**29/12/2023 Why Azure Databricks?**

**What is Azure Databricks and how is it related to Spark?**

Databricks is the implementation of Apache Spark on Azure. With fully managed Spark clusters, it is used to process large workloads of data and also helps in data engineering, data exploring and also visualizing data using Machine learning

While I was working on databricks, I find this analytic platform to be extremely developer-friendly and flexible with ease to use APIs like Python, R, etc. To explain this a little more, say you have created a data frame in Python, with Azure Databricks, you can load this data into a temporary view and can use Scala, R or SQL with a pointer referring to this temporary view. This allows you to code in multiple languages in the same notebook. This was just one of the cool features of it

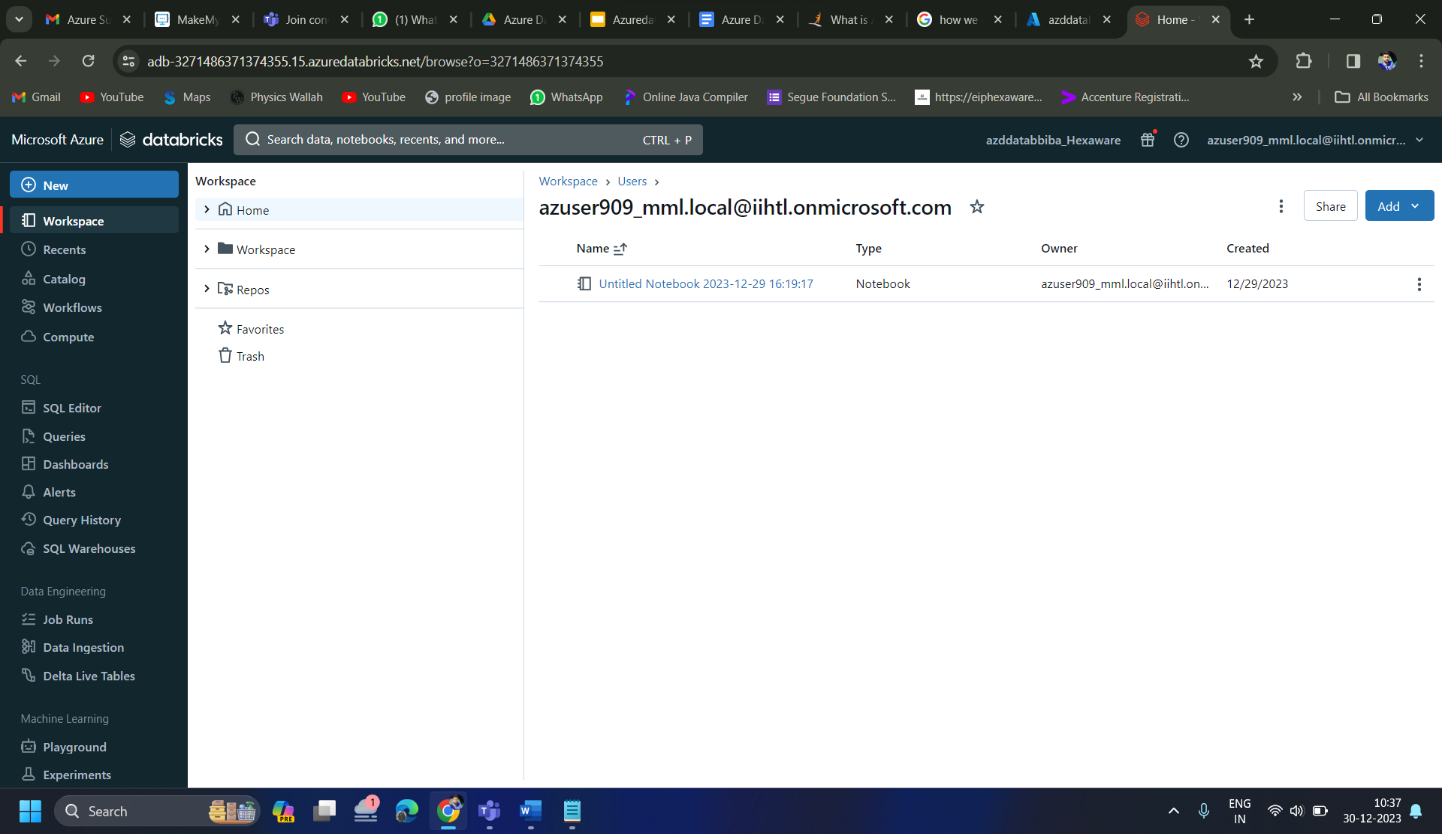
**Why Azure Databricks?**

Evidently, the adoption of Databricks is gaining importance and relevance in a big data world for a couple of reasons. Apart from multiple language support, this service allows us to integrate easily with many Azure services like Blob Storage, Data Lake Store, SQL Database and BI tools like Power BI, Tableau, etc. It is a great collaborative platform letting data professionals share clusters and workspaces, which leads to higher productivity.

**Create a notebook in the Spark cluster**

A notebook in the spark cluster is a web-based interface that lets you run code and visualizations using different languages**.**

Once the cluster is up and running, you can create notebooks in it and also run Spark jobs. In the Workspace tab on the left vertical menu bar, click Create and select Notebook



Run some pyspark command on notebook created on azure

