

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 & - Databricks

Azure + databricks --- > cloud + 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

HCI → Hyper Converged Infrastructure  
DCs →

Why cloud → Speed →

Scale  
Economics

PaaS → SQL, Storage

IaaS →

SaaS →

Core Cloud Services:

VMs, Containers

- Azure compute options

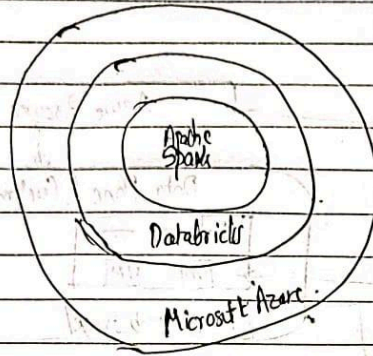
- Azure data storage options

- Azure networking options

- Azure IoT

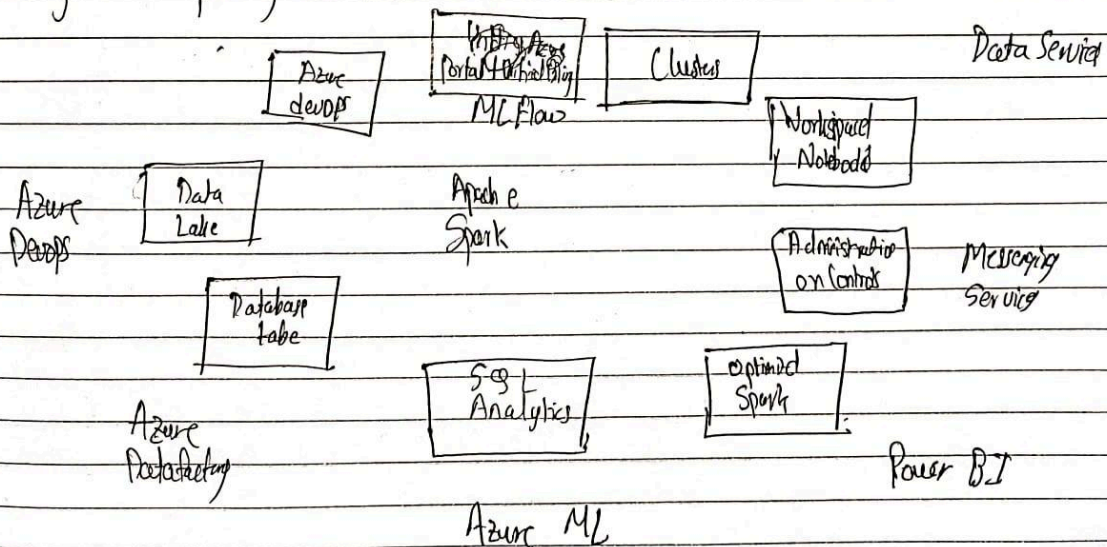
Azure Stack:

Azure Databricks :: =>



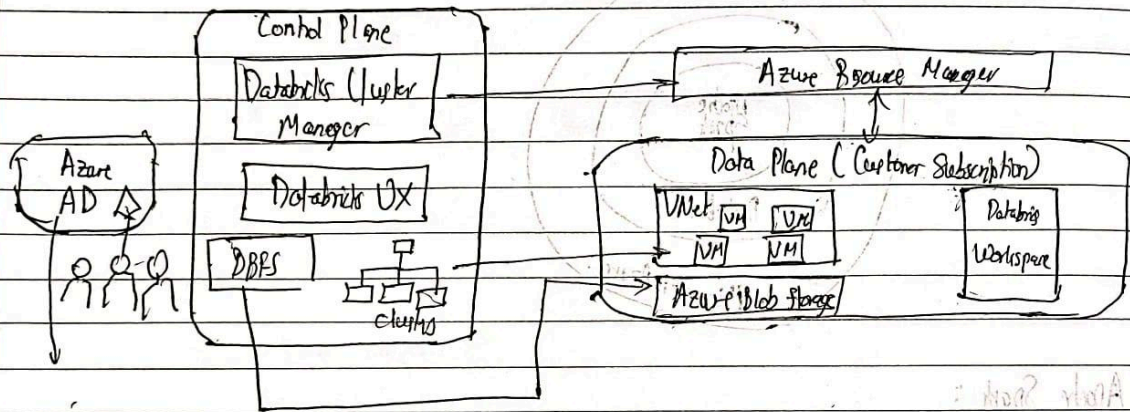
Apache Spark :

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



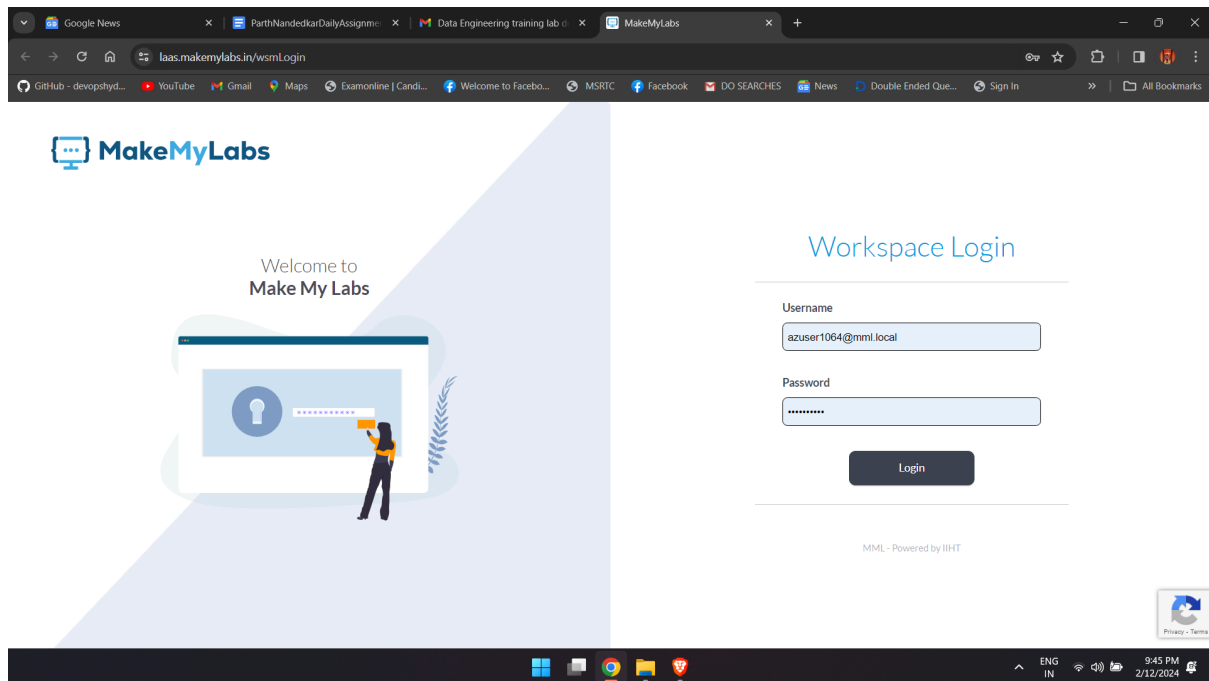


# Azure Databricks Architecture



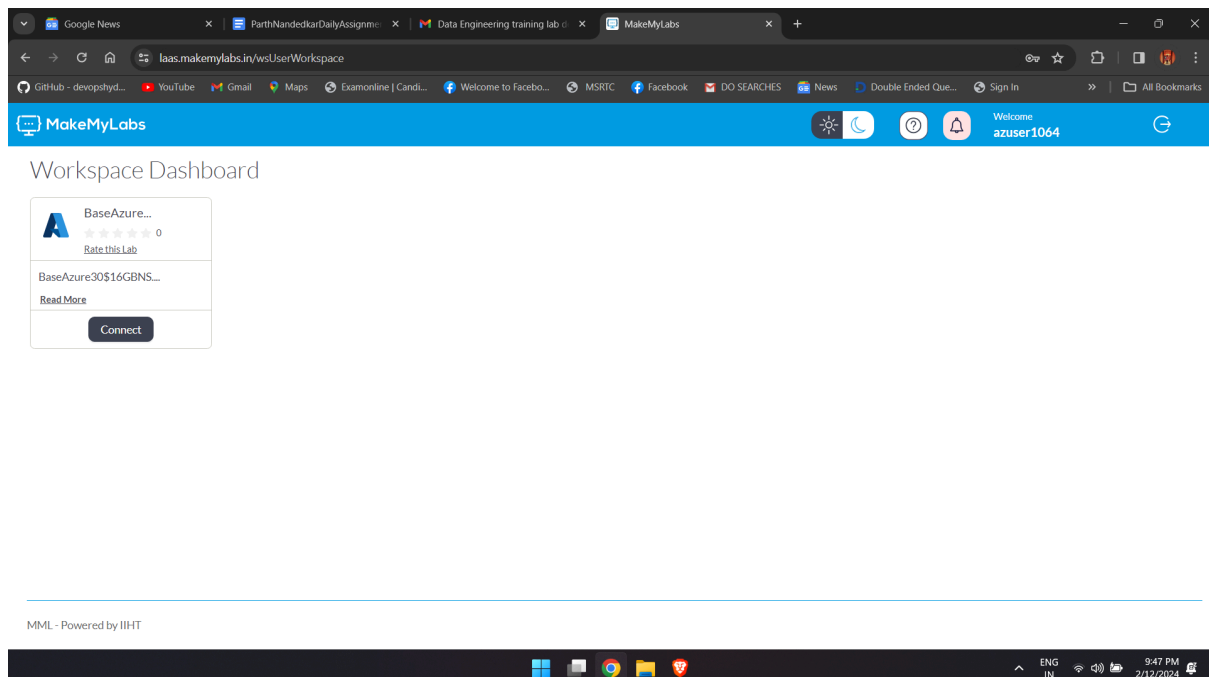
AD → Active Directory

## Setting up Azure Lab :



Homepage preview where we need to add mail address and password to login.

## Workspace Dashboard :



## Lab creation page :

The screenshot shows a web browser window with the MakeMyLabs interface. The page title is "Machine Catalog > BaseAzure30\$16GBNS-9d5ea271-fbd6-4856-8003-27cc482fa5c5-1". The URL is "laas.makemylabs.in/userMachineCatalog/creditBasedLabDetails/0294507f-57b5-493b-8977-85217c692f72". The page has three tabs: "Access Information", "Usage Details", and "Extension Details". The "Access Information" tab is active. It displays the lab name "BaseAzure30\$16GBNS-9d5ea271-fbd6-4856-8003-27cc482fa5c5-1" with a rating of 4.5 stars. Below the name, it shows "Validity 45", "Available Amount 30", and "Available Duration -". The "Lab Start" is "2024-02-05" and "Lab End" is "2024-03-21". The "Description" is "BaseAzure30\$16GBNS". A note states: "Note: To login to Azure Lab, please copy the Auth URL below and paste it into a Private/Incognito browser window." The "Auth Url" is "https://portal.azure.com" with a "Refresh Credentials" button. The "Username" is "azuser1064\_mml.local@iitl.onmicrosoft.com" and the "Password" is masked with dots. A "Session Rules" section is at the bottom, and a "Session Expires In: 0h:0m:0s" timer is visible. The Windows taskbar at the bottom shows the time as 9:48 PM on 2/12/2024.

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 :

The screenshot displays the Microsoft Azure portal dashboard. At the top, there's a navigation bar with the 'Microsoft Azure' logo and a search bar. Below this, the 'Azure services' section features a row of icons for 'Create a resource', 'Azure Databricks', 'All resources', 'Quickstart Center', 'Virtual machines', 'App Services', 'Storage accounts', 'SQL databases', 'Azure Cosmos DB', and 'More services'. The 'Resources' section has tabs for 'Recent' and 'Favorite', with a table listing recent resources. The 'Navigate' section includes icons for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. The 'Tools' section is currently empty. The bottom of the screen shows a Windows taskbar with the Start button, task view, and several open applications, along with system icons for language, network, and time.

**Azure services**

Create a resource, Azure Databricks, All resources, Quickstart Center, Virtual machines, App Services, Storage accounts, SQL databases, Azure Cosmos DB, More services

**Resources**

Recent Favorite

Name	Type	Last Viewed
Workspace1064	Azure Databricks Service	4 hours ago
Test1064	Resource group	4 hours ago

See all

**Navigate**

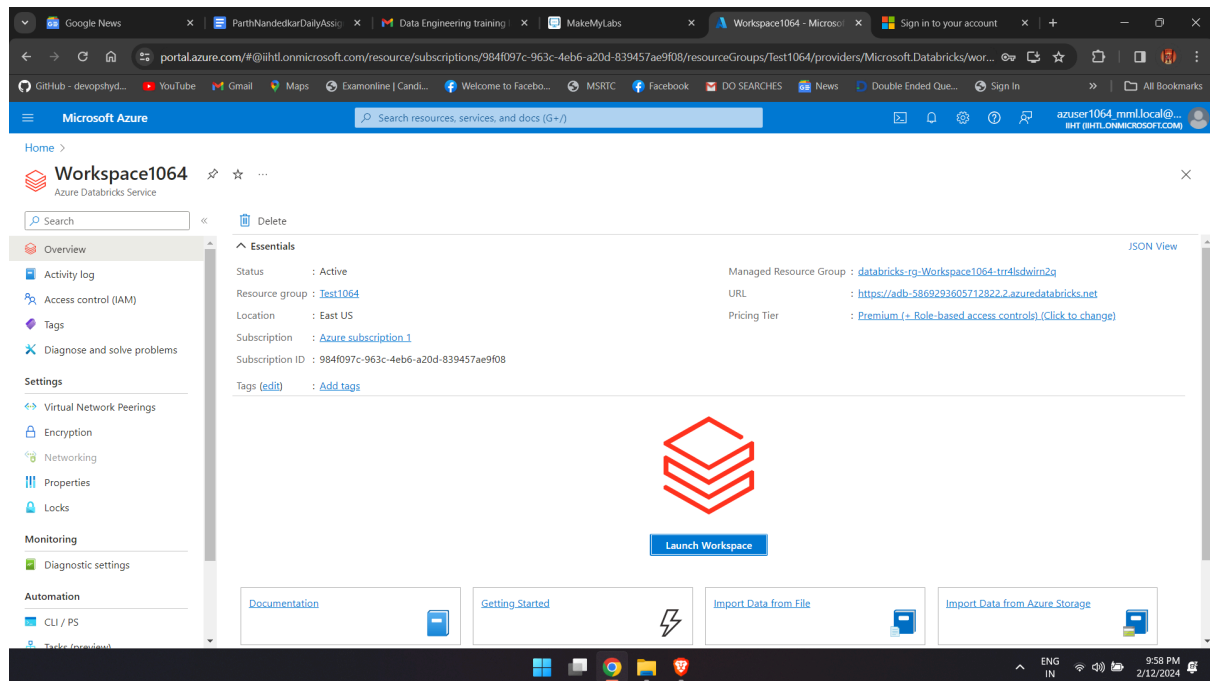
Subscriptions, Resource groups, All resources, Dashboard

**Tools**

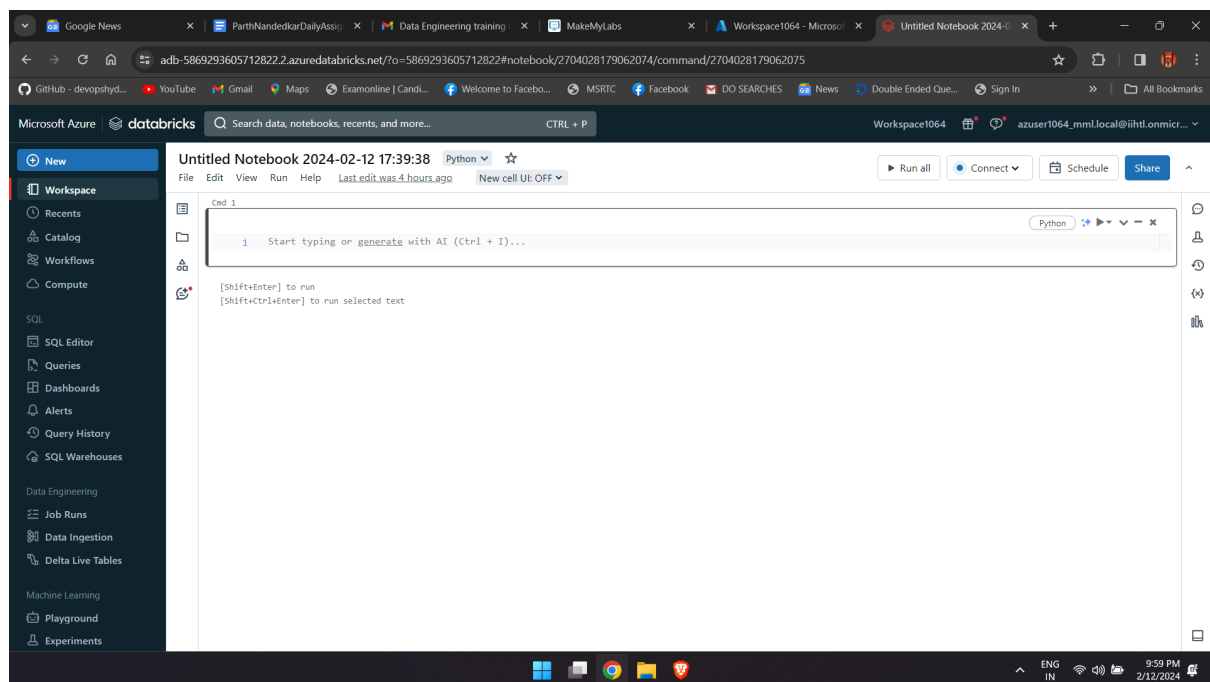
https://portal.azure.com/#create/hub

9:54 PM 2/12/2024

## 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.