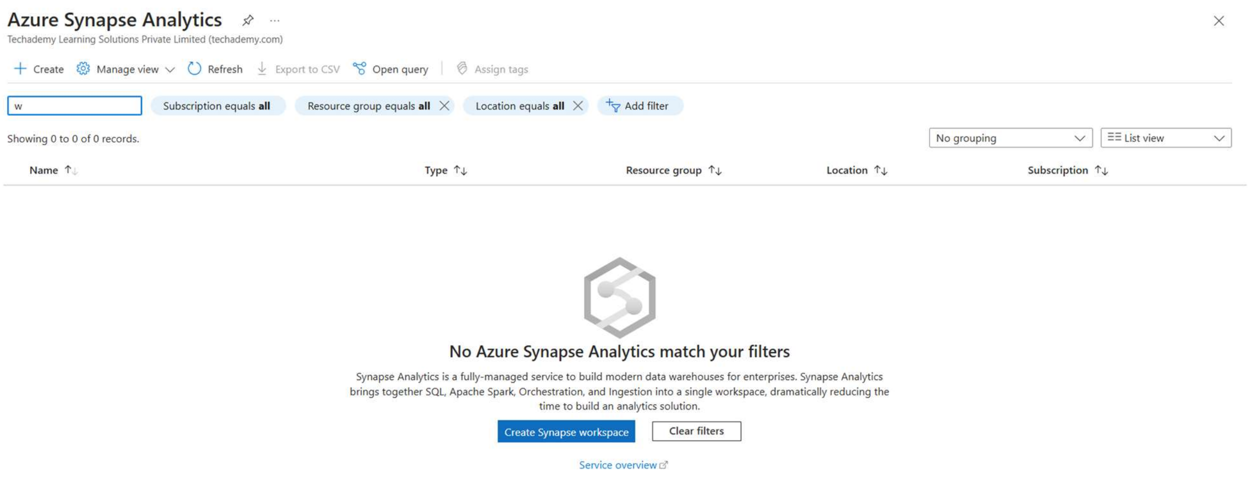
Buiild an etl pipline with azure synapse with dataflow running on it

Read me: page 1 to 5 – steps to do. Page 6 to 8 Conclusion

Create Synapse workspace

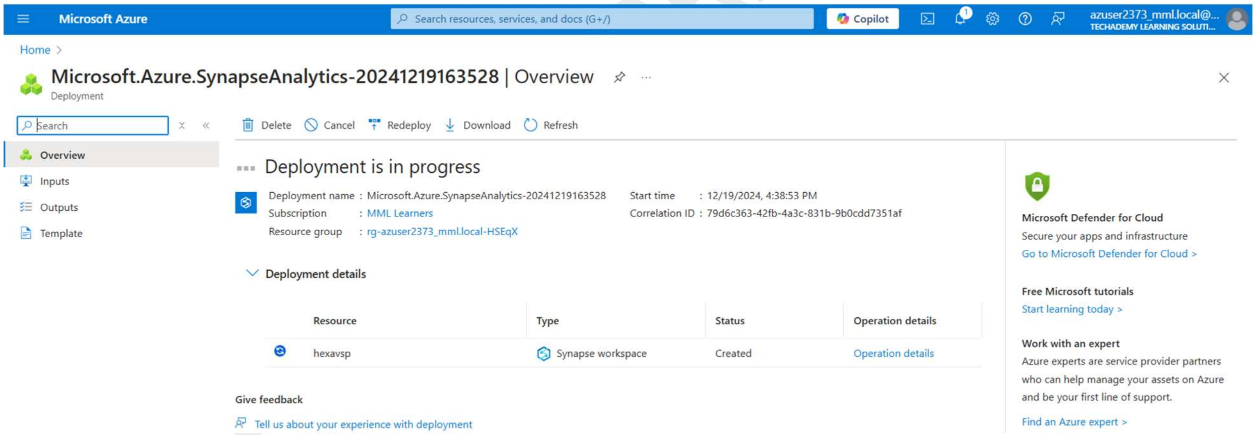


Filter: w

Showing 0 to 0 of 0 records.

No Azure Synapse Analytics match your filters

Synapse Analytics is a fully-managed service to build modern data warehouses for enterprises. Synapse Analytics brings together SQL, Apache Spark, Orchestration, and Ingestion into a single workspace, dramatically reducing the time to build an analytics solution.



Microsoft.Azure.SynapseAnalytics-20241219163528 | Overview

Deployment is in progress

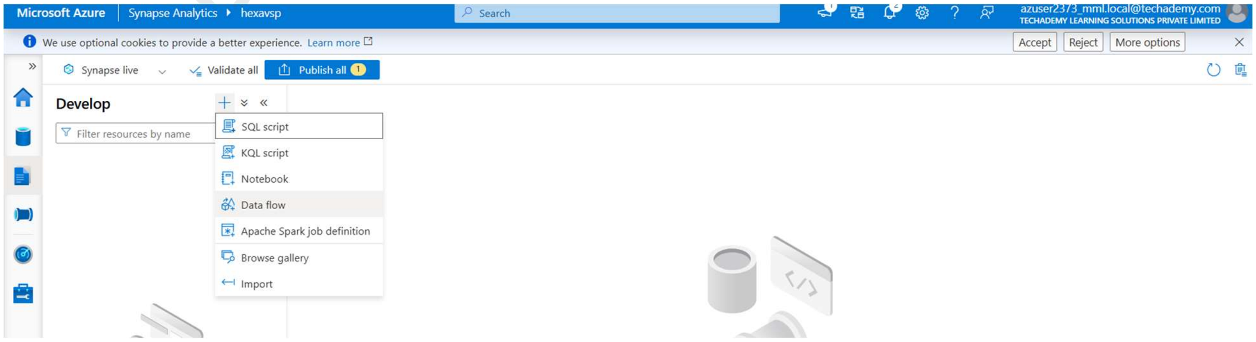
Deployment name : Microsoft.Azure.SynapseAnalytics-20241219163528
Subscription : MML Learners
Resource group : rg-azuser2373_mml.local-HSEqX
Start time : 12/19/2024, 4:38:53 PM
Correlation ID : 79d6c363-42fb-4a3c-831b-9b0cdd7351af

Deployment details

Resource	Type	Status	Operation details
hexavsp	Synapse workspace	Created	Operation details

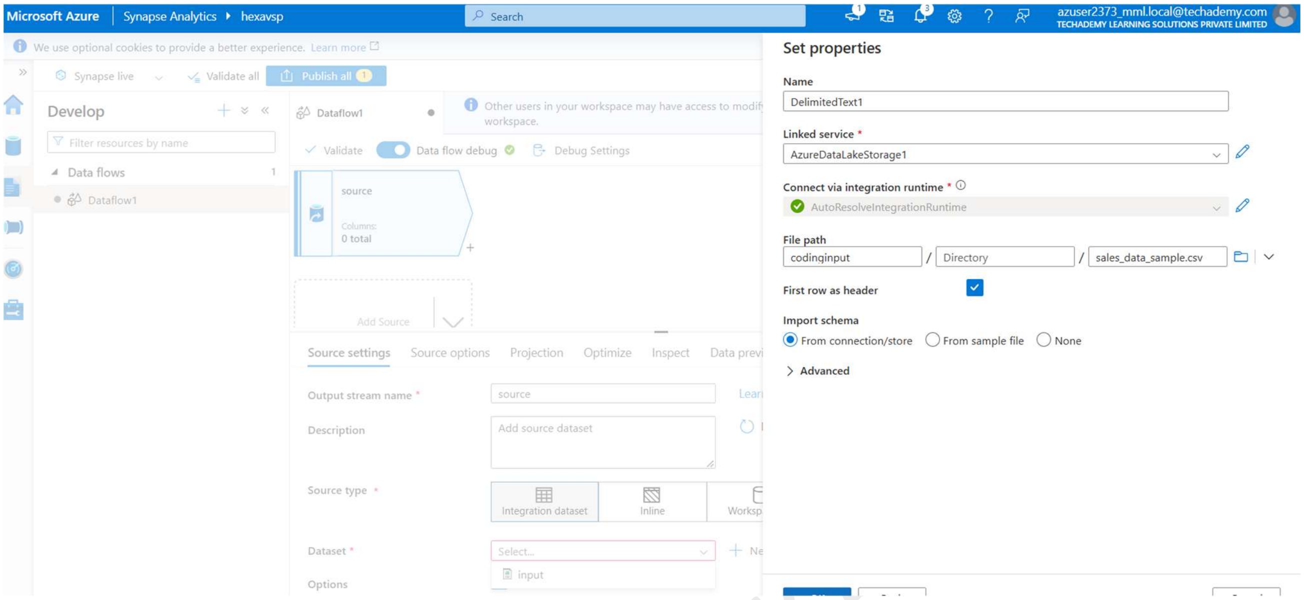
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

Go to develop tab and create dataflow



We use optional cookies to provide a better experience.

Add source


Microsoft Azure | Synapse Analytics › hexavsp
Search
azuser2373_mml.local@techademy.com
TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED
Develop
Data flows
Dataflow1
source
Add Source
Source settings
Source options
Projection
Optimize
Inspect

Set properties

Name: DelimitedText1

Linked service *: AzureDataLakeStorage1

Connect via integration runtime *: AutoResolveIntegrationRuntime

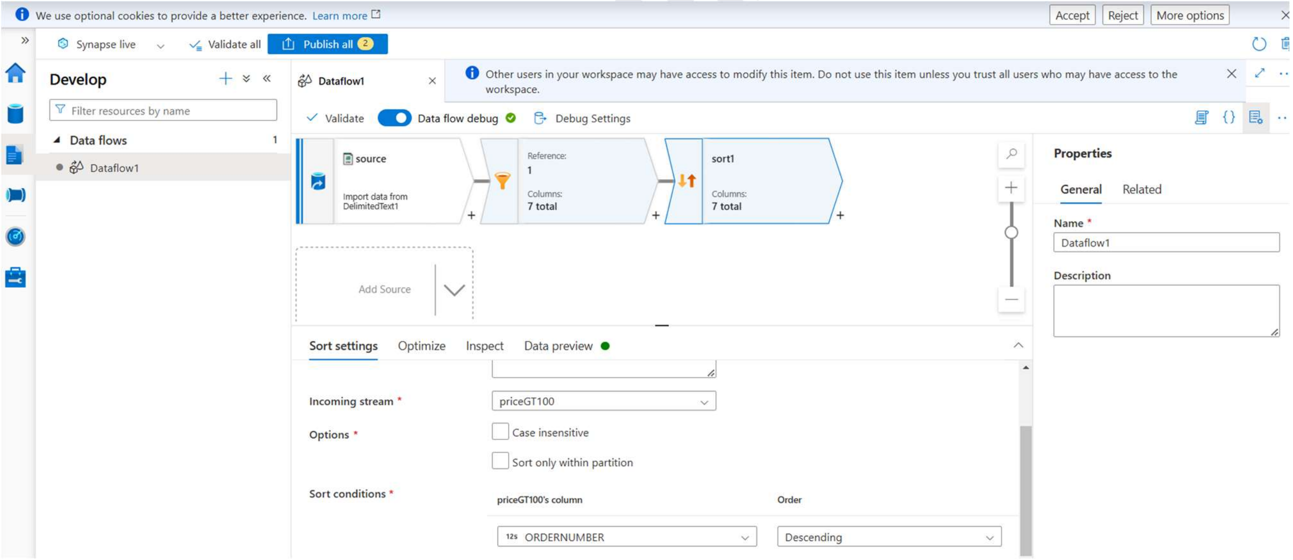
File path: codinginput / Directory / sales_data_sample.csv

First row as header ☑

Import schema: ● From connection/store ○ From sample file ○ None

› Advanced

Give filter condition for data transformation and sort for data transformation



Incoming stream *: priceGT100

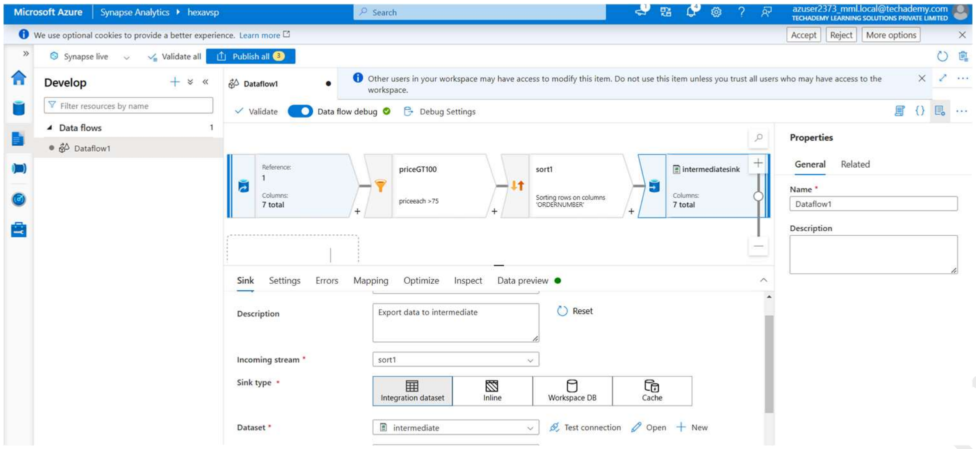
Options *: □ Case insensitive □ Sort only within partition

Sort conditions *:

priceGT100's column	Order
ORDERNUMBER	Descending

Name *: Dataflow1

Validate and publish



We use optional cookies to provide a better experience. Learn more

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Description: Export data to intermediate

Incoming stream *: sort1

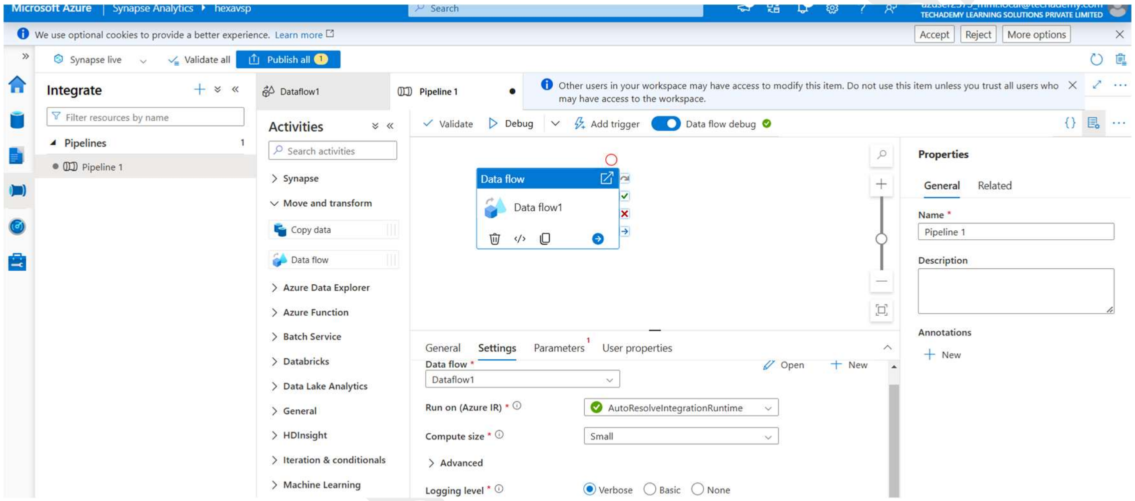
Sink type *: Integration dataset

Dataset *: intermediate

Name *: Dataflow1

Description:

Create pipeline drag and drop dataflow from move and transform and the created dataflow



We use optional cookies to provide a better experience. Learn more

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Data flow *: Dataflow1

Run on (Azure IR) *: AutoResolveIntegrationRuntime

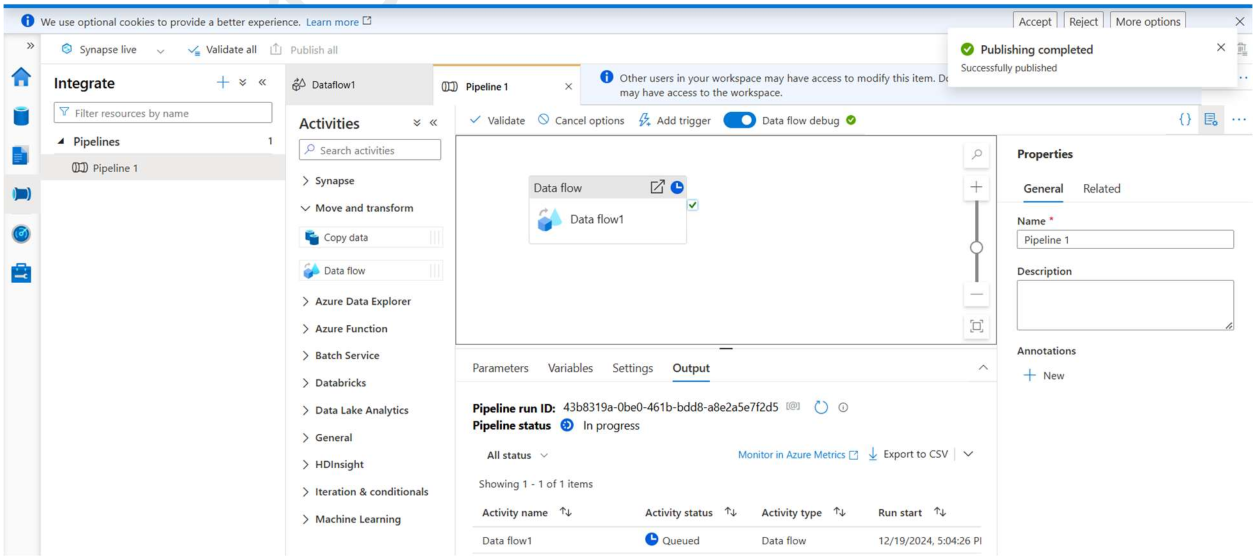
Compute size *: Small

Logging level *: ● Verbose ○ Basic ○ None

Name *: Pipeline 1

Description:

Click on debug to run the pipeline



We use optional cookies to provide a better experience. Learn more

Publishing completed
Successfully published

Other users in your workspace may have access to modify this item. Do
may have access to the workspace.

Pipeline run ID: 43b8319a-0be0-461b-bdd8-a8e2a5e7f2d5

Pipeline status: In progress

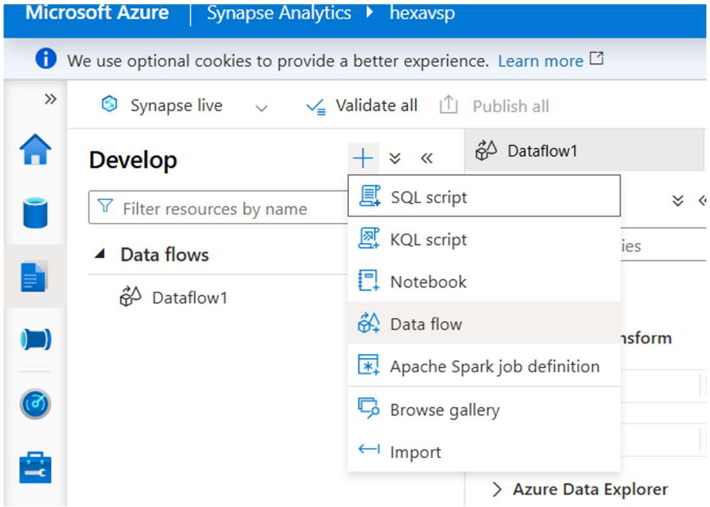
Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start
Data flow1	Queued	Data flow	12/19/2024, 5:04:26 PI

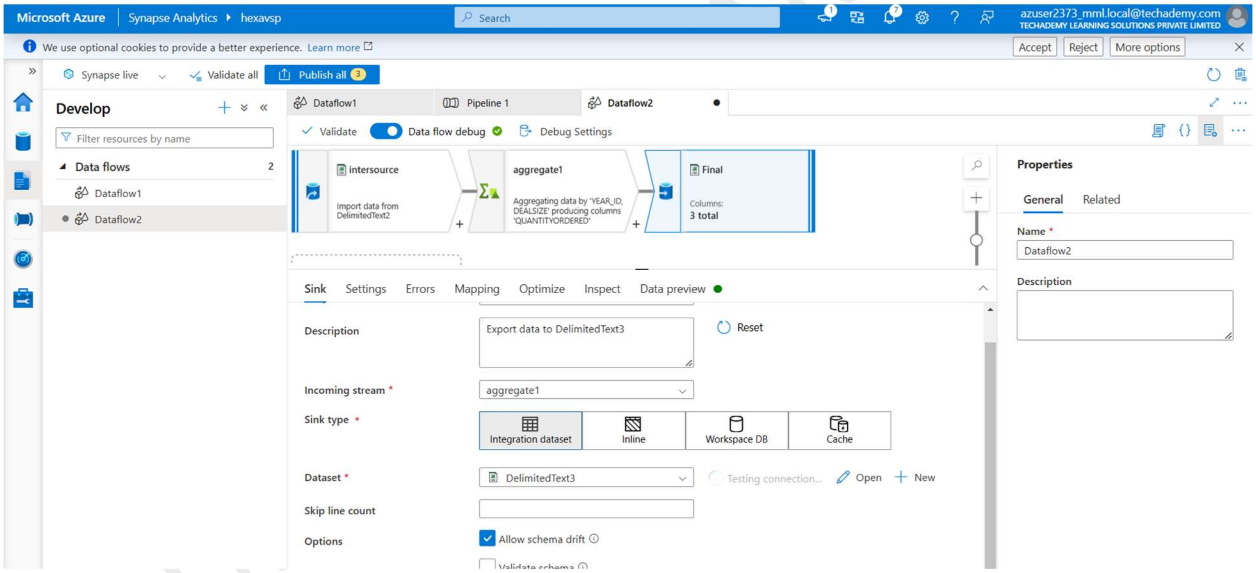
Name *: Pipeline 1

Description:

Create new data flow to see the final results



Aggregate it by year and deal size and calculate the quantity ordered.



Description: Export data to DelimitedText3

Incoming stream *: aggregate1

Sink type *: Integration dataset | Inline | Workspace DB | Cache

Dataset *: DelimitedText3

Skip line count:

Options: ☑ Allow schema drift ⓘ

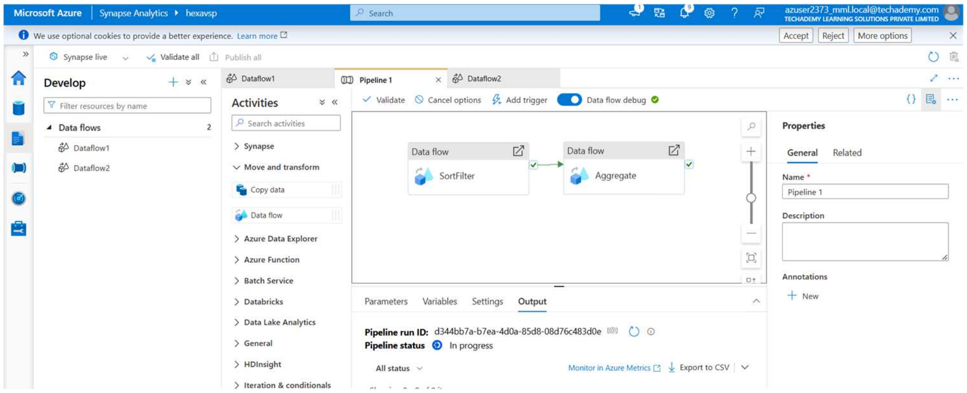
☐ Validate schema ⓘ

Name *: Dataflow2

Description:

Add this dataflow in pipeline and run the pipeline to see the results.

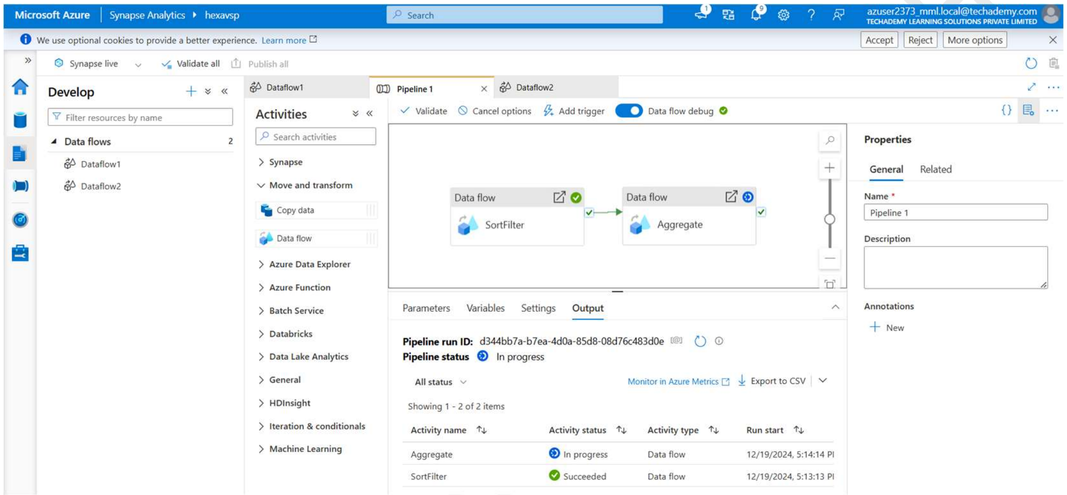
Pipeline started



Pipeline run ID: d344bb7a-b7ea-4d0a-85d8-08d76c483d0e

Pipeline status In progress

Sort and filter dataflow succeeded now aggregation dataflow is running.



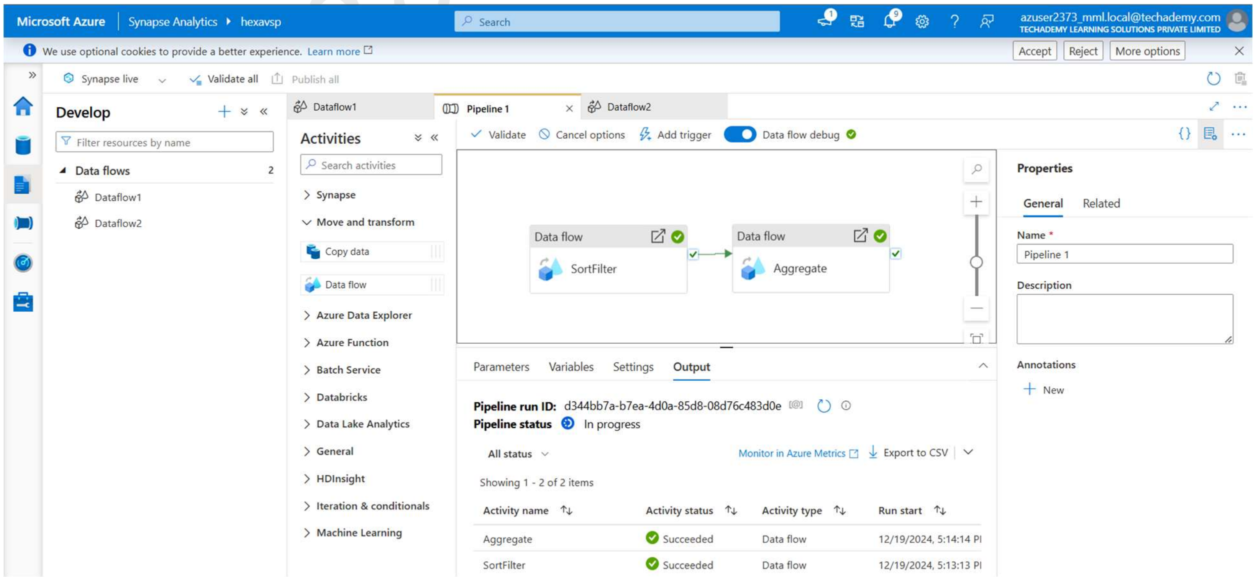
Pipeline run ID: d344bb7a-b7ea-4d0a-85d8-08d76c483d0e

Pipeline status In progress

Showing 1 - 2 of 2 items

Activity name	Activity status	Activity type	Run start
Aggregate	In progress	Data flow	12/19/2024, 5:14:14 PI
SortFilter	✓ Succeeded	Data flow	12/19/2024, 5:13:13 PI

Both ran successfully



Pipeline run ID: d344bb7a-b7ea-4d0a-85d8-08d76c483d0e

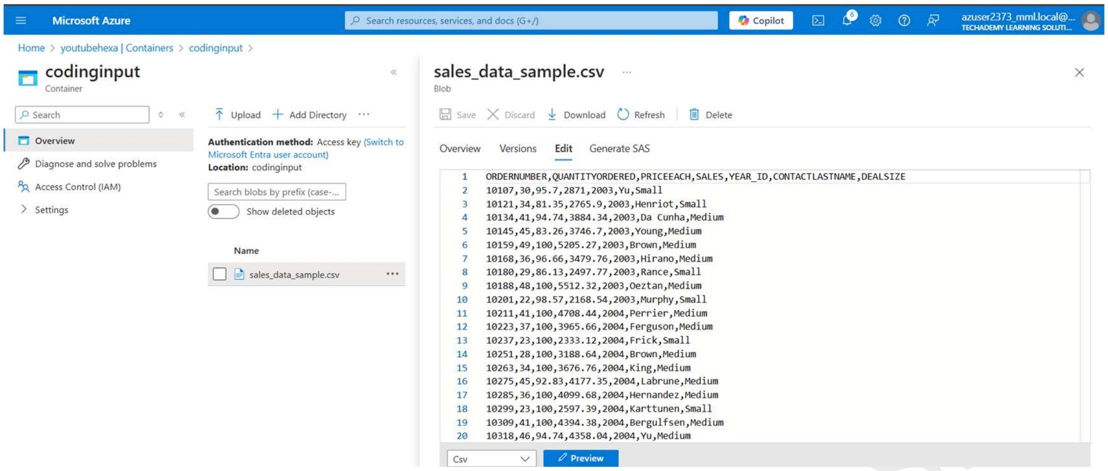
Pipeline status In progress

Showing 1 - 2 of 2 items

Activity name	Activity status	Activity type	Run start
Aggregate	✓ Succeeded	Data flow	12/19/2024, 5:14:14 PI
SortFilter	✓ Succeeded	Data flow	12/19/2024, 5:13:13 PI

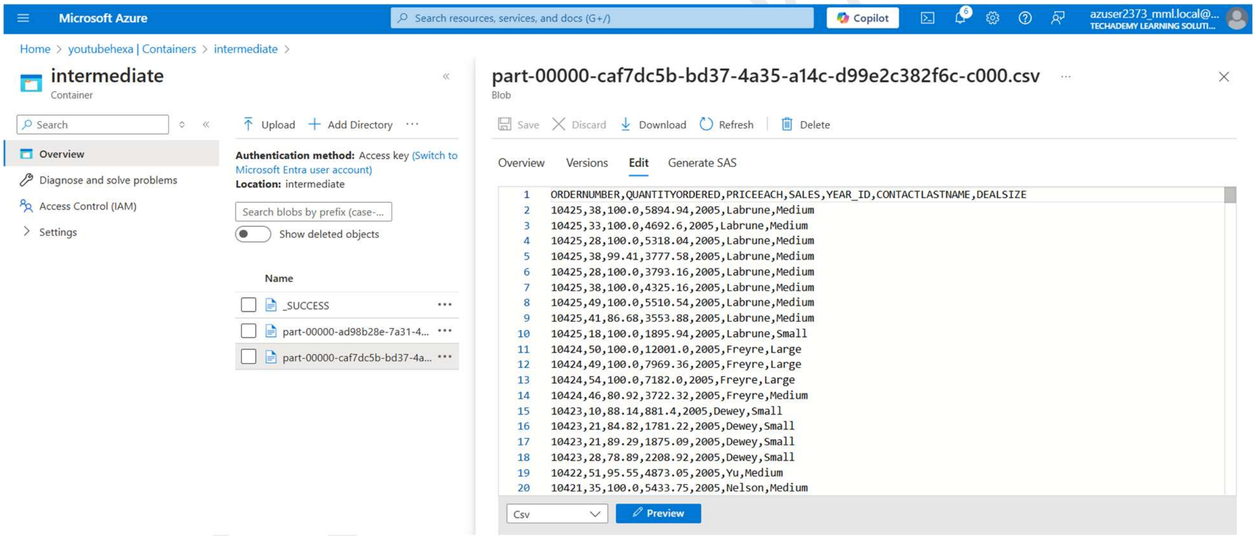
CONCLUSION

Input data in datalake



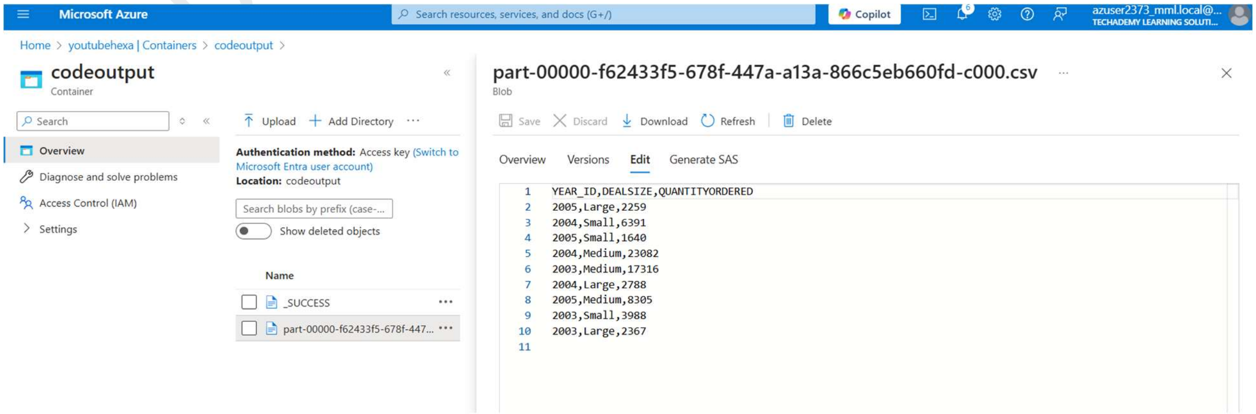
```
ORDERNUMBER,QUANTITYORDERED,PRICEEACH,SALES,YEAR_ID,CONTACTLASTNAME,DEALSIZE
10107,30,95.7,2871,2003,Yu,Small
10121,34,81.35,2765.9,2003,Henriot,Small
10134,41,94.74,3884.34,2003,Da Cunha,Medium
10145,45,83.26,3746.7,2003,Young,Medium
10159,49,100,5205.27,2003,Brown,Medium
10168,36,96.66,3479.76,2003,Hirano,Medium
10180,29,86.13,2497.77,2003,Rance,Small
10188,48,100,5512.32,2003,Oeztan,Medium
10201,22,98.57,2168.54,2003,Murphy,Small
10211,41,100,4708.44,2004,Perrier,Medium
10223,37,100,3965.66,2004,Ferguson,Medium
10237,23,100,2333.12,2004,Frick,Small
10251,28,100,3188.64,2004,Brown,Medium
10263,34,100,3676.76,2004,King,Medium
10275,45,92.83,4177.35,2004,Labrune,Medium
10285,36,100,4099.68,2004,Hernandez,Medium
10299,23,100,2597.39,2004,Karttunen,Small
10309,41,100,4394.38,2004,Bergulfsen,Medium
10318,46,94.74,4358.04,2004,Yu,Medium
```

Intermediate result data sorted and filtered in datalake


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
azuser2373_mml.local@...
TECHADEMY LEARNING SOLUTI...
Home > youtubehexa | Containers > intermediate >
intermediate
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Upload
Add Directory
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: intermediate
Search blobs by prefix (case-...
Show deleted objects
Name
_SUCCESS
part-00000-ad98b28e-7a31-4...
part-00000-caf7dc5b-bd37-4a...
part-00000-caf7dc5b-bd37-4a35-a14c-d99e2c382f6c-c000.csv
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Edit
Generate SAS
Csv
Preview

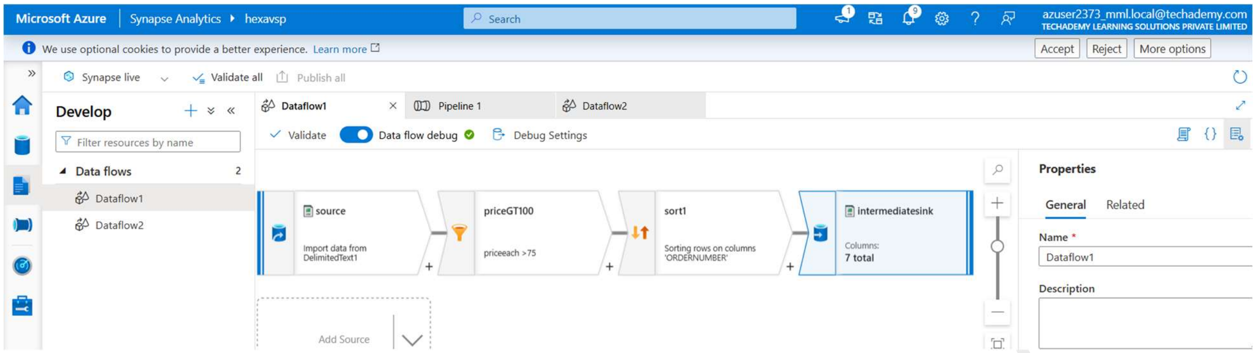
```
ORDERNUMBER,QUANTITYORDERED,PRICEEACH,SALES,YEAR_ID,CONTACTLASTNAME,DEALSIZE
10425,38,100.0,5894.94,2005,Labrune,Medium
10425,33,100.0,4692.6,2005,Labrune,Medium
10425,28,100.0,5318.04,2005,Labrune,Medium
10425,38,99.41,3777.58,2005,Labrune,Medium
10425,28,100.0,3793.16,2005,Labrune,Medium
10425,38,100.0,4325.16,2005,Labrune,Medium
10425,49,100.0,5510.54,2005,Labrune,Medium
10425,41,86.68,3553.88,2005,Labrune,Medium
10425,18,100.0,1895.94,2005,Labrune,Small
10424,50,100.0,12001.0,2005,Freyre,Large
10424,49,100.0,7969.36,2005,Freyre,Large
10424,54,100.0,7182.0,2005,Freyre,Large
10424,46,80.92,3722.32,2005,Freyre,Medium
10423,10,88.14,881.4,2005,Dewey,Small
10423,21,84.82,1781.22,2005,Dewey,Small
10423,21,89.29,1875.09,2005,Dewey,Small
10423,28,78.89,2208.92,2005,Dewey,Small
10422,51,95.55,4873.05,2005,Yu,Medium
10421,35,100.0,5433.75,2005,Nelson,Medium
```

Result – Aggregated values


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
azuser2373_mml.local@...
TECHADEMY LEARNING SOLUTI...
Home > youtubehexa | Containers > codeoutput >
codeoutput
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Upload
Add Directory
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: codeoutput
Search blobs by prefix (case-...
Show deleted objects
Name
_SUCCESS
part-00000-f62433f5-678f-447...
part-00000-f62433f5-678f-447a-a13a-866c5eb660fd-c000.csv
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Edit
Generate SAS

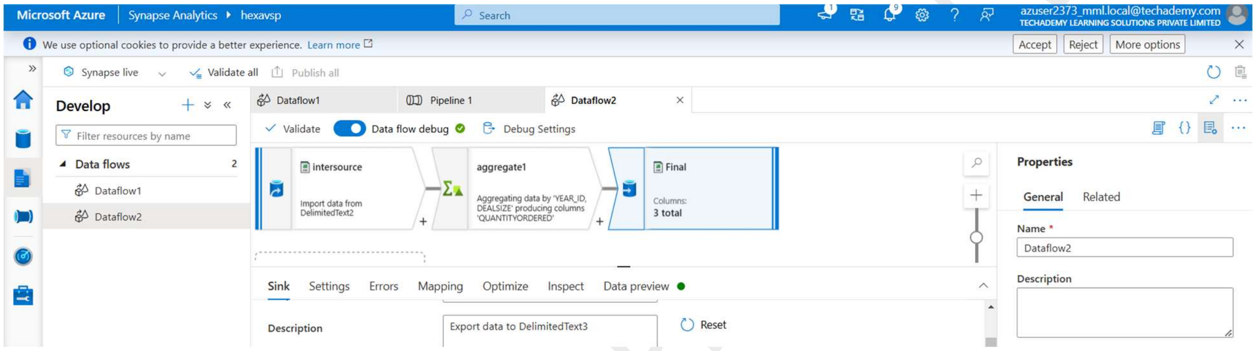
```
YEAR_ID,DEALSIZE,QUANTITYORDERED
2005,Large,2259
2004,Small,6391
2005,Small,1640
2004,Medium,23082
2003,Medium,17316
2004,Large,2788
2005,Medium,8305
2003,Small,3988
2003,Large,2367

```

DATAFLOW 1



Name * Dataflow1

DATAFLOW2

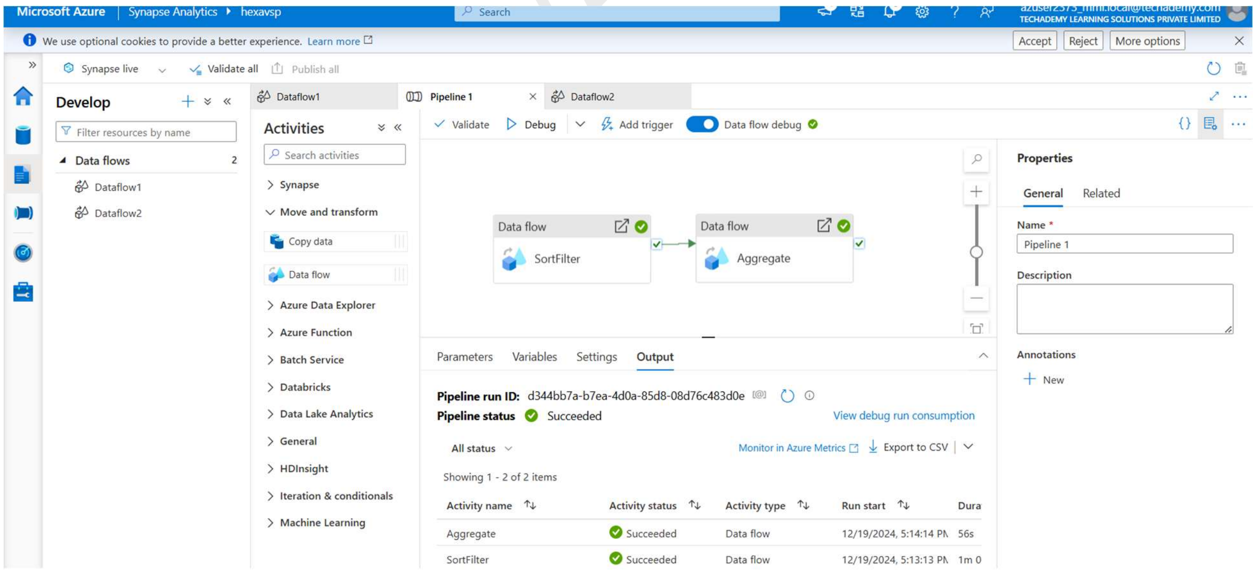


Description: Export data to DelimitedText3

Name * Dataflow2

Description

PIPELINE



Name * Pipeline 1

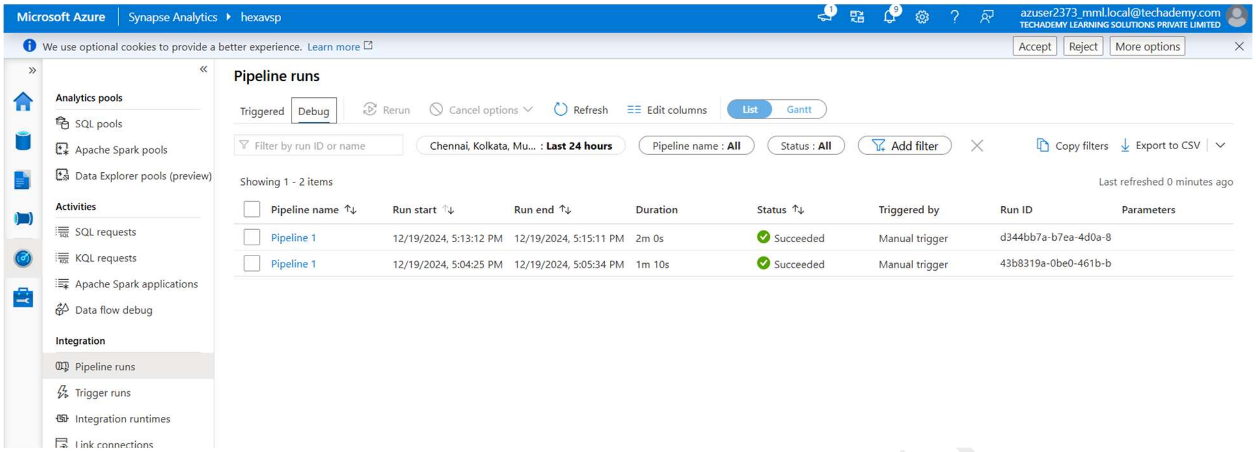
Pipeline run ID: d344bb7a-b7ea-4d0a-85d8-08d76c483d0e

Pipeline status ✓ Succeeded

Showing 1 - 2 of 2 items

Activity name	Activity status	Activity type	Run start	Dura
Aggregate	✓ Succeeded	Data flow	12/19/2024, 5:14:14 PN	56s
SortFilter	✓ Succeeded	Data flow	12/19/2024, 5:13:13 PN	1m 0

MONITORING PIPELINE RUNS



Pipeline runs

Showing 1 - 2 items

Last refreshed 0 minutes ago

□	Pipeline name	Run start	Run end	Duration	Status	Triggered by	Run ID	Parameters
□	Pipeline 1	12/19/2024, 5:13:12 PM	12/19/2024, 5:15:11 PM	2m 0s	✓ Succeeded	Manual trigger	d344bb7a-b7ea-4d0a-8	
□	Pipeline 1	12/19/2024, 5:04:25 PM	12/19/2024, 5:05:34 PM	1m 10s	✓ Succeeded	Manual trigger	43b8319a-0be0-461b-b	