DevOps Coding Challenge

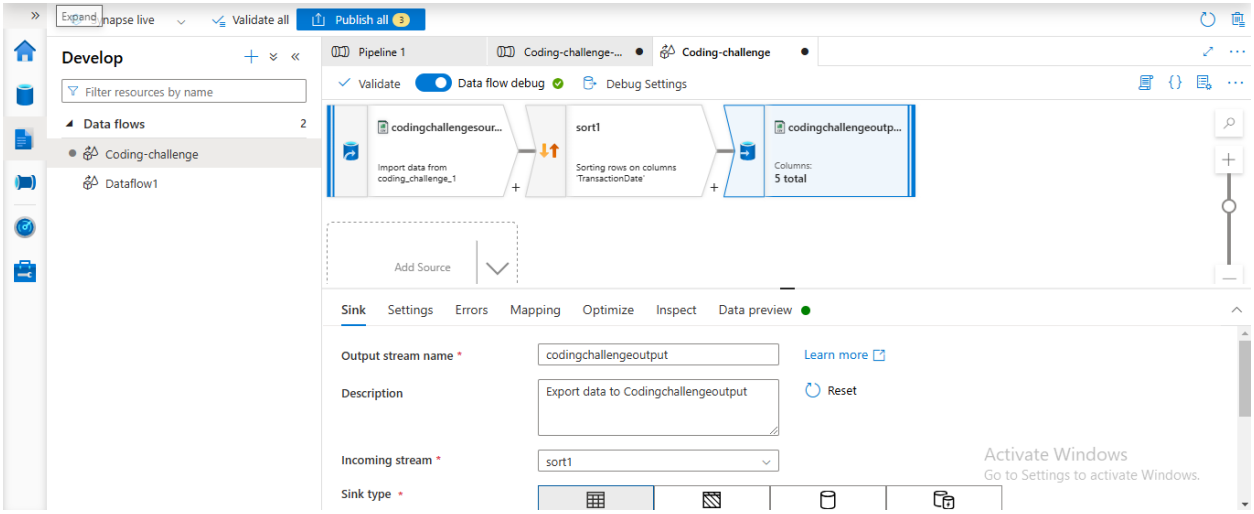
-Sarthak Niranjan Kulkarni (Maverick)

- sarthakkul2311@gmail.com - (+91) 93256 02791

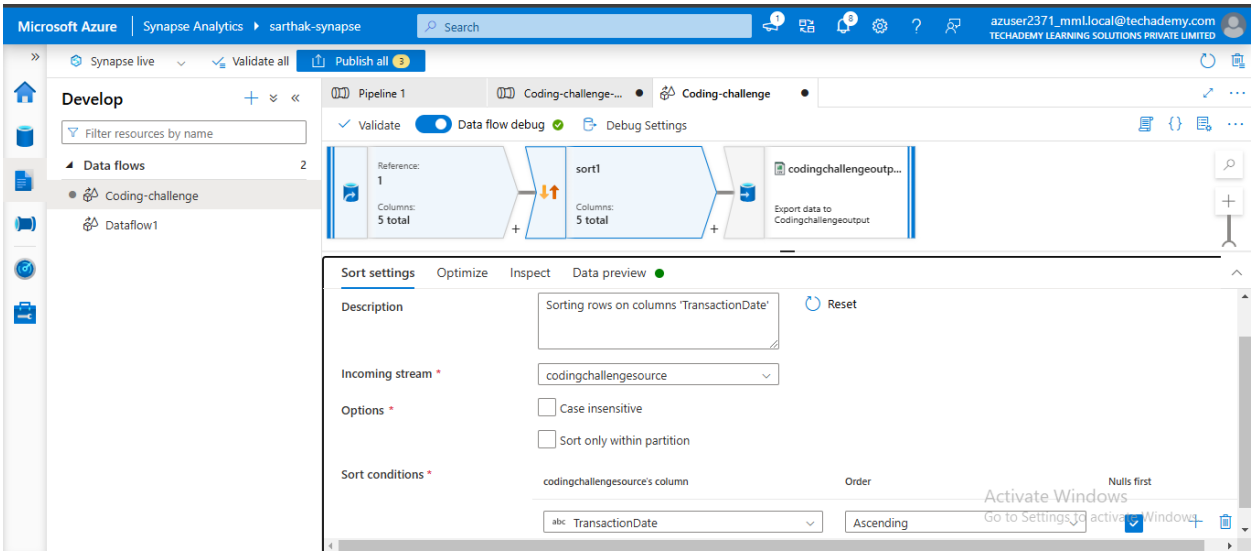
19/12/2024 (Thursday)

Task 1 in Pipeline

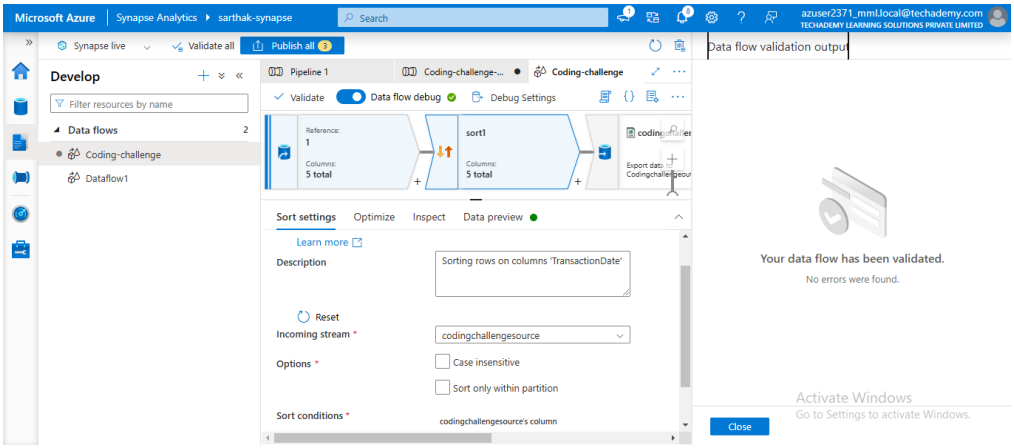
1. Setup Dataflow



2. Configure Sort Settings

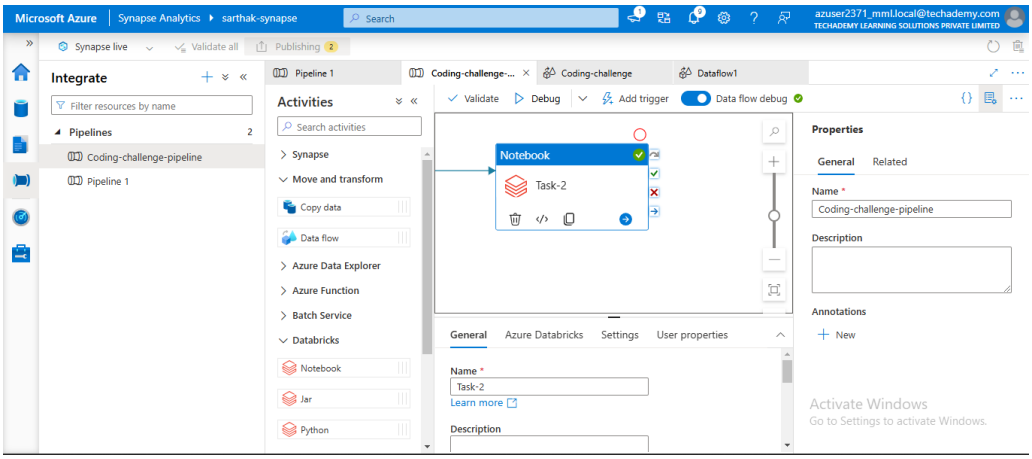


3. Validate the whole dataflow

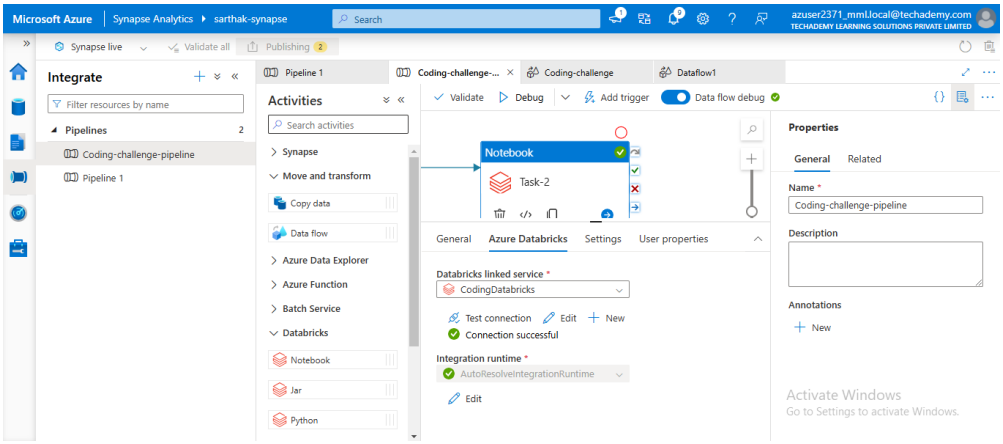


Task 2 in Pipeline

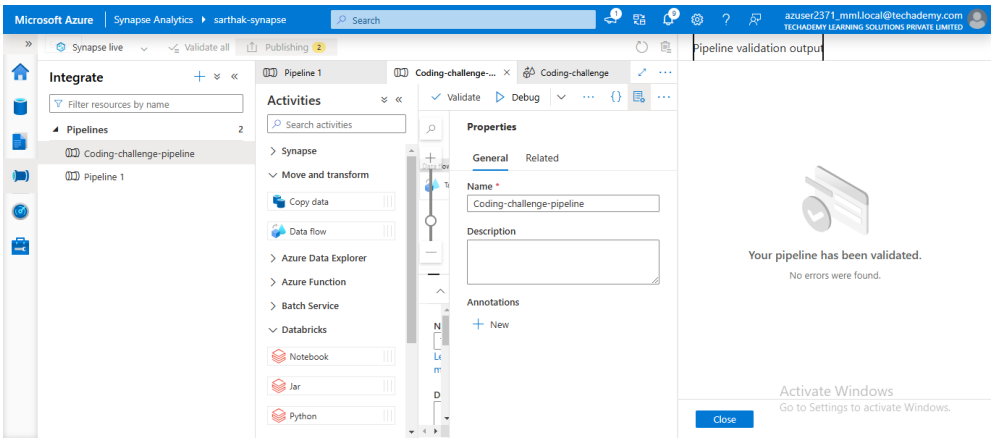
1. Insert Databricks Notebook



2. Connect the Azure Databricks notebook

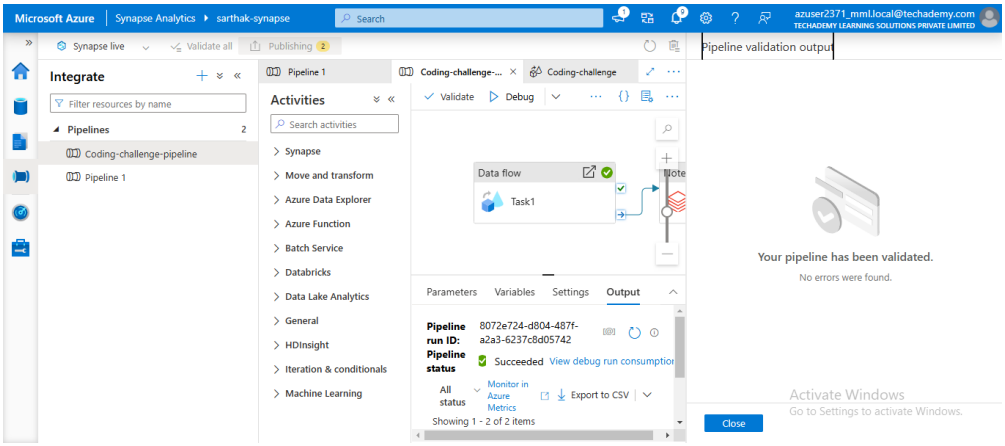


3. Then validate the notebook

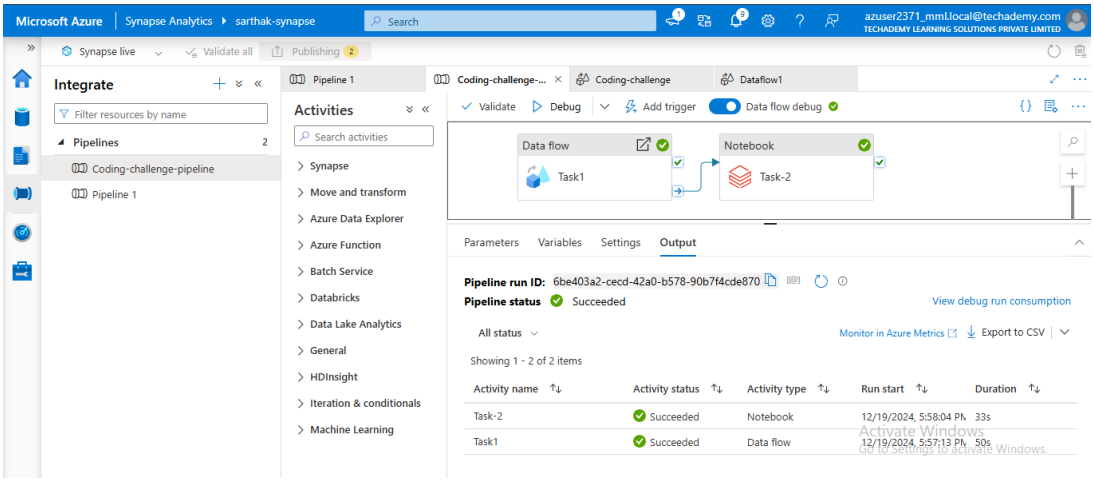


Deploying Pipeline:

1. Validate the whole Pipeline



2. Debug the pipeline



Summary:

Azure Dataflow:

Azure Synapse Data Flow is a cloud-based data transformation service that enables users to visually design and execute data transformations at scale. It allows data engineers to orchestrate complex ETL processes using a drag-and-drop interface, integrating seamlessly with Azure Synapse Analytics to process and transform data across various sources.

Azure Synapse Pipeline:

An Azure Synapse Pipeline is a data orchestration tool that enables the automation of data workflows, allowing users to define, schedule, and manage data movement and transformation tasks. It integrates various data sources, processing services, and destinations, enabling seamless end-to-end data workflows within the Azure Synapse environment.

Azure Databricks Notebook in Synapse:

Azure Databricks Notebooks can be integrated with Azure Synapse Analytics to provide a collaborative environment for advanced analytics and data engineering workflows. This integration allows users to run Spark-based data transformations and machine learning models directly within Synapse pipelines, enabling seamless end-to-end data processing and analytics.