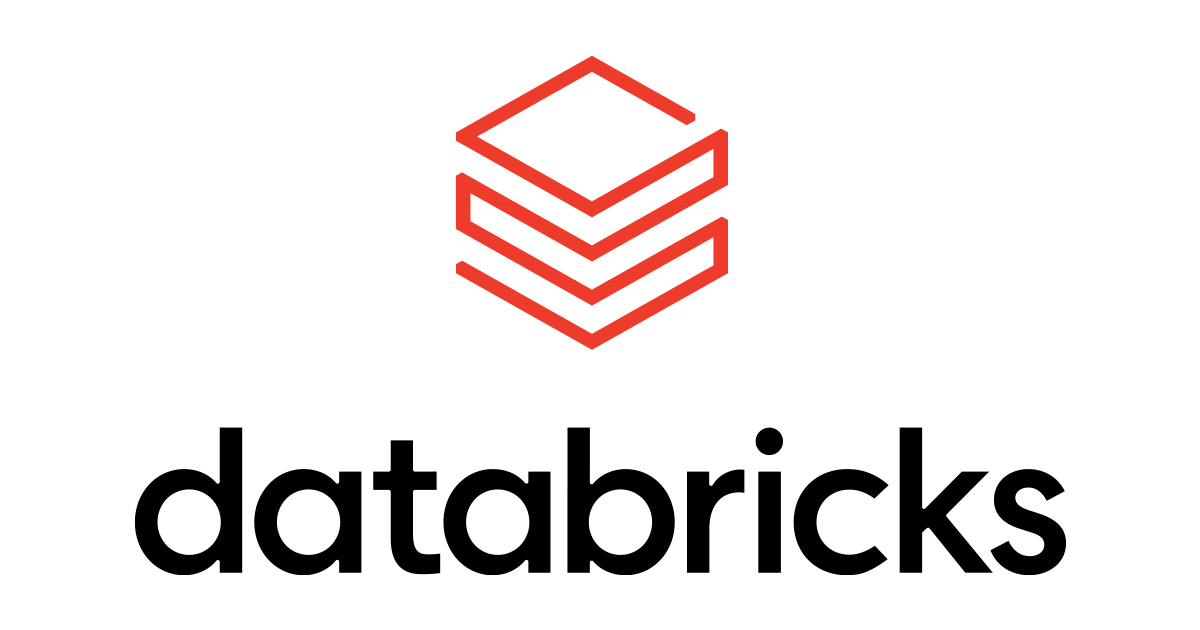
# System Requirements Specification Index

For

**Data Bricks Netflix shows data ingestion and analysis using pyspark .**

Version 1.0



IIHT Pvt. Ltd.

[fullstack@iiht.com](mailto:fullstack@iiht.com)

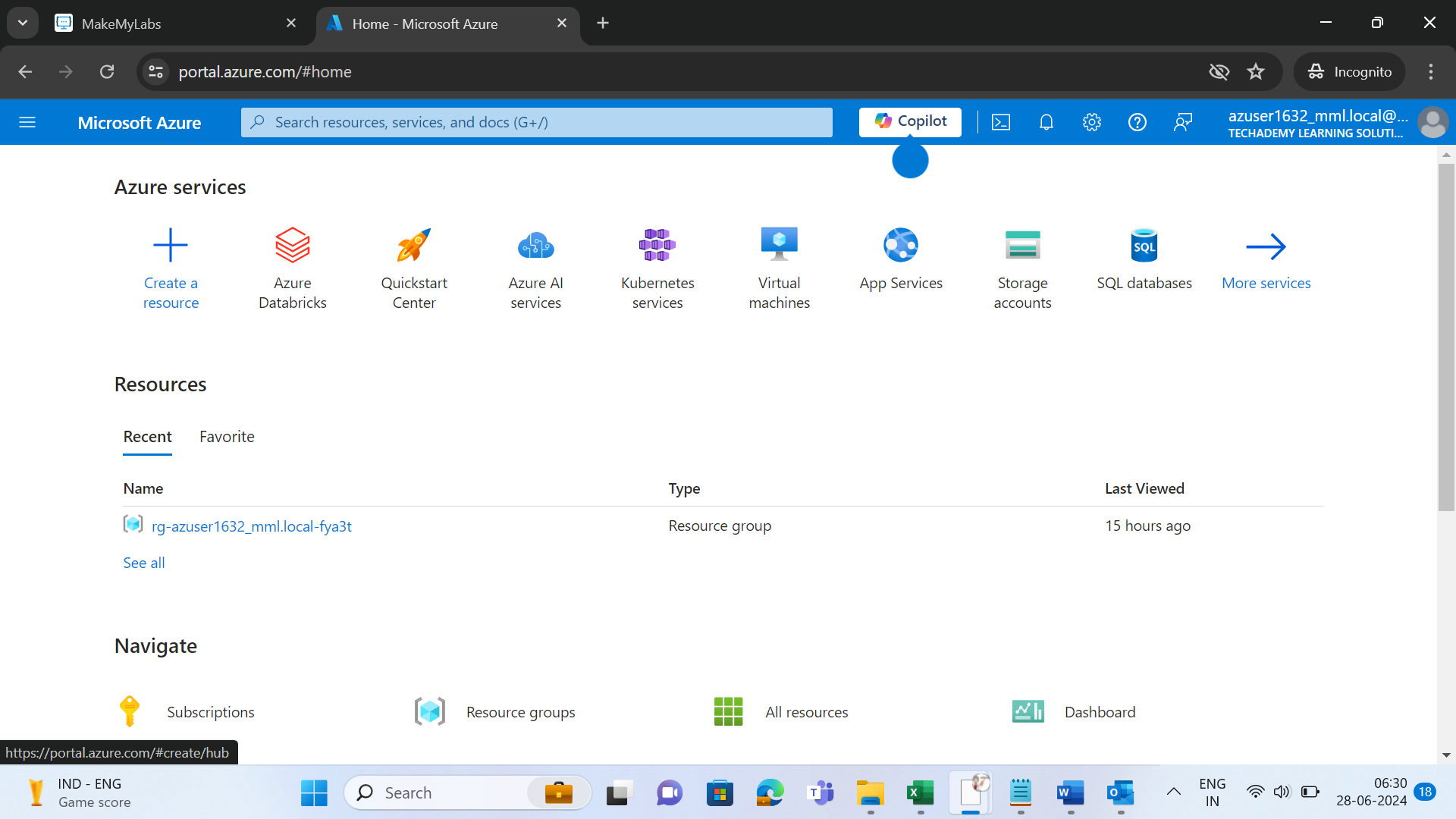
Problem Statement : **DataBricks Netflix shows data ingestion and analysis using pyspark .**

.

Description : Use relevant methods operations to perform specified activities which are given in the instructions.

ABC Media Inc. is a global media company that operates a streaming platform similar to Netflix, offering a diverse catalog of movies and TV shows to millions of subscribers worldwide. To enhance user experience and optimize content offerings, ABC Media Inc. aims to leverage advanced data analytics to understand preferences, and content performance, most watched . By leveraging Databricks with PySpark, ABC Media Inc. transforms complex data into actionable insights, gaining a competitive edge in the global streaming industry

* Steps to login in the Azure account 🡪 Azure Databricks🡪 create workspace 🡪 Create cluster 🡪
* Use the given dataset to create a table in the workspace



* Click on new and create a new notebook

A screenshot of a computer

Description automatically generated

Once you have imported the table check for respective rows and columns with their respective datatypes are imported correctly.

A screenshot of a computer

Description automatically generated

Dataset link

https://github.com/IIHTDevelopers/Azuredatabricksdatasets.git

Problems

1. How can one access the table according to the instructions given take screenshot whenever necessary ?
2. How would you retrieve and display the counts of movies and TV shows from the dataset?
3. How would you handle null values in the database, specifically using "NA" as a replacement?
4. How would you filter the database to count the number of movies and TV shows?
5. How would you query the database to display the types of genres available?
6. How could you create a bar graph using the genre data extracted from the table?

## Execution Steps to Follow:

1. Open the Azure dashboard console search for Databricks
2. Import the dataset from the public repository
3. Perform all the query respective to the question provided
4. Take screenshots of the query execution
5. Upload the code to the Github

## -----x-----