

P4: Work with the database

Team Members		
1	Sriram Ravindran	001003684
2	Rishab Shekar Gunashekar	001040578
3	Nirmiti Mahendra Patil	002103927

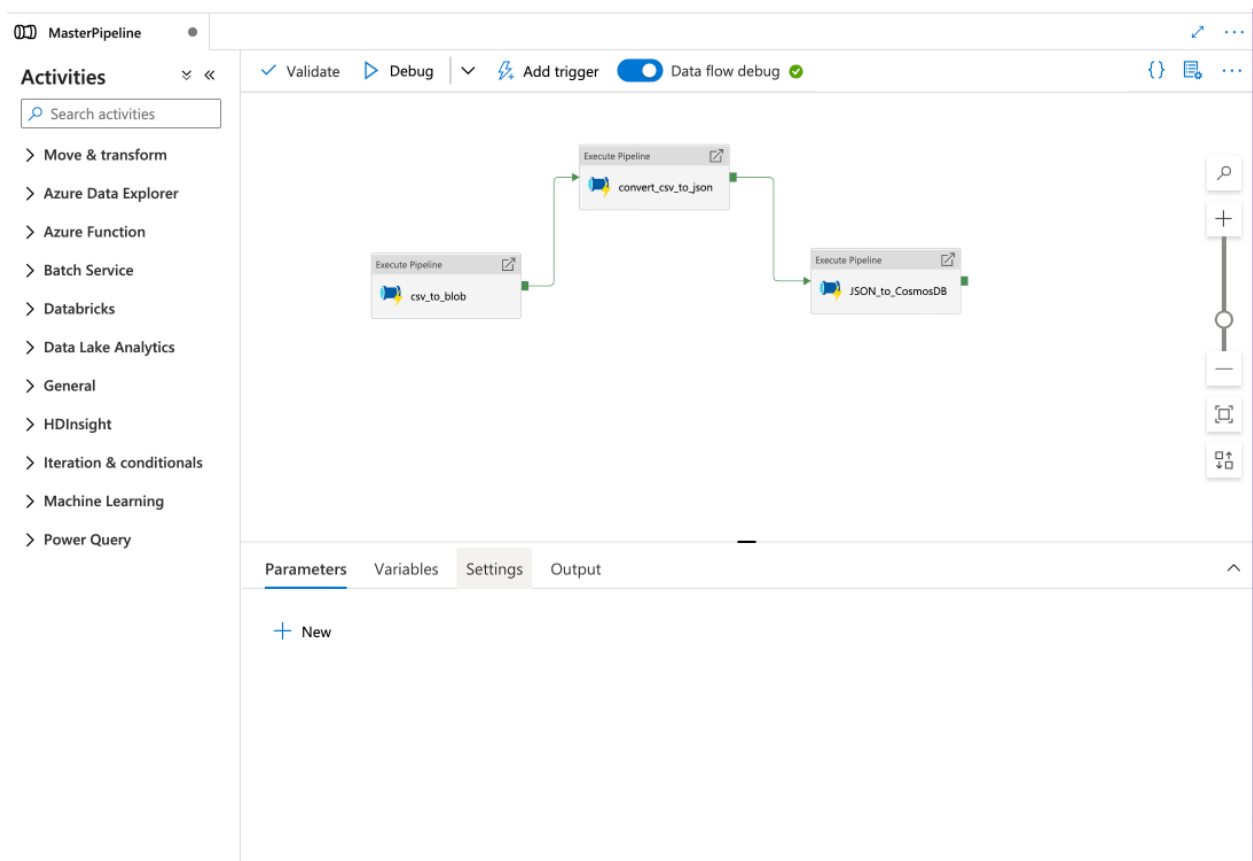
Creating Master Pipeline:

The Master pipeline contains 3 individual pipelines for the below tasks:

Pipeline-1: To do API fetch using HTTP service and staging the data in csv format in Azure Blob Storage.

Pipeline-2: Converting CSV data to JSON data by creating a dataflow.

Pipeline-3: JSON Data Ingestion in Cosmos DB



Pipeline Automation using Azure Data Factory Trigger:

Workflow: Creating a trigger to automate and End-to-End ETL Pipeline every day at 7.31 PM (EST). The API refresh happens every day at 7.30 PM EST. Hence the pipeline start time is kept 1 minute ahead of it.

New trigger

Schedule

Start date * ⓘ
04/11/2022 02:28:47

Time zone * ⓘ
Eastern Time (US & Canada) (UTC-5)
 ⓘ This time zone observes daylight savings. Trigger will auto-adjust for one hour difference.

Recurrence * ⓘ
Every 1 Day(s)

Advanced recurrence options
Execute at these times ⓘ
Hours 15 X
Minutes 31 X
Schedule execution times
15:31
☒ Specify an end date
End On * ⓘ
04/30/2022 15:32:47
Annotations
+ New
Start trigger ⓘ
☒ Start trigger on creation

OK Cancel

Data Refresh:

The **upsert** in the write behavior of JSON to Cosmos DB Pipeline does the update if the data already exists or inserts a new data if the data doesn't exist.

Activities

Search activities

Move & transform

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

Validate

Validate copy runtime

Debug

Add trigger

Copy data

json_to_cosmosDB

General

Source

Sink

Mapping

Settings

User properties

Sink dataset *

CosmosDbSqlApiCollection1

Open

New

Learn more

Write behavior

Upsert

Write batch timeout

Write batch size

Max concurrent connections