Custom Training

Day 8

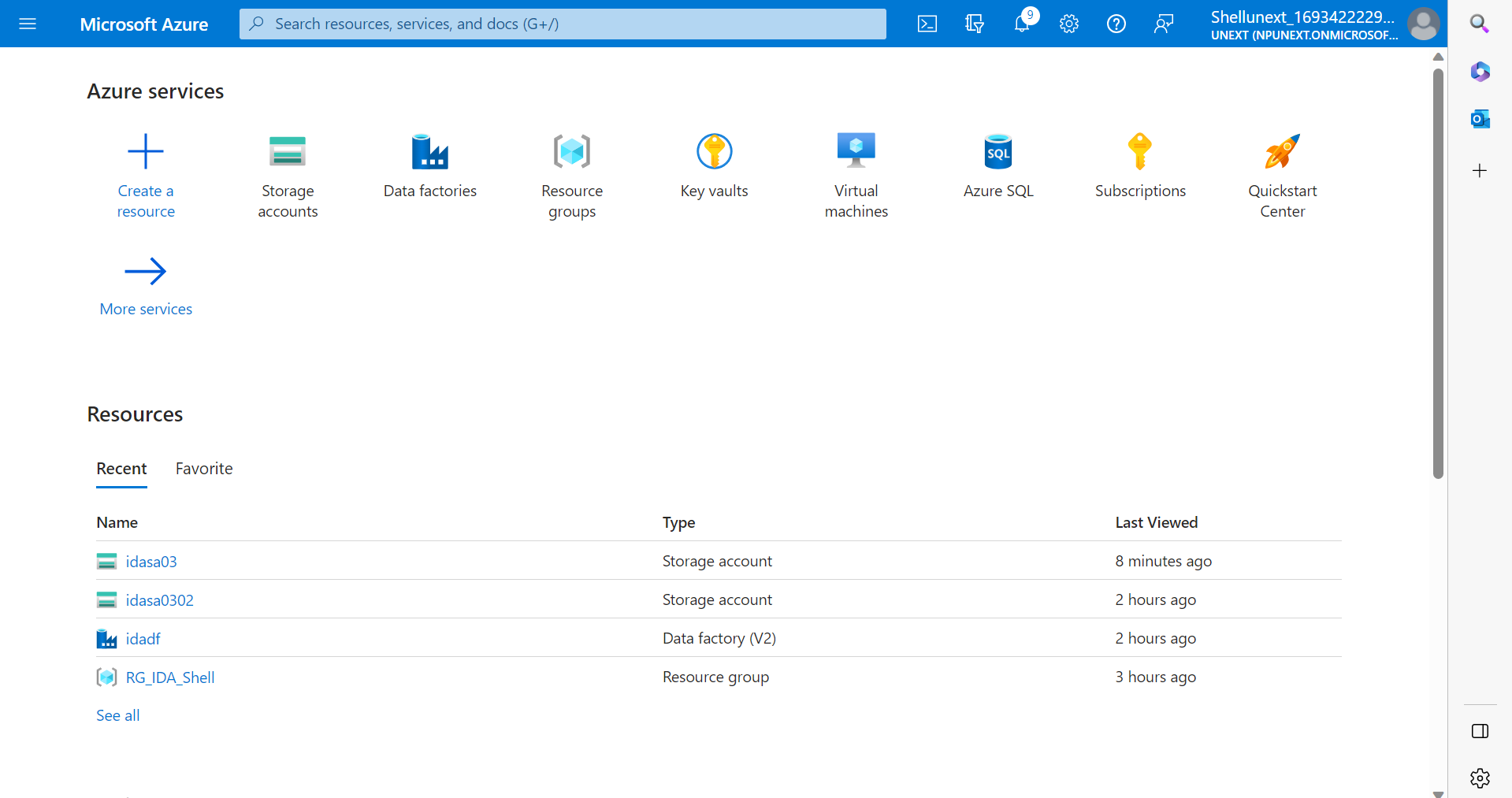
**Azure Data Factory**

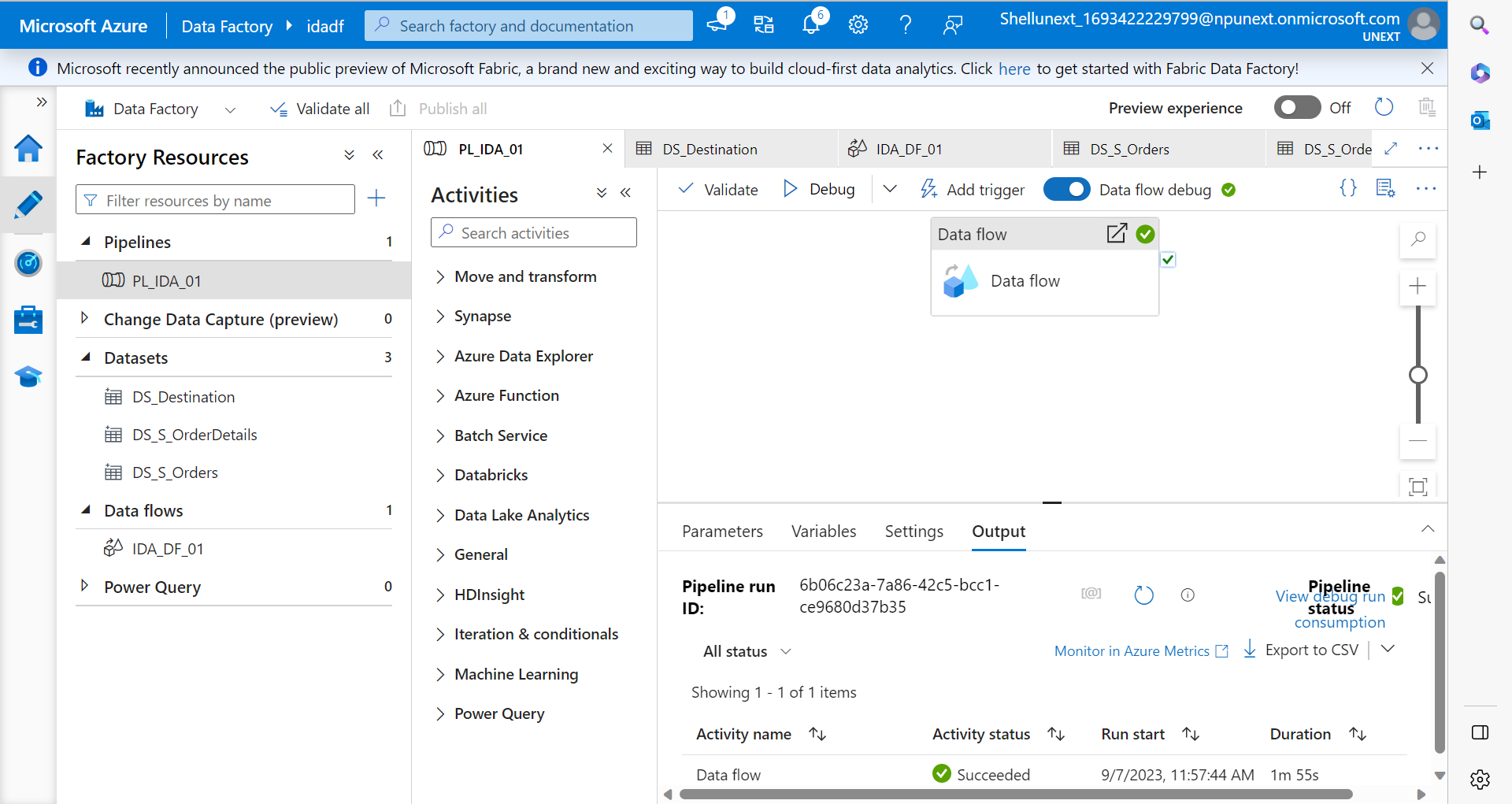
Data Flow: Used for data transformation.

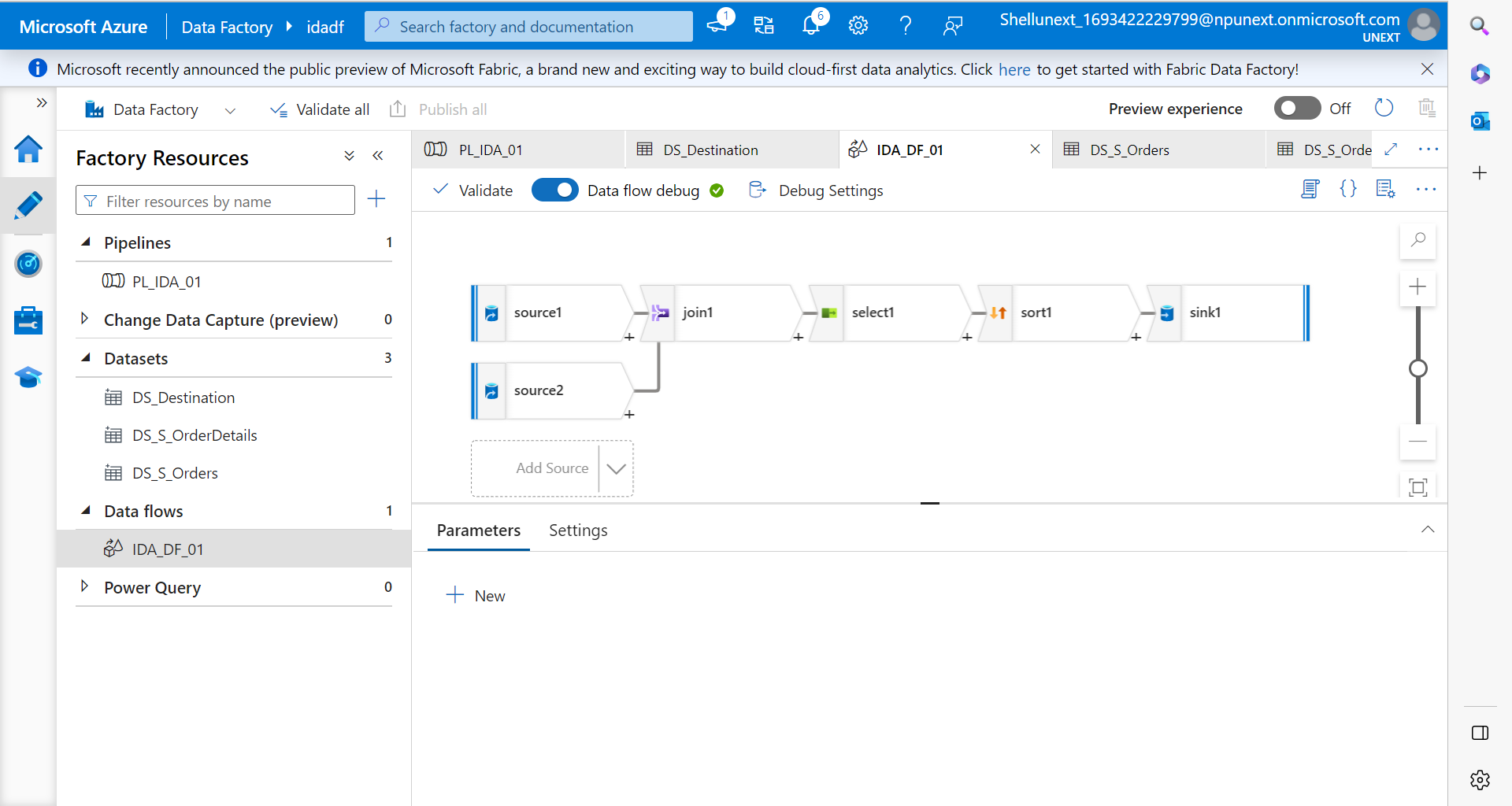
To do the transformation in Azure Data Factory it needs the cluster in the backend for spark.

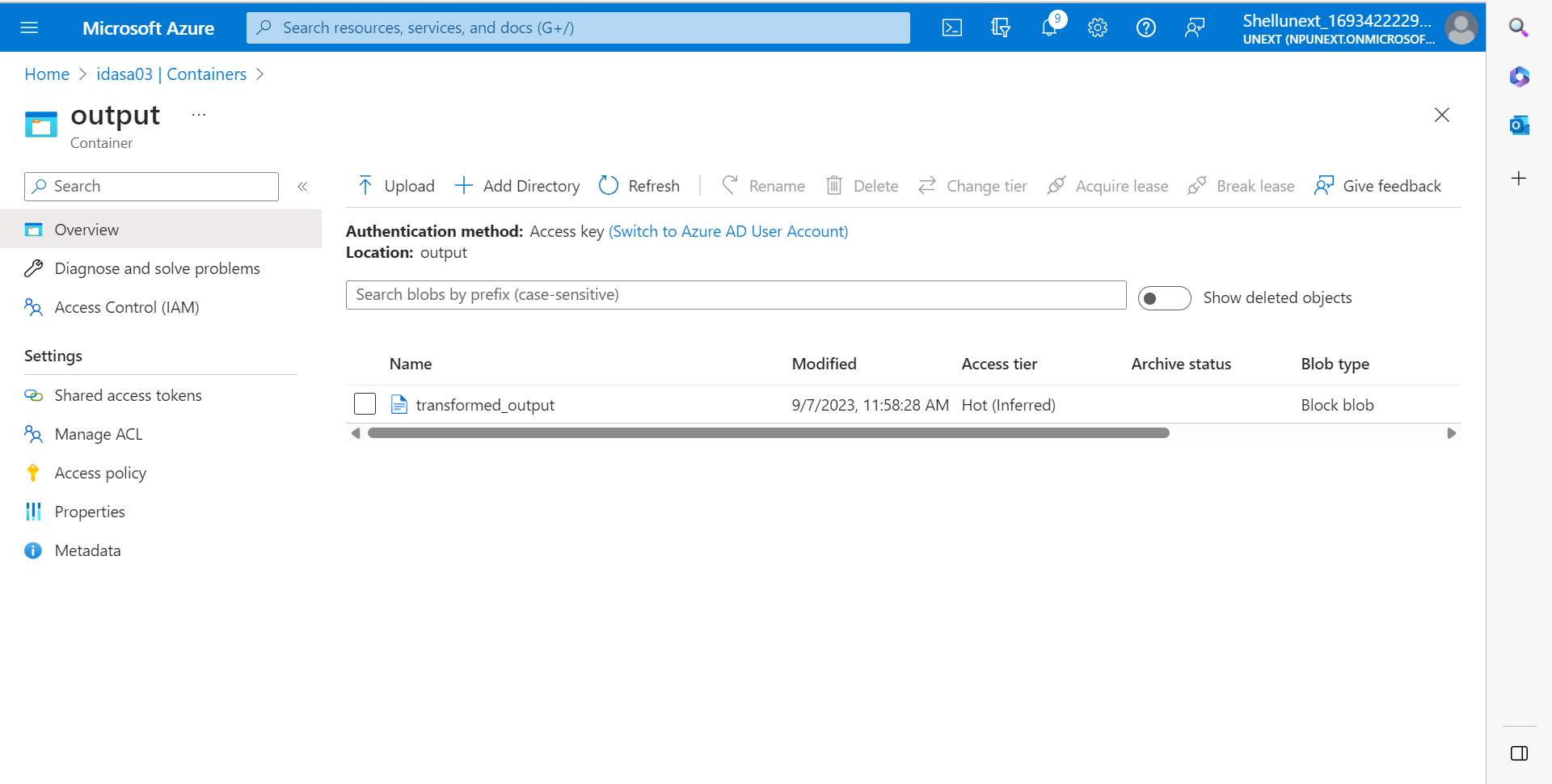
For performing some basic transformations on the data we can use ADF but for complex and advanced transformations we use Databricks.

Creation of dataflow for taking the input data from Azure blob storage(idasa0302) and storing it in ADLS(idasa03) after performing the transformations:

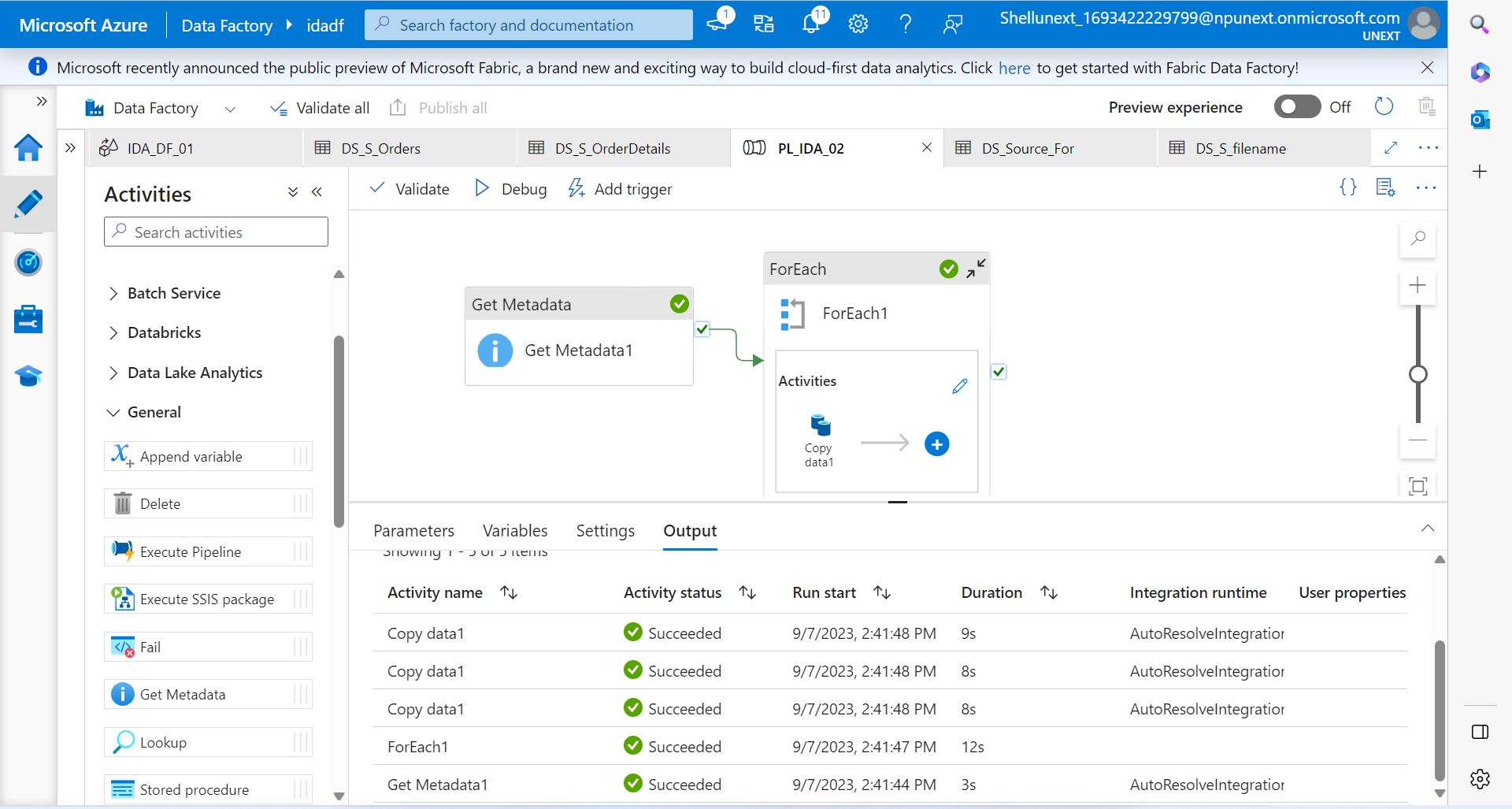


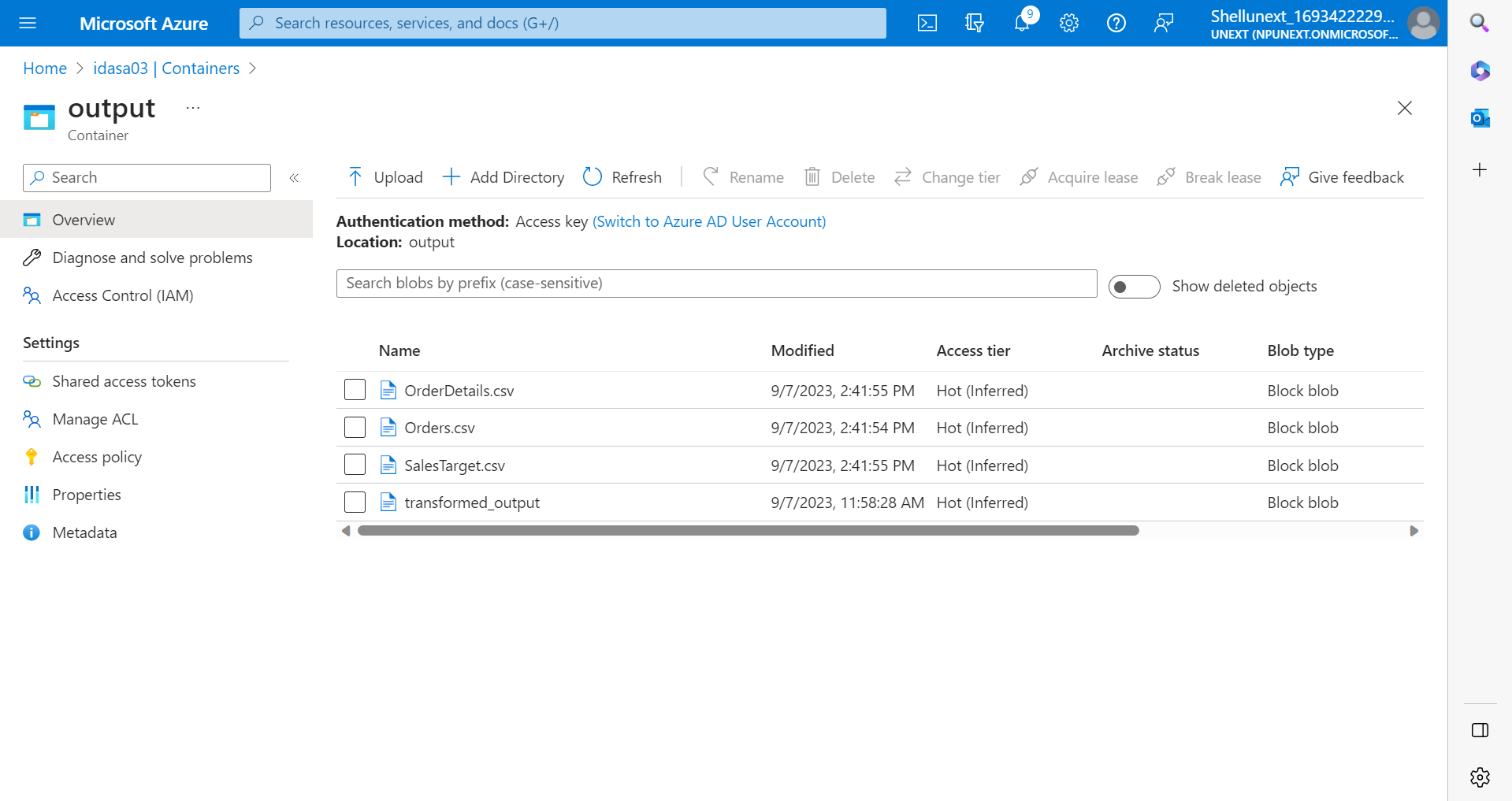




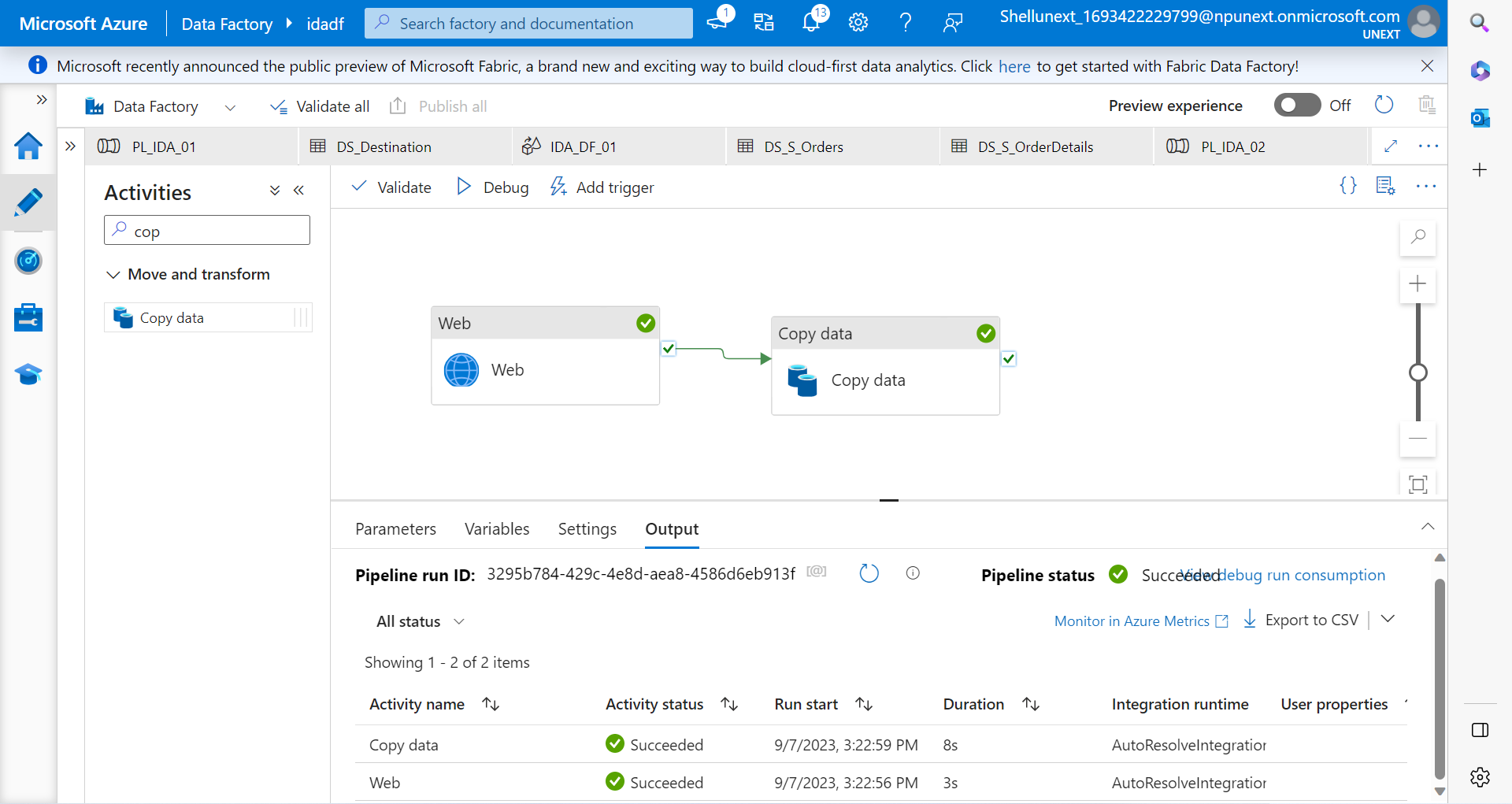


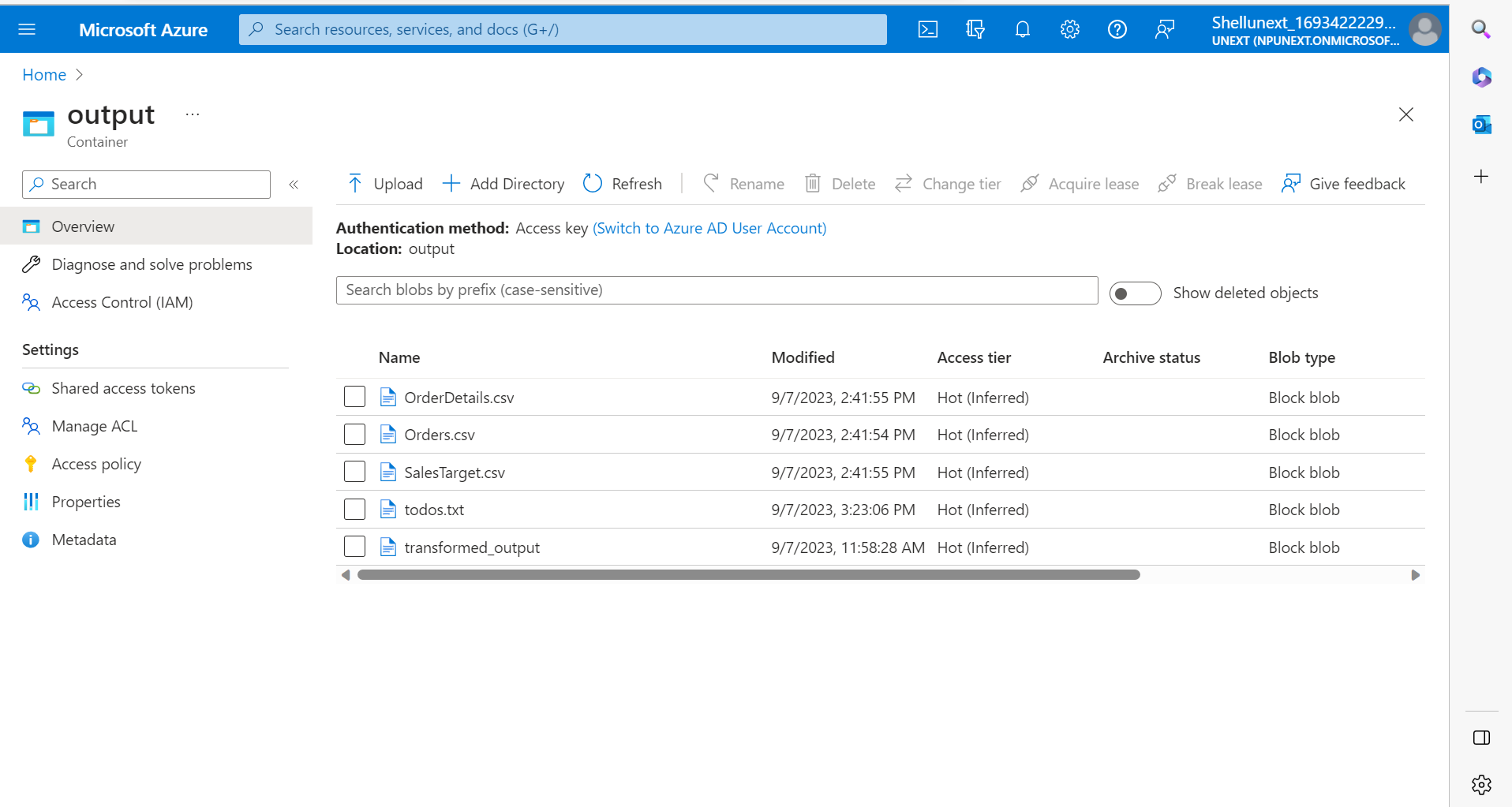
Task: To get the names of the files inside sales container and perform for each activity to copy data of each file into ADLS container.



****

API Calls:





CDC (Change Data Capture):

We can use timestamp to note the times whenever the new records gets inserted/deleted/updated in the table. So now for only the changed data we can perform the pipeline activity.