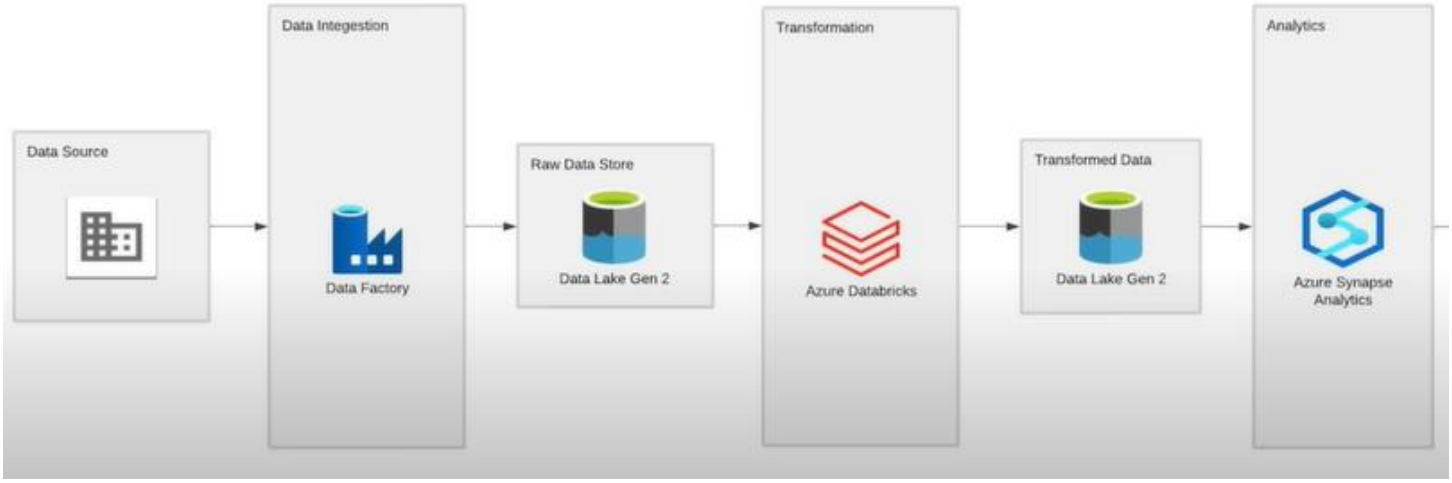
Tokyo Olympic End-to-End Data Engineering Project

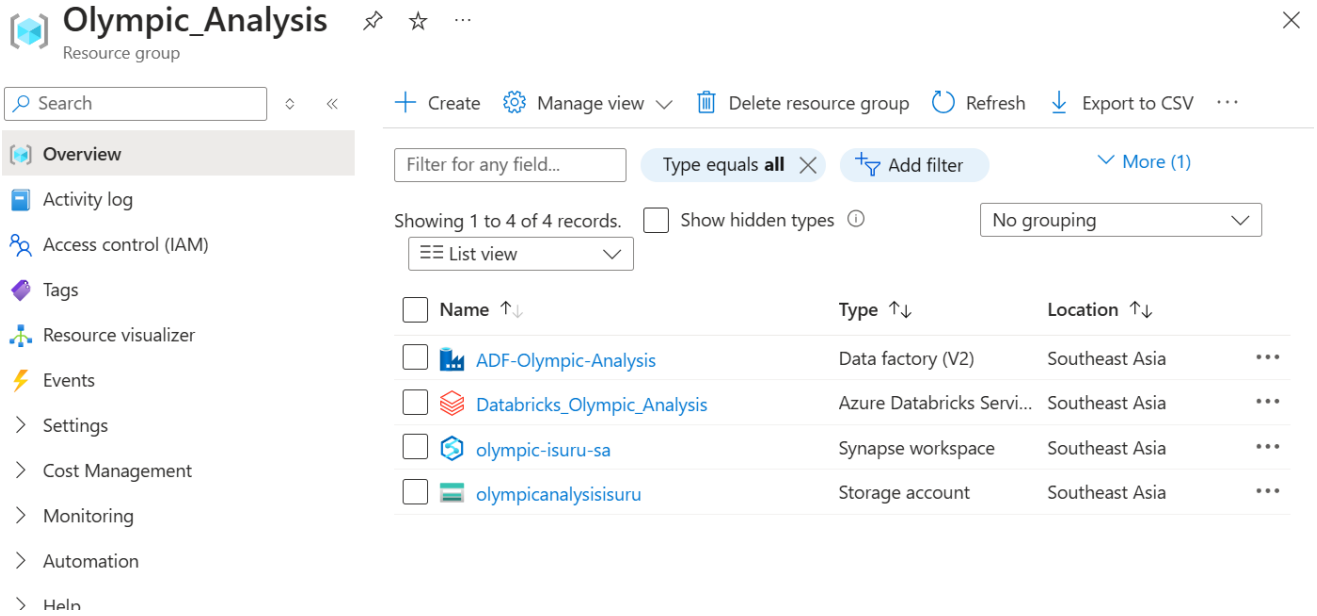


❖ Resources Used

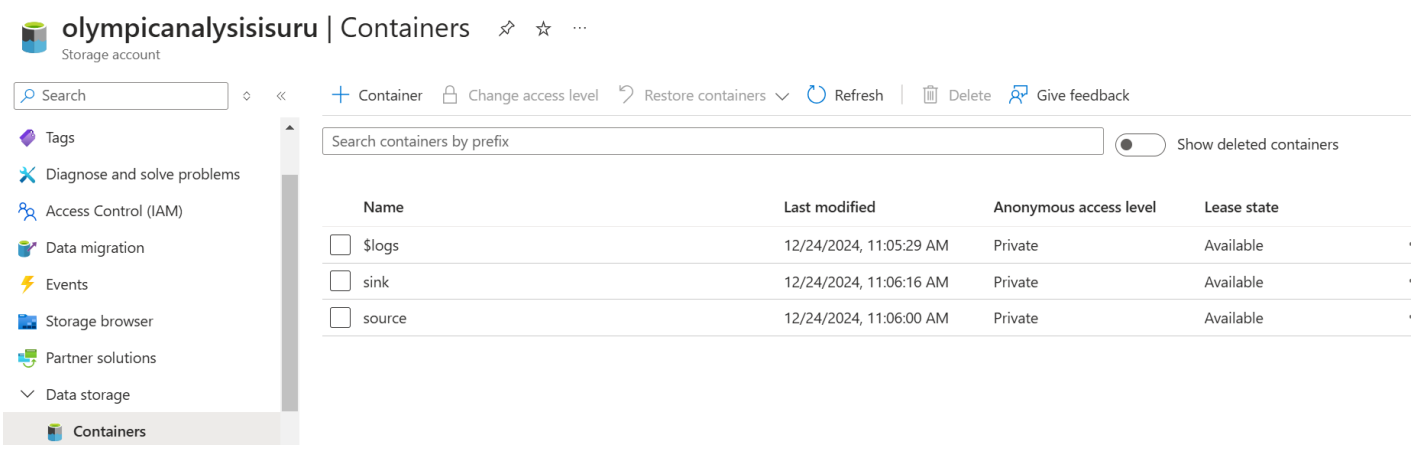
- ➢ Azure Data Lake
- ➢ Azure Data Factory
- ➢ Databricks
- ➢ Synapse Analytics

❖ Steps Followed

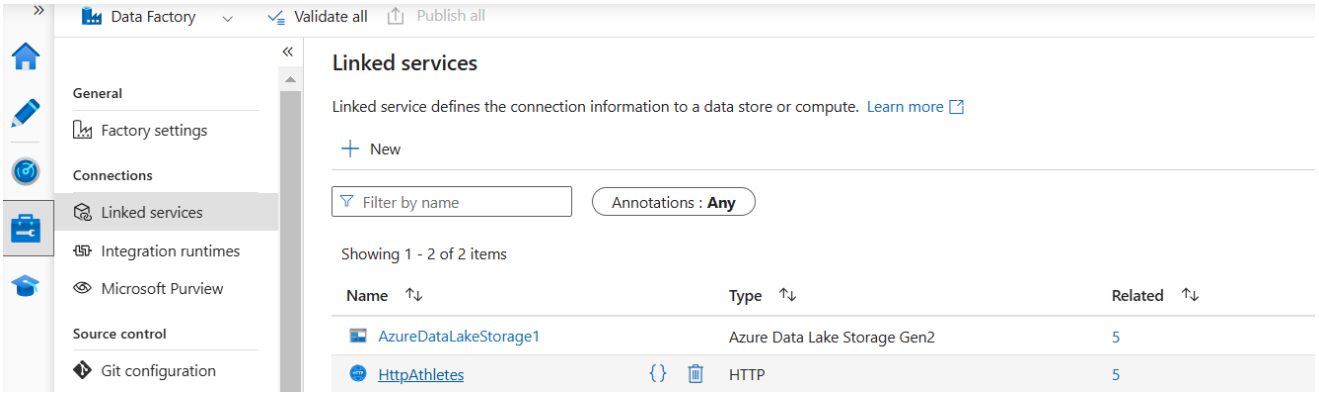
- Creating Resources Group



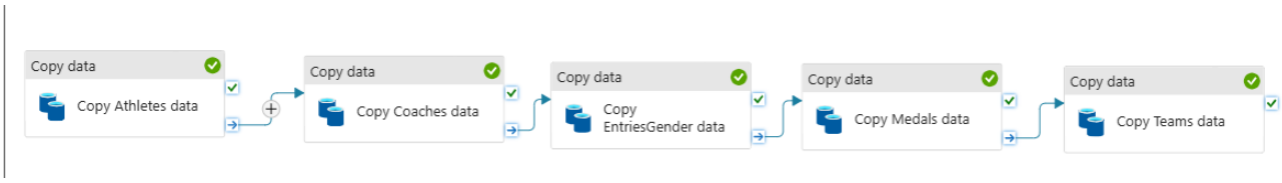
- Creating a data lake using the storage account
    - Creating two containers
        - Source
        - Sink

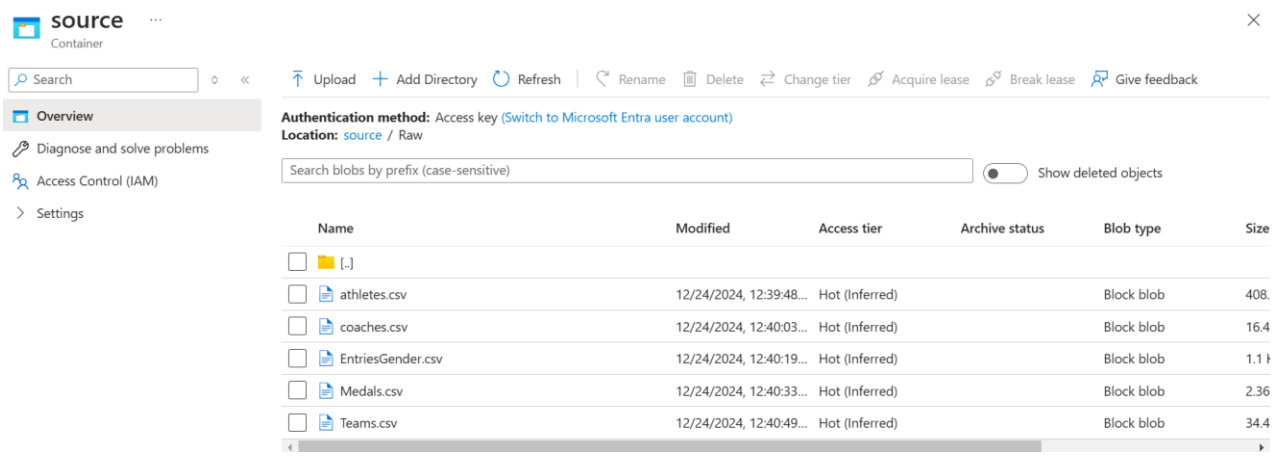


- Creating ADF
  - Creating two link services
    - Data lake
    - Git Hub

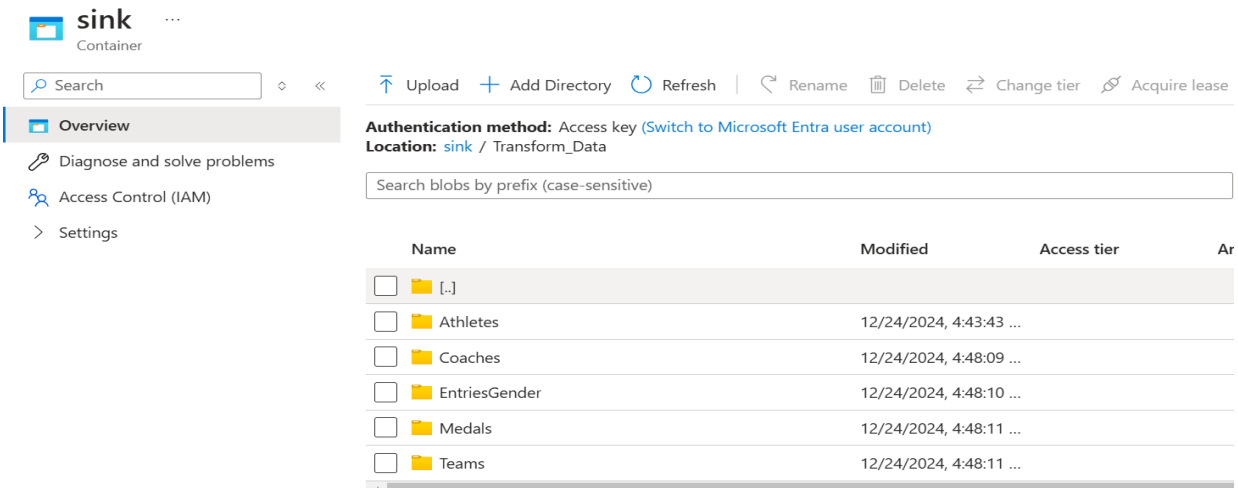


  - Creating copy activities for extracting data from the git hub repository to the data lake → Source (inside the raw folder).
  - Creating copy activity for each data set to load the data set into the data lake.





- Create Databricks workspace
    - Create cluster (Single Node)
    - Create a workspace and make a notebook for transformation
    - Create a mount to link Azure Data Lake and Databricks.
    - Do the transformation and finally load that data into the data lake.



- Create a Synapse analytics workspace
    - Create Database
    - After making tables for each data set inside the data lake sink container.
    - Create SQL Script for analyse the data

- Databricks Notebook : Olympic Transformation.ipynb
- Synapse Analytics SQL Query : Synpase Analytics SQL Script