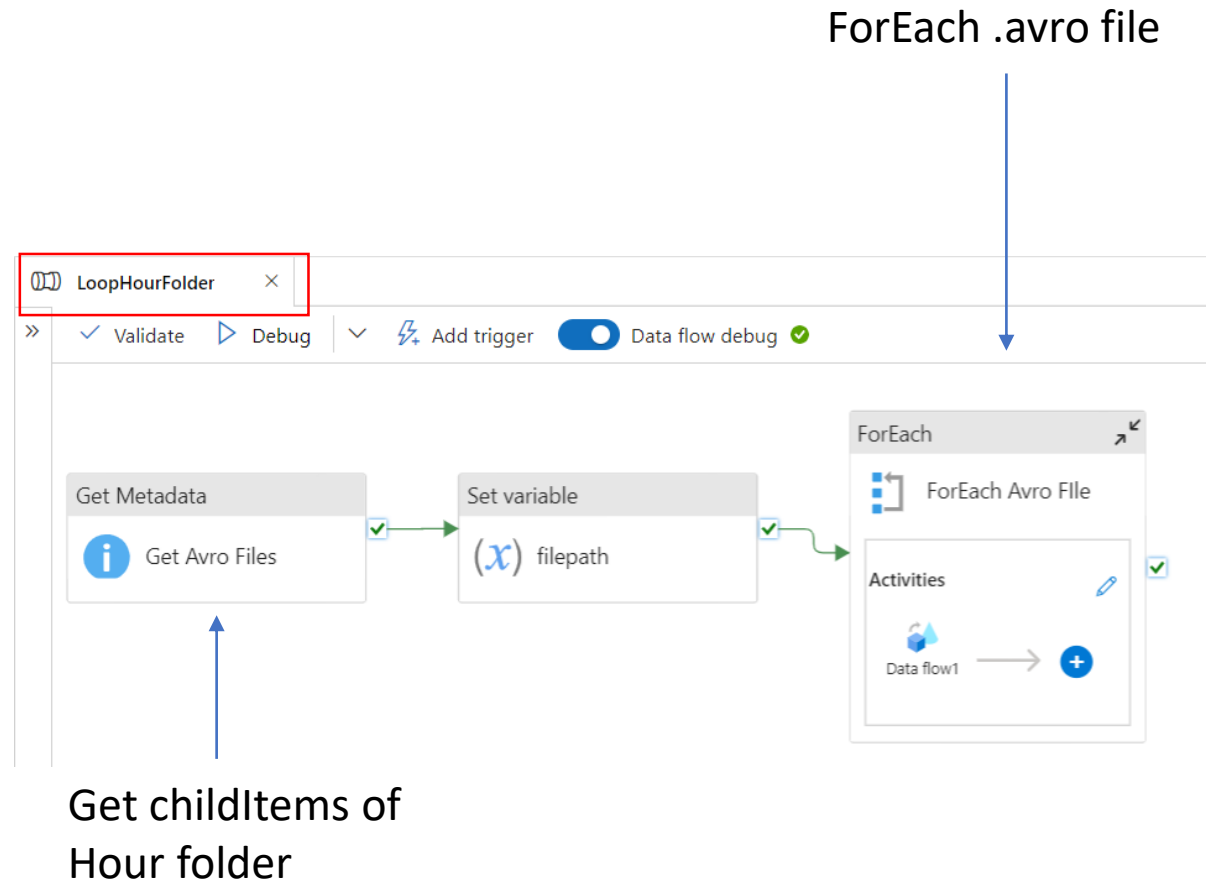
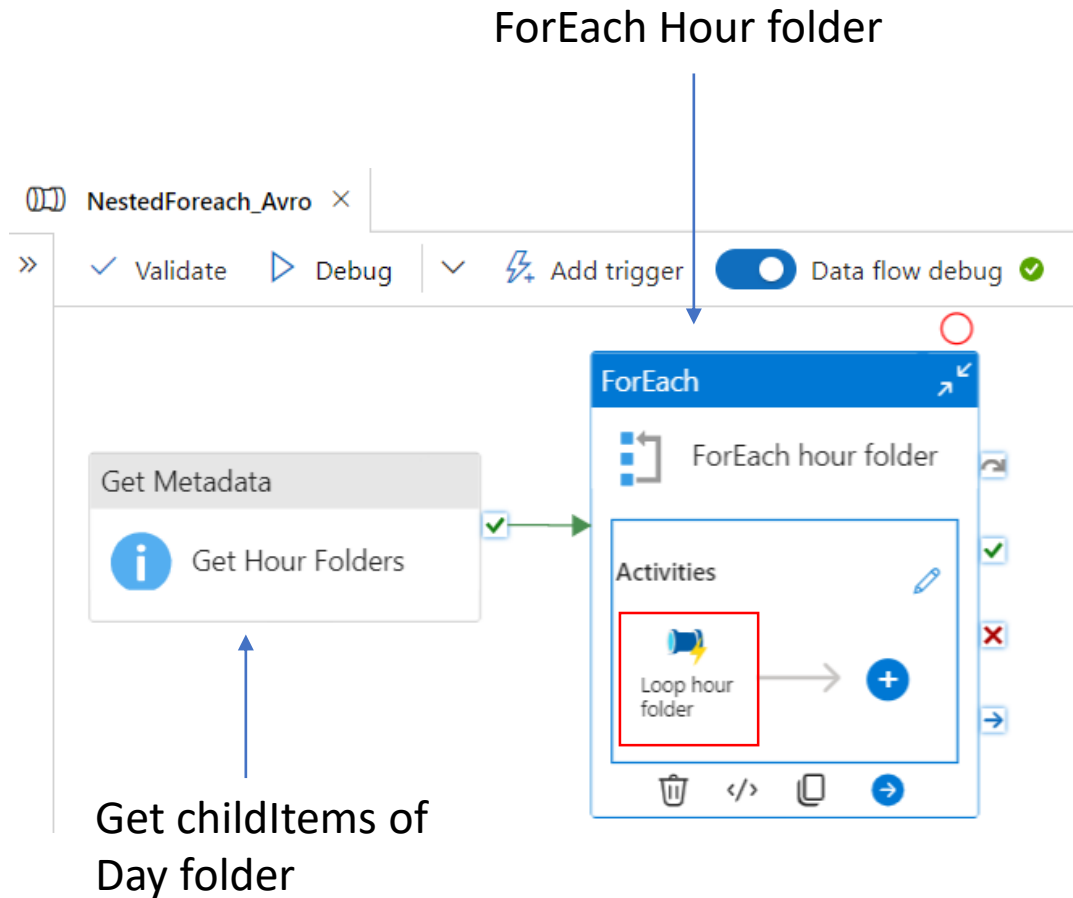
- ☐ Year/Month/Day – from parameters
- ☐ Foreach Hour Folder -> get Avro Files
- ☐ Foreach Avro file -> process & generate a CSV file

How to use Nested Foreach?



Azure Data Factory does not support nested Foreach activities.

How to use Nested Foreach?



Tip: Use Parameters, Global
Parameters and Variables

Parameters, Global Parameters and Variables

Parameters -> defined at the pipeline level, and cannot be modified during a pipeline run

- `@pipeline().parameters.SourceContainer`

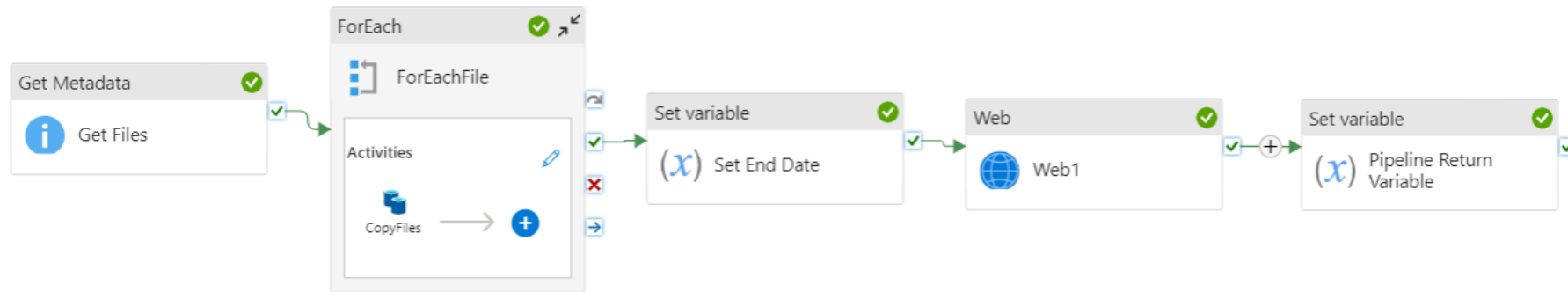
Global Parameters -> constants across a data factory that can be consumed by a pipeline in any expression

- `@pipeline().globalParameters.SendEmail`

Variables -> values that can be set and modified during a pipeline run

- `variables('filepath')`

Parameters, Global Parameters and Variables



Parameters Variables Settings Output

+ New | Delete

<input type="checkbox"/>	Name	Type	Default value	
<input type="checkbox"/>	SourceContainer	String	demo4	
<input type="checkbox"/>	SourceDirectory	String	input	

Tip: Lookup activity used for
insert/update/delete/create
table operations

Lookup activity with CREATE TABLE

create table demo1(id int)

Table is created, but Lookup activity fails.

Error: The specified SQL Query is not valid. It could be caused by that the query doesn't return any data.

create table demo2(id int)

select 1 as abcd

Table is created and Lookup activity runs successfully.

Lookup activity with CREATE SCHEMA



```
create schema demo
```

```
select 1 as abcd
```

Schema is not created, and Lookup activity fails.

Error: Incorrect syntax near the keyword 'select'

```
declare @sql nvarchar(100)
```

```
set @sql = 'create schema demo'
```

```
exec sp_executesql @sql
```

```
select 1 as abcd
```

Schema is created and Lookup activity runs successfully.

Tip: Enable isolated development & testing without impacting existing work

Disable activities within a pipeline:

- Safe testing & debugging
- Ensure that no accidental changes are made
- Facilitate analysis of one activity's performance

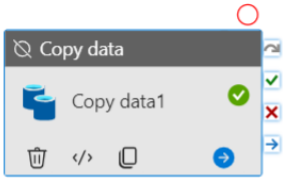
Clone a pipeline:

- Safe experimentation
- Parallel development
- Allows adhoc changes
- Training purposes



How to disable an activity?

Activity state -> Active / Inactive



The screenshot shows the 'Copy data' activity in the Azure Data Factory portal. The activity is highlighted with a red circle. Below the activity, the 'General' tab is selected, showing the 'Activity state (preview)' set to 'Inactive'.

General Source Sink Mapping Settings User properties

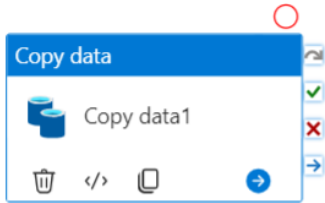
Description

Activity state (preview) ^① ☐ Active ☒ Inactive

Mark activity as

Timeout ^①

Retry ^①



The screenshot shows the 'Copy data' activity in the Azure Data Factory portal. The activity is highlighted with a red circle. Below the activity, the 'General' tab is selected, showing the 'Activity state (preview)' set to 'Active'.

General Source Sink Mapping Settings User properties

Name * [Learn more](#) ^①

Description

Activity state (preview) ^① ☒ Active ☐ Inactive

How to clone a pipeline?

The image illustrates the process of cloning a pipeline in the Azure Data Factory (ADF) interface. It is divided into two panels: the initial state on the left and the state after cloning on the right.

Left Panel (Initial State):

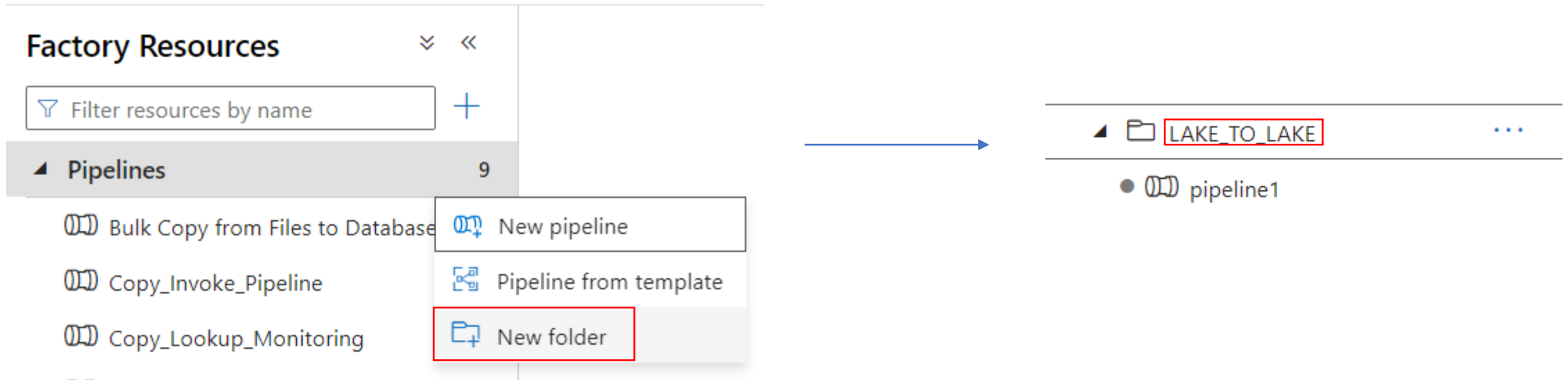
- Factory Resources:** A sidebar with a search bar "Filter resources by name" and a list of resources. Under the "Pipelines" category (indicated by 4 items), "pipeline1" is selected.
- Activities:** A panel showing a search bar "Search activities" and a list of activity categories: "Move & transform", "Synapse", "Azure Data Explorer", and "Azure Function".
- Context Menu:** A right-click context menu is open over "pipeline1". The "Clone" option, represented by a document icon, is highlighted. A blue arrow points from this "Clone" option to the cloned pipeline in the right panel.

Right Panel (After Cloning):

- Factory Resources:** The sidebar now shows 5 pipelines. The list includes "CopyFiles_OneFolderToAnother", "ExecStoredProcedureWithParams", "NestedForeachLoop", "pipeline1", and a newly added "pipeline1_copy1" which is highlighted with a dark grey background and a small black dot to its left.

Tip: Organize your pipelines in folders

How to organize pipelines in folders?



Tip: Always debug your
pipelines before publishing

How to debug a data flow?

Turn on Data flow debug -> Data preview

The screenshot shows the Databricks Dataflow interface. At the top, there's a tab labeled 'dataflow1'. Below it, a toolbar includes 'Validate' (checked), 'Data flow debug' (toggle on), and 'Debug Settings'. The main area displays a data flow diagram with three steps: 'source1' (Import data from stdemodataa001_LS), 'derivedColumn1' (Creating/updating the columns 'name, age, email, is_subscribed, address.street, address.city, address.country, ...'), and 'flatten1' (Unrolling arrays from orders to with columns 'name, email, orders, filename'). The flow ends at 'sink1' (Columns: 6 total). Below the diagram, a tab bar shows 'Sink', 'Settings', 'Errors', 'Mapping', 'Optimize', 'Inspect', and 'Data preview' (selected). The 'Data preview' tab displays a table with 2 rows of data. The table has columns: name, email, order_id, product, quantity, price, and filename. The first row shows 'Mark Doe' with email 'mark.doe@ex...', order_id 'ORD123458', product 'Widget2', quantity '52', price '9.98', and filename 'output csv/2023/07...'. The second row shows 'Mark Doe' with email 'mark.doe@ex...', order_id 'ORD789018', product 'Gadget2', quantity '2', price '29.95', and filename 'output csv/2023/07...'. Above the table, there's a summary bar showing 'Number of rows' and counts for 'INSERT', 'UPDATE', 'DELETE', 'UPSERT', 'LOOKUP', 'ERROR', and 'TOTAL' (2).

name	email	order_id	product	quantity	price	filename
Mark Doe	mark.doe@ex...	ORD123458	Widget2	52	9.98	output csv/2023/07...
Mark Doe	mark.doe@ex...	ORD789018	Gadget2	2	29.95	output csv/2023/07...

How to debug a pipeline?

Use Debug -> Output

Trigger test runs of the current pipeline without publishing your changes to the service

Copy to clipboard

```
"dataset": {  
  "referenceName": "AVRO_DS",  
  "type": "DatasetReference",  
  "parameters": {  
    "year": "2023",  
    "month": "07",  
    "day": 23  
  }  
},
```

Pipeline run ID: b8d5311f-3115-441d-b97a-db9f591ffbdc Pipeline status: Succeeded

Showing 1 - 4 of 4 items

Activity name	Activity status	Activity type	Run start	Duration	Log
Loop hour folder	✓ Succeeded	Execute Pipeline	7/23/2023, 12:41:44 PM	1m 3s	
Loop hour folder	✓ Succeeded	Execute Pipeline	7/23/2023, 12:40:27 PM	1m 17s	
ForEach1	✓ Succeeded	ForEach	7/23/2023, 12:40:26 PM	2m 22s	
Get Hour Folder	✓ Succeeded	Get Metadata	7/23/2023, 12:40:24 PM	2s	

Get Hour Folder

Succeeded

Get Metadata

Tip: When unsure about the input/output format of an activity, try running it in Debug mode.

How to get activity input/output details?

Run in Debug Mode

The screenshot shows the 'Output' window of a development tool. At the top, there's a 'Copy to clipboard' button and a link 'Learn more on output details'. The main content area displays JSON-like output for an activity. The 'errors' field is highlighted with a red box, showing an array with one object. This object contains 'Code': 22301 and 'Message': 'Failure happened on 'Source' side. ErrorCode=SqlOperationFailed, Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException, Message=A database operation failed with the following error: 'Invalid object name'. Below the output, a table lists activities. The 'Products' activity is highlighted, and its status is 'Failed', indicated by a red 'X' icon. The 'Copy data' button is also visible.

Output

Copy to clipboard [Learn more on output details](#)

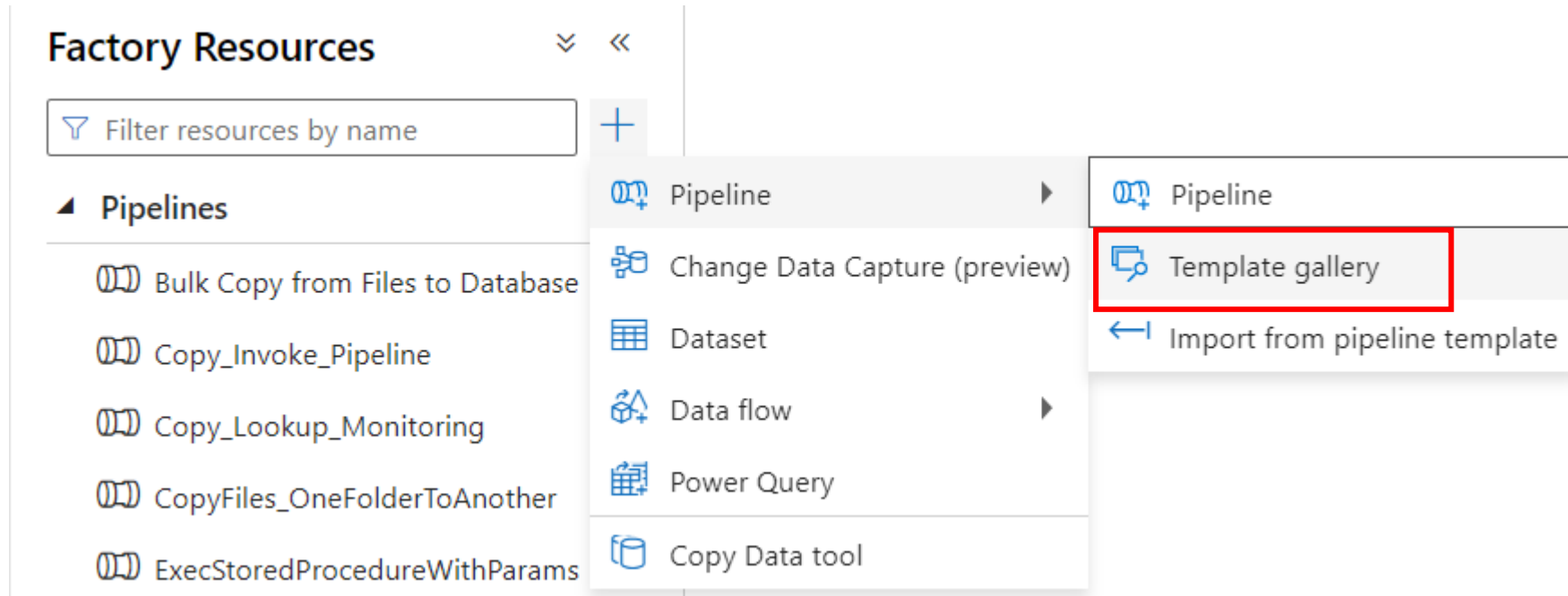
"copyDuration": 6,
"throughput": 0,
"errors":
{
 "Code": 22301,
 "Message": "Failure happened on 'Source' side.
ErrorCode=SqlOperationFailed, Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException, Message=A database operation failed with the following error: 'Invalid object name"

Products [Copy] [Refresh] [Info] [Failed] Copy data

```
string(activity('Products').  
output.errors[0].Message
```

Tip: Create ADF pipelines from
templates whenever possible

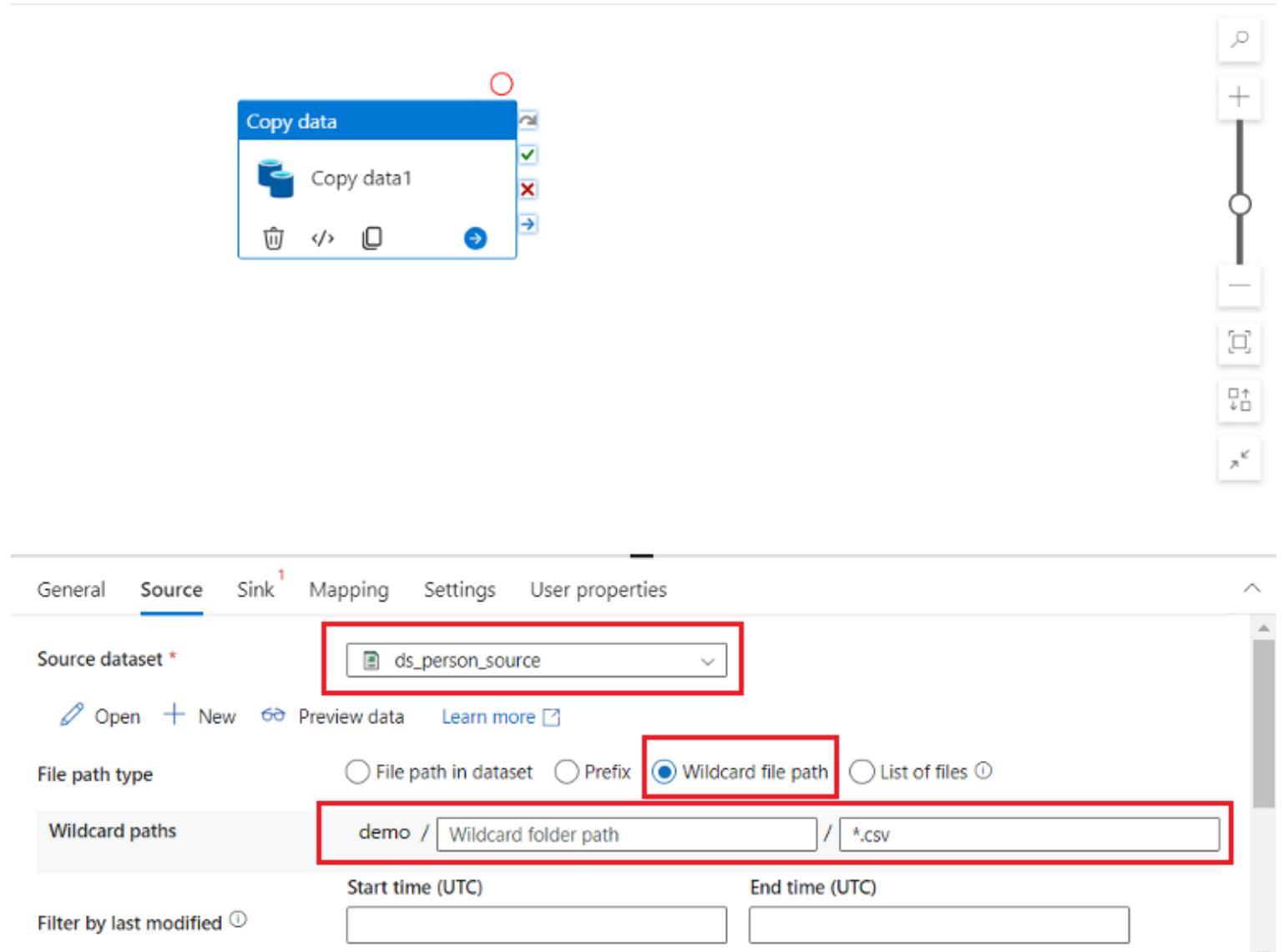
Create ADF pipelines from templates



Tip: Always use Copy data activity for data movement scenarios and basic transformations

How to use Copy data activity?

Scenario: copy .csv files from demo container



How to use Copy data activity?

Scenario: copy .csv files from demo container

Input file structure

Name	Modified
<input type="checkbox"/>  [..]	
<input type="checkbox"/>  2023	

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



Location: [demo / input / 2023 / 06](#)

Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/>  [..]	
<input type="checkbox"/>  16	
<input type="checkbox"/>  17	
<input type="checkbox"/>  18	



Location: [demo / input / 2023 / 06 / 16](#)

Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/>  [..]	
<input type="checkbox"/>  person_20230616.csv	6/24/2023, 7:31:27 PM


Location: [demo / input / 2023 / 06 / 17](#)

Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/>  [..]	
<input type="checkbox"/>  person_20230617.csv	6/24/2023, 7:31:46 PM

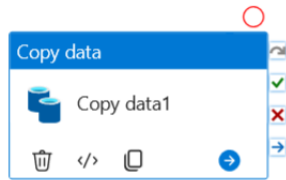
Location: [demo / input / 2023 / 06 / 18](#)

Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/>  [..]	
<input type="checkbox"/>  person_20230618.csv	6/24/2023, 7:31:58 PM

How to use Copy data activity?

Scenario: copy .csv files from demo container using Preserve Hierarchy as Copy behavior



General Source **Sink** Mapping Settings User properties

Sink dataset * [Open](#)

Copy behavior ⓘ

Max concurrent connections ⓘ

Block size (MB) ⓘ

Metadata ⓘ [+ New](#)

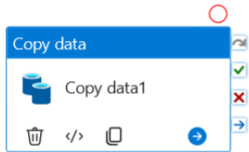
Location: [demo](#) / [output](#) / [2023](#) / [06](#) / [16](#)

☐ Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> person_20230616.csv	6/24/2023, 7:45:48 PM	Hot (Inferred)		Block blob	67 B

How to use Copy data activity?

Scenario: copy .csv files from demo container using Flatten Hierarchy as Copy behavior



General Source **Sink** Mapping Settings User properties

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

Copy behavior Flatten hierarchy

Max concurrent connections

Block size (MB)

Location: [demo](#) / [output](#)

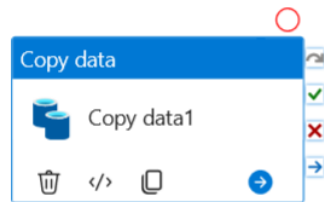
Search blobs by prefix (case-sensitive)

☐ Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> [-]					
<input type="checkbox"/> data_9136028a-309f-42e6-9982-36fdeb146abf_8ce94fa5-8e89-4a6d-a74c-8dba34f2f096.csv	6/24/2023, 7:36:52 PM	Hot (Inferred)		Block blob	67 B
<input type="checkbox"/> data_9136028a-309f-42e6-9982-36fdeb146abf_b3635294-71a7-4774-b94b-27e6ccd1f228.csv	6/24/2023, 7:36:52 PM	Hot (Inferred)		Block blob	67 B
<input type="checkbox"/> data_9136028a-309f-42e6-9982-36fdeb146abf_e2e3c433-5211-4faf-8eeb-30a1ae97f074.csv	6/24/2023, 7:36:52 PM	Hot (Inferred)		Block blob	67 B

How to use Copy data activity?

Scenario: copy .csv files from demo container using Merge Files as Copy behavior



General Source **Sink** Mapping Settings User properties

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

Copy behavior ① Merge files

Max concurrent connections ①

Block size (MB) ①

Metadata ① + New

Location: [demo](#) / output

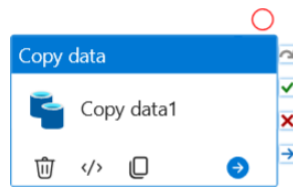
☐ Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> data_bf914172-3ab0-4927-8407-fd427b4609f2_2f85c153-1641-42cf-8878-ce931d4886c6.csv	6/24/2023, 7:52:43 PM	Hot (Inferred)		Block blob	280 B

How to use Copy data activity?

Scenario: copy .csv files from demo container using Merge Files as Copy behavior

Trick: change the autogenerated name -> just add filename.csv in the dataset!



General Source **Sink** Mapping Settings User properties

Sink dataset * [Open](#) [New](#)


Copy behavior ①

Max concurrent connections ①

Block size (MB) ①

Metadata ① [New](#)

[Save](#)

 DelimitedText
ds_person_dest

Connection Schema Parameters

Linked service * [Test connection](#) [Edit](#) [New](#) [Learn more](#)

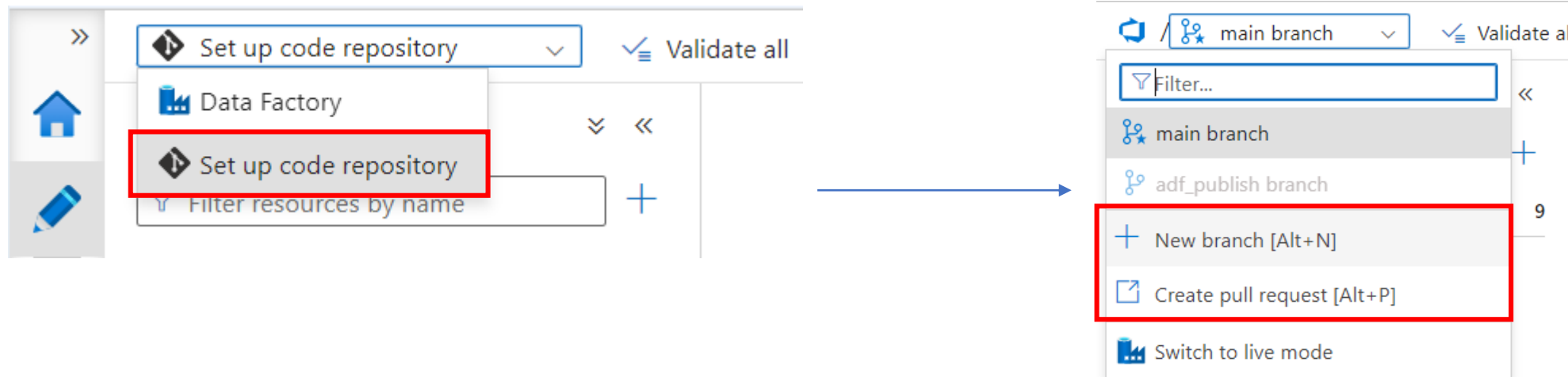
File path * / /

Tip: Set up code repository

A thick, hand-drawn style orange line underlining the text.

How to set up code repository?

Author -> Set up code repository



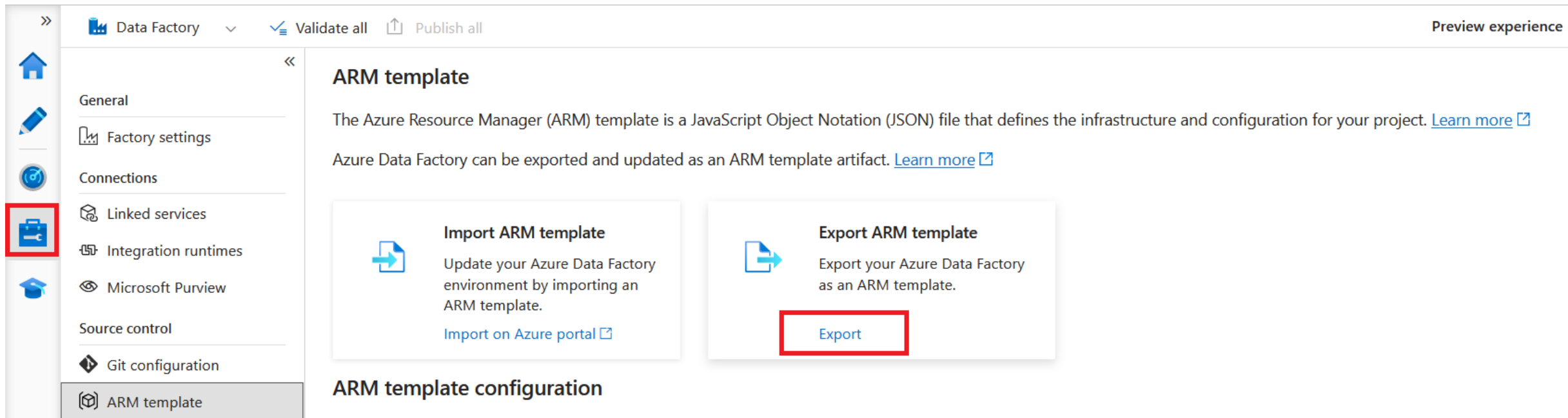
Main branch – collaboration branch

adf_publish – created automatically, contains the code in .json format (ARM templates)

Tip: Always use ARM
templates when you need to
transfer pipelines between
different ADF instances

How to export and import ARM templates?

Export -> Manage tab -> Export



The screenshot shows the Azure Data Factory ARM template management interface. On the left, a sidebar contains navigation icons and labels: Home, Factory settings, Connections, Linked services, Integration runtimes, Microsoft Purview, Source control, Git configuration, and ARM template (highlighted with a red box). The main content area is titled 'ARM template' and includes a description of ARM templates and links to learn more. Below the description are two cards: 'Import ARM template' and 'Export ARM template'. The 'Export ARM template' card has a red box around the 'Export' button. At the bottom, the 'ARM template configuration' section is partially visible.

>> Data Factory Validate all Publish all Preview experience

ARM template

The Azure Resource Manager (ARM) template is a JavaScript Object Notation (JSON) file that defines the infrastructure and configuration for your project. [Learn more](#)

Azure Data Factory can be exported and updated as an ARM template artifact. [Learn more](#)

Import ARM template
Update your Azure Data Factory environment by importing an ARM template.
[Import on Azure portal](#)

Export ARM template
Export your Azure Data Factory as an ARM template.
[Export](#)

ARM template configuration

How to export and import ARM templates?

Import own template – Load ARMTemplateForFactory.json

Custom deployment

Deploy from a custom template

Select a template

Basics

Review + create

Automate deploying resources with Azure Resource Manager. select a template below to get started. [Learn more about ARM templates](#)



Build your own template in the editor

Home > Custom deployment > Edit template

Edit your Azure Resource Manager template

+ Add resource ↑ Quickstart template **↑ Load file** ↓ Download

Parameters (4)

Variables (1)

Resources (12)

- [concat(parameters('factoryName'), '/ds_person_source')] (Microsoft.DataFactory/factories)
- [concat(parameters('factoryName'), '/ds_person_dest')] (Microsoft.DataFactory/factories)
- [concat(parameters('factoryName'), '/DelimitedText1')] (Microsoft.DataFactory/factories)
- [concat(parameters('factoryName'), '/AzureSQL')] (Microsoft.DataFactory/factories)
- [concat(parameters('factoryName'), '/customerOrders')] (Microsoft.DataFactory/factories)
- [concat(parameters('factoryName'), '/demo')] (Microsoft.DataFactory/factories)
- [concat(parameters('factoryName'), '/linkedService1')] (Microsoft.DataFactory/factories)

```
1 {
2   "$schema": "http://schema.management.azure.com/schemas/2015-01-01/deploymentTemplate.json",
3   "contentVersion": "1.0.0.0",
4   "parameters": {
5     "factoryName": {
6       "type": "string",
7       "metadata": "Data Factory name",
8       "defaultValue": "adf-demo-prod"
9     },
10    "demo_connectionString": {
11      "type": "secureString",
12      "metadata": "Secure string for 'connectionString' of 'demo'"
13    },
14    "linkedService1_connectionString": {
15      "type": "secureString",
16      "metadata": "Secure string for 'connectionString' of 'linkedService1'"
17    },
18    "AzureSqlDatabase1_connectionString": {
19      "type": "secureString",
20      "metadata": "Secure string for 'connectionString' of 'AzureSqlDatabase1'"
21    }
22  },
23   "variables": {
24     "factoryId": "[concat('Microsoft.DataFactory/factories/', parameters('factoryName'))]"
25   },
26   "resources": [
```

Change with the name of your current ADF!

Save Discard


How to export and import ARM templates?

Import ARMTemplateForFactory.json & ARMTemplateParametersForFactory.json

[Home](#) >


Custom deployment ...

Deploy from a custom template


 New! Deployment Stacks let you manage the lifecycle of your deployments. Try it now →

Select a template **Basics** Review + create

Template

 Customized template 
12 resources

ARMTemplateForFactory.json

 Edit template

ARMTemplateParametersForFactory.json

 Edit parameters

 Visualize

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ

Azure subscription 1

Resource group * ⓘ

rg-demo-dev-001

[Create new](#)

Instance details

Region * ⓘ

(Europe) West Europe ✓

Factory Name

adf-demo-dev ✓

Demo_connection String *

..... ✓



[Previous](#)

[Next](#)

[Review + create](#)

[Home](#) > [Custom deployment](#) >

Edit parameters ...

 Load file  Download

```
1 {
2   "$schema": "https://schema.management.azure.com/schemas/2015-01-01/deploymentParameters.json#",
3   "contentVersion": "1.0.0.0",
4   "parameters": {
5     "factoryName": {
6       "value": "adf-demo-dev" Change with the name
7                               of your current ADF!
8     },
9     "demo_connectionString": {
10      "value": ""
11    },
12     "linkedService1_connectionString": {
13      "value": ""
14    },
15     "AzureSqlDatabase1_connectionString": {
16      "value": ""
17    }
18  }
```


How to export and import ARM templates?

Review & Create -> Check your ADF imported pipelines


[Home](#) >

Custom deployment

Deploy from a custom template

Select a template Basics **Review + create**

Summary

 Customized template
12 resources

Terms

[Azure Marketplace Terms](#) | [Azure Marketplace](#)

By clicking "Create," I (a) agree to the applicable legal terms associated with the offering; (b) authorize Microsoft to charge or bill my current payment method for the fees associated the offering(s), including applicable taxes, with the same billing frequency as my Azure subscription, until I discontinue use of the offering(s); and (c) agree that, if the deployment involves 3rd party offerings, Microsoft may share my contact information and other details of such deployment with the publisher of that offering.

Microsoft assumes no responsibility for any actions performed by third-party templates and does not provide rights for third-party products or services. See the [Azure Marketplace Terms](#) for additional terms.

Deploying this template will create one or more Azure resources or Marketplace offerings. You acknowledge that you are responsible for reviewing the applicable pricing and legal terms associated with all resources and offerings deployed as part of this template. Prices and associated legal terms for any Marketplace offerings can be found in the [Azure Marketplace](#); both are subject to change at any time prior to deployment.

Neither subscription credits nor monetary commitment funds may be used to purchase non-Microsoft offerings. These purchases are billed separately.

[Previous](#) [Next](#) [Create](#)

Deployment error:

ERROR DETAILS



The Resource 'Microsoft.DataFactory/factories/adf-demo-dev' under resource group 'rg-demo-dev-001' was not found. For more details please go to <https://aka.ms/ARMResourceNotFoundFix> (Code: ResourceNotFound)

Thank you!

