Project Title

**Data Analysis of Olympic Dataset using Azure Services.**

**Sponsor: Dear Future Ready Talent,**

**I would like to express my heartfelt gratitude for your invaluable guidance and support throughout our project. Your sponsorship has been instrumental in our success, and we truly appreciate your contributions.**

**Project Overview:**

In this project, we will leverage Azure Data Factory, an Azure-based data pipeline tool, to extract data from an API source. We will then orchestrate a data flow that loads this data into Azure Data Lake Storage. Initially, the data will be ingested into the Data Lake Storage. Subsequently, we will employ Azure Databricks to write Spark code, facilitating data transformation. The transformed data will then be loaded back into the Data Lake Storage. Lastly, we will harness the power of Power BI for in-depth data analysis and visualization.

**Project Synopsis:**

What is the core idea of your project?

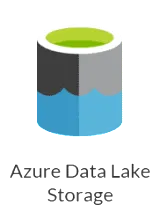
The core idea of your project is to collect data from various datasets (Excel sheets), transfer them to the cloud using Azure Data Factory, and then perform data transformations to facilitate data analysis in various forms.

In essence, your project aims to enable efficient and scalable data extraction, migration, and transformation processes for the purpose of in-depth data analysis and insights.

Services used are as follows:

Azure Data Factory simplifies ETL processes and streamlines data movement and transformation, making data integration efficient and accessible.





Azure Data Lake Gen 2 merges data lake features with Azure Blob Storage for efficient storage, analysis of extensive structured and unstructured data, offering enhanced performance, security, and advanced analytics.

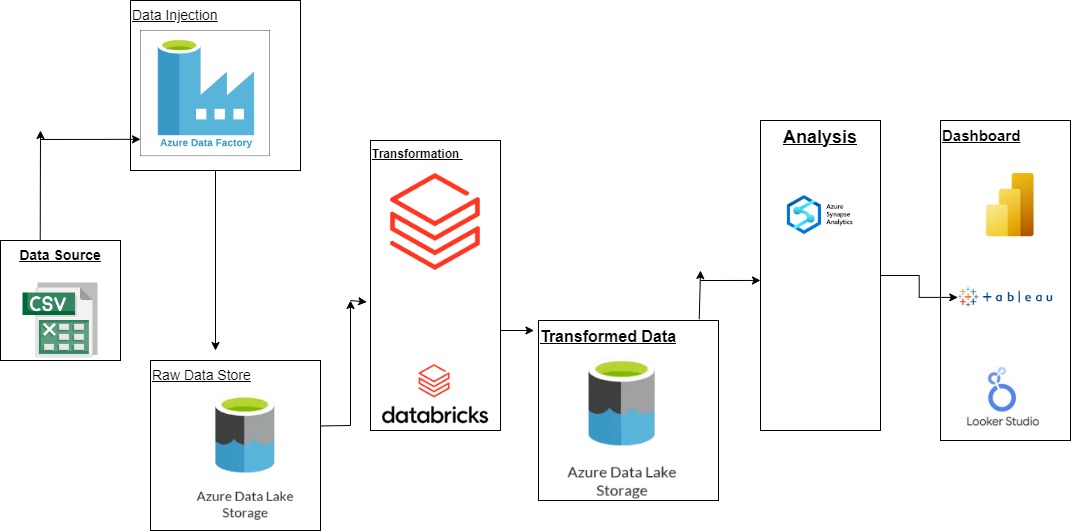


Databricks is a unified platform built on Apache Spark, facilitating collaboration between data engineers and data scientists for streamlined big data processing and machine learning tasks.

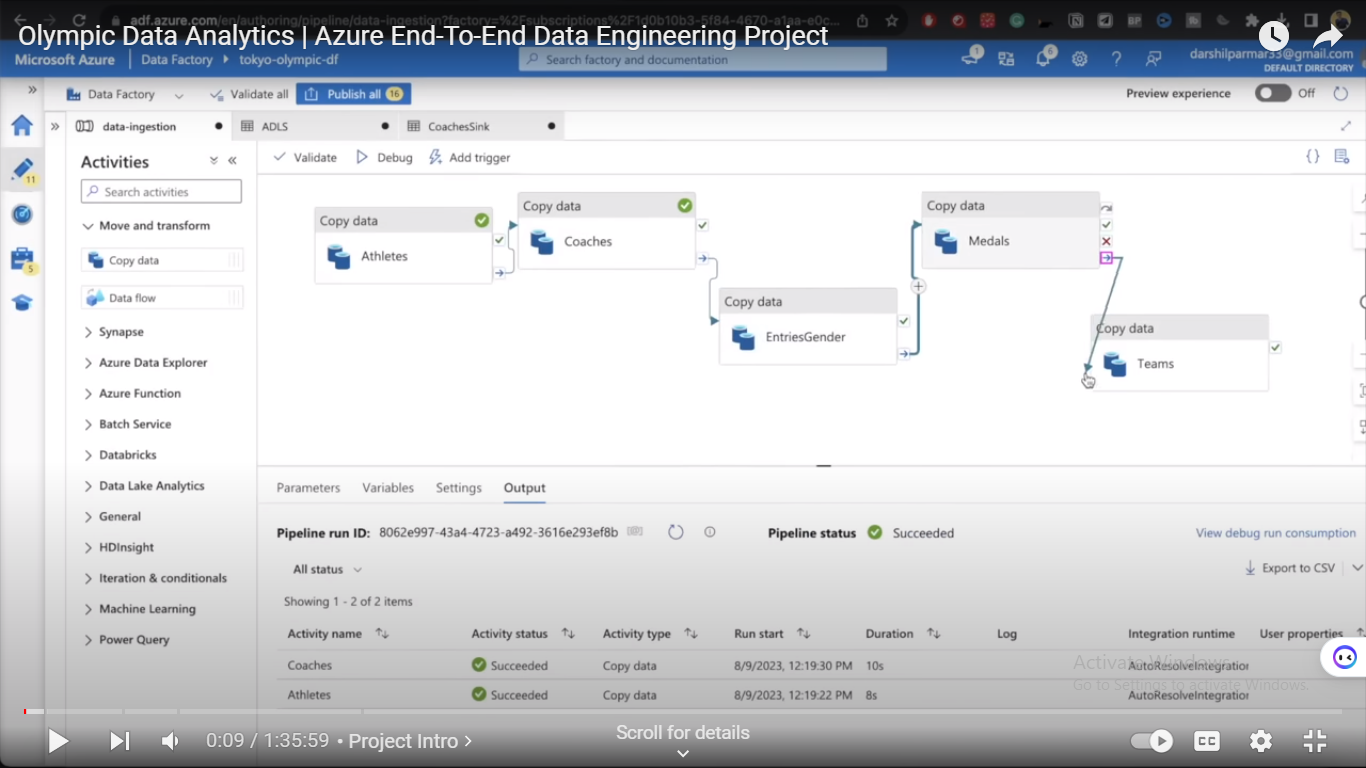


Power BI is a powerful business intelligence tool by Microsoft, designed for data visualization, reporting, and interactive insights to support data-driven decision-making.

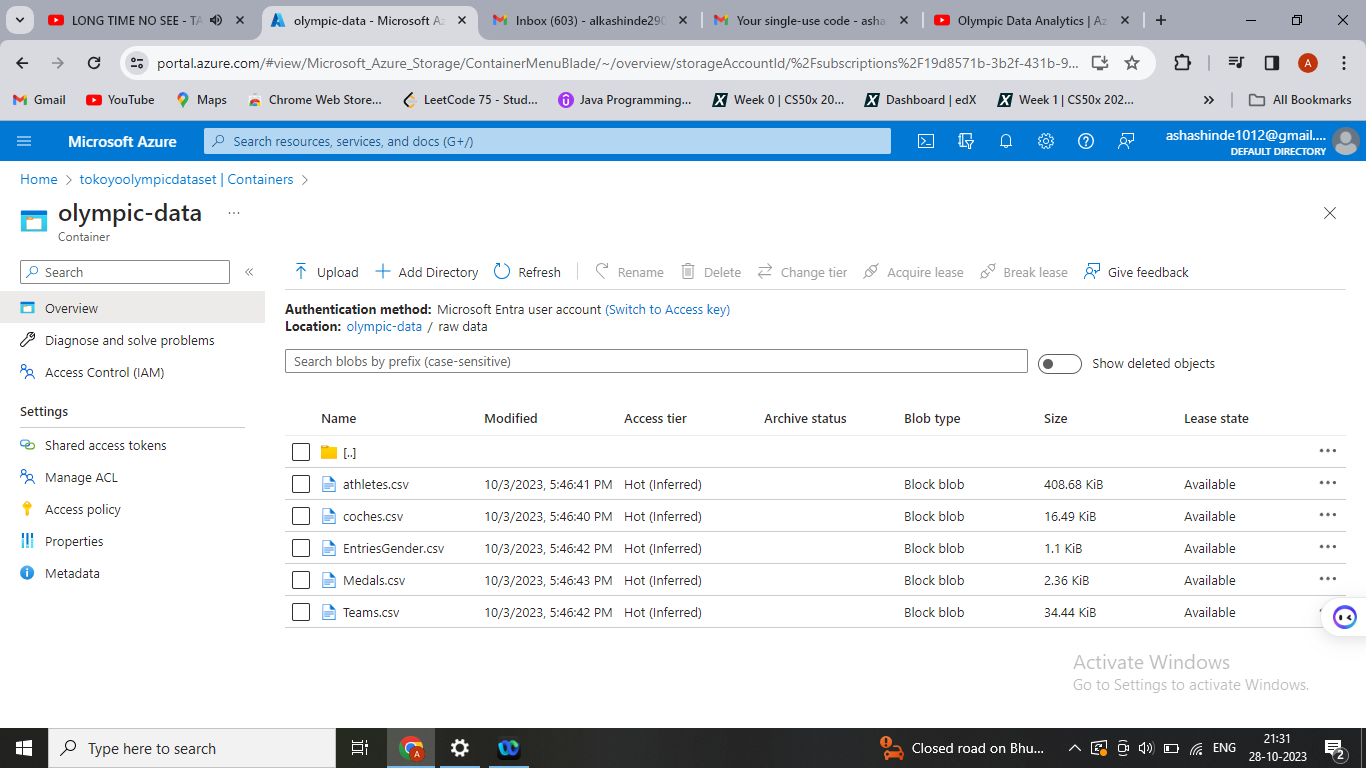
Data Flow Diagram for this project is as follows:



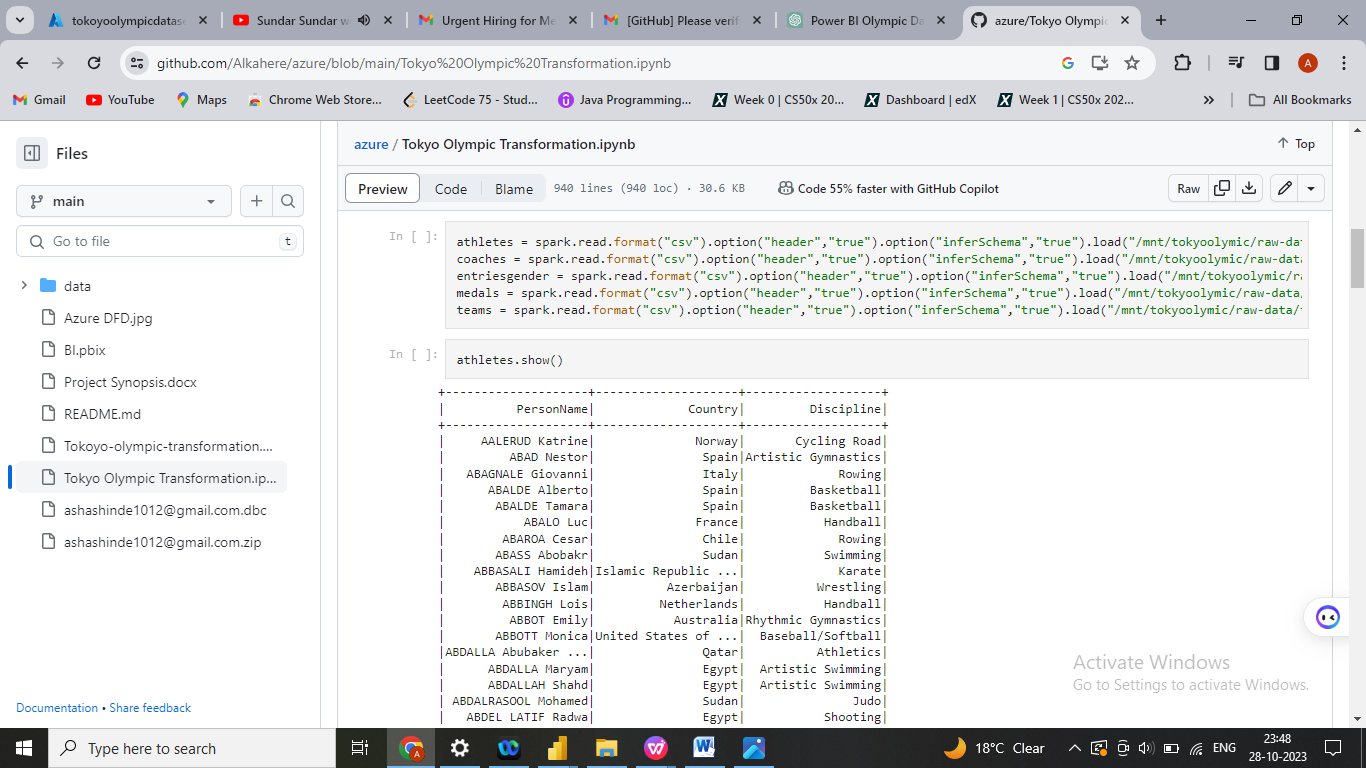
1. Creating a pipeline and storing into Azure Data Factory



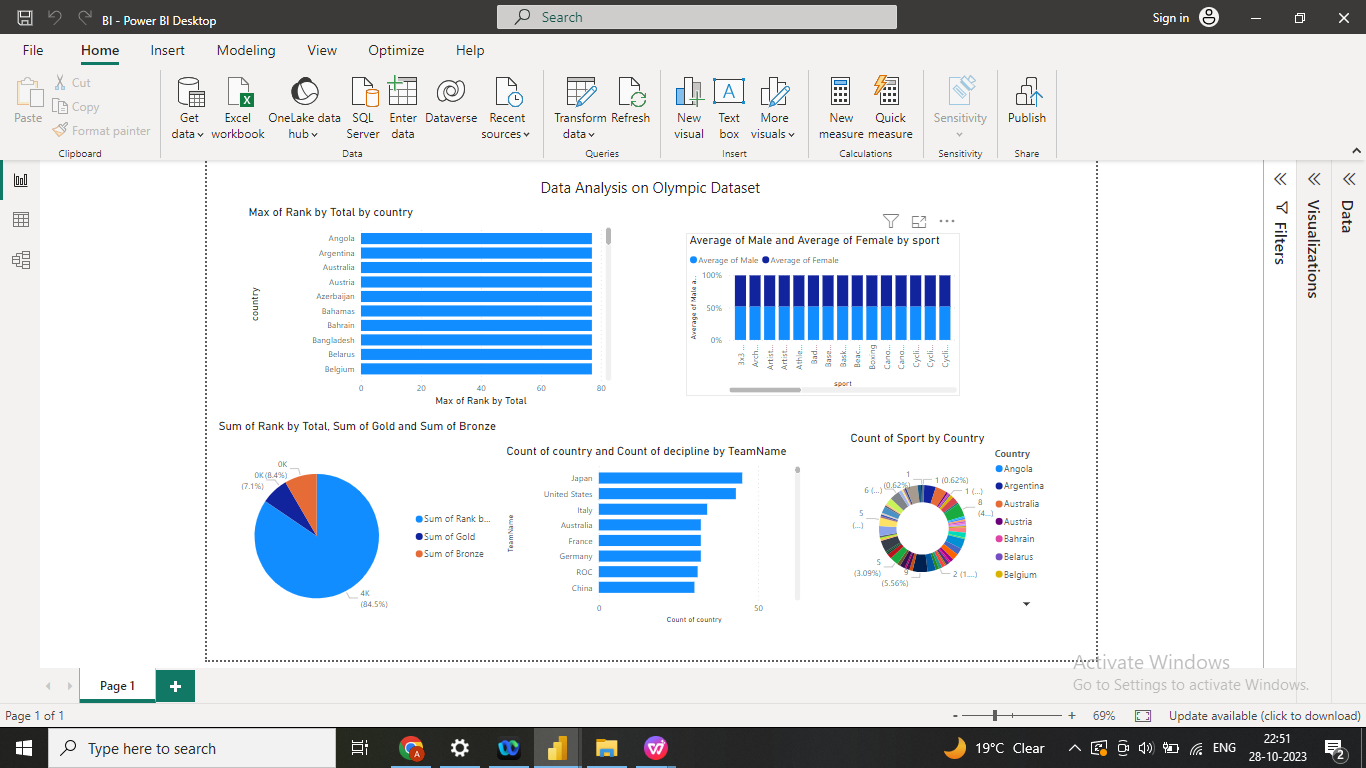
1. Loading all the data into Azure Gen 2 Cloud Storage.



1. Perform Transformation on data using PySpark in Azure Databricks.



1. Perform Analytics on the transformed data using Power BI.



Github Link : https://github.com/Alkahere/azure