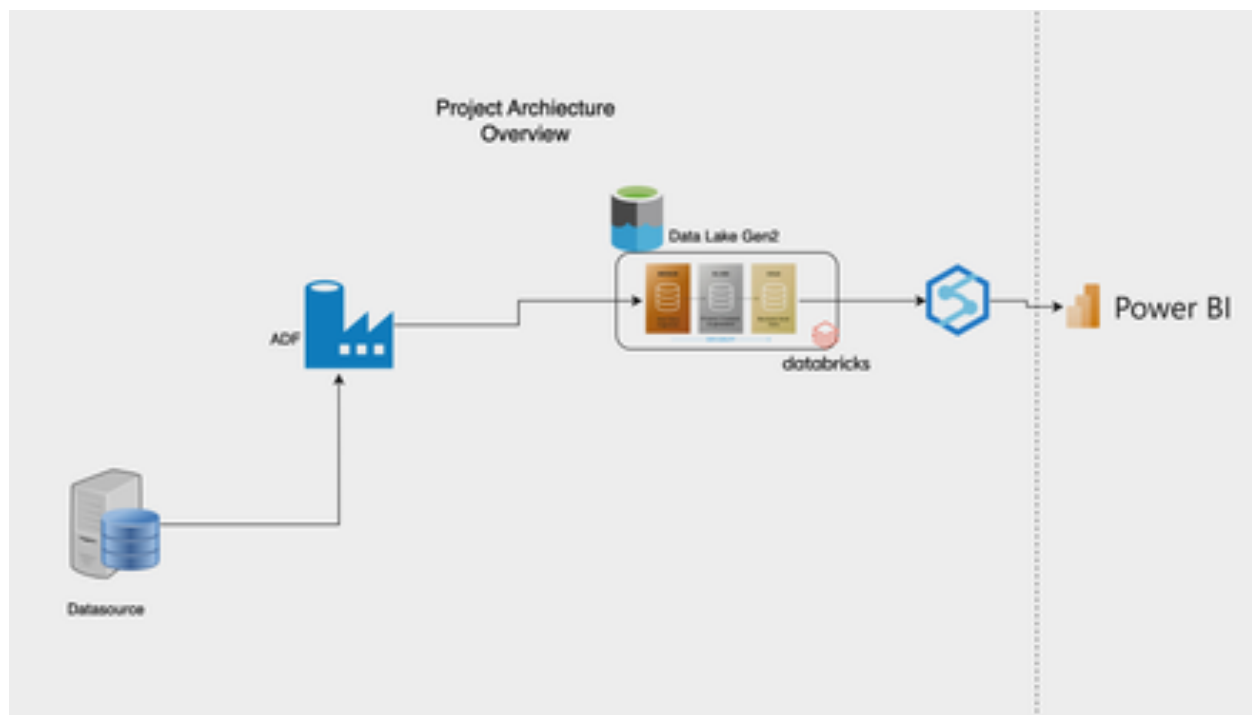


Project 9th

Tools used in this project

- Azure data factory
- Azure synapse
- Azure databricks
- Azure data lake
- Powerbi

Project Overview Archiecture



The project involves creating a comprehensive solution by ingesting tables from an on-premises SQL Server database using Azure Data Factory and storing the data in Azure Data Lake. Azure Databricks is then utilized to transform the raw data into its cleanest form. The clean data is loaded into Azure Synapse Analytics, and Microsoft Power BI is used to connect with Azure Synapse Analytics to develop an interactive dashboard.

Dataset

2021 Olympics in Tokyo

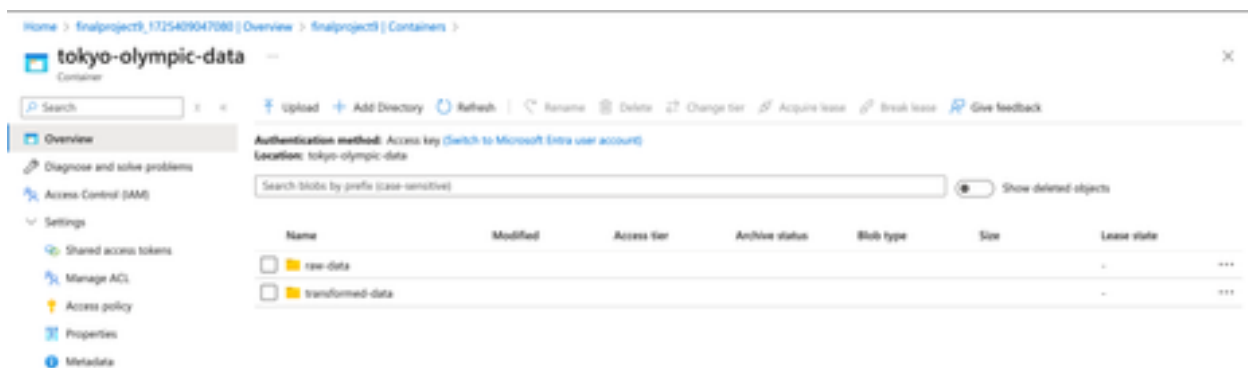
Data about Athletes, Teams, Coaches, Events



Dataset Link - [Kaggle](#)

Setting-up the Azure Data Lake Storage

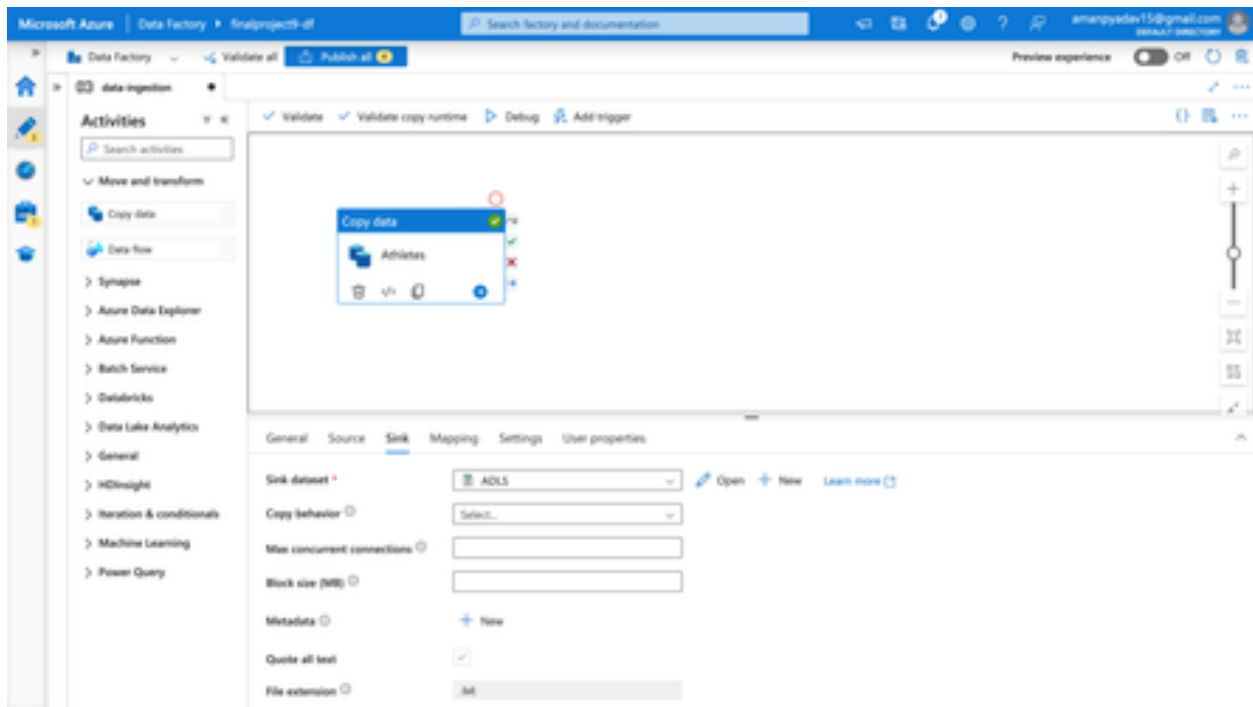
First we will create the container name **olympic-data** and inside our container we will create two directory **raw-data** and **transformed-data**. Inside raw-data directory we will upload the entire data-set as it is. Inside transformed data we will transform our dataset from raw-data directory and move in this directory.



Setting up Azure Data Factory

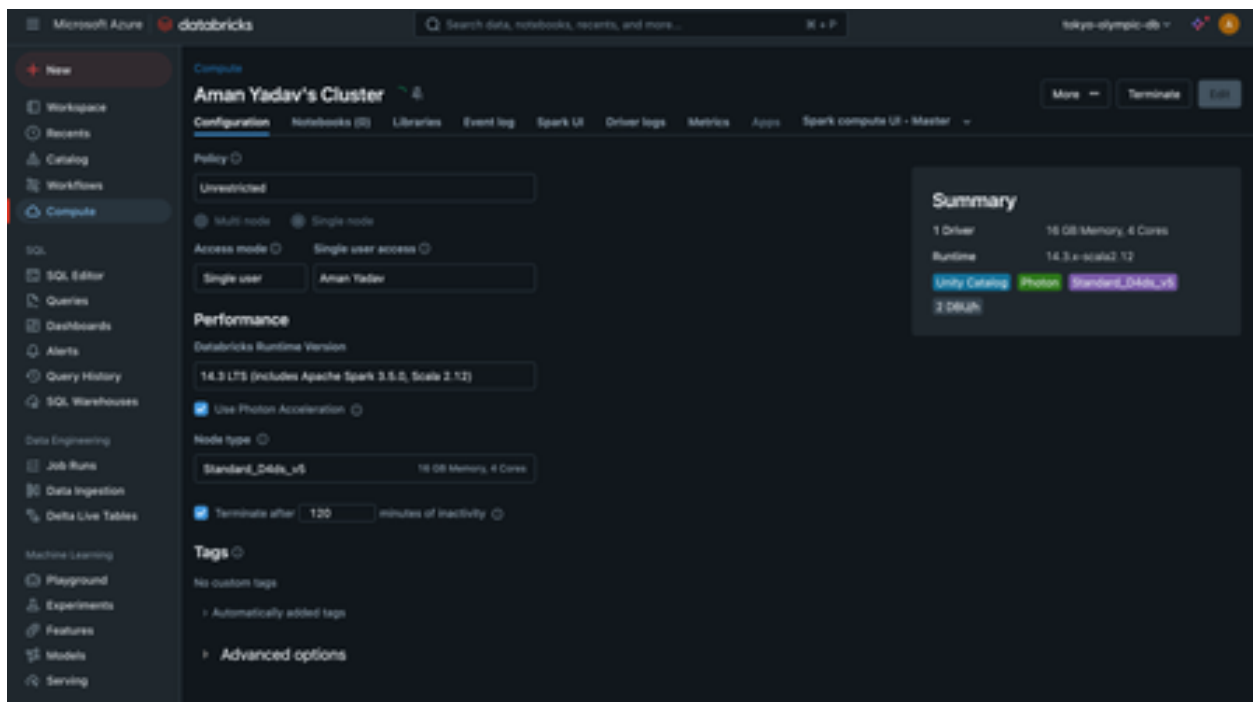
First we will create a pipeline to perform copy activity from the github to raw-data.

Creating Linked Service in which source is an HTTP file and sink is our azure blob storage.

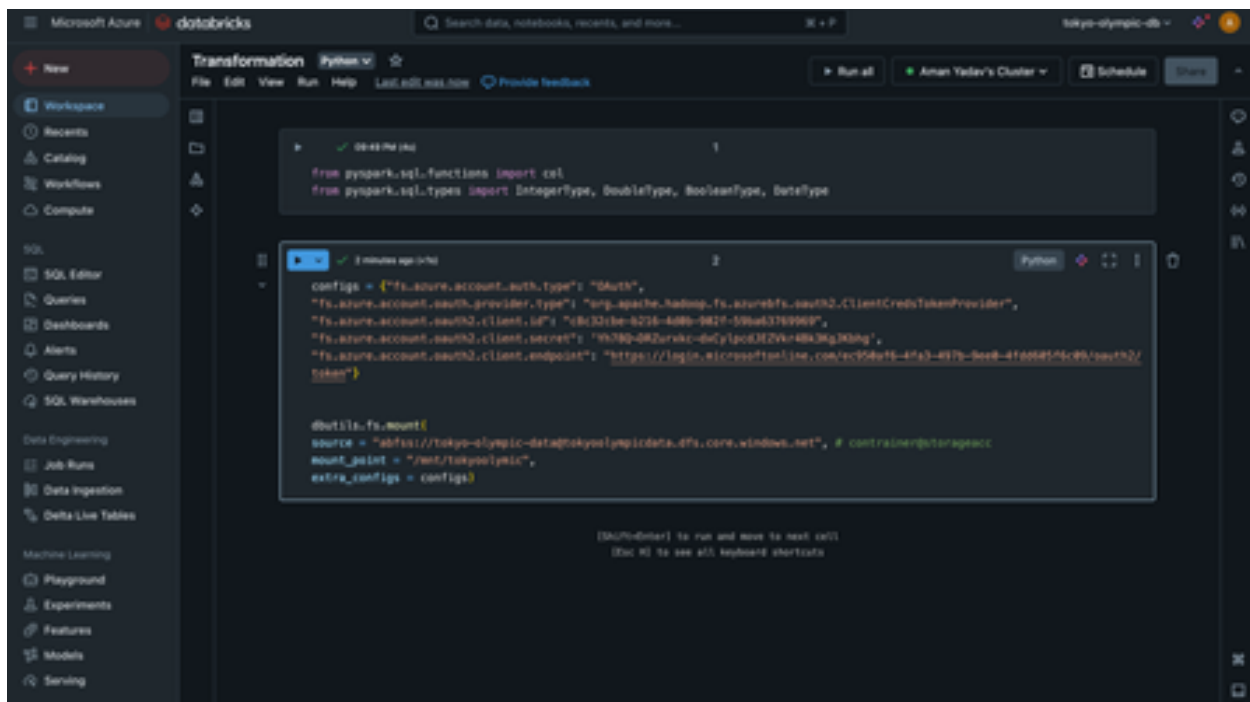


Setting Up Databricks to do transformation

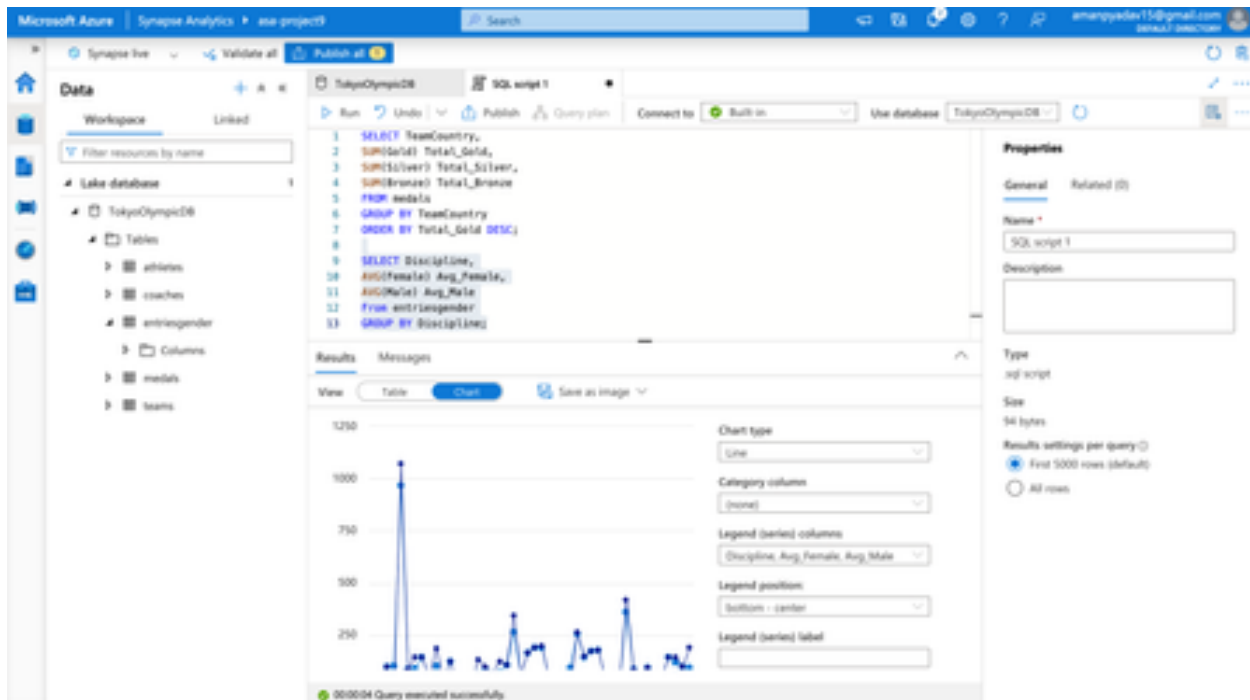
Creating cluster to run the pyspark.



Creating a notebook in databricks and creating a data connection between databricks and datalake



Azure Synapse



Azure VM

Creating and configuring up virtual machine using azure VM.

Instance details

Virtual machine name * ⓘ

myVM

Region * ⓘ

(US) West US 3

Availability options ⓘ

No infrastructure redundancy required

Security type ⓘ

Trusted launch virtual machines

[Configure security features](#)

Image * ⓘ

Windows Server 2022 Datacenter: Azure Edition - x64 Gen2

[See all images](#) | [Configure VM generation](#)

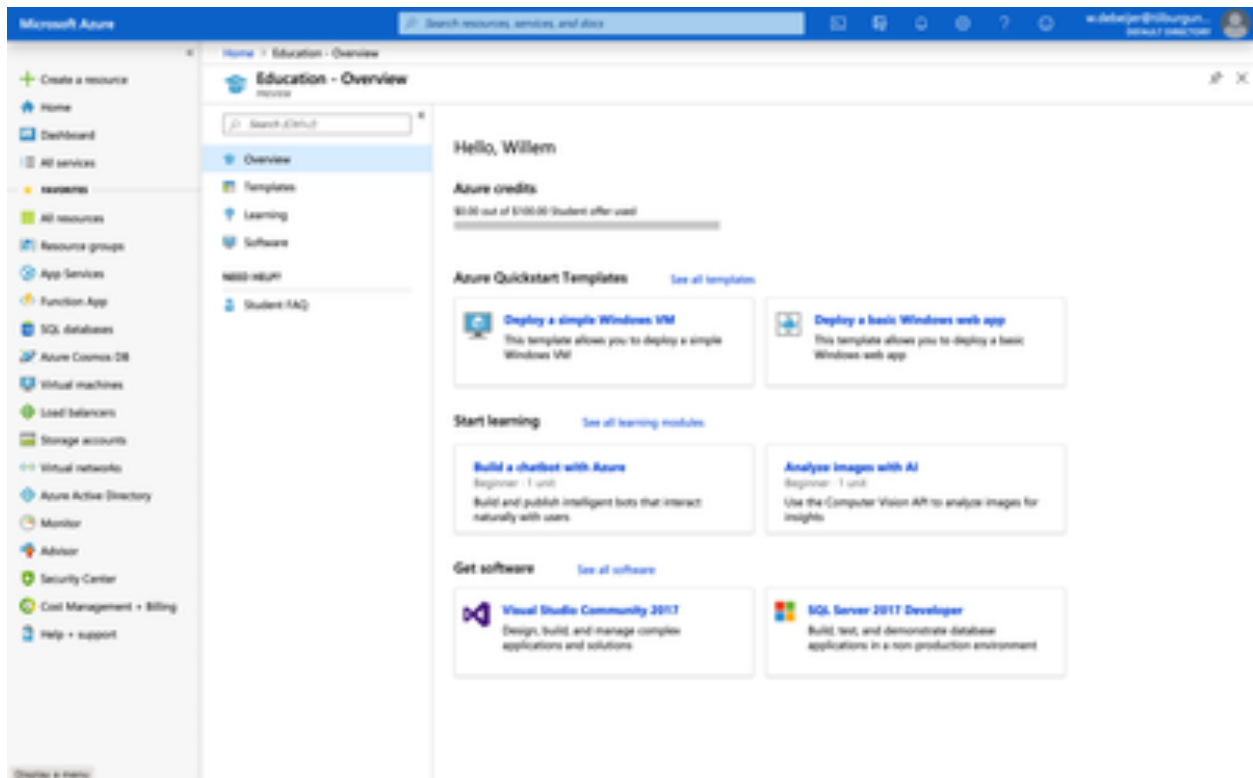
VM architecture ⓘ

☐ Arm64

☒ x64

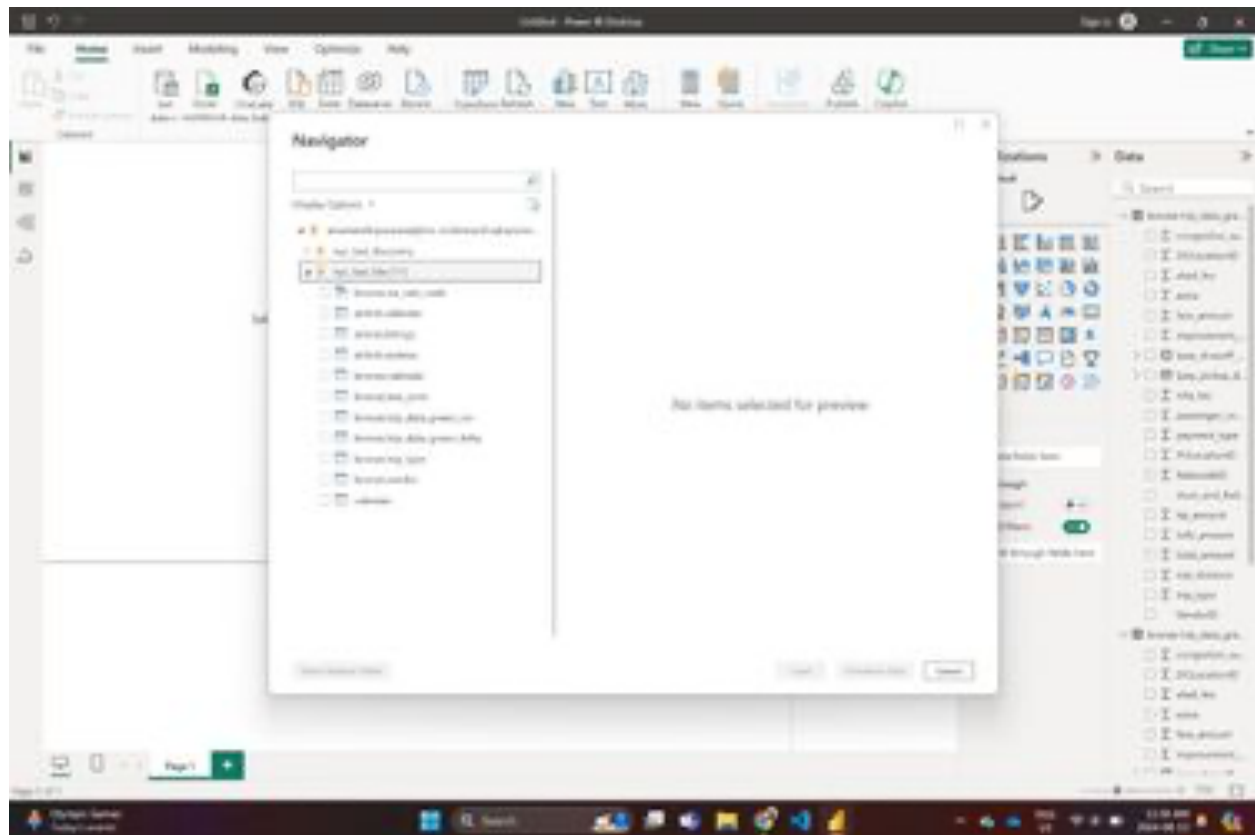
i Arm64 is not supported with the selected image.

Deploying VM



PowerBI

Connecting Synapse Analytics with PowerBI



Loading the dataset

Challenges Faced

Challenges	Solution
Unable to create cluster in databricks	-Requested more core permission through microsoft.
Unable to use power BI due to system os incapability	-Used microsoft VM