Custom Training

Day 7

**Azure Data Factory**

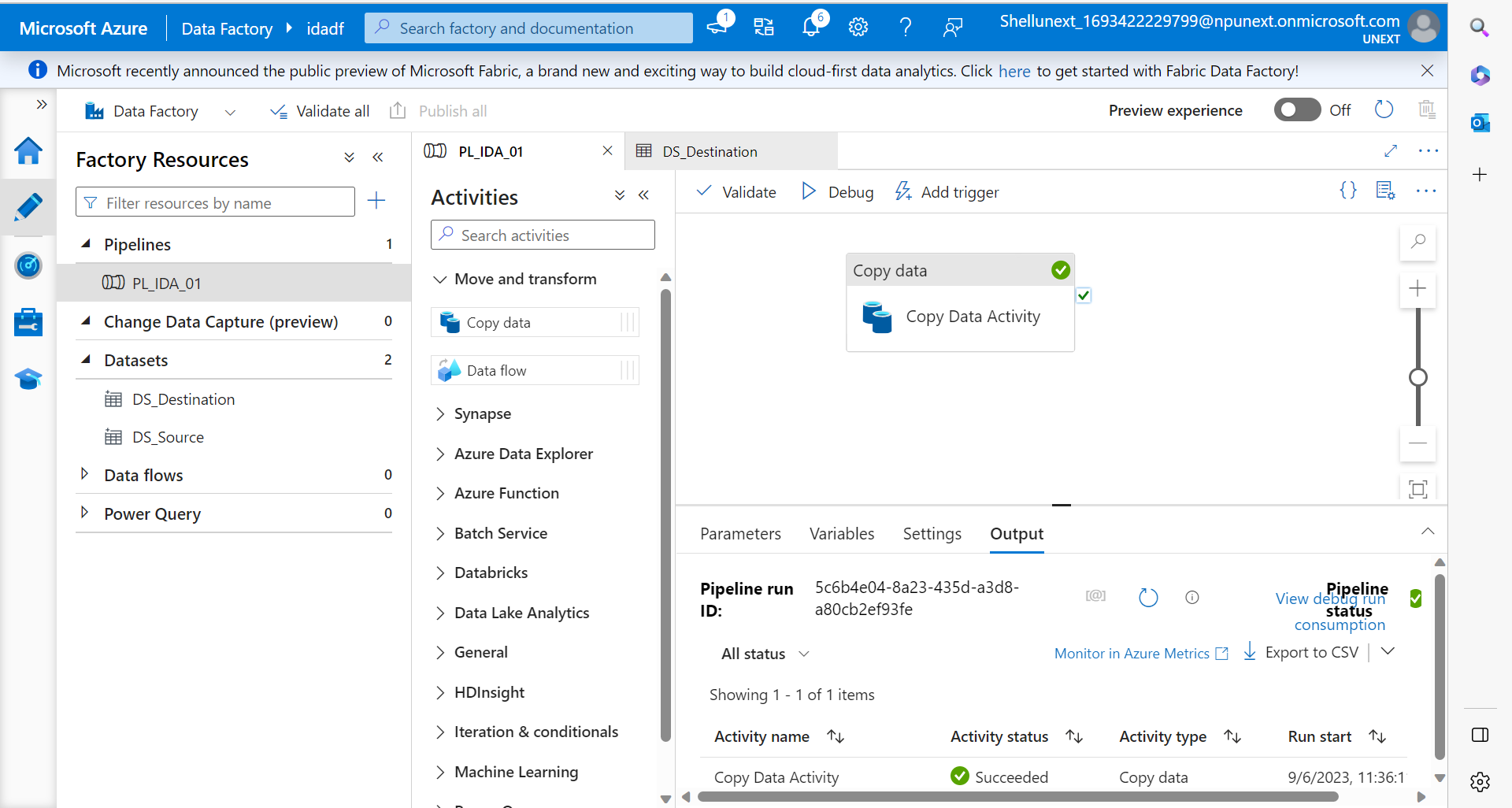
General Activities:

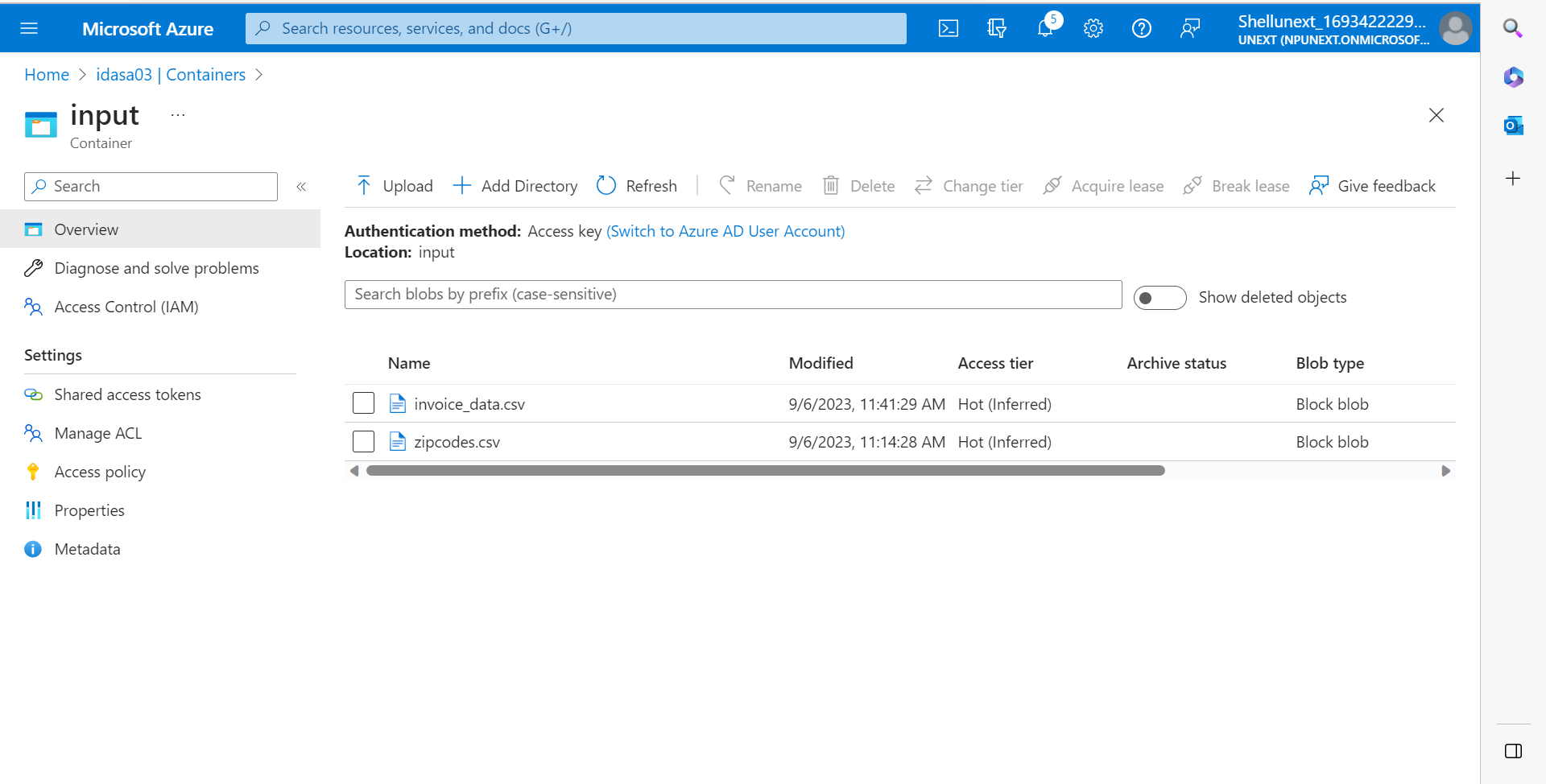
1. Append Variable: Supported only in array.
2. Set Variable: It is used to set/change the value of a variable.
3. Delete: Delete the values

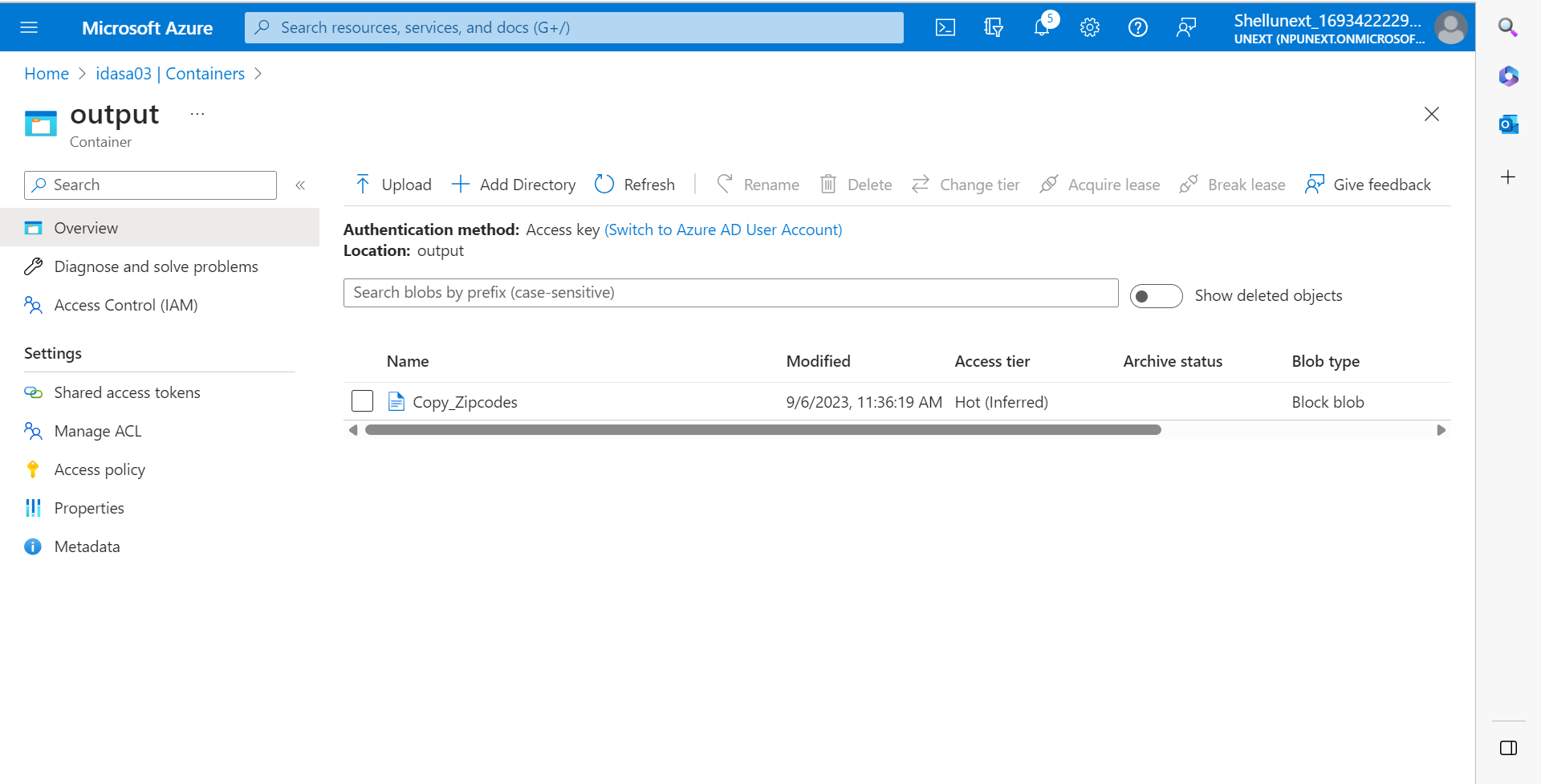
Within a single pipeline we cannot have more than 40 activities. For this we can have execute pipeline activity. This is known as chained pipelines.

1. Lookup: To get the information about the table we can use Lookup activity.
2. System Variables: Used for audit purpose. When the pipeline runs and completes. We can create audit files and tables.
3. MetaData: To get the information about the data or files or folders like the size, name,etc.
4. Filter: To filter out the data based on some conditions.
5. Web Hook: To call an API, and perform crud operations.
6. Wait: To wait for sometime between the execution of two activities.

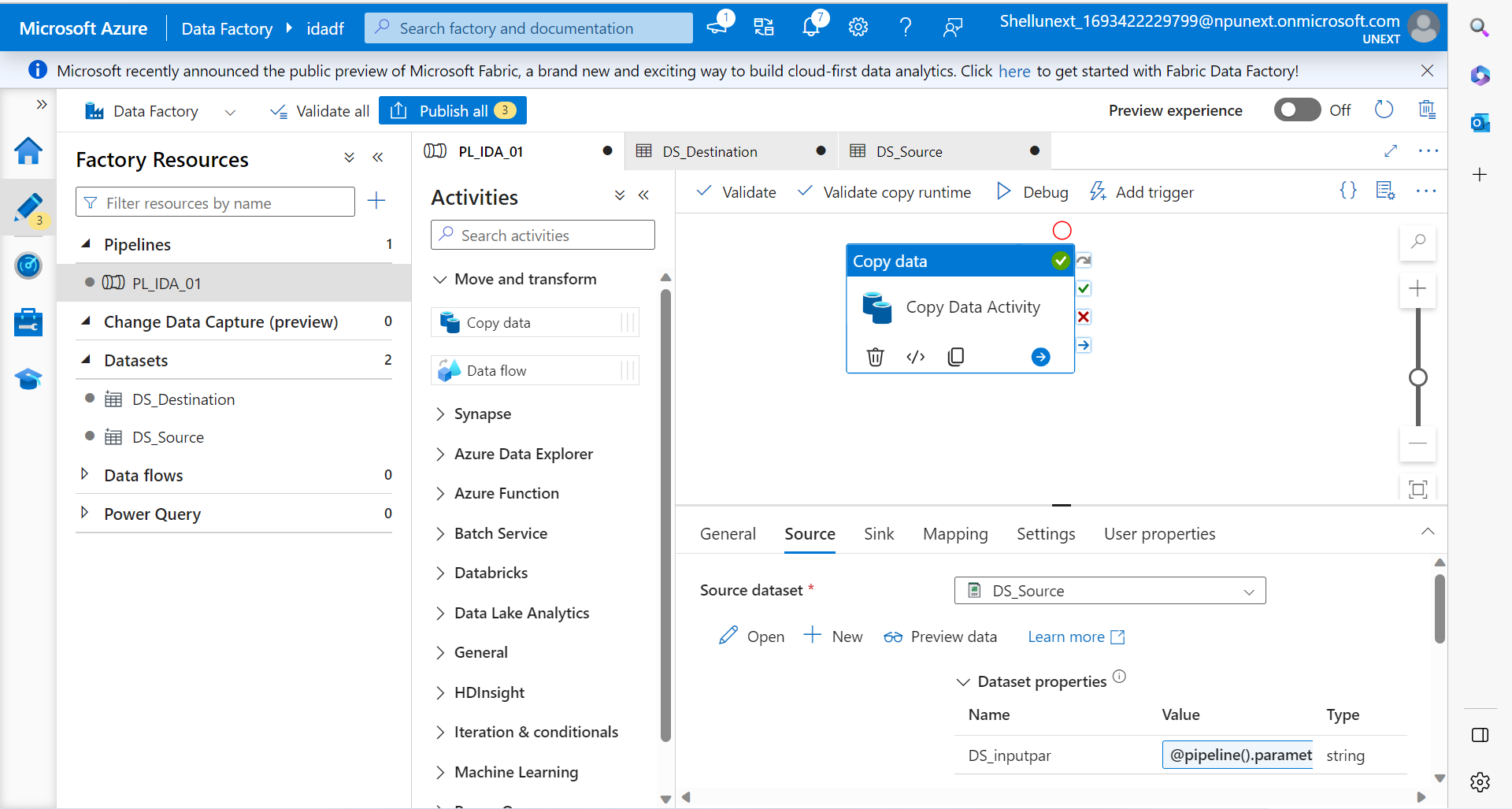
Pipeline creation without parameters:

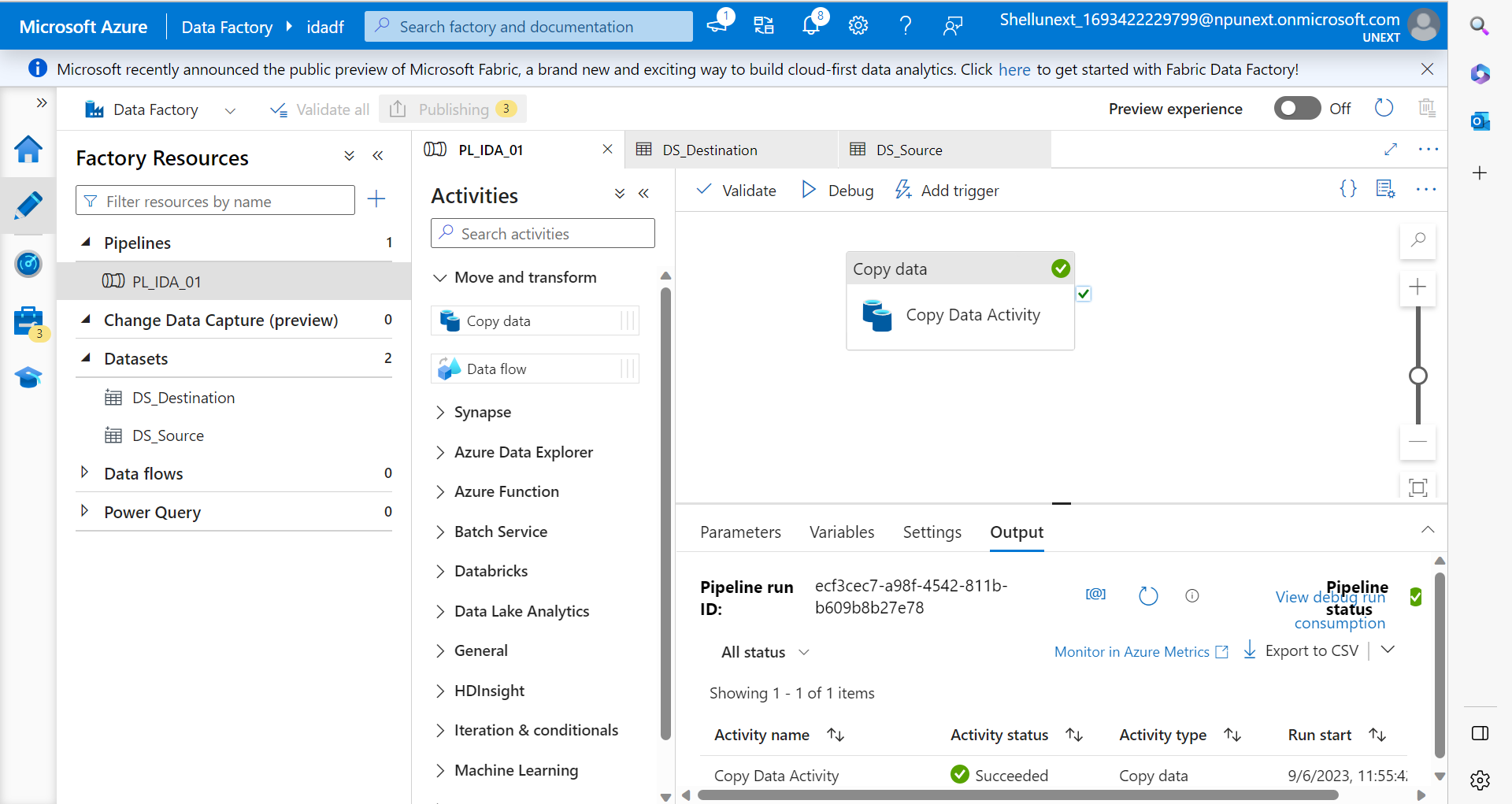


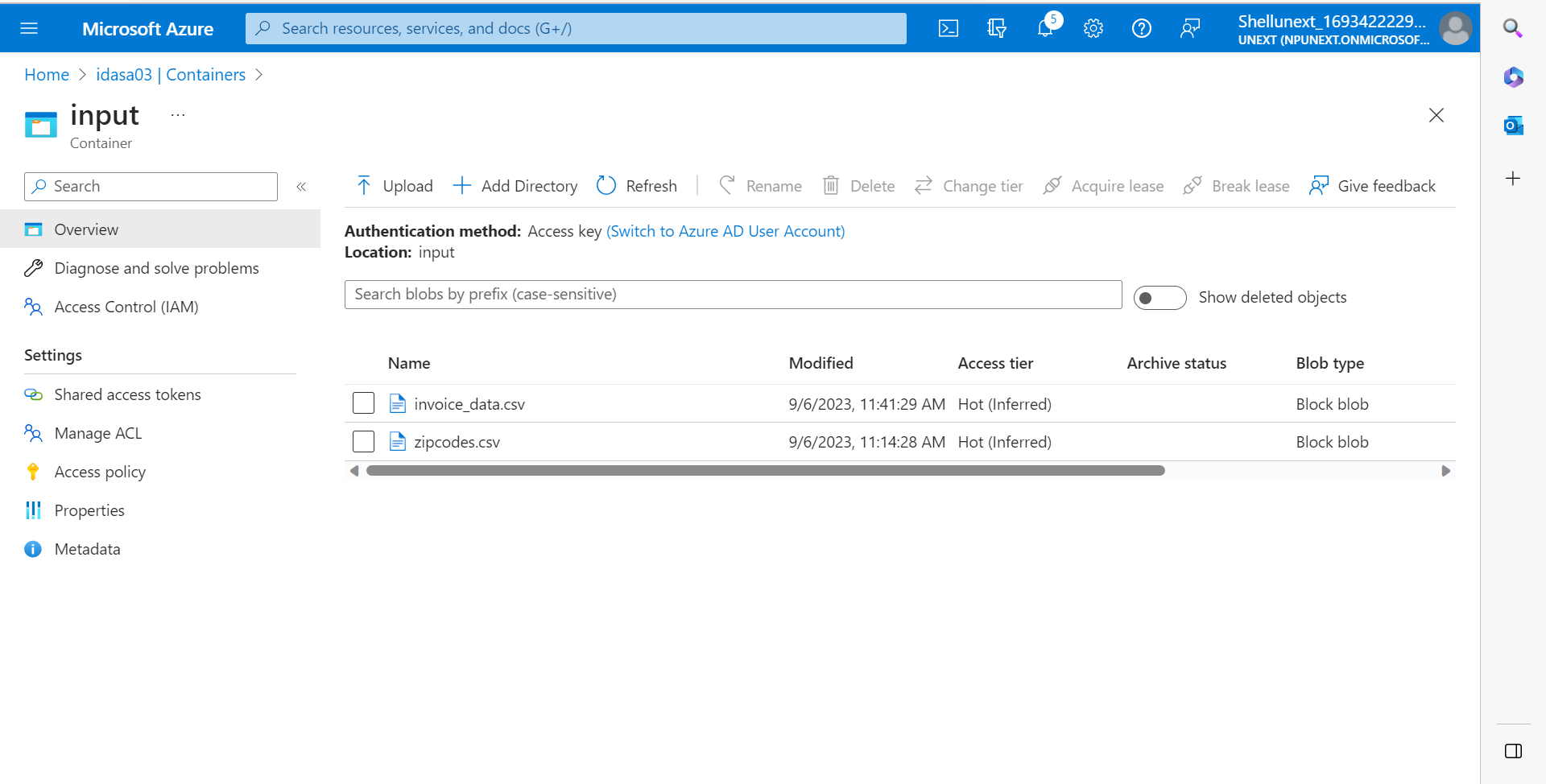


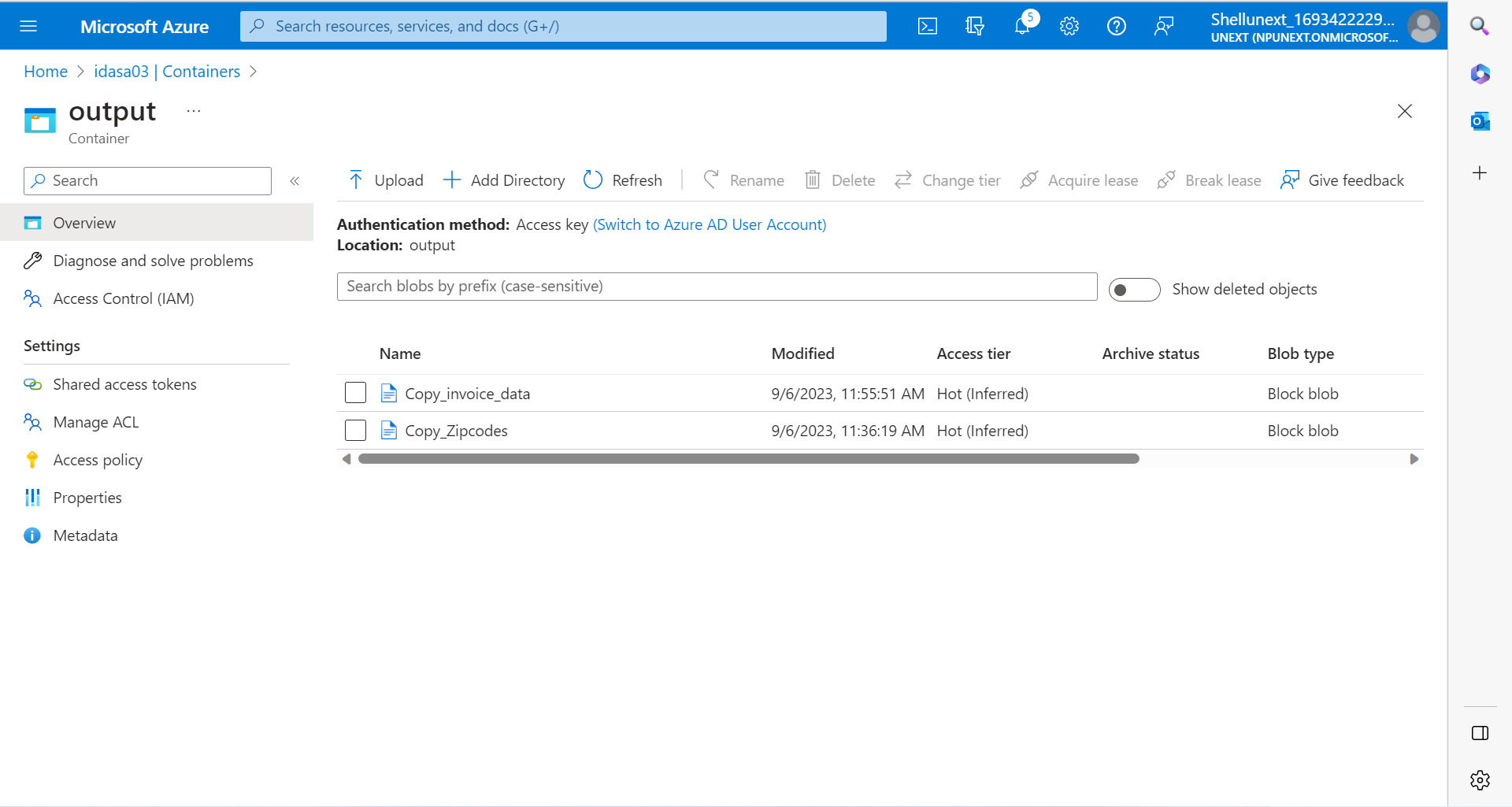


Pipeline with Parameters:

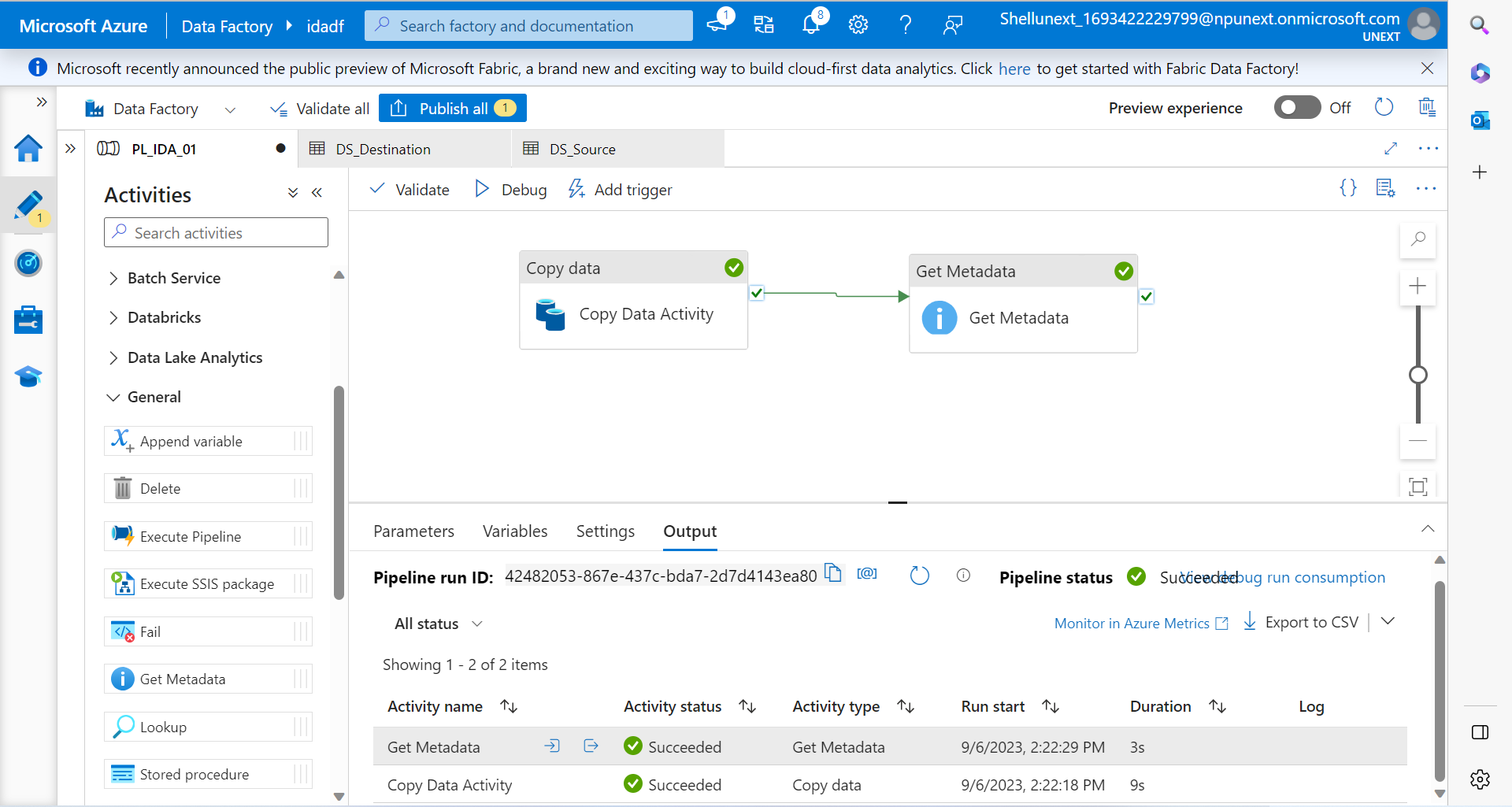


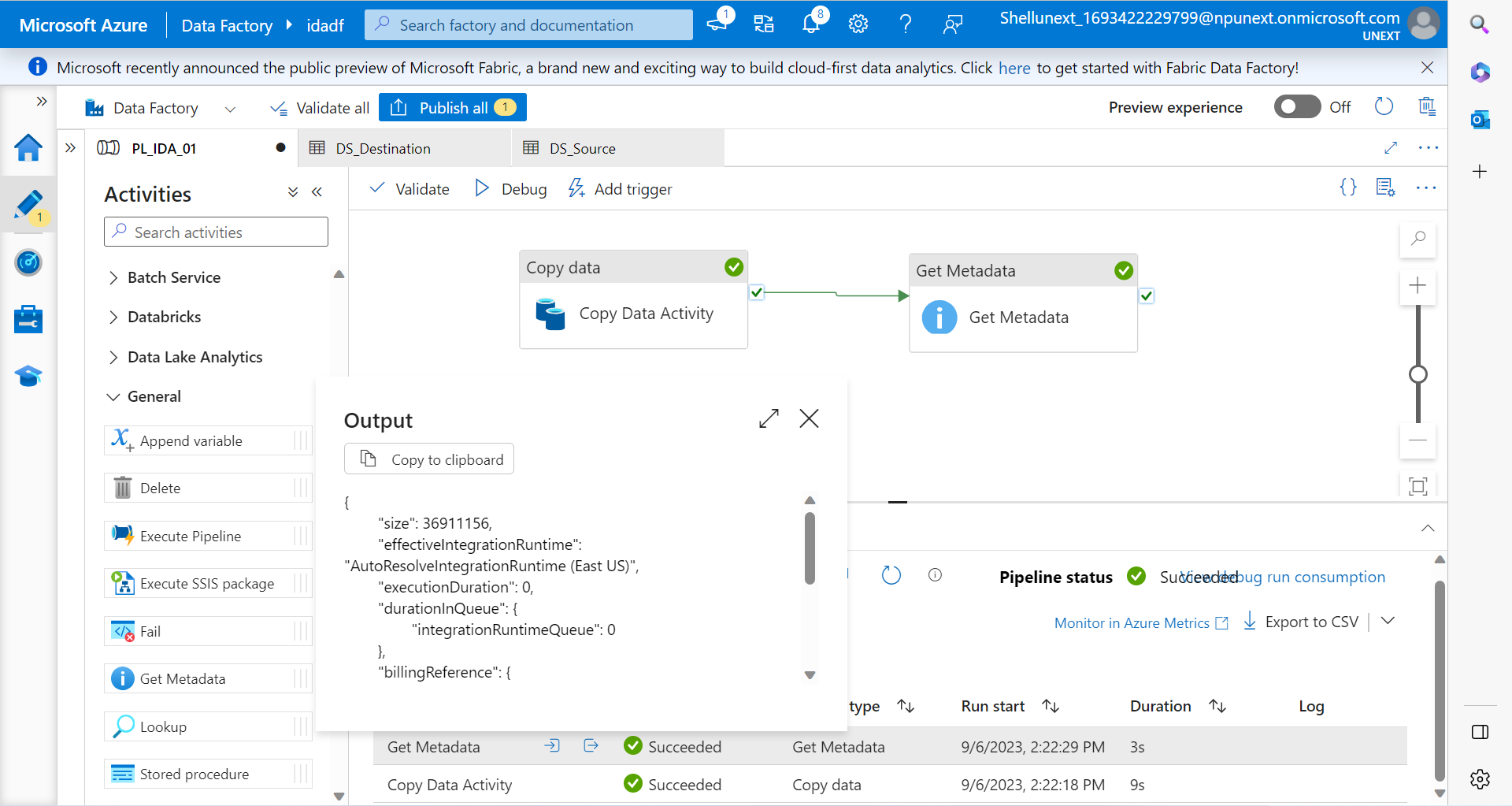


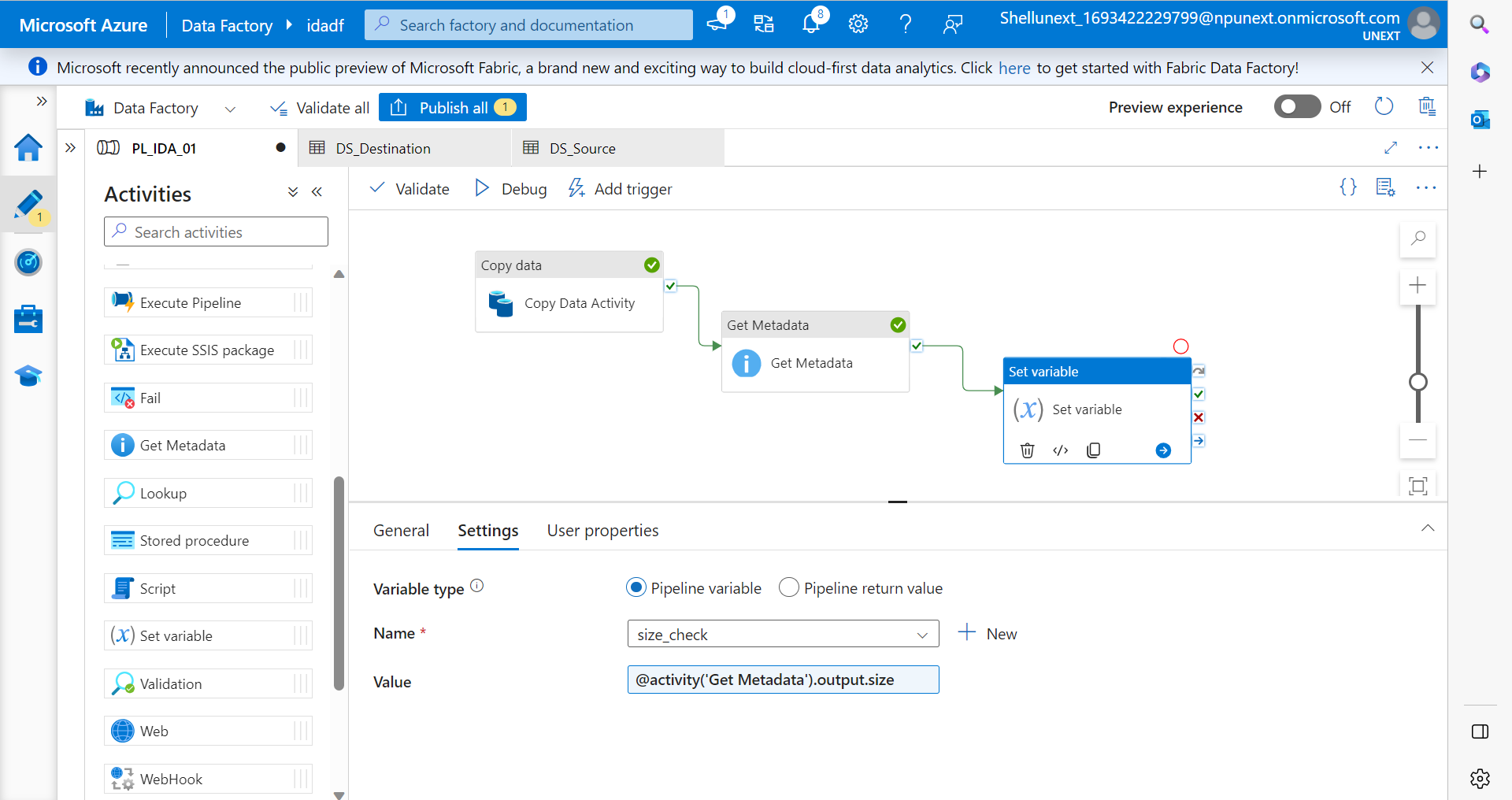


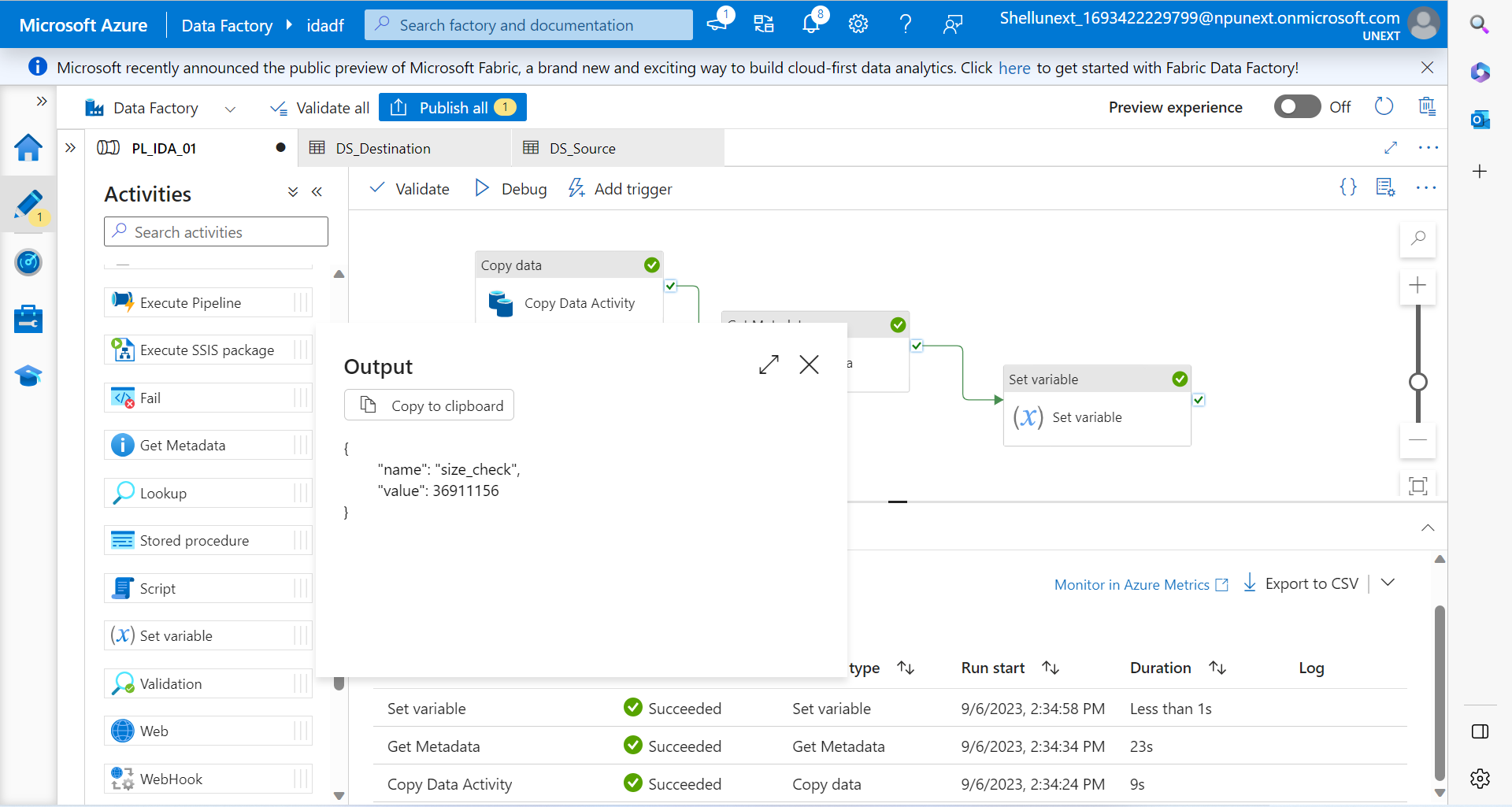


To get the meta data of the input and output:









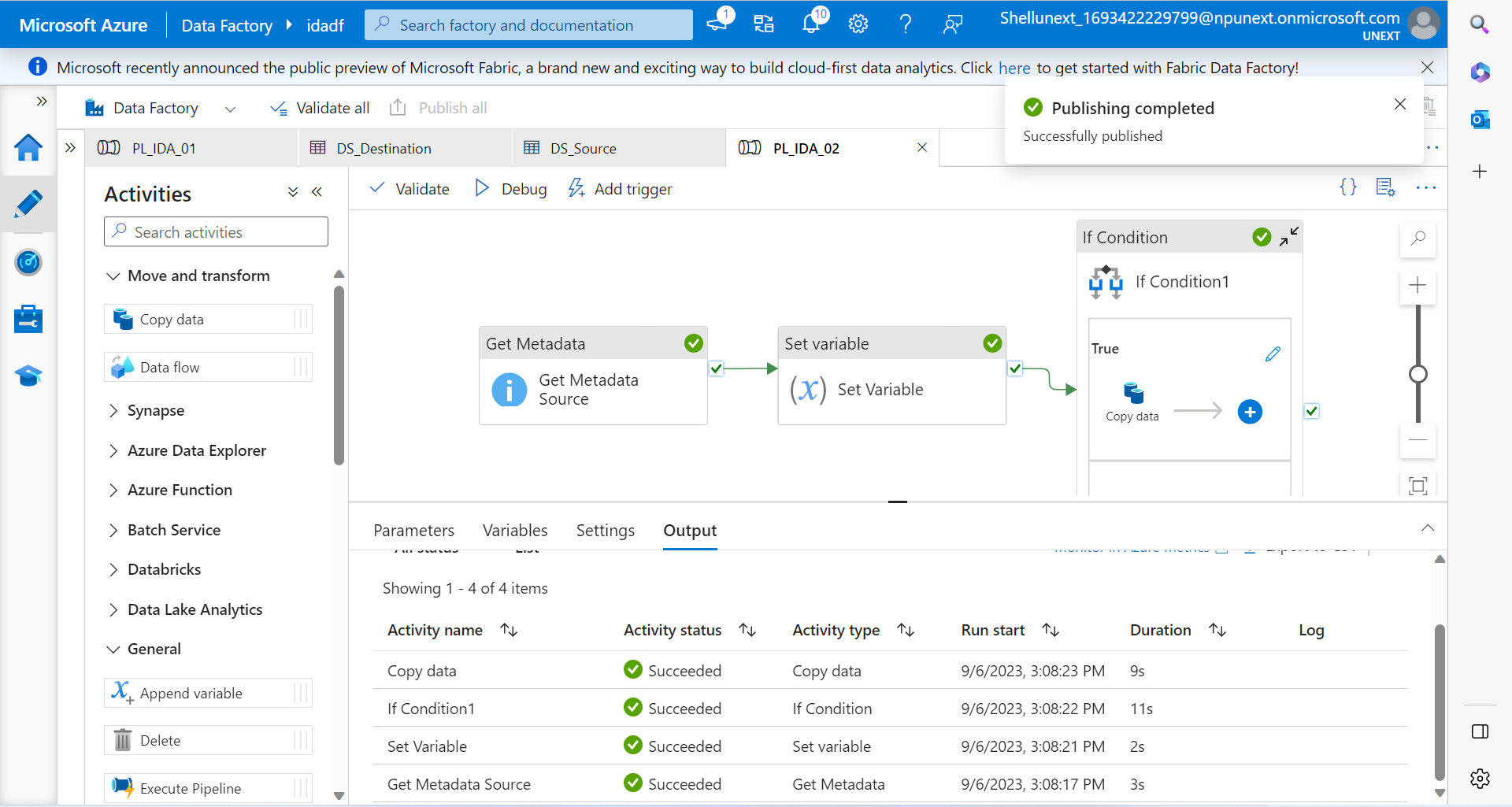
Trigger: To schedule a pipeline based on some event we use triggers.

There are three types of triggers:

1. Schedule Trigger: We schedule the trigger to run the pipelines on wall-clock schedule. This trigger supports periodic and advanced calendar options. For example: the trigger supports intervals like “weekly” or “Monday” at 5 pm and Thursday at 9 pm.
2. Tumbling Trigger: This type of trigger fires at a periodic time interval from a specified start time, while retaining state. It has one-to-one relationship with a pipeline and can only refer a single pipeline. For example: It can be used for time-based partitioning of data where we can split the data into discrete time intervals.
3. Storage Events Trigger: This trigger is fired when a particular event occurs on the storage. It responds to events in Azure Blob Storage or Data Lake Storage, such as file creation, modification, or deletion. For example: We can create trigger for real-time data processing which will execute the data pipeline on any changes in data stores.

Monitor: To monitor the information about the trigger runs, pipeline runs, etc. We can filter the monitored logs based on success or failure and various other filters.

If Condition Task:



There will be more than one file in real world so we will not create if conditions for each and every file. This is a drawback of the above approach.

ForEach Activity:

We can process and check each and every file that was the disadvantage above in if-else condition.