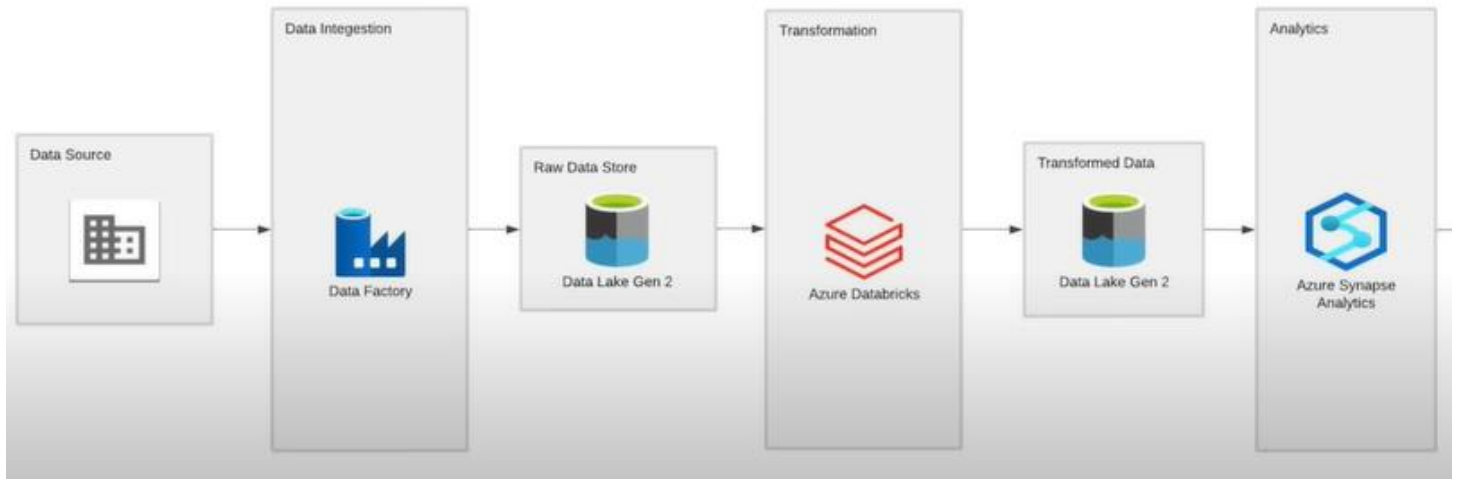


Tokyo Olympic End-to-End Data Engineering Project



❖ Resources Used

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

❖ Steps Followed

- Creating Resources Group

The screenshot shows the 'Olympic_Analysis' resource group overview in the Azure portal. The left sidebar contains navigation links: Overview (selected), Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings, Cost Management, Monitoring, Automation, and Help. The main content area displays a table of resources within the group. At the top, there are filters and controls: 'Filter for any field...', 'Type equals all', 'Add filter', and 'More (1)'. Below these, it says 'Showing 1 to 4 of 4 records.' and 'List view'. The table has columns for Name, Type, and Location. The resources listed are ADF-Olympic-Analysis (Data factory (V2)), Databricks_Olympic_Analysis (Azure Databricks Servi...), olympic-isuru-sa (Synapse workspace), and olympicanalysisisuru (Storage account). All resources are located in Southeast Asia.

Name	Type	Location
ADF-Olympic-Analysis	Data factory (V2)	Southeast Asia
Databricks_Olympic_Analysis	Azure Databricks Servi...	Southeast Asia
olympic-isuru-sa	Synapse workspace	Southeast Asia
olympicanalysisisuru	Storage account	Southeast Asia

- Creating a data lake using the storage account

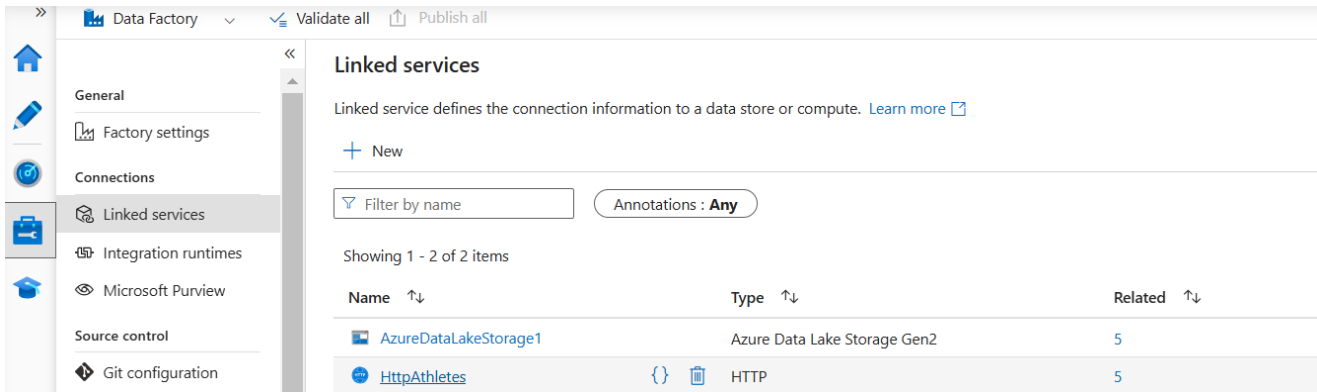
- Creating two containers

- Source
 - Sink

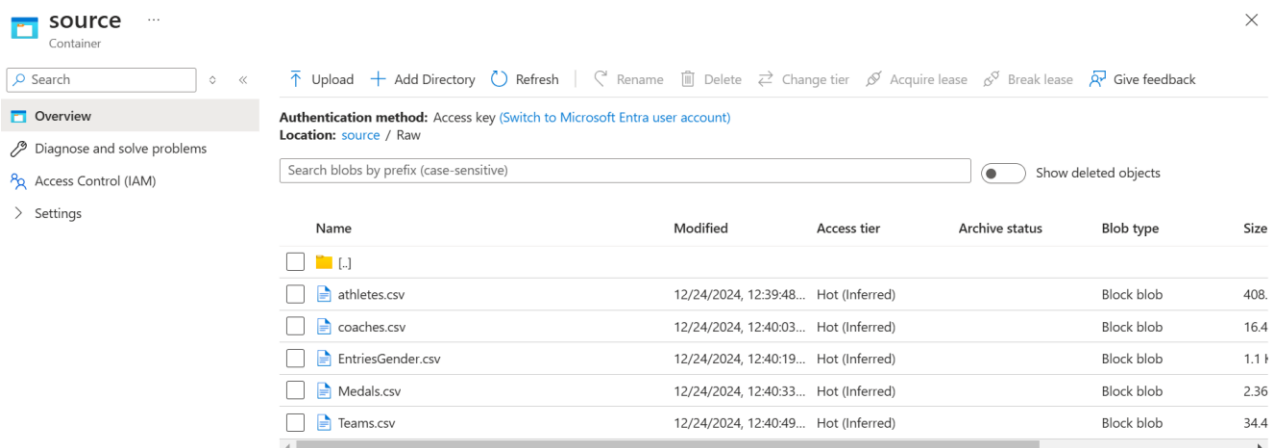
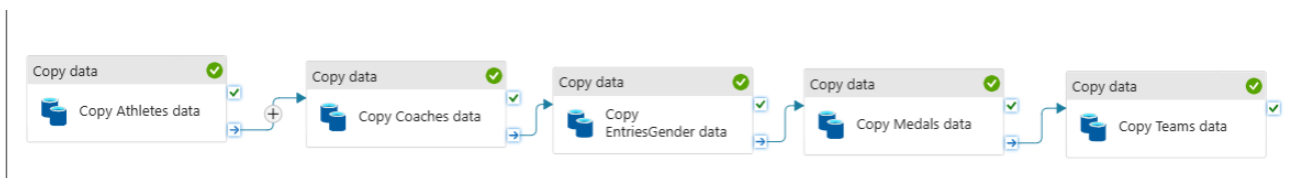
The screenshot shows the 'olympicanalysisisuru' storage account containers page in the Azure portal. The left sidebar contains navigation links: Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Partner solutions, Data storage, and Containers (selected). The main content area displays a table of containers. At the top, there are controls: '+ Container', 'Change access level', 'Restore containers', 'Refresh', 'Delete', and 'Give feedback'. Below these, there is a search bar 'Search containers by prefix' and a toggle for 'Show deleted containers'. The table has columns for Name, Last modified, Anonymous access level, and Lease state. The containers listed are \$logs, sink, and source, all created on 12/24/2024 at 11:05:29 AM, 11:06:16 AM, and 11:06:00 AM respectively, with Private access level and Available lease state.

Name	Last modified	Anonymous access level	Lease state
\$logs	12/24/2024, 11:05:29 AM	Private	Available
sink	12/24/2024, 11:06:16 AM	Private	Available
source	12/24/2024, 11:06:00 AM	Private	Available

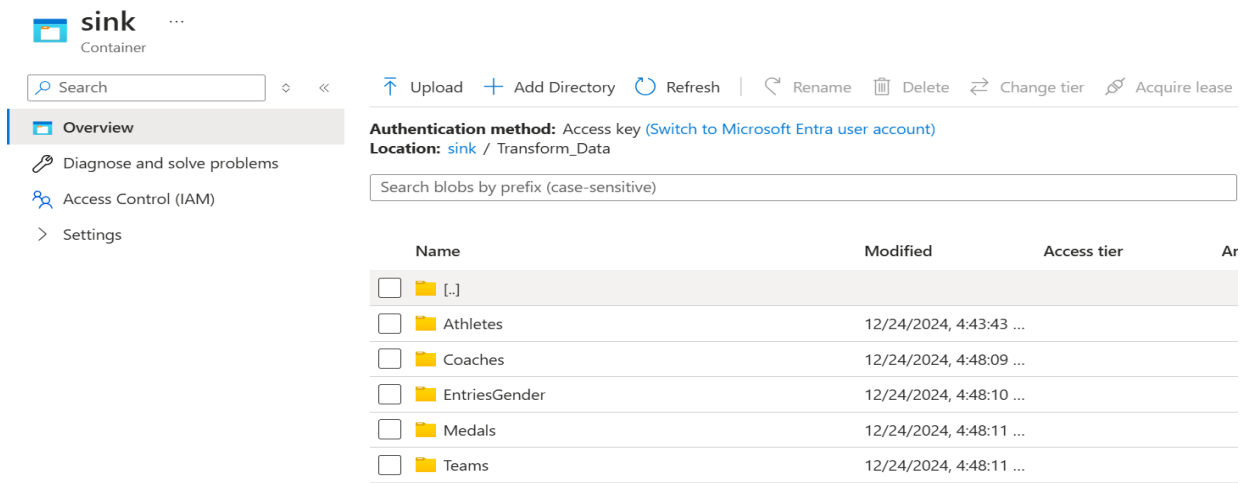
- Creating ADF
 - Creating two link services
 - Data lake
 - Git Hub



- 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 : [Synapse Analytics SQL Script](#)