

DevOps Coding Challenge

-Sarthak Niranjan Kulkarni (Maverick)

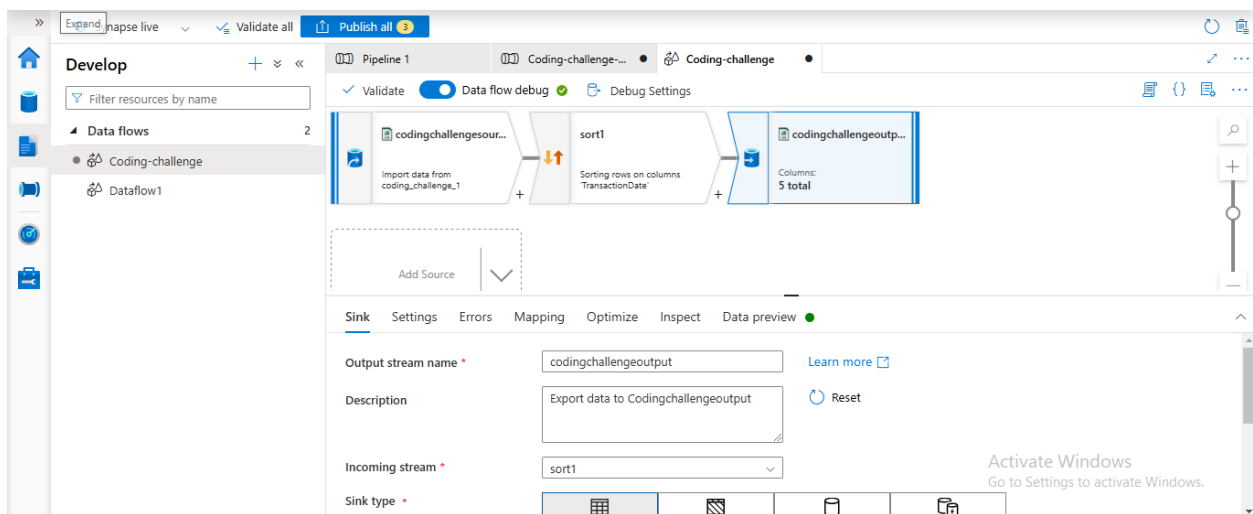
- sarthakkul2311@gmail.com

- (+91) 93256 02791

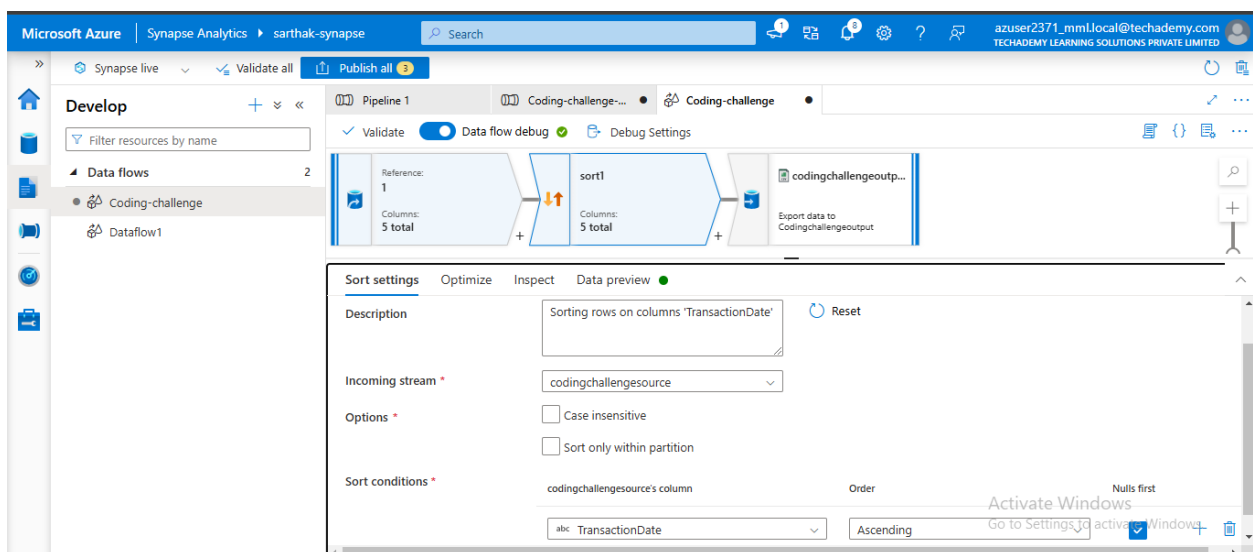
19/12/2024 (Thursday)

Task 1 in Pipeline

1. Setup Dataflow



2. Configure Sort Settings



3. Validate the whole dataflow

The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar is set to 'Develop' and shows a list of data flows, including 'Coding-challenge' and 'Dataflow1'. The main area displays a data flow diagram with a 'sort1' task. Below the diagram, the 'Sort settings' panel is visible, showing the description 'Sorting rows on columns TransactionDate' and the incoming stream 'codingchallengesource'. The 'Validate' button is highlighted. On the right, a 'Data flow validation output' window is open, displaying a green checkmark and the message 'Your data flow has been validated. No errors were found.' Below this, there is an 'Activate Windows' watermark.

Task 2 in Pipeline

1. Insert Databricks Notebook

The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar is set to 'Integrate' and shows a list of pipelines, including 'Coding-challenge-pipeline' and 'Pipeline 1'. The main area displays a pipeline diagram with a 'Notebook' task. Below the diagram, the 'Properties' panel is visible, showing the 'Name' field set to 'Task-2' and the 'Description' field. The 'Activities' panel on the left shows a list of activities, including 'Notebook', 'Jar', and 'Python'. The 'Notebook' activity is selected, and its properties are displayed in the 'Properties' panel.

2. Connect the Azure Databricks notebook

The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar is set to 'Integrate' and shows a list of pipelines, including 'Coding-challenge-pipeline' and 'Pipeline 1'. The main area displays a pipeline diagram with a 'Notebook' task. Below the diagram, the 'Properties' panel is visible, showing the 'Name' field set to 'Coding-challenge-pipeline' and the 'Description' field. The 'Activities' panel on the left shows a list of activities, including 'Notebook', 'Jar', and 'Python'. The 'Notebook' activity is selected, and its properties are displayed in the 'Properties' panel. The 'Databricks linked service' is set to 'CodingDataBrics', and the 'Integration runtime' is set to 'AutoResolveIntegrationRuntime'.

3. Then validate the notebook

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, the 'Integrate' tab is active, displaying a list of pipelines. The 'Activities' pane on the right shows a search bar and a list of activities including Synapse, Move and transform, Copy data, Data flow, Azure Data Explorer, Azure Function, Batch Service, Databricks, Notebook, Jar, and Python. The 'Properties' pane on the right shows the 'General' tab for a pipeline named 'Coding-challenge-pipeline'. The 'Pipeline validation output' pane on the far right displays a green checkmark and the message: 'Your pipeline has been validated. No errors were found.' Below this message is an 'Activate Windows' watermark and a 'Close' button.

Deploying Pipeline:

1. Validate the whole Pipeline

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, the 'Integrate' tab is active, displaying a list of pipelines. The 'Activities' pane on the right shows a search bar and a list of activities including Synapse, Move and transform, Copy data, Data flow, Azure Data Explorer, Azure Function, Batch Service, Databricks, Notebook, Jar, and Python. The 'Pipeline validation output' pane on the far right displays a green checkmark and the message: 'Your pipeline has been validated. No errors were found.' Below this message is an 'Activate Windows' watermark and a 'Close' button.

2. Debug the pipeline

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, the 'Integrate' tab is active, displaying a list of pipelines. The 'Activities' pane on the right shows a search bar and a list of activities including Synapse, Move and transform, Copy data, Data flow, Azure Data Explorer, Azure Function, Batch Service, Databricks, Notebook, Jar, and Python. The 'Pipeline debug' pane on the far right displays a green checkmark and the message: 'Your pipeline has been validated. No errors were found.' Below this message is an 'Activate Windows' watermark and a 'Close' button.

Activity name	Activity status	Activity type	Run start	Duration
Task-2	Succeeded	Notebook	12/19/2024, 5:58:04 PM	33s
Task1	Succeeded	Data flow	12/19/2024, 5:57:13 PM	50s

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.
