Name : Parth Nandedkar
Date : 12 Feb 2024
Topics : PySpark
Batch : Data Engineering Batch-1

Handwritten Notes :

12/02/24

Databricks hosted on Cloud platform (azure, aws)

Databricks hosted on opensource -- Databricks.com

pyspark ϕ - Databricks

Azure + databricks --- > Azcloud + databricks → cloud databricks.

Databricks hosted on Cloud platform.

before AI

Virtualization :
     Everything we can create in a virtual space (VMs).

Azure :
     Azure is Microsoft's cloud computing platform.

     Azure is a continually expanding set of cloud services that help your organization meet your current & future business challenges

     Azure gives you the freedom to build, manage & deploy applications on a massive global network using your favourite tools and frameworks.

IaaS	PaaS	SaaS
Infrastructure as-a service	Platform	Software as-a service

11/2/21

Why cloud →   Speed

                Scale

                Economics

Paas →   SQL, Storage

Iaas →

Saas →

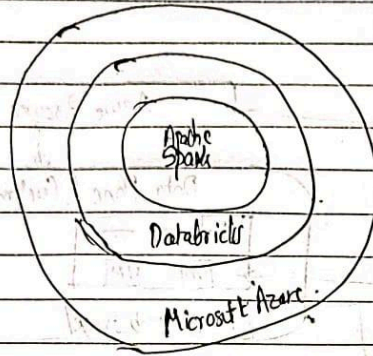Core Cloud Services :

     (VMs, Containers

- Azure compute Option

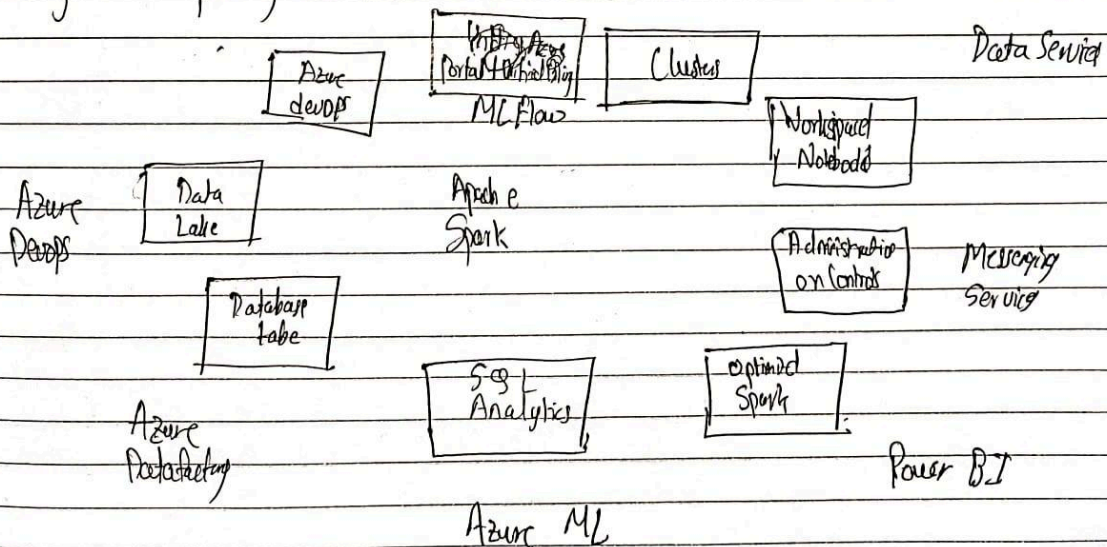- Azure data storage options

- Azure Networking options
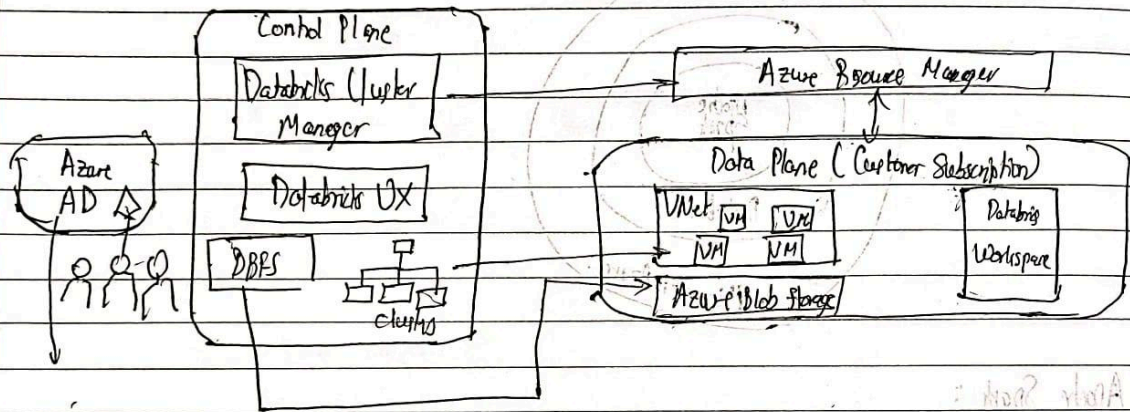
- Azure IOT.

Azure Stack :

Azure Databricks : =>

(Concentric circles, inner to outer: Apache Spark, Databricks, Microsoft Azure)

Apache Spark :

Apache Spark is a lightning – fast unified analytics engine for big data processing & ML.

Azure devops	Unify Azure Portal + Virtual Bin ML Flow	Clusters	Data Service
Azure Devops	Data Lake	Apache Spark	Workspace Notebooks
	Database table		Administration Control Messaging Service
Azure Datafactry	SQL Analytics	optimized Spark	
			Power BI
	Azure ML		

Azure Databricks Architecture

Control Plane
- Databricks Cluster Manager
- Databricks UX
- DBFS
- Clusters

Azure AD

Azure Resource Manager

Data Plane (Customer Subscription)
- VNet
  - VM
  - VM
  - VM
  - VM
- Azure Blob Storage
- Databricks Workspace

AD → Active Directory

Setting up Azure Lab :
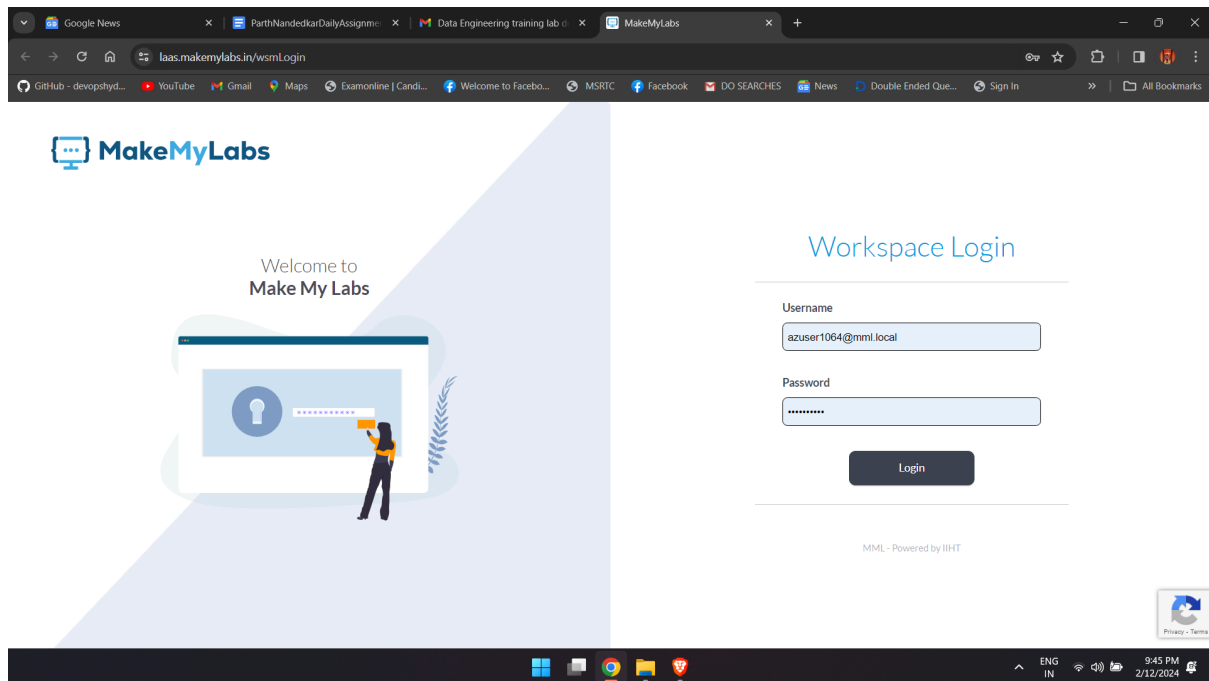
Homepage preview where we need to add mail address and password to login.

Workspace Dashboard :

Lab creation page :

After clicking on connect we get this page where Authentication URL mail id and temporary password is provided .

If we log in to the session it will be only valid for 60 mins after that we need to generate a new password and login again.

Azure Dashboard :

Creating Workspace for Azure Databricks :
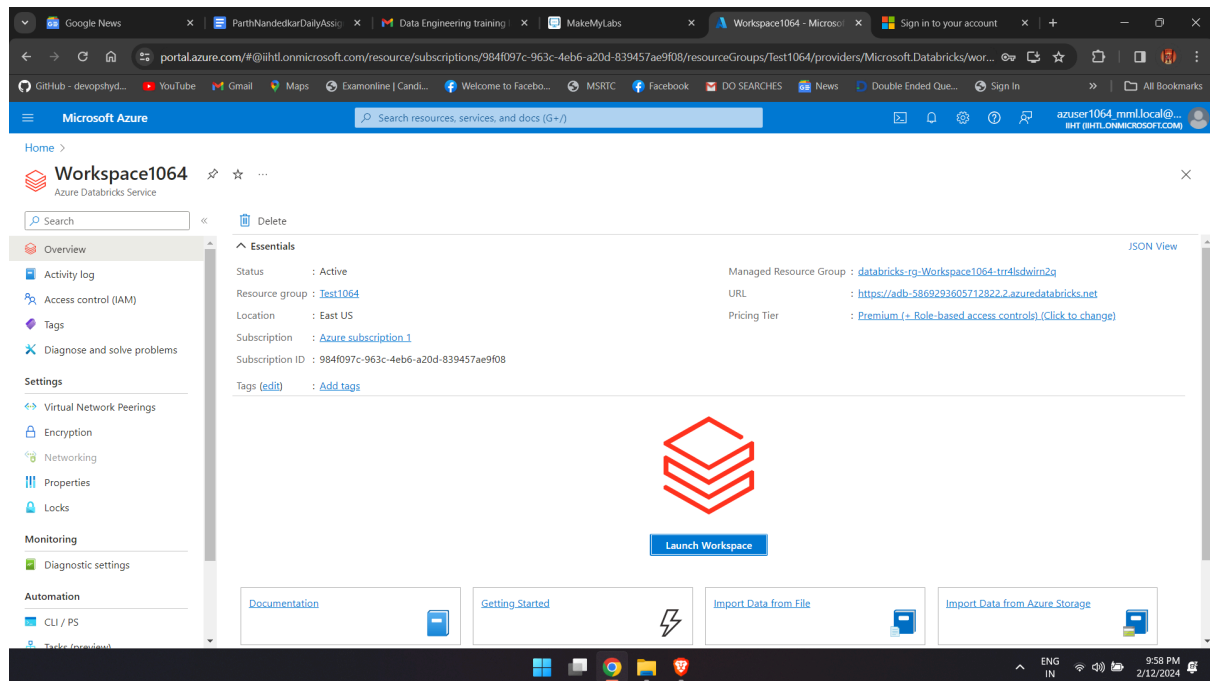
Above image shows the workspace that we created for using Azure Data Bricks as we click on Launch Workspace the following databricks UI will be opened in a new tab .
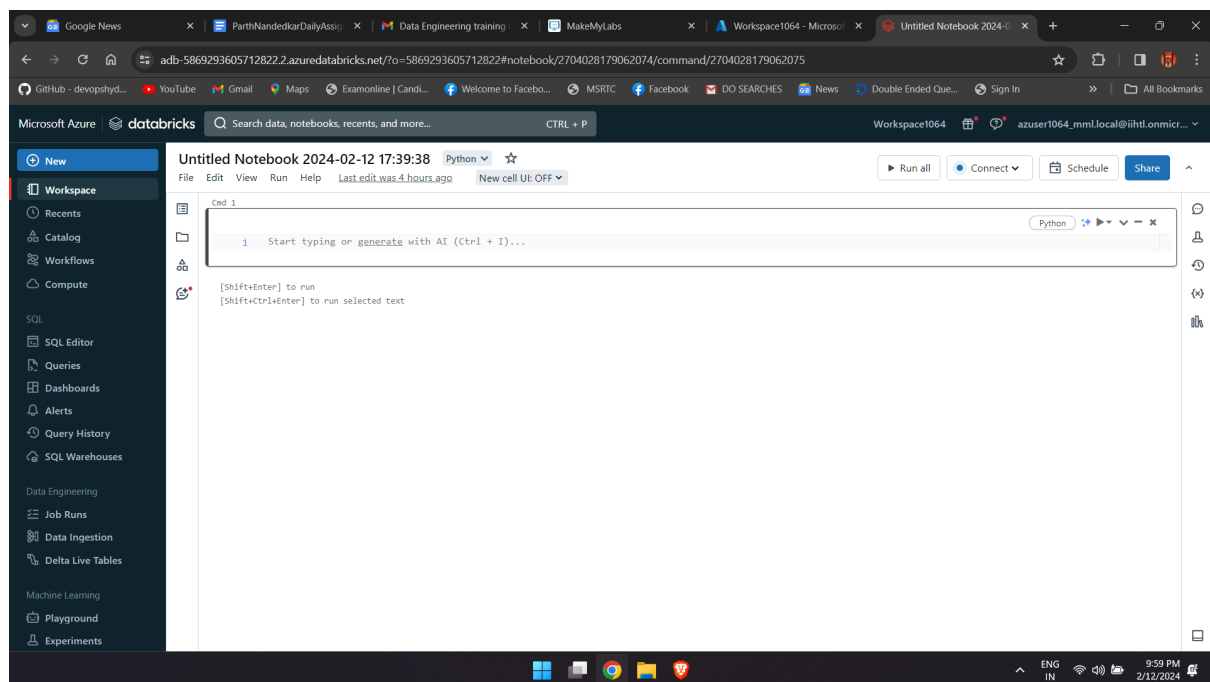
As we ran commands on Databricks we can run multiple commands on Azure Databricks as it was a community edition, this lab provides many more functionalities so that we can work on data easily using many VMs or clusters.

As it is a very powerful cloud tool many heavy softwares can be ran using this Azure cloud.