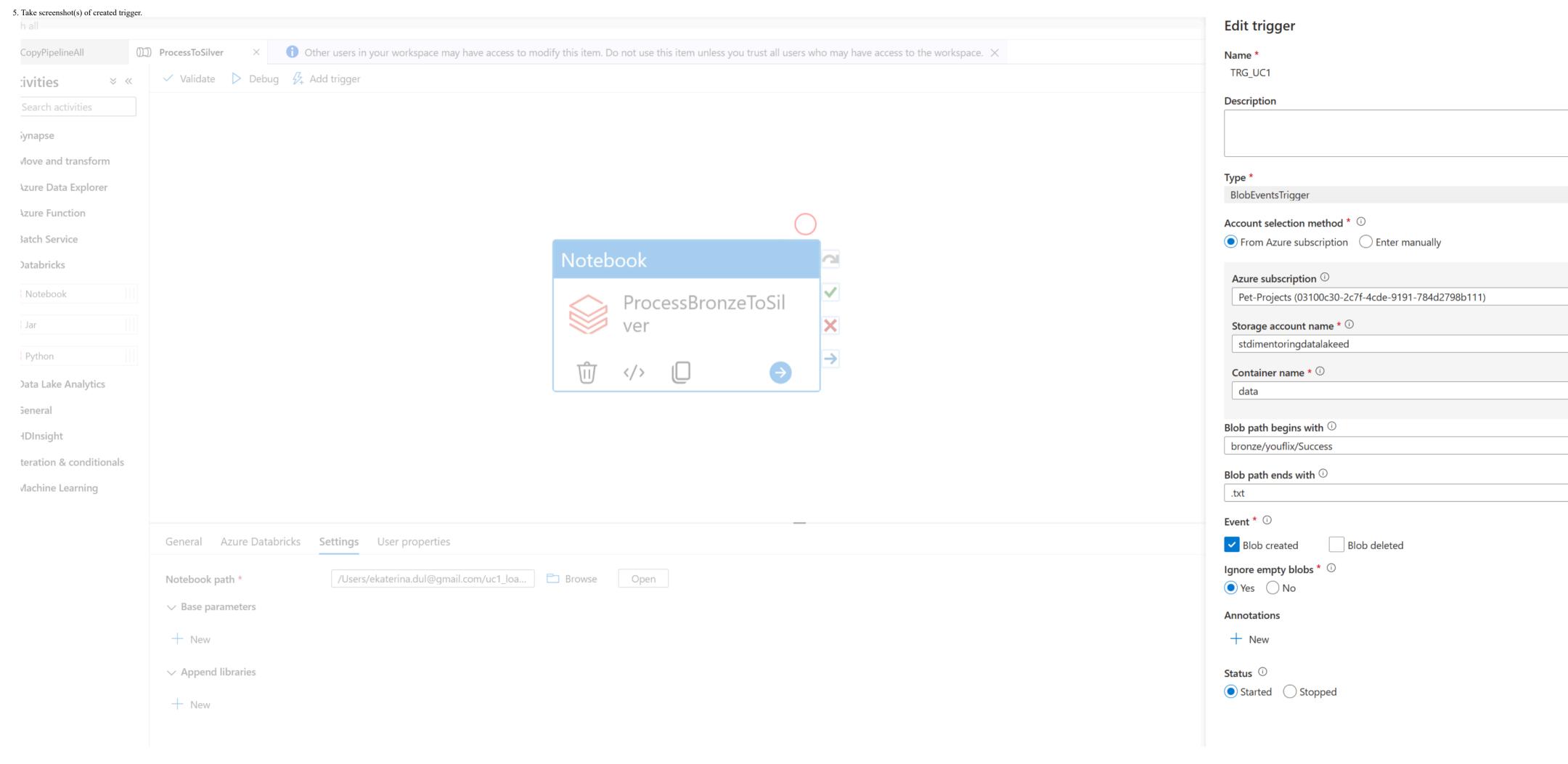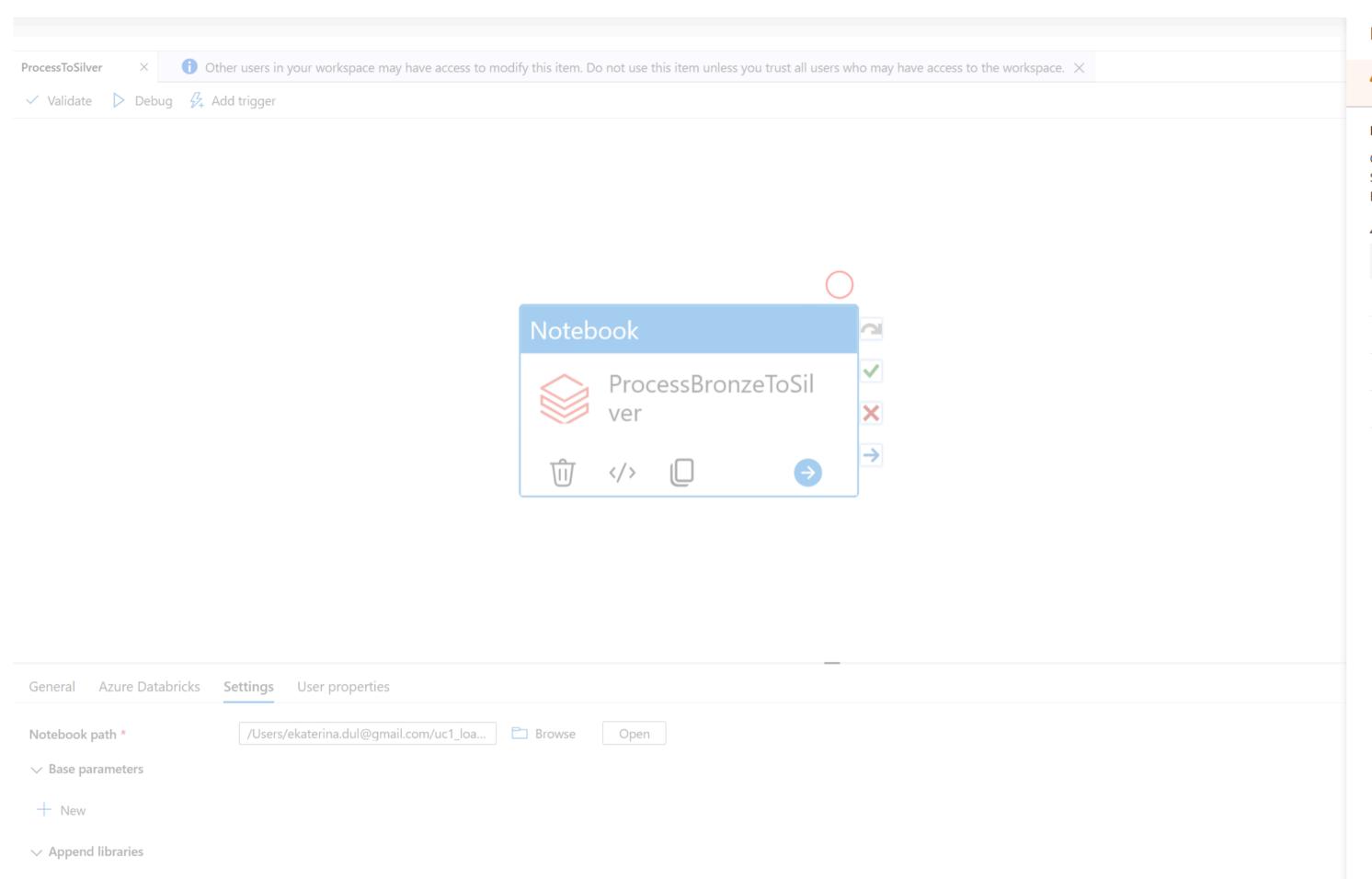# Task 1.3 Orchestration with Storage Event Trigger

Run Scenario:

1. Before starting the run scenario, clear â€œbronze/youflixâ€ and â€œsilver/youflixâ€ directories.
2. Delete 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.
3. Drop `YouFlix` database from your Databricks workspace.
4. 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`.

5. Take screenshot(s) of created trigger.

h all

| CopyPipelineAll | ProcessToSilver ✕ | ⓘ 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. ✕ |

:tivities ≫ «

Search activities

Synapse

Move and transform

Azure Data Explorer

Azure Function

Batch Service

Databricks

| Notebook

| Jar

| Python

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

✓ Validate   ▷ Debug   ⚡ Add trigger

Notebook

ProcessBronzeToSilver

🗑   </>   ⧉   →

General   Azure Databricks   **Settings**   User properties

Notebook path *   /Users/ekaterina.dul@gmail.com/uc1_loa...   📁 Browse   Open

∨ Base parameters

+ New

∨ Append libraries

+ New

Edit trigger

Name *

TRG_UC1

Description

Type *

BlobEventsTrigger

Account selection method * ⓘ

⦿ From Azure subscription   ◯ Enter manually

Azure subscription ⓘ

Pet-Projects (03100c30-2c7f-4cde-9191-784d2798b111)

Storage account name * ⓘ

stdimentoringdatalakeed

Container name * ⓘ

data

Blob path begins with ⓘ

bronze/youflix/Success

Blob path ends with ⓘ

.txt

Event * ⓘ

☑ Blob created   ☐ Blob deleted

Ignore empty blobs * ⓘ

⦿ Yes   ◯ No

Annotations

+ New

Status ⓘ

⦿ Started   ◯ Stopped

Validate Debug Add trigger

Refresh

Notebook

ProcessBronzeToSil
ver

General Azure Databricks **Settings** User properties

Notebook path *

/Users/ekaterina.dul@gmail.com/uc1_loa... Browse Open

Base parameters

+ New

Append libraries

+ New

6. Execute pipeline from task 1.1 manually.

Data preview

⚠ Make sure you have specific filters. Configuring filters that are too broad can match a large number of files created/deleted and may significantly impact your cost.

Event Trigger Filters

Container name:	**data**
Starts with:	**bronze/youflix/Success**
Ends with:	**.txt**

4 blobs matched in "data"   Refresh

	Blob name
1	bronze/youflix/Success_device.txt
2	bronze/youflix/Success_subscription.txt
3	bronze/youflix/Success_user.txt
4	bronze/youflix/Success_user_subscription_device.txt

7. When it's completed and pipeline from task 1.2 is triggered, take screenshot(s) from Monitor tab with "Triggered by" column is visible for the triggered pipeline.

Integrate

Filter resources by name

▼ Pipelines 2

    CopyPipelineAll

    ● ProcessToSilver

CopyPipelineAll | ProcessToSilver ●

Activities

Search activities

> Synapse

> Move and transform

> Azure Data Explorer

> Azure Function

> Batch Service

▼ Databricks

    ▦ Notebook

    ▦ Jar

    ▦ Python

> Data Lake Analytics

> General

> HDInsight

> Iteration & conditionals

> Machine Learning

✓ Validate   ▷ Debug   Add trigger

Notebook
ProcessBronzeToSilver

General   Azure Databricks   **Settings**   User properties

Notebook path *   /Users/ekaterina.dul@gmail.com/uc1_loa...   📁 Browse   Open

∨ Base parameters

    + New

∨ Append libraries

    + New

rg-di-synapse-ed

Pipeline runs

Triggered Debug   ⟳ Rerun   ⊘ Cancel options ∨   ⟳ Refresh   ☰ Edit columns   List Gantt

Filter by run ID or name   Tbilisi (UTC+4) : **Last 24 hours**   Pipeline name : **ProcessToSilver**   Status : **All**   Runs : Latest runs   Triggered by : **All**   ⊻ Add filter   ✕

Showing 1 - 1 items

☐ Pipeline name ↑↓	Run start ↑↓	Run end ↑↓	Duration	Triggered by	Status ↑↓	Run	Parameters
☐ ProcessToSilver	1/14/2025, 11:58:09 PM	--	3s	TRG_UC1	◐ In progress	Original	

8. In Synapse Workspace, navigate to Data section, find in Linked tab your container, open "silver/youflix" and check number of rows for each delta table using SQL query.

```
-- This is auto-generated code
SELECT
    count(*)
FROM
    OPENROWSET(
        BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/silver/youflix/youflix_subscription/*.snappy.parquet',
        FORMAT = 'PARQUET'
    ) AS [result]
```

Results | Messages

View Table Chart   Export results

(No column name)

15

```
-- This is auto-generated code
SELECT
    count(*)
FROM
    OPENROWSET(
        BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/silver/youflix/youflix_device/*.snappy.parquet',
        FORMAT = 'PARQUET'
    ) AS [result]
```

Results | Messages

View Table Chart   Export results

(No column name)

180

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

CopyPipelineAll | data | SQL script 4 | SQL script 5 | SQL script 6

```
-- This is auto-generated code
SELECT
    count(*)
FROM
    OPENROWSET(
        BULK 'https://stdimentoringdatalakeed.dfs.core.windows.net/data/silver/youflix/youflix_user_subscription_device/*.snappy.parquet',
        FORMAT = 'PARQUET'
    ) AS [result]
```

Results | Messages

View Table Chart   Export results

(No column name)

79938

```
EXEC youflix_internal.sp_youflix_tables_insert_update 200000, 0;
```

Messages

```
Nothing to update in YOUFLIX.user table.
Moving users to the main YOUFLIX schema.
200000 users moved to the main YOUFLIX schema.
902678 devices moved to the main YOUFLIX schema.

Completion time: 2025-01-14T23:49:27.5502046+04:00
```

10. Connect to MS SQL Server YouFlix database and run the following command: sql EXEC youflix_internal.sp_youflix_tables_insert_update 200000, 0;

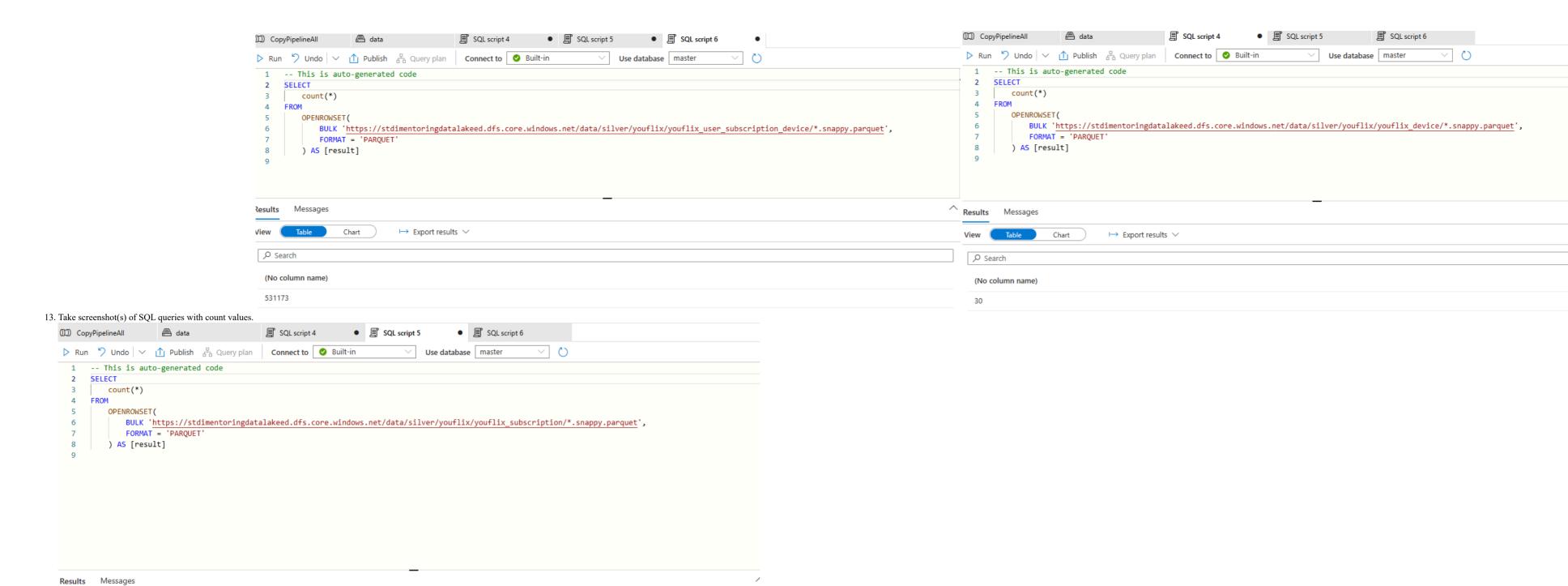11. Execute pipeline from task 1.1 manually and wait until triggered pipeline from 1.2 is finished.

12. In Synapse Workspace, navigate to Data section, find in Linked tab your container, open "silver/youflix" and check number of rows for each delta table 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/silver/youflix/youflix_user_subscription_device/*.snappy.parquet',
7            FORMAT = 'PARQUET'
8        ) AS [result]
9
```

Results   Messages

View [Table] [Chart]   ⇥ Export results ⌄

🔍 Search

(No column name)

531173

▷ 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/silver/youflix/youflix_device/*.snappy.parquet',
7            FORMAT = 'PARQUET'
8        ) AS [result]
9
```

Results   Messages

View [Table] [Chart]   ⇥ Export results ⌄

🔍 Search

(No column name)

30

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

▷ 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/silver/youflix/youflix_subscription/*.snappy.parquet',
7            FORMAT = 'PARQUET'
8        ) AS [result]
9
```

Results   Messages

View [Table] [Chart]   ⇥ Export results ⌄

🔍 Search

(No column name)

3