Azure Data Engineer: Batch 6

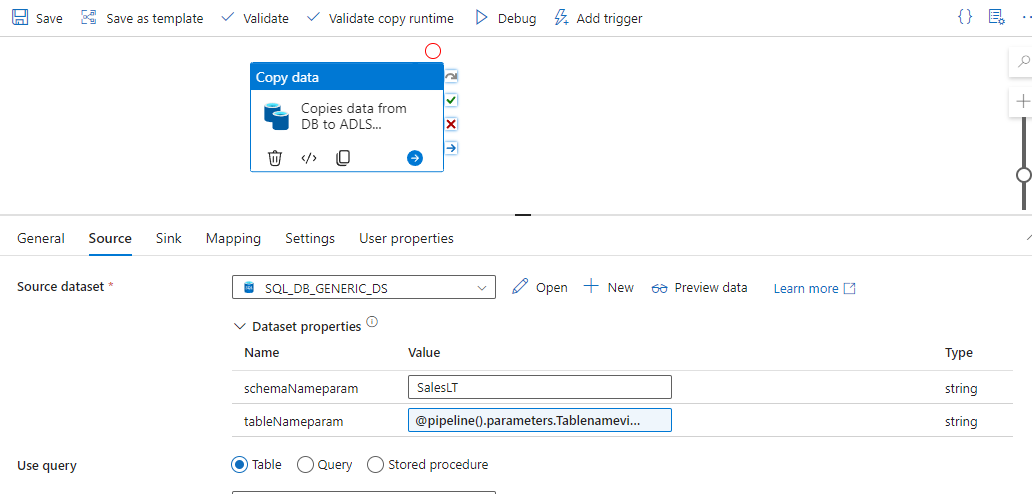
(October 2023)

**Submitted by: Roopmathi Gunna on Dec 8,2023**

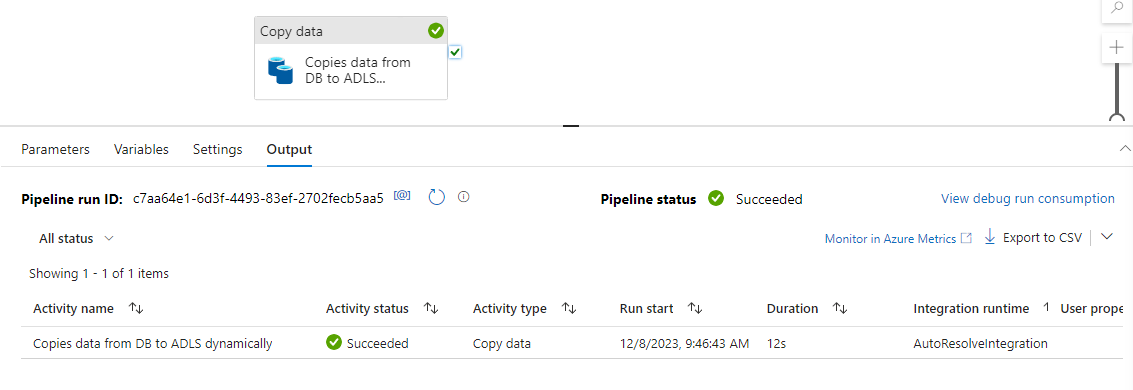
# Assignment 11: Pipeline Parameters and Set Variables

Q1. Create a pipeline to pull the data from the Az SQL database to ADLS. However, while creating the pipeline ensure that Table name and folder is dynamic in nature. For example, if I ask to run pipeline to pull the data from customer table it without any change pipeline should pull the data for customer table and save it in Customer folder in ADLS. Incase if ask to do same for Product table it changes table and folder accordingly without making any change in pipeline. Hint: use pipeline Param

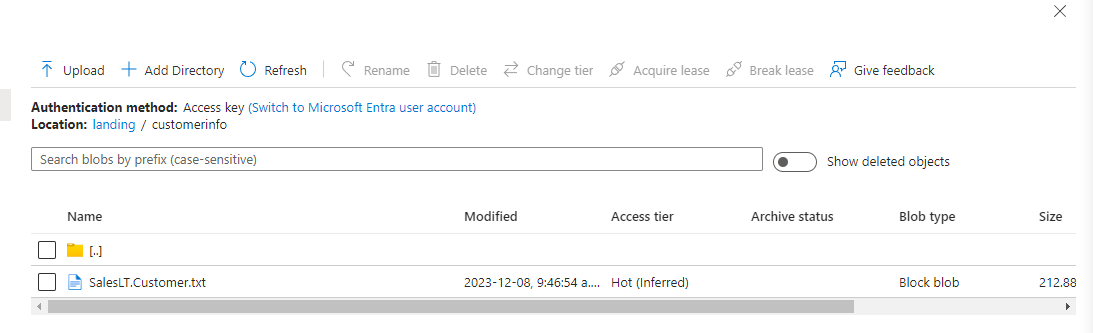
Here is how to set-up the Copy Activity- Source Settings



**Here is the output:**



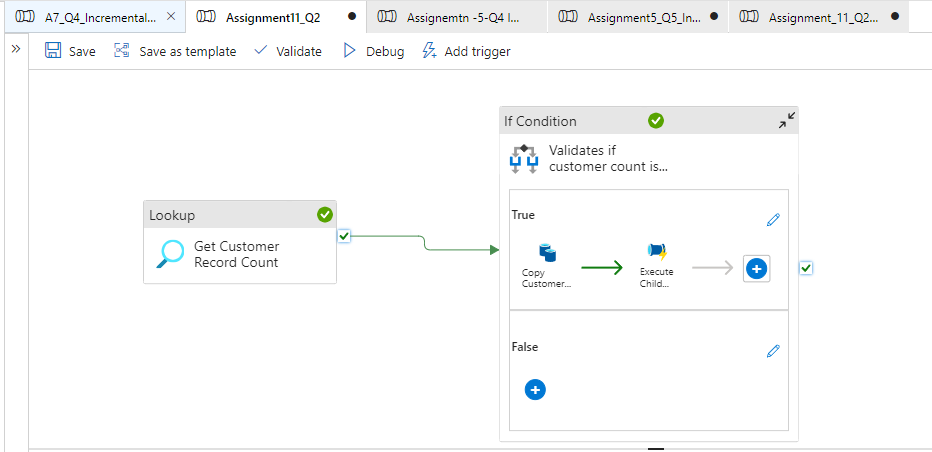
Let’s validate if the data has been copied:



Customer Data is copied successfully in our ADLS location provide in the pipeline parameter.

Q2. Create a pipeline to copy customer data from SQL DB to ADLS only if record count is more than 500. Once a data gets copy it should call a child pipeline (which will copy the product data from table if customer record count is > 600). Design the pipeline in such a manner that it will pass the Customer count to the child product pipeline via Pipeline parameter.

This is the pipeline I have built which as Parent Pipeline( To copy Customer table) and Child pipeline(To copy product table)



Inside the True condition of If activity – we have passed the record count as a pipeline parameter

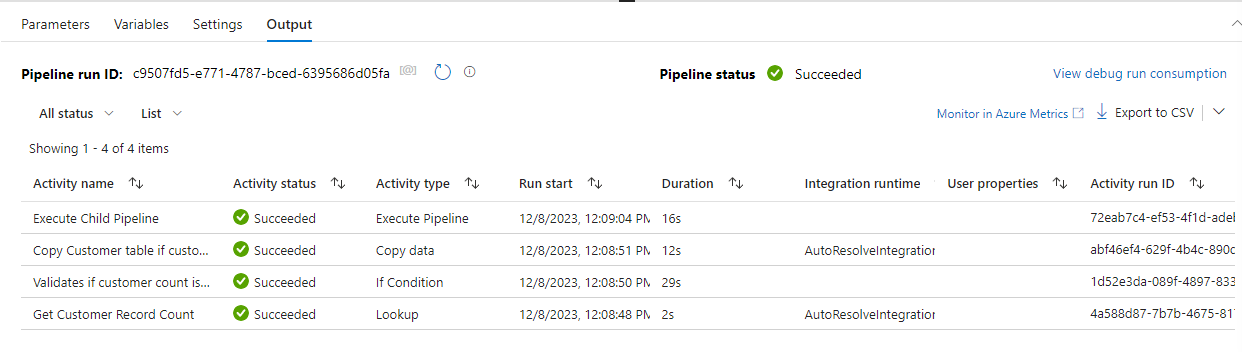
A screenshot of a computer

Description automatically generated

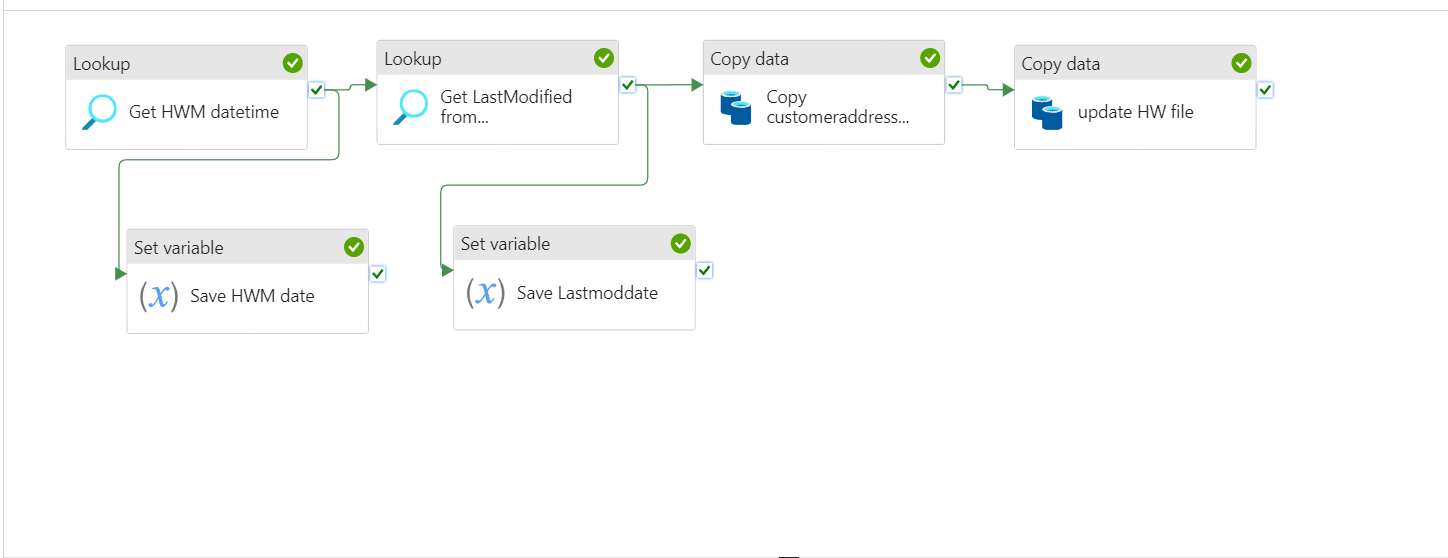
A screenshot of a computer

Description automatically generated

Al steps executed successfully:



Q3. Create the incremental pipeline to pull the CustomerAddress data. But this time use the Set variable activity to hold the last Modified date and highwater mark date. Hint: (So far we are directly using the output of lookup activity. Lets store the lookup activity output to variable and then use it.)



This how you configure each of the Set Variables :

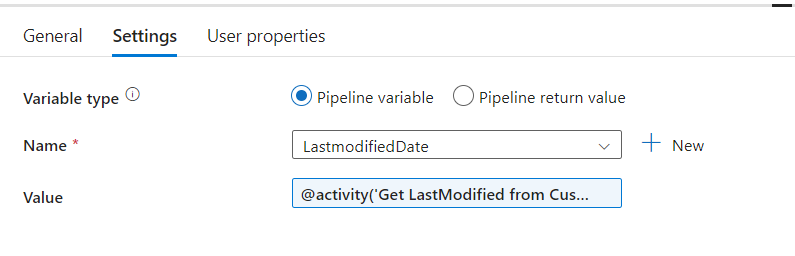
Lets see the config of the HW\_date variable

A screenshot of a computer

Description automatically generated

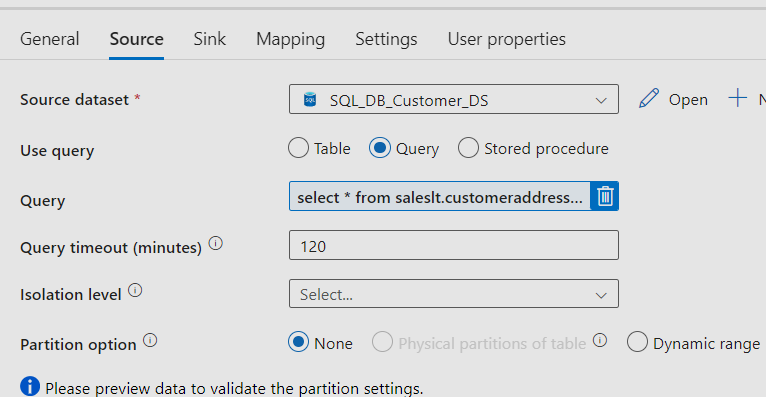
Dynamic expression : @activity('Get HWM datetime').output.firstRow.hw\_date

Let look now at the configuration for the Lastmodifieddate variable



@activity('Get LastModified from CustomerAddress Table').output.firstRow.lastdate

This is the configuration of the copy activity



select \*

from saleslt.customeraddress

where modifieddate > '@{variables('Hw\_date')}'

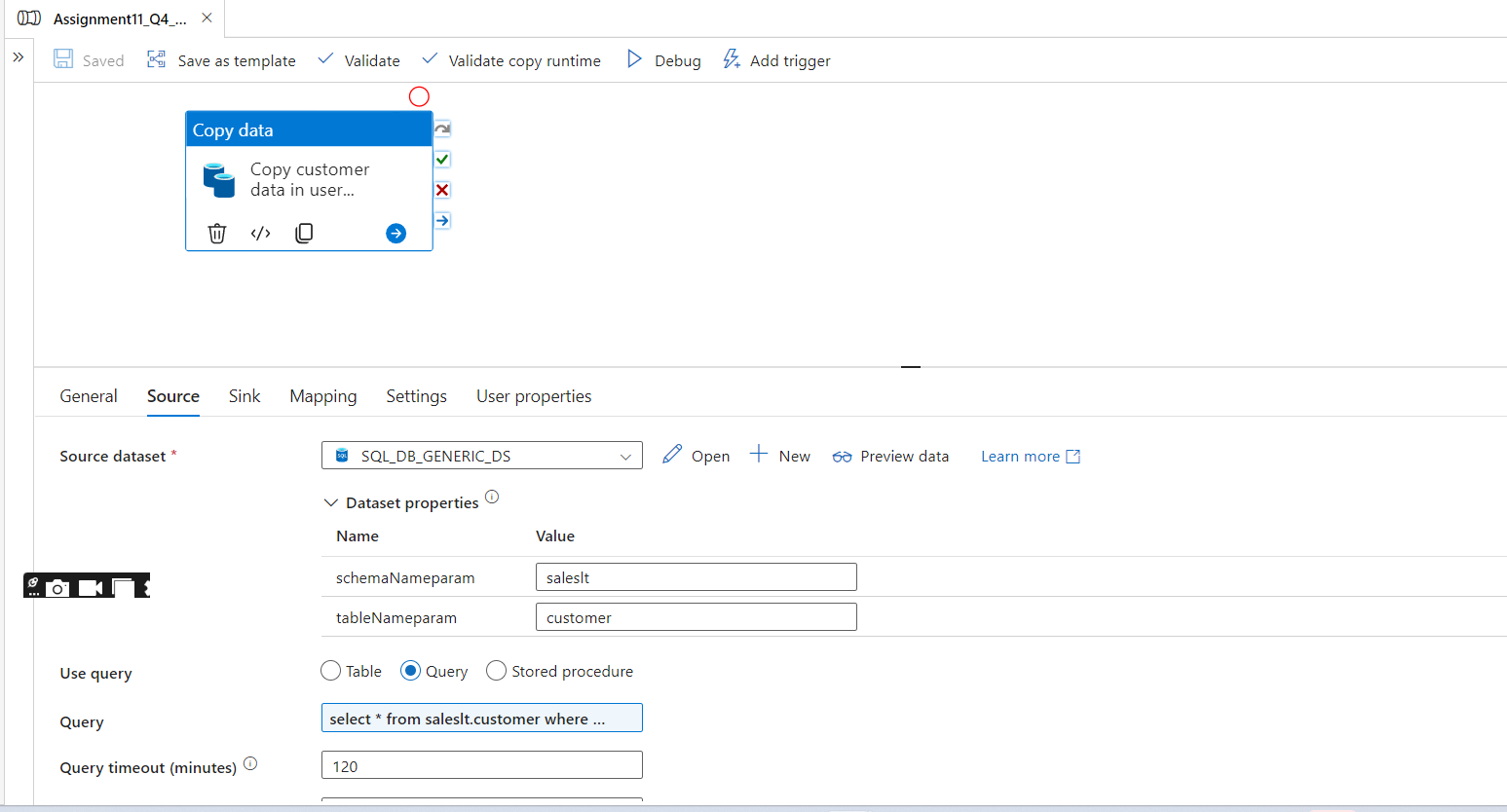
    and modifieddate <= '@{variables('LastmodifiedDate')}'

And this is our Output:

A screenshot of a computer

Description automatically generated

Q4. Create a pipeline which will pull the data for customer table from Azure SQL Db to ADLS. But this time data needs to be pulled based on the date range provided at the runtime. Hint: use pipeline Param.



Dynamic content in the Query is this :

select \*

from saleslt.customer

where modifieddate >= '@{pipeline().parameters.Startdate}'

  and modifieddate <= '@{pipeline().parameters.EndDate}'

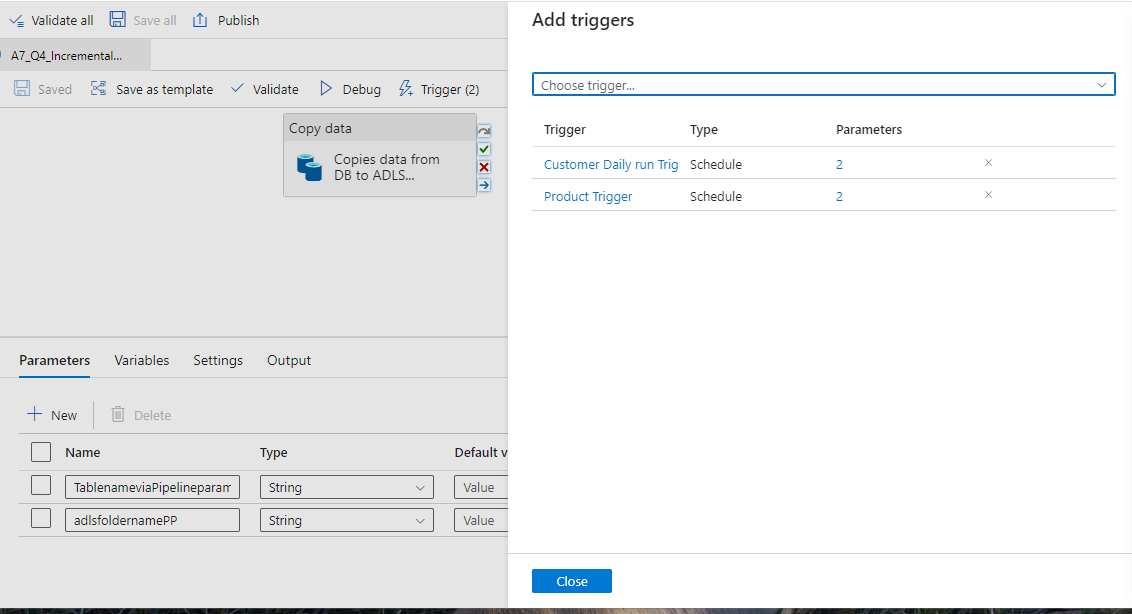
These are the pipeline parameters used:

A screenshot of a computer

Description automatically generated

Q5. Create the two triggers for Q1 pipeline one which is used to pull Product data and other for CustomerAddress. (Daily Once trigger)

Here I have added two triggers one for customer table and one for Product Table that triggers the same Q1 pipeline but copies different tables as per this question.



These are the triggers set-up :

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated