Azure Data Factory

Lab: Trigger a pipeline using Tumbling window trigger

Pre-requisites:

- Azure Pass subscription
- Azure Data Lake Storage Gen2 storage account
- Azure SQL Database

Lab Objective:

After completing this lab, you will be able to:

- Copy data from SQL Database to Data Lake Storage
- Create and use Tumbling window trigger in Azure Data Factory.

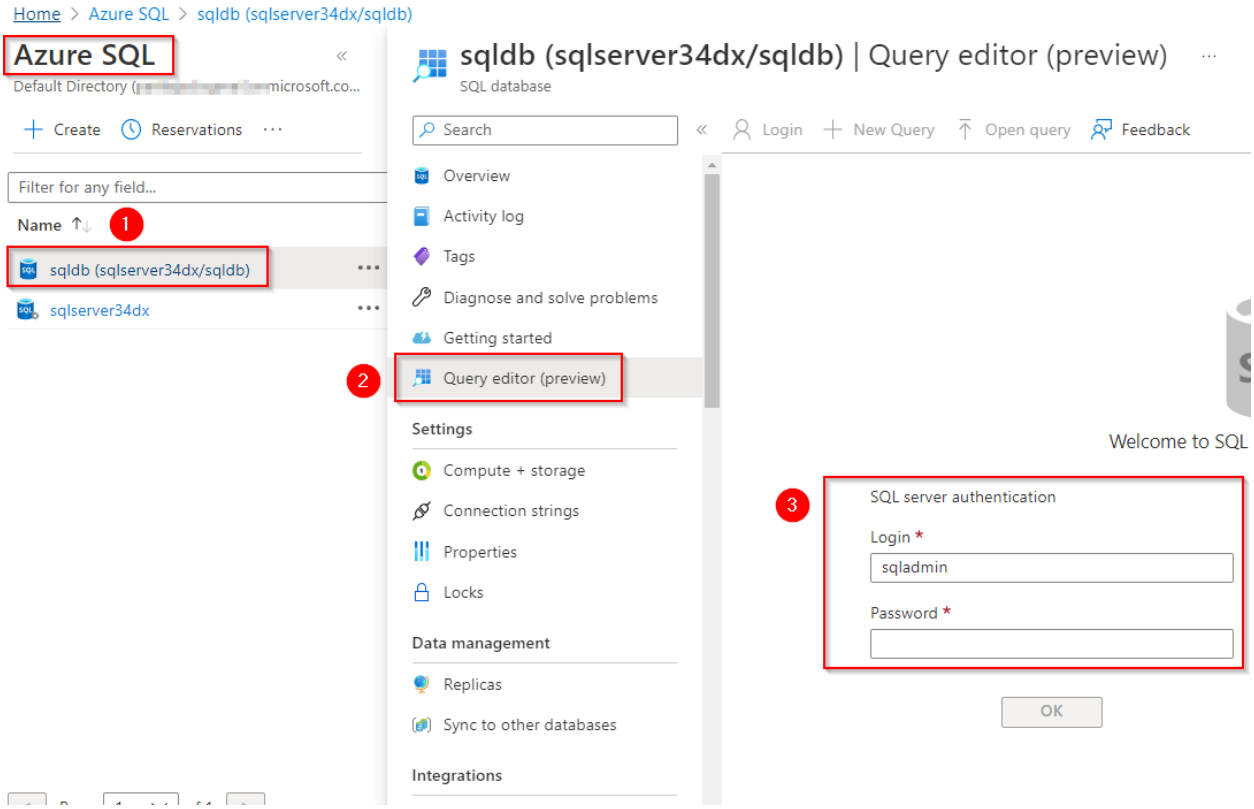
Exercise 1: Create a Tumbling window trigger to execute the pipeline.

The main task for this exercise are as follows:

- Create a copy activity to copy data from SQL DB to Data Lake.
- Create a Tumbling Window trigger to run the 2 pipelines at a time.

Task 1: Create a table in SQL DB

1. Open query editor in SQL DB



2. Add following code to create a table

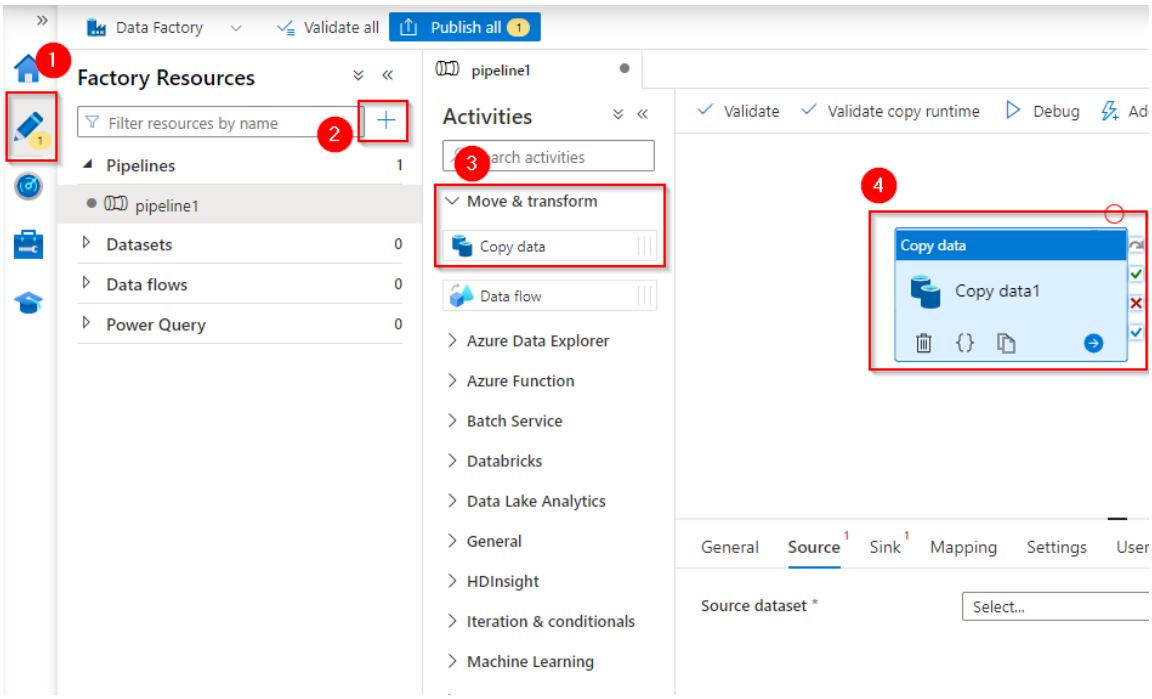
```
CREATE TABLE [dbo].[cars2] (
        [Make] [nvarchar](4000)  NULL,
        [Model] [nvarchar](4000)  NULL,
        [Type] [nvarchar](4000)  NULL,
        [Origin] [nvarchar](4000)  NULL,
        [DriveTrain] [nvarchar](4000)  NULL,
        [Length] [int]  NULL,
        [ManufactureDate] [datetime2](7)  NULL
)
```

3. Insert following rows

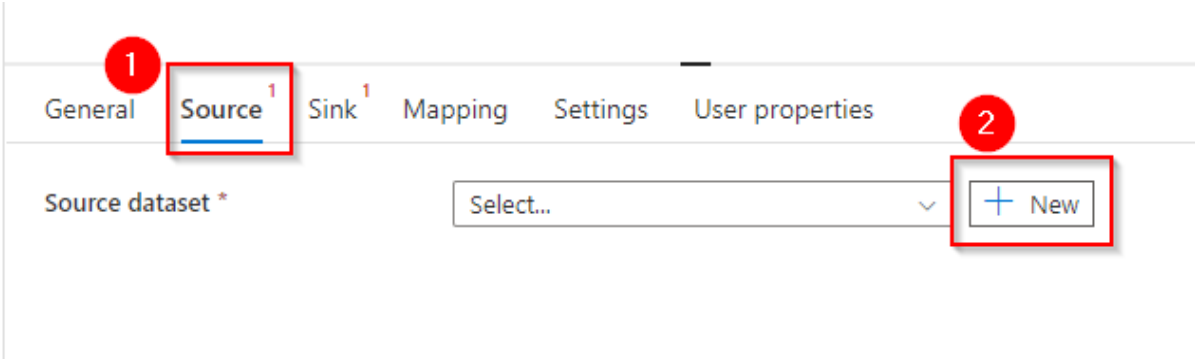
```
INSERT INTO cars2 VALUES('2001','MAX-1','MIN','ASIA','XXX',2889, '2012-01-01T00:02:00')
INSERT INTO cars2 VALUES('2002','MAX-2','MIN','ASIA','XXX',4889, '2012-01-01T00:03:00')
INSERT INTO cars2 VALUES('2003','MAX-3','MIN','ASIA','XXX',5589, '2012-01-01T00:05:00')
INSERT INTO cars2 VALUES('2005','MAX-4','MIN','ASIA','XXX',7879, '2012-01-01T00:07:00')
INSERT INTO cars2 VALUES('2006','MAX-5','MIN','ASIA','XXX',9889, '2012-01-01T00:09:00')
INSERT INTO cars2 VALUES('2004','MAX-1','MIN','ASIA','XXX',5589, '2012-01-01T00:04:00')
INSERT INTO cars2 VALUES('2005','MAX-2','MIN','ASIA','XXX',4889, '2012-01-01T00:06:00')
INSERT INTO cars2 VALUES('2007','MAX-3','MIN','ASIA','XXX',5839, '2012-01-01T00:05:00')
INSERT INTO cars2 VALUES('2008','MAX-4','MIN','ASIA','XXX',7189, '2012-01-01T00:07:00')
INSERT INTO cars2 VALUES('2009','MAX-5','MIN','ASIA','XXX',9789, '2012-01-01T00:01:00')
INSERT INTO cars2 VALUES('2011','MAX-1','MIN','ASIA','XXX',2889, '2012-01-01T00:08:00')
INSERT INTO cars2 VALUES('2012','MAX-2','MIN','ASIA','XXX',4879, '2012-01-01T00:08:00')
INSERT INTO cars2 VALUES('2013','MAX-3','MIN','ASIA','XXX',5399, '2012-01-01T00:02:00')
INSERT INTO cars2 VALUES('2025','MAX-4','MIN','ASIA','XXX',7489, '2012-01-01T00:11:00')
INSERT INTO cars2 VALUES('2036','MAX-5','MIN','ASIA','XXX',9869, '2012-01-01T00:12:00')
INSERT INTO cars2 VALUES('2021','MAX-1','MIN','ASIA','XXX',2449, '2012-01-01T00:15:00')
INSERT INTO cars2 VALUES('2042','MAX-2','MIN','ASIA','XXX',4969, '2012-01-01T00:12:00')
INSERT INTO cars2 VALUES('2023','MAX-3','MIN','ASIA','XXX',5259, '2012-01-01T00:13:00')
INSERT INTO cars2 VALUES('2065','MAX-4','MIN','ASIA','XXX',7819, '2012-01-01T00:14:00')
INSERT INTO cars2 VALUES('2026','MAX-5','MIN','ASIA','XXX',9579, '2012-01-01T00:14:00')
```

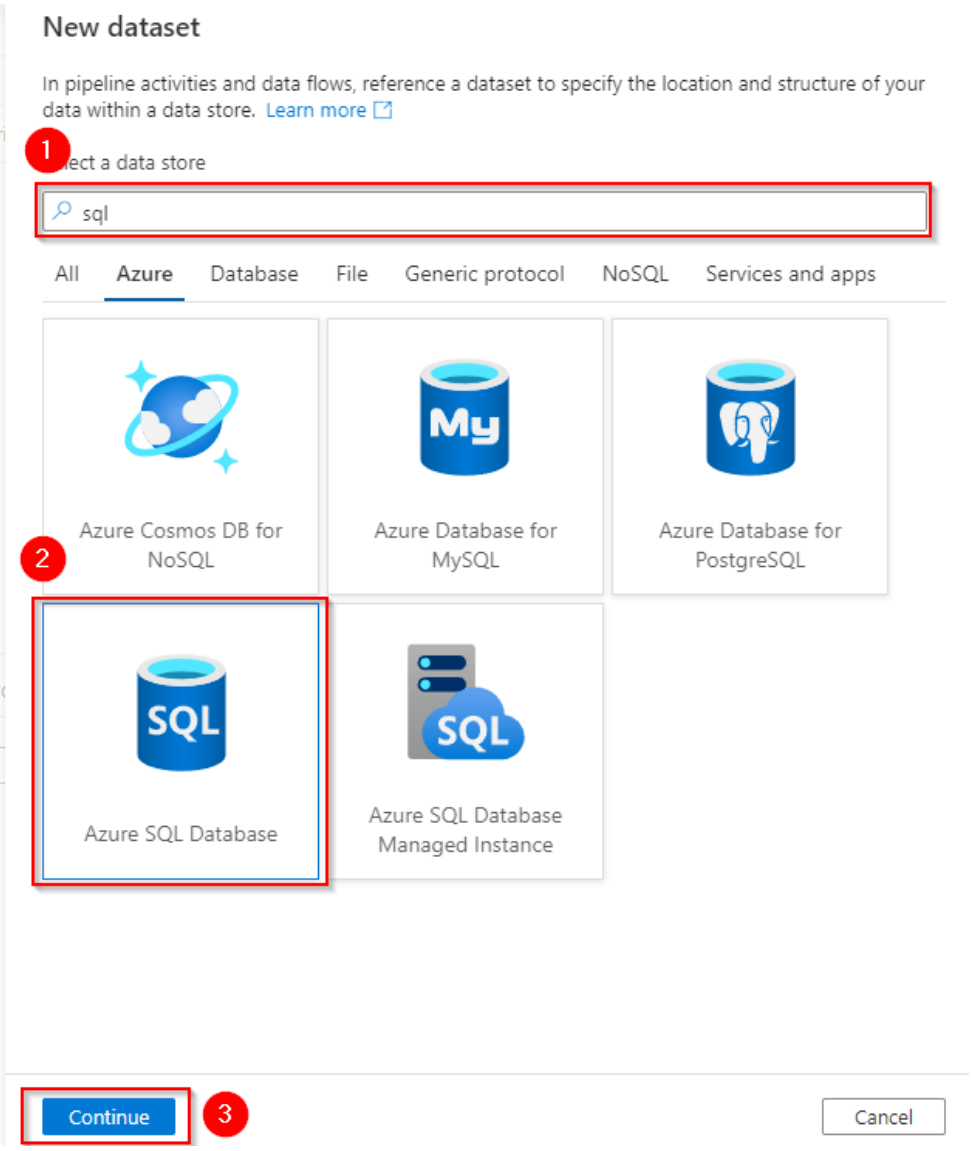
Task 2: Create a copy activity to copy data from SQL Table

1. Add copy activity to pipeline canvas



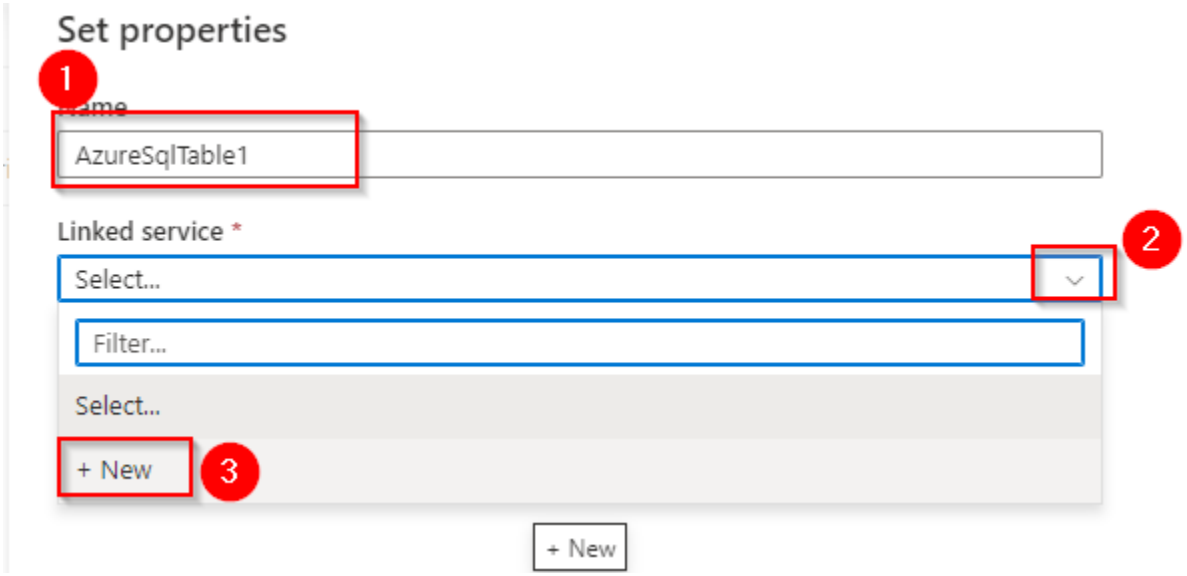
2. Source dataset: Add new SQL DB data source. (Create linked service if not exists for SQL DB)



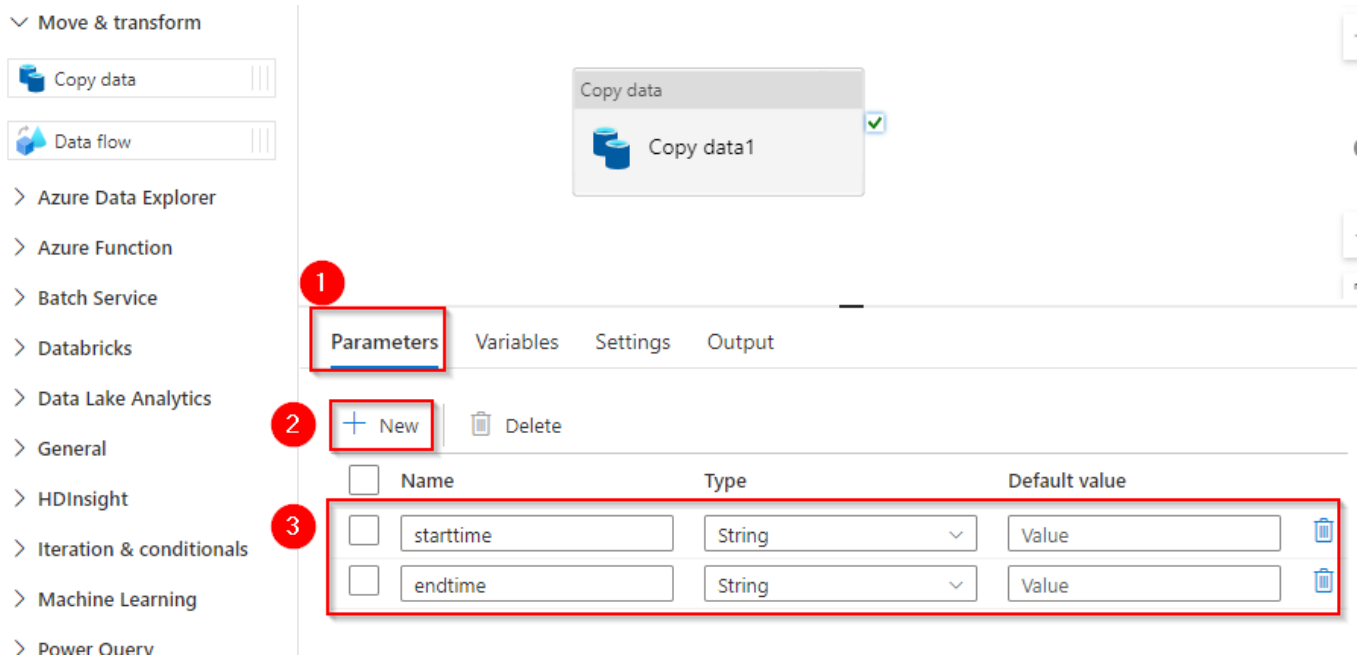


3. Select the linked Service and select the dbo.cars2 table,

Note: Create a linked service if not exists otherwise skip this step.

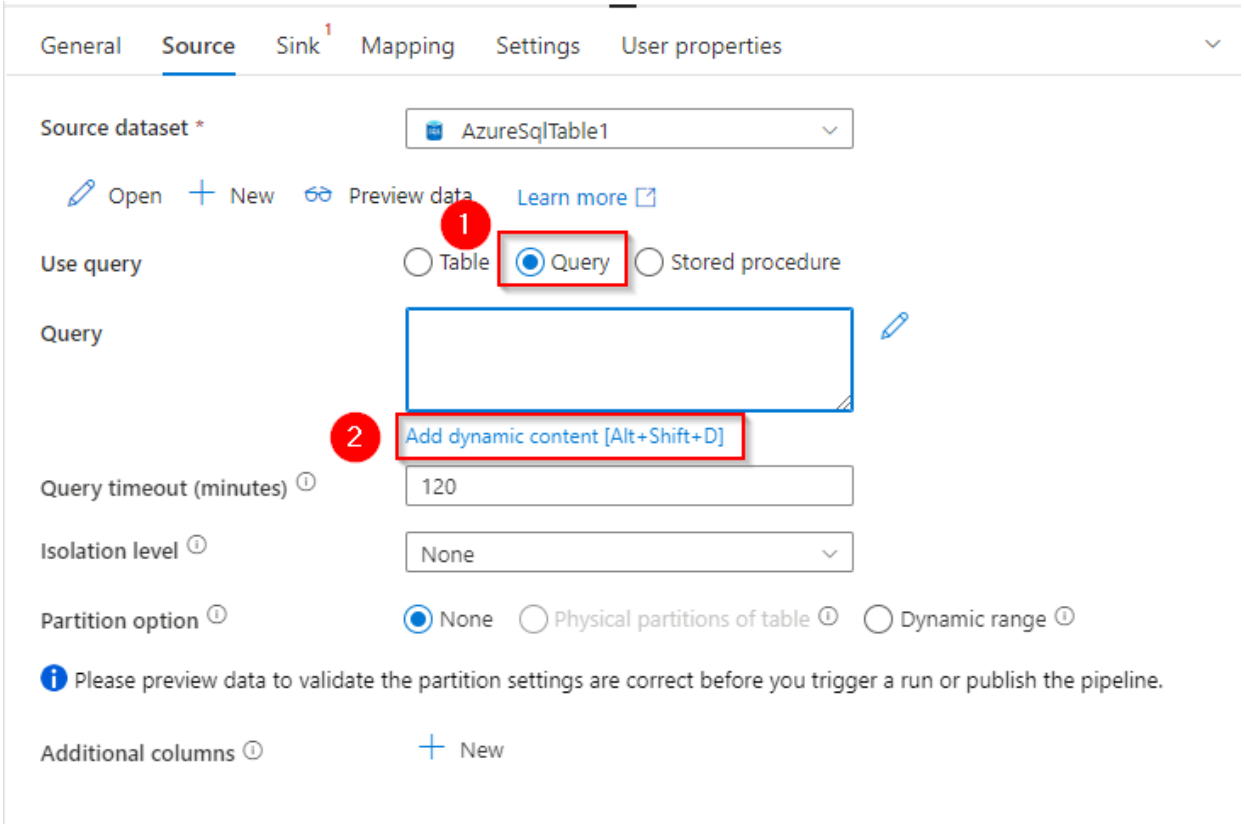


4. Create two pipeline parameter as follows:
   a. starttime
   b. endtime

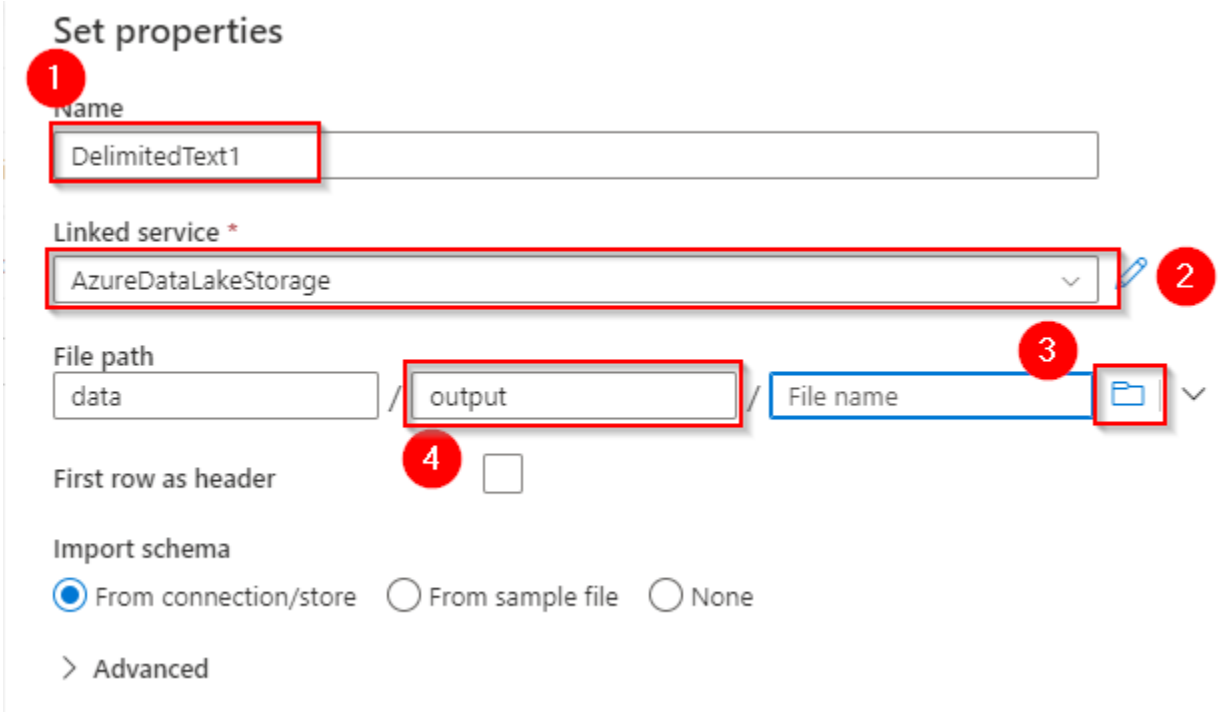


5. Configure the source of copy activity as show in figure:
   Use query: Query
   Query: Click Add dynamic content and add the following query

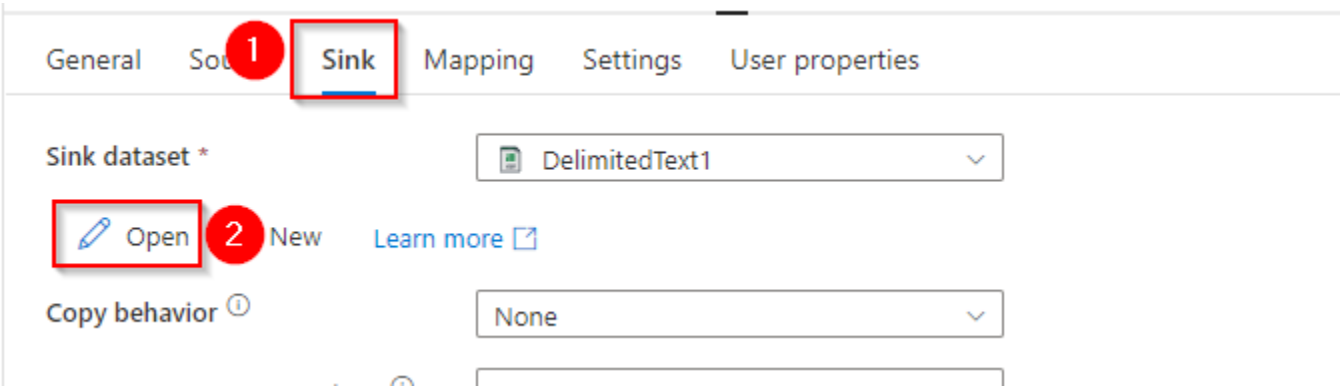
   select * from cars2 where manufacturedate > '@{pipeline().parameters.starttime}' and manufacturedate <= '@{pipeline().parameters.endtime}'



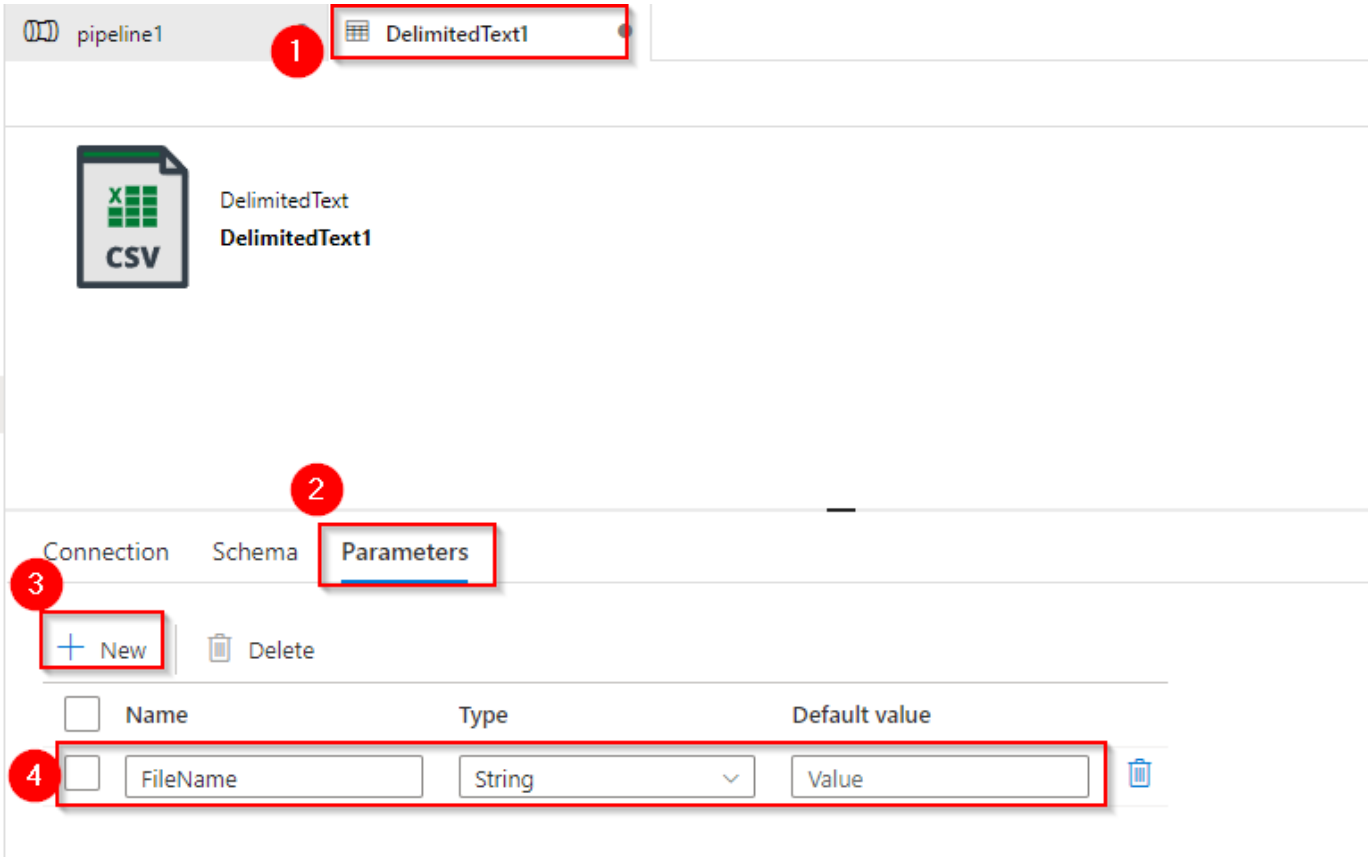
6. Configure the sink of copy activity, select new dataset.
   Type: Azure Data Lake Gen2 Storage Account
   File Format: Delimited Text
   Linked Service: Select from the drop down if exists other wise create a new one (2).
   File Path: Select data container (3) and enter directory name output (4). Leave file name empty. Click Ok.



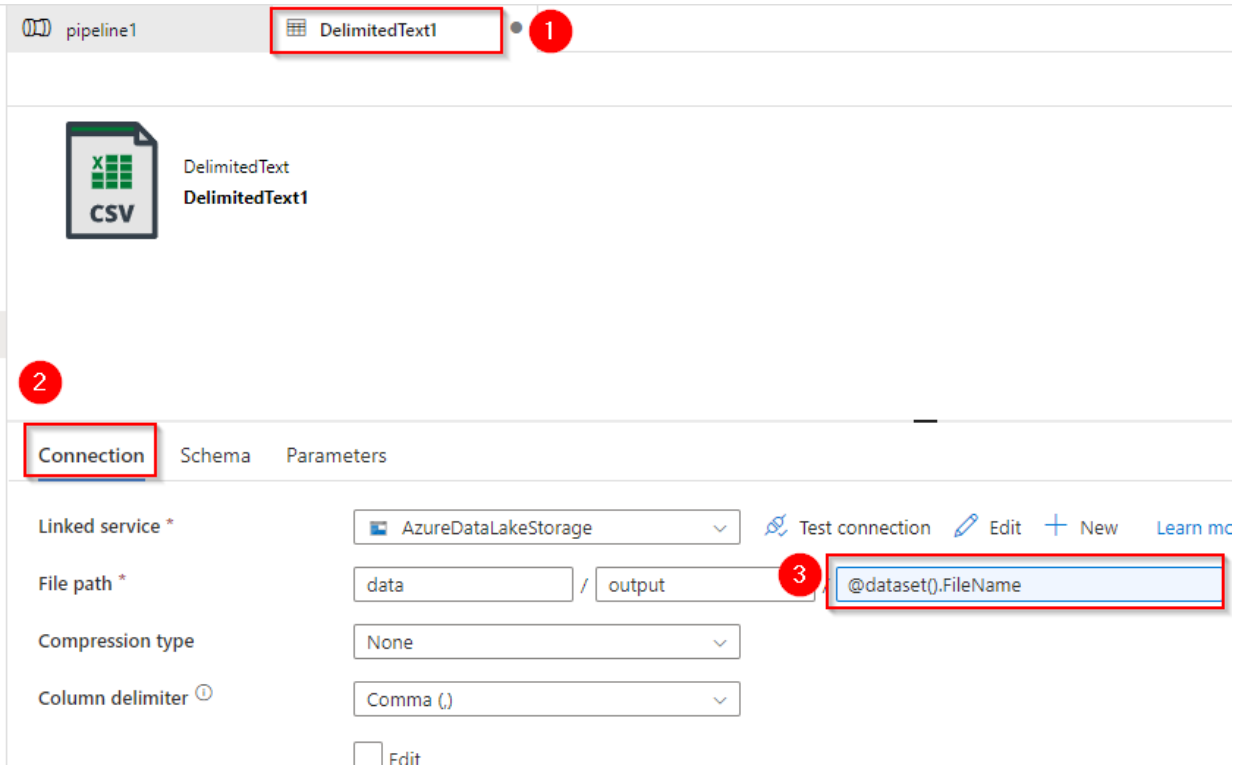
7. Open sink dataset (created in previous step)



8. Add dataset parameter**, FileName**.

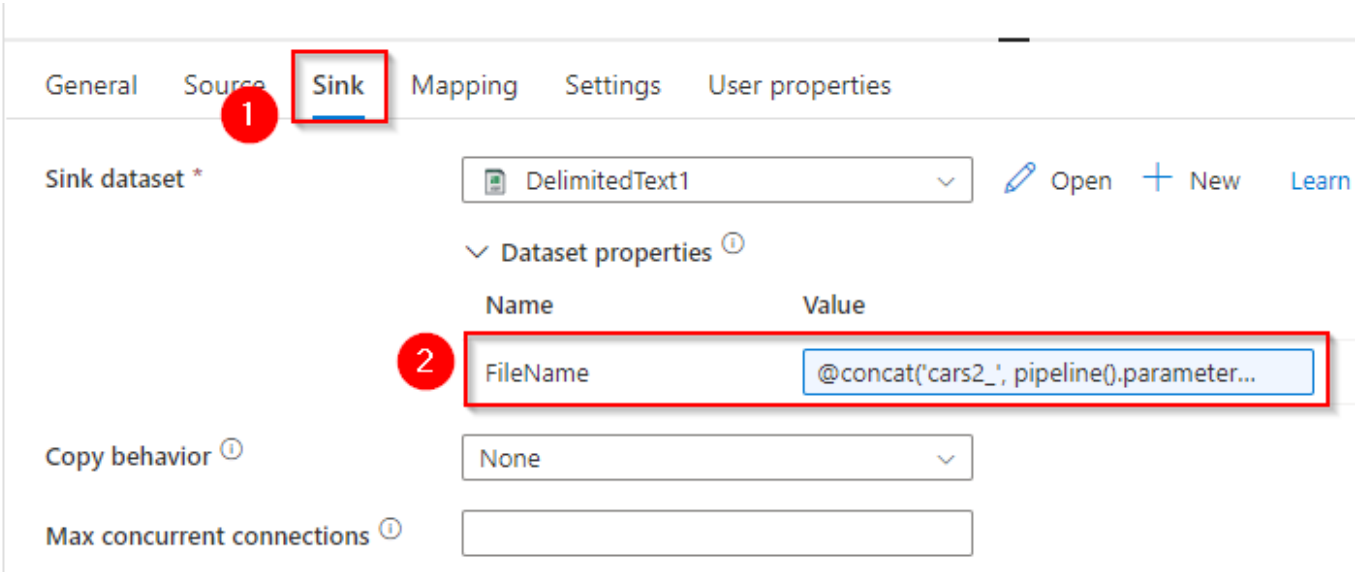


9. In dataset, go to connection tab and give the file name as show:

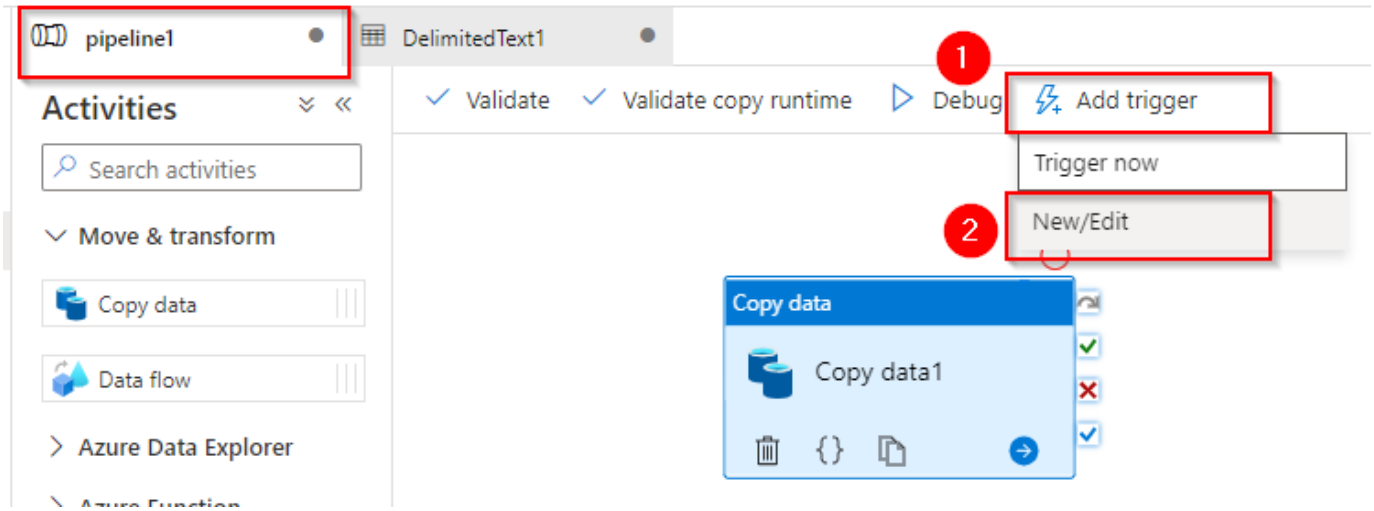


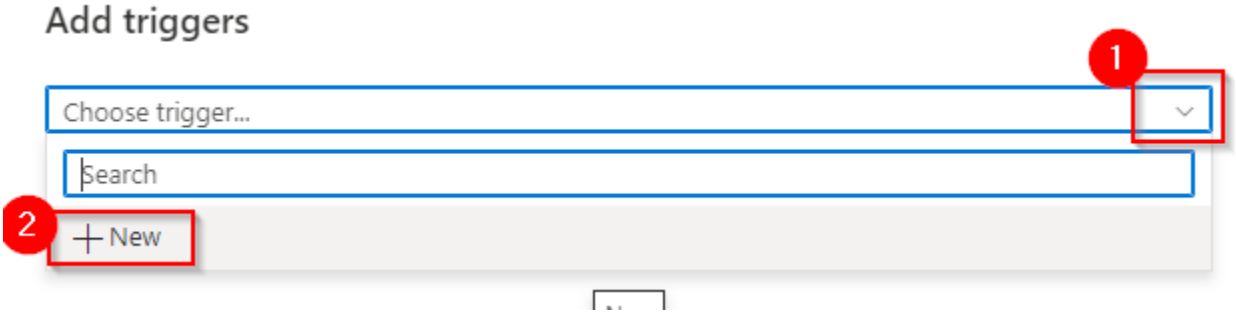
10. Go back the sink configuration in pipeline and enter the following using add dynamic content for **FileName** properties of Dataset.

```
@concat('cars2_', pipeline().parameters.endtime, '.csv')
```



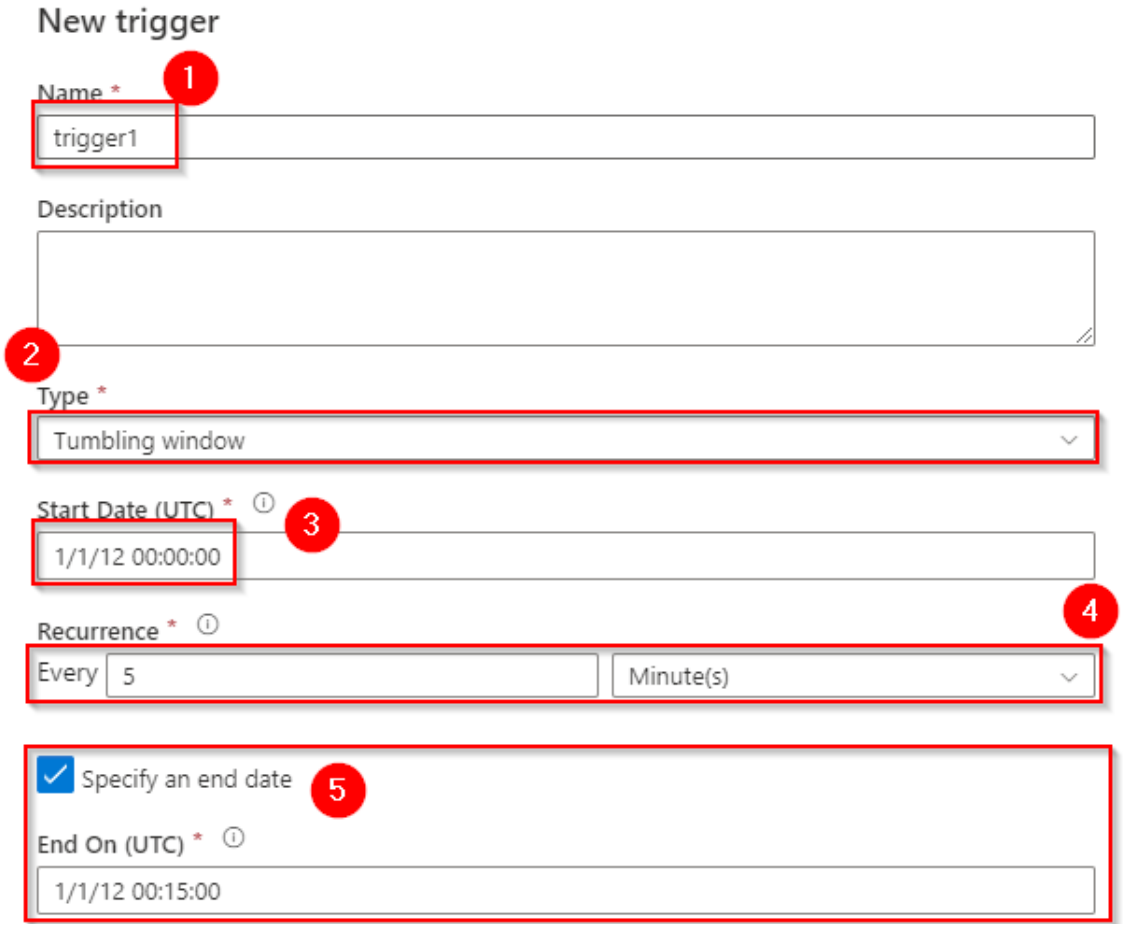
11. Add a trigger to pipeline



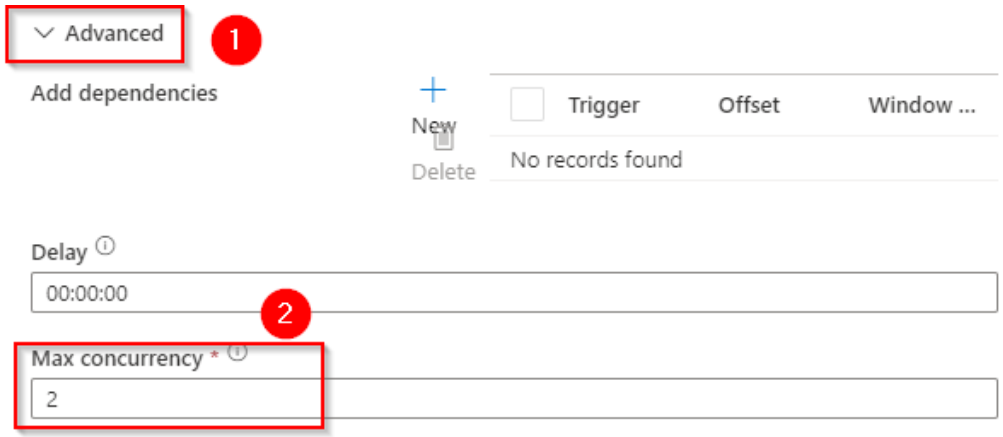


12. Create new Tumbling window trigger as follows:

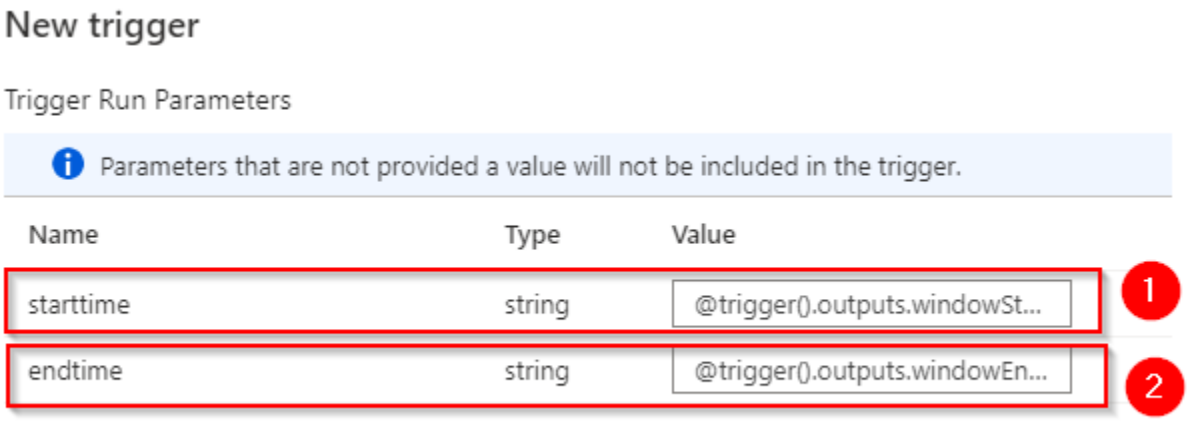
Trigger Name: trigger1
Type: Tumbling Window
Start Date: 1/1/12 00:00:00
End Date: 1/1/12 00:15:00



In Advanced: Max Concurrency: 2

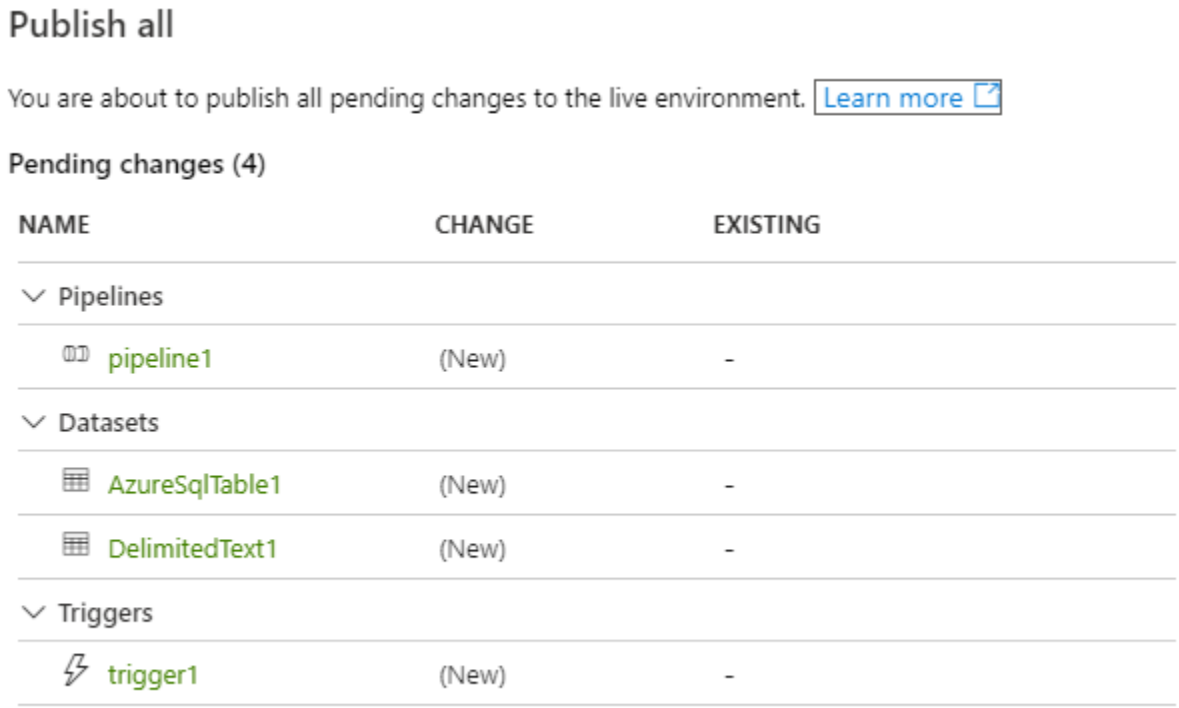


13. Click ok, on the next page enter Trigger Run Parameter
    strarttime: @trigger().outputs.windowStartTime
    endtime: @trigger().outputs.windowEndTime



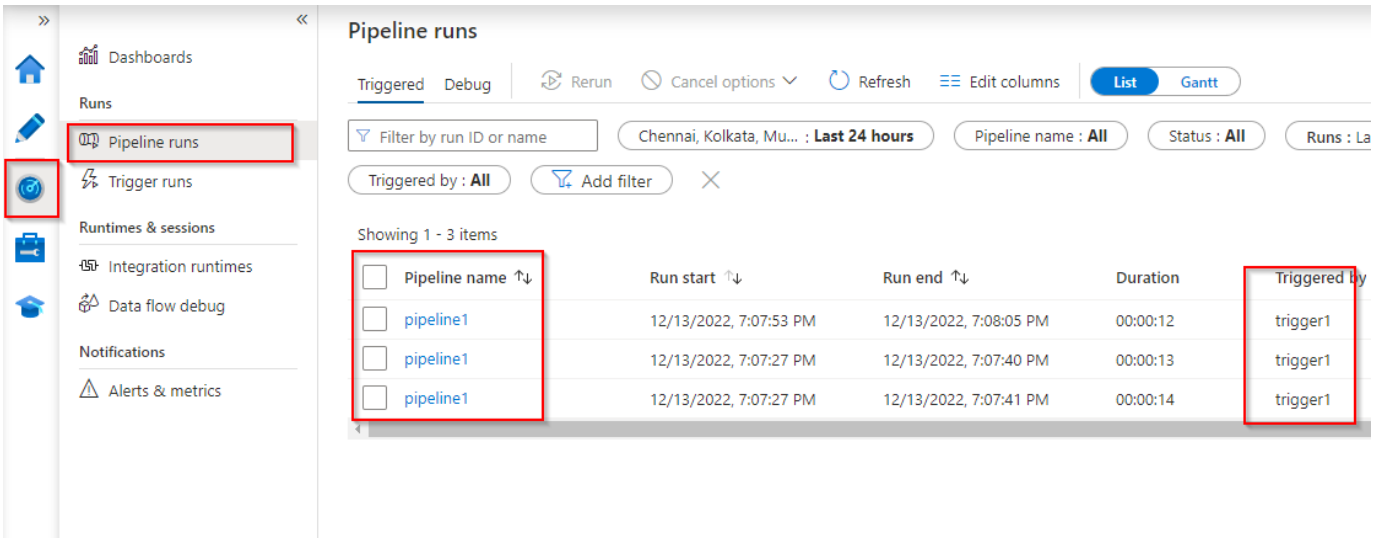
Then click Ok.

14. Click on Publish all to save the pipeline, dataset, linked service etc.



Click on Publish.

15. When success fully published, go to Monitor Tab and see the pipeline run.



16. Verify the files in the storage account.