Day-4 (Data Factory)

Thursday, September 7, 2023   9:53 AM

HANDS ON - ACTIVITY 1 (DATAFLOW)

Create a resource group ——————————➤ Create a storage account ——————➤ Input (create container) ----------------> Movies.csv (upload to container) (link)
Output (create container)

Create azure data factory

Manage
(Link Service) ⟷ Create Dataset I/P
DSInput
(input/movies.csv)

Create Dataset O/P
DSOutput
(output/movies.csv)
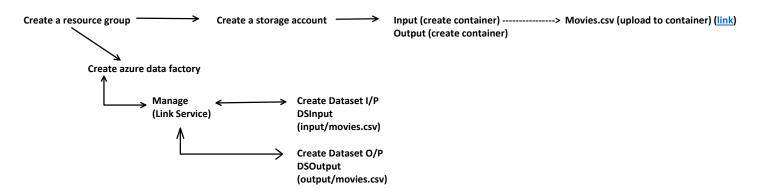
- Create a Pipeline
- Add Dataflow activity (name: Act_DataFlow)
- Go to settings
    ○ Open data flow - to perform transformation
    ○ Select data flow debug - creates a cluster for data processing
    ○

Add source	- load the dataset : DSInput - click "check connection" for checking retrieval status - Name the source (as moviesdataset)
After adding source, create a filter transformation	Filter on -> write expression toInteger(Year) >= 2000 && toInteger(Year) <= 2010
Add sink	- Load the dataset : DSOutput

- Validate all, Publish all
- Select created pipeline and run

Dataflow can take the data and perform some operation and then the result data can be stored in another store (more customizable)
Copy activity just moves the data from one container to another without any transformation

ACTIVITY - 2 (INTEGRATE DIFF SOURCE TO ADF)

Azure Map
- Authentication code
- Endpoint url to fetch data

- Create a pipeline
- Add 'web' activity
    • Go to setting of the activity
    • Add the url of the azure map
    • Select GET method