

Task 1.1 Synapse Pipeline with Copy Data Activity

Run Scenario:

- 1. Before starting the run scenario, clear bronze/youflix folder and re-create YouFlixDB database from scratch using DeploymentScript.sql script. It is necessary to get rid of the results of your previous execution and testing activities.

eed | Storage browser

stdimentoringdatalakeed

Favorites

Recently viewed

Blob containers

File shares

Queues

Tables

YouFlixWatermark

View all

+ Add entity

Refresh

Delete

Edit columns

Tables > YouFlixWatermark

Authentication method: Access key (Switch to Microsoft Entra user account)

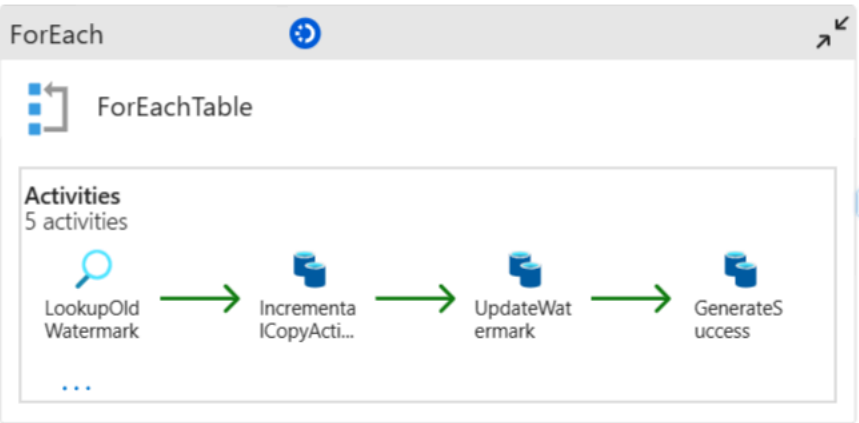
Add filter

Advanced filters

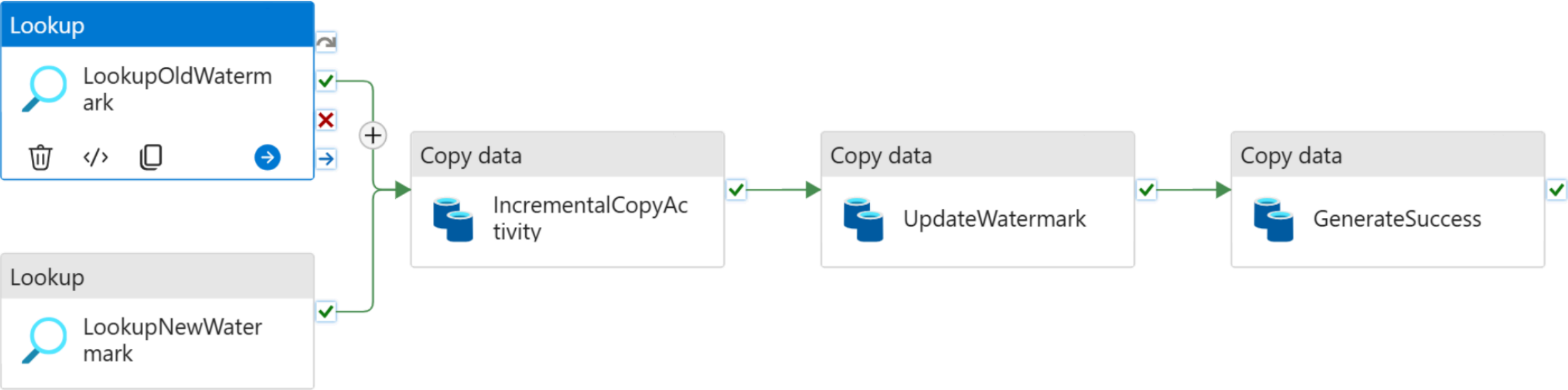
Showing all 4 items

	PartitionKey	RowKey	Timestamp	Watermark	
<input type="checkbox"/>	device	device	2024-12-28T19:42:17.22...	2000-01-01T00:00:00Z	...
<input type="checkbox"/>	subscription	subscription	2024-12-28T19:42:46.44...	2000-01-01T00:00:00Z	...
<input type="checkbox"/>	user	user	2024-12-28T19:43:03.08...	2000-01-01T00:00:00Z	...
<input type="checkbox"/>	user_subscription_device	user_subscription_device	2024-12-28T19:43:29.96...	2000-01-01T00:00:00Z	...

- 2. Go to data lake stdimentoringdatalakexx and proceed to Storage browser, then click on Tables and edit each of entity by setting watermark value to 2000-01-01T00:00:00.00Z.



3. Take screenshot(s) of your pipeline.
PipelineCopyAllTables > ForEachTable



General Settings User properties

Source dataset *

WatermarkTable

Open

New

Preview data

Learn more

First row only

Use query

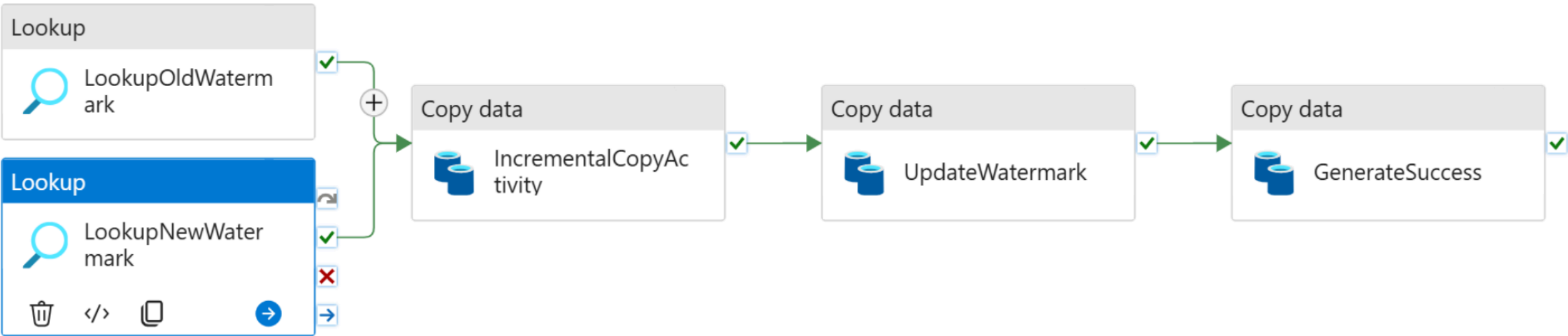
Table

Query

Query

PartitionKey eq '@(item().TABLE_NAME)

Ignore table not found



General

Settings

User properties

Source dataset *

SqlServerAllTables

Open

New

Preview data

Learn more

First row only

☒

Use query

☐ Table

☒ Query

☐ Stored procedure

Query

select MAX(created_timestamp) as New...

Query timeout (minutes)

120

Isolation level

Select...

Partition option

☒ None

☐ Physical partitions of table

☐ Dynamic range

Please preview data to validate the partition settings.

✓ Validate

✓ Validate copy runtime

▶ Debug

⚡ Add trigger

PipelineCopyAllTables > ForEachTable

Lookup

LookupOldWatermark

✓

Copy data

IncrementalCopyActivity

✓

Lookup

LookupNewWatermark

✓

General

Source

Sink

Mapping

Settings

User properties

Source dataset *

SqlServerAllTables

Open

New

Preview

Use query

Table

Query

Stored procedure

Query

select * from youflix.@{item().TABLE_NAME}

Query timeout (minutes)

120

Isolation level

Select...

Partition option

None

Physical partitions of table

Dynamic range

Please preview data to validate the partition settings.

Additional columns

New

Pipeline expression builder

Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

select

*

from youflix.@{item().TABLE_NAME}

where

created_timestamp > '@{activity('LookupOldWatermark').output.firstRow.Watermark}'

and created_timestamp <= '@{activity('LookupNewWatermark').output.firstRow.NewWatermarkvalue}'

Clear contents

ForEach iterator

Activity outputs

Parameters

System variables

...

Search

ForEachTable

Current item

OK

Cancel

Source query: ``sql select

◦ from youflix.@{item().TABLE_NAME} where created_timestamp > cast('@{activity('LookupOldWatermark').output.firstRow.Watermark}' as datetime2) and created_timestamp <= cast('@{activity('LookupNewWatermark').output.firstRow.NewWatermarkvalue}' as datetime2) ``

PipelineCopyAllTables x

✓ Validate

✓ Validate copy runtime

▶ Debug

⚡ Add trigger

{} {}

PipelineCopyAllTables > ForEachTable

Lookup

LookupOldWatermark

✓

Copy data

IncrementalCopyActivity

✓

Copy data

UpdateWatermark

✓

Copy data

GenerateSuccess

✓

Lookup

LookupNewWatermark

✓

General

Source

Sink

Mapping

Settings

User properties

Sink dataset *

DataLakeAllTables

Open

New

Learn more

Copy behavior

Preserve hierarchy

Max concurrent connections

2

Block size (MB)

Metadata

New

Quote all text

✓

File extension

@concat('/youflix_', item().TABLE_NAME)

Max rows per file

Lookup

LookupOldWatermark

✓

Copy data

IncrementalCopyActivity

✓

Copy data

UpdateWatermark

✓

Copy data

GenerateSuccess

✓

Lookup

LookupNewWatermark

✓

General

Source

Sink

Mapping

Settings

User properties

Sink dataset *

WatermarkTable

Open

New

Learn more

Insert type

Replace

Partition key value selection

Use sink column

Partition key column

PartitionKey

Row key value selection

Use sink column

Row key column

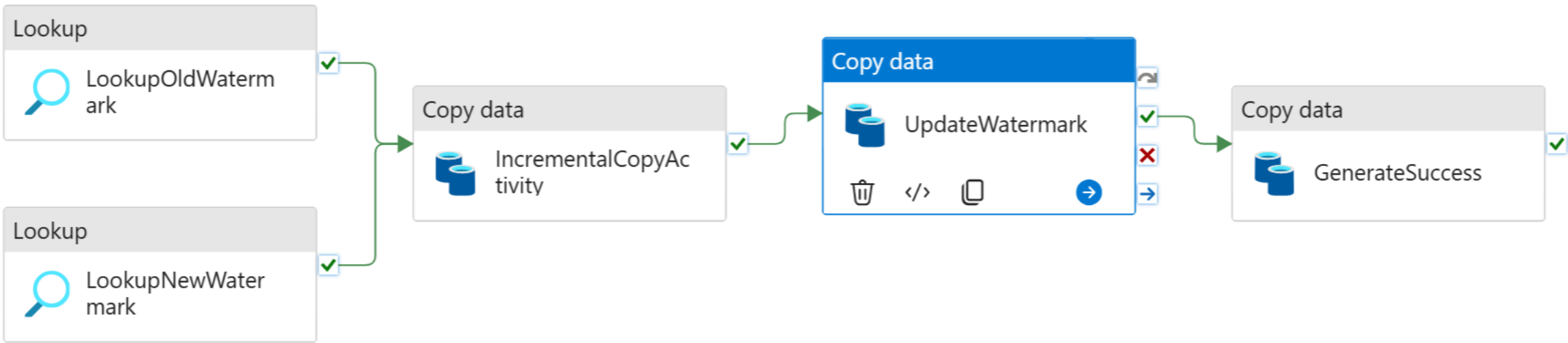
RowKey

Write batch timeout

Write batch size

10000

Max concurrent connections



General

Source

Sink

Mapping

Settings

User properties

Sink dataset *

WatermarkTable

Open

New

Learn more

Insert type

Replace

Partition key value selection

Use sink column

Partition key column

PartitionKey

Row key value selection

Use sink column

Row key column

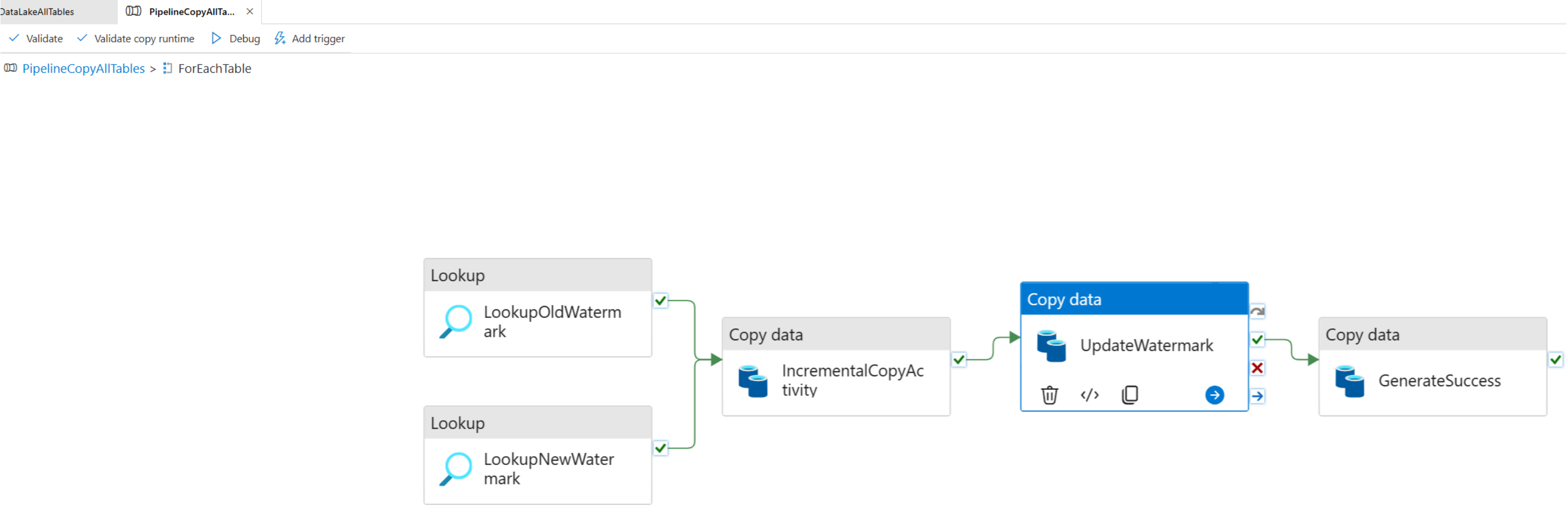
RowKey

Write batch timeout

Write batch size

10000

Max concurrent connections



GeneralSourceSinkMappingSettingsUser properties

> Type conversion settings

← Import schemas

🔗 Preview source

+ New mapping

🗑 Clear

↺ Reset

🗑 Delete

<input type="checkbox"/>	Source	Type	Destination	Type		
<input type="checkbox"/>	TableName	abc varchar	→ PartitionKey	abc String	+	🗑
<input type="checkbox"/>	RowKey	abc varchar	→ RowKey	abc String	+	🗑
<input type="checkbox"/>	NewWatermark	📅 datetime	→ Watermark	📅 DateTimeOffset	+	🗑

✓ Validate

✓ Validate copy runtime

▶ Debug

⚡ Add trigger

PipelineCopyAllTables > ForEachTable

Lookup

LookupOldWatermark

Lookup

LookupNewWatermark

Copy data

IncrementalCopyActivity

General

Source

Sink

Mapping

Settings

User properties

Sink dataset *

DataLakeAllTables

🔗 Open

+ New

Learn more

Copy behavior

ⓘ

Preserve hierarchy

Max concurrent connections

ⓘ

Block size (MB)

ⓘ

Metadata

ⓘ

+ New

Quote all text

ⓘ

✓

File extension

ⓘ

@concat('Success_', item().TABLE_NAME, '.txt')

🗑

Pipeline expression builder

🔗

Add dynamic content below using any combination of expressions, functions and system variables.

@concat('Success_', item().TABLE_NAME, '.txt')

Clear contents

ForEach iterator

Activity outputs

Parameters

System variables

...

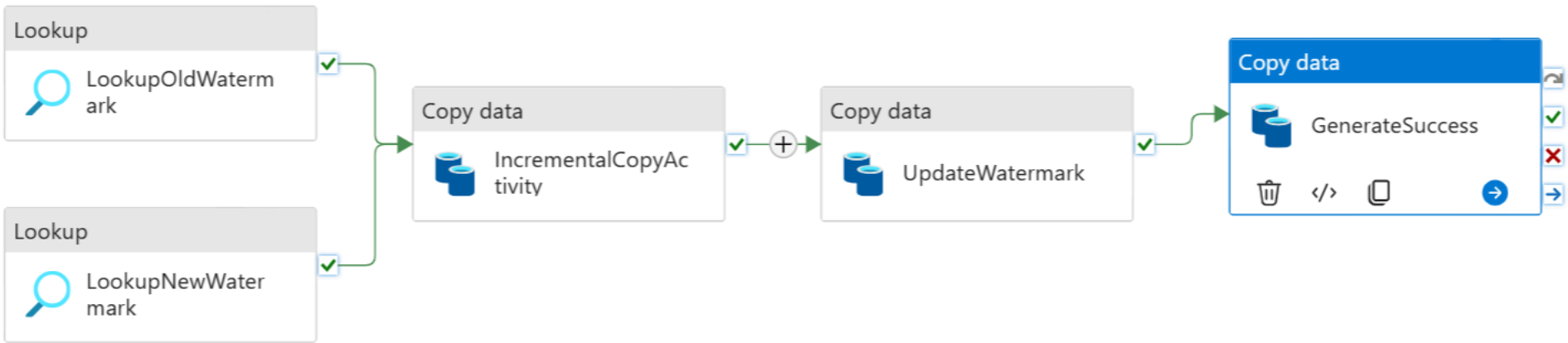
🔍 Search

ForEachTable

Current item

OK

Cancel



GeneralSourceSinkMappingSettingsUser properties

Source dataset *

SqlServerAllTables

OpenNewPreview dataLearn more

Use query

TableQueryStored procedure

Query *

select 'SUCCESSFUL UPLOAD'

Edit

Query timeout (minutes)

120

Isolation level

Select...

Partition option

NonePhysical partitions of tableDynamic range

Please preview data to validate the partition settings.

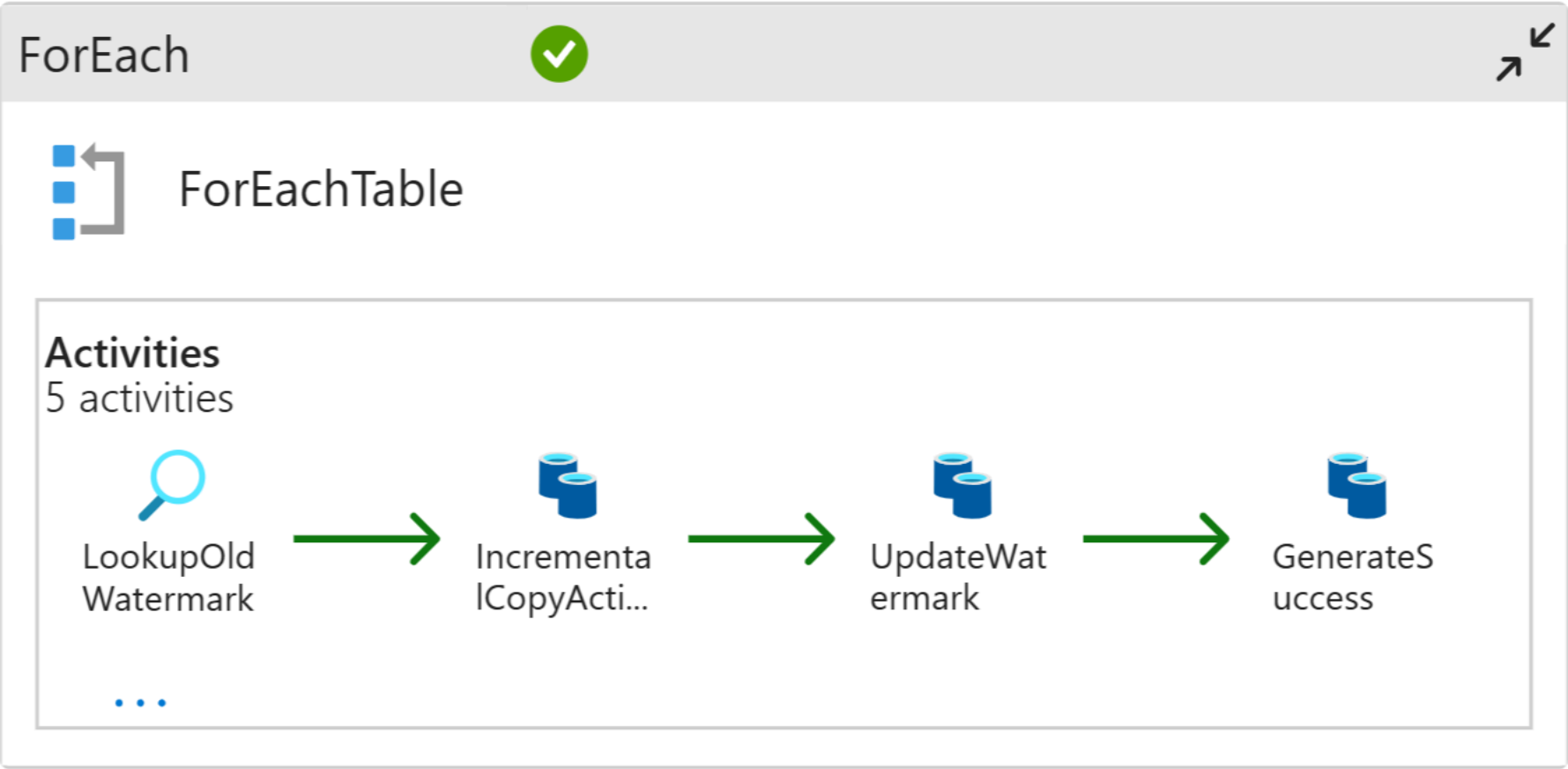
Additional columns

New

4. Go to your created pipeline and execute it manually.

[All pipeline runs](#) > ✔ PipelineCopyAllTables - Activity runs

Rerun Cancel Refresh Update pipeline List Gantt



data

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

Authentication method: Microsoft Entra user account (Switch to Access key)

Location: data / bronze / youflix

Search blobs by prefix (case-sensitive)

Show deleted object:

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> [.]					
<input type="checkbox"/> youflix_device	12/26/2024, 5:19:41 ...				
<input type="checkbox"/> youflix_subscription	12/26/2024, 5:20:02 ...				
<input type="checkbox"/> youflix_subscription_user_device	12/26/2024, 5:20:29 ...				
<input type="checkbox"/> youflix_user	12/26/2024, 5:19:27 ...				
<input type="checkbox"/> Success_device.txt	1/1/2025, 12:35:14 PM	Hot (Inferred)		Block blob	0 B
<input type="checkbox"/> Success_subscription.txt	1/1/2025, 12:38:12 PM	Hot (Inferred)		Block blob	0 B
<input type="checkbox"/> Success_user_subscription_device.txt	1/1/2025, 12:38:12 PM	Hot (Inferred)		Block blob	0 B
<input type="checkbox"/> Success_user.txt	1/1/2025, 12:38:12 PM	Hot (Inferred)		Block blob	0 B

6. Take screenshot(s) of the target folder.

data

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

Authentication method: Microsoft Entra user account (Switch to Access key)

Location: data / bronze / youflix / youflix_device

Search blobs by prefix (case-sensitive)

youflix_device20250114044550.csv

data

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

Authentication method: Microsoft Entra user account (Switch to Access key)

Location: data / bronze / youflix / youflix_subscription

Search blobs by prefix (case-sensitive)

youflix_subscription20250114044833.csv

7. Go inside each of 4 folders and take screenshot(s) of files generated.

data

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

Authentication method: Microsoft Entra user account (Switch to Access key)

Location: data / bronze / youflix / youflix_user_subscription_device

Search blobs by prefix (case-sensitive)

youflix_user_subscription_device20250114044939.csv

8. In Synapse Workspace, navigate to Data section, find in Linked tab your container, open bronze/youf1ix and select top 10 rows for each file using SQL queries.

RunUndoPublishQuery planConnect toBuilt-inUse databasemaster

```
1  -- This is auto-generated code
2  SELECT
3    TOP 100 *
4  FROM
5    OPENROWSET(
6      BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_device/youflix_device20250114044550.csv',
7      FORMAT = 'CSV',
8      PARSE_VERSION = '2.0'
9    ) AS [result]
10
```

ResultsMessages

ViewTableChartExport results

Search

C1	C2	C3	C4	C5
device_id	device_name	device_type	device_os	created_timestamp
1	HP	PC	Windows	2021-01-01 00:00:00.0000000
2	Lenovo	PC	Linux	2021-01-01 00:00:00.0000000
3	iMac	PC	macOS	2021-01-01 00:00:00.0000000
4	Raspberry Pi	PC	Linux	2021-01-01 00:00:00.0000000
5	MSI	PC	Windows	2021-01-01 00:00:00.0000000
6	Dell	PC	Windows	2021-01-01 00:00:00.0000000
7	HP	Laptop	Windows	2021-01-01 00:00:00.0000000

```
1  -- This is auto-generated code
2  SELECT
3    TOP 100 *
4  FROM
5    OPENROWSET(
6      BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_subscription/youflix_subscription20250114044833.csv',
7      FORMAT = 'CSV',
8      PARSE_VERSION = '2.0'
9    ) AS [result]
10
```

ResultsMessages

ViewTableChartExport results

Search

C1	C2	C3	C4	C5	C6	C7
subscription_id	subscription_name	subscription_type	subscription_video_quality	subscription_m...	created_timesta...	expiration_time...
1	Basic subscripion with two allow...	BASIC	HD	2	2021-01-01 00:...	9999-01-01 00:...
2	Standard subscripion with four al...	STANDARD	FullHD	4	2021-01-01 00:...	9999-01-01 00:...
3	Premium subscripion with seven ...	PREMIUM	4k	7	2021-01-01 00:...	9999-01-01 00:...

9. Take screenshot(s) of SQL queries and result output.

2 SELECT
3 TOP 100 *
4 FROM
5 OPENROWSET(
6 BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_user_subscription_device/youflix_user_subscription_device20250114044833.csv',
7 FORMAT = 'CSV',
8 PARSE_VERSION = '2.0'
9) AS [result]
10

ResultsMessages

ViewTableChartExport results

Search

C1	C2	C3	C4	C5
user_subscription_device_id	user_id	subscription_id	device_id	created_timestamp
1	1	3	1	2024-12-25 20:40:09.1630000
2	1	3	19	2024-12-25 20:40:09.1630000
3	2	1	2	2024-12-25 20:40:09.1630000
4	2	1	29	2024-12-25 20:40:09.1630000
5	3	3	15	2024-12-25 20:40:09.1630000
6	4	3	22	2024-12-25 20:40:09.1630000
7	4	3	7	2024-12-25 20:40:09.1630000

10. In Synapse Workspace, navigate to Data section, find in Linked tab your container, open bronze/youf1ix and check number of rows for each file using SQL query.

```
1 -- This is auto-generated code
2 SELECT
3   count(*)
4 FROM
5   OPENROWSET(
6     BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_subscription/youflix_subscription20250114044833.csv',
7     FORMAT = 'CSV',
8     PARSER_VERSION = '2.0'
9   ) AS [result]
10
```

ResultsMessages

ViewTableChartExport results

Search

C1	C2	C3	C4	C5
user_subscription_device_id	user_id	subscription_id	device_id	created_timestamp
1	1	3	1	2024-12-25 20:40:09.1630000
2	1	3	19	2024-12-25 20:40:09.1630000
3	2	1	2	2024-12-25 20:40:09.1630000
4	2	1	29	2024-12-25 20:40:09.1630000
5	3	3	15	2024-12-25 20:40:09.1630000
6	4	3	22	2024-12-25 20:40:09.1630000
7	4	3	7	2024-12-25 20:40:09.1630000

11. Take screenshot(s) of SQL queries with count values.

RunUndoPublishQuery plan

Connect toBuilt-inUse database

master

```
1 -- This is auto-generated code
2 SELECT
3   count(*)
4 FROM
5   OPENROWSET(
6     BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_device/youflix_device20250114044550.csv',
7     FORMAT = 'CSV',
8     PARSER_VERSION = '2.0'
9   ) AS [result]
10
```

ResultsMessages

ViewTableChartExport results

Search

(No column name)

31

eed | Storage browser

>+ Add entityRefreshDeleteEdit columns

Tables > YouFlixWatermark

Authentication method: Access key (Switch to Microsoft Entra user account)

Add filterAdvanced filters

Showing all 4 items

<input type="checkbox"/>	PartitionKey	RowKey	Timestamp	Watermark	
<input type="checkbox"/>	device	device	2025-01-14T17:40:55.19...	2021-01-01T00:00:00Z	...
<input type="checkbox"/>	subscription	subscription	2025-01-14T17:41:05.24...	2021-01-01T00:00:00Z	...
<input type="checkbox"/>	user	user	2025-01-14T17:41:14.89...	2021-01-01T00:00:00Z	...
<input type="checkbox"/>	user_subscription_device	user_subscription_device	2025-01-14T17:41:25.11...	2021-01-01T00:00:00Z	...

12. Go to your Azure Table and take a screenshot of new watermark values.

13. Without changing source data, execute your pipeline one more time.

14. In Synapse Workspace, navigate to Data section, find in Linked tab your container, open bronze/youflix and check number of rows for each newly loaded file using SQL query.

Run Undo Publish Query plan Connect to Built-in Use database master

```
1 -- This is auto-generated code
2 SELECT
3     count(*)
4 FROM
5     OPENROWSET(
6         BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_user_subscription_device/youflix_user_subscription_device',
7         FORMAT = 'CSV',
8         PARSE_VERSION = '2.0'
9     ) AS [result]
10
```

Results Messages

View Table Chart Export results

Search

(No column name)

1

CopyPipelineAll data SQL script 1 SQL script 2 SQL script 3

Run Undo Publish Query plan Connect to Built-in Use database master

```
1 -- This is auto-generated code
2 SELECT
3     count(*)
4 FROM
5     OPENROWSET(
6         BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_device',
7         FORMAT = 'CSV',
8         PARSE_VERSION = '2.0'
9     ) AS [result]
10
```

Results Messages

View Table Chart Export results

Search

(No column name)

1

15. Take screenshot(s) of SQL queries with count values.

CopyPipelineAll data SQL script 1 SQL script 2 SQL script 3

Run Undo Publish Query plan Connect to Built-in Use database master

```
1 -- This is auto-generated code
2 SELECT
3     count(*)
4 FROM
5     OPENROWSET(
6         BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_subscription/you',
7         FORMAT = 'CSV',
8         PARSE_VERSION = '2.0'
9     ) AS [result]
10
```

Results Messages

View Table Chart Export results

Search

(No column name)

1

YouflixDB Execute

SQLQuery1.sql - EP...Ekaterina_Dul (51)*

```
EXEC youflix_internal.sp_youflix_tables_insert_update 10000, 15;
```

150 %

Messages

15: updating users in YOUFLIX.user table.
Moving users to the main YOUFLIX schema.
10000 users moved to the main YOUFLIX schema.
52984 devices moved to the main YOUFLIX schema.

Completion time: 2025-01-14T22:13:36.5832918+04:00

16. Connect to MS SQL Server YouFlixDB database and run the following command: sql EXEC youflix_internal.sp_youflix_tables_insert_update 10000, 15;

17. Execute pipeline manually again.

18. In Synapse Workspace, navigate to Data section, find in Linked tab your container, open bronze/youflix and check number of rows for each newly loaded file using SQL query.

CopyPipelineAlldataSQL script 1SQL script 2SQL script 3

RunUndoPublishQuery planConnect toBuilt-inUse databasemaster

```
1  -- This is auto-generated code
2  SELECT
3    count(*)
4  FROM
5    OPENROWSET(
6      BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/bronze/youflix/youflix_user_subscript:
7      FORMAT = 'CSV',
8      PARSE_VERSION = '2.0'
9    ) AS [result]
10
```

ResultsMessages

ViewTableChartExport results

Search

(No column name)

26597

19. Take screenshot(s) of SQL queries with count values.