Microsoft Azure

* Everything on Azure portal is a service.
* In this training, Azure DevOps organization is used.
* Databricks is hosted on Azure cloud platform. (or) it can be any platform (like AWS, GCP etc.,)
* If hosted on Azure cloud platform, it is called as Hosting the services on Azure cloud platform.

Azure + databricks ⟶ cloud + Databricks

↓

cloud Databricks

* Databricks hosted on cloud platform (AZURE)

cloud :- cluster of servers in a n/w diagram has several overlapping circles, which resembles cloud

Microsoft Azure :-

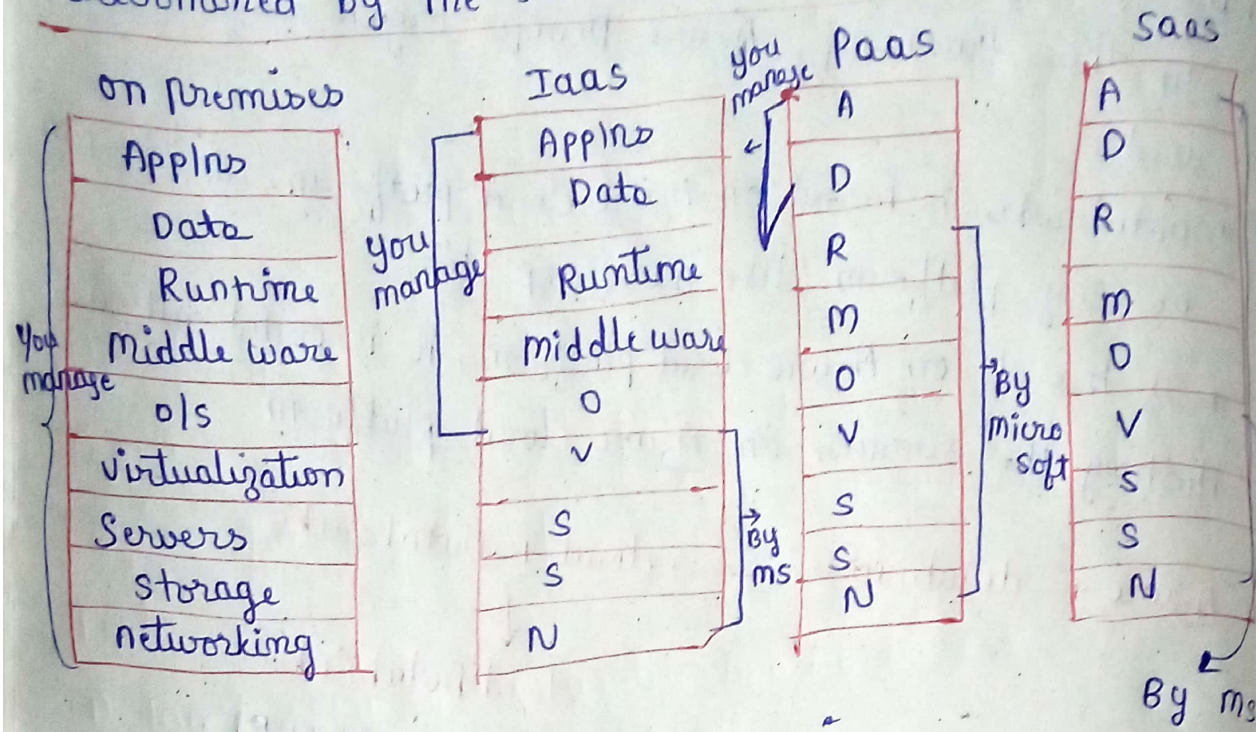
* Azure is Microsoft's cloud computing platform
* Azure is continually expanding set of cloud services that help your organization meet your current & future business challenges.
* Azure gives you freedom to build, manage & deploy appln on a massive global n/w using your favourite tools & frameworks

Types of Services / Service models

1) Iaas ⇒ SQL virtual machine
2) Paas
3) Saas ⇒ Gmail, drive.

Cloud Services (or) Cloud Models

All contains same services. The type of model is determined by the ~~SO~~ AAAS who is managing servi



Why cloud ? Speed, Scale, Economics

1) Rapidly setup environments to drive business priorities
2) Scale to meet peak demands
3) Increase daily activities, efficiency & reduced cost

Platform services

1) Media & CDN
2) Integration
3) Compute Services
4) Appln platform
5) Developer Services
6) Data
7) Intelligence
8) Analytics & IOT

Infrastructure services

1) Compute
2) Storage
3) Networking

Azure data centers

Azure has more global regions than any other cloud provider - offering the scale needed to bring applns closer to users around world. 54 regions worldwide 140 available in 140 countries

Core cloud Services

1) Azure Compute options
Virtual machines, containers, Azure App Service, Serverless Computing

2) Azure data storage options - Azure SQL Database, Azure Cosmos DB, A BLOB, A^TM Data Lake, A files, disk storage

3) Azure networking options

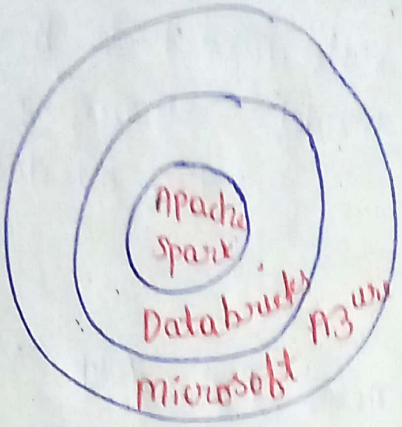
4) Azure IOT

Azure Stack

* It is a portfolio of products that extend Azure services and capabilities to your environment of choice - from data center to edge locations & remote offices

* the portfolio enables hybrid and edge computing applns to be built, deployed and run consistently across location boundaries, providing choice & flexibility to address your diverse workloads

Introduction to Azure Databricks



Apache Spark runs on Databricks and is hosted on the Microsoft Azure cloud platform

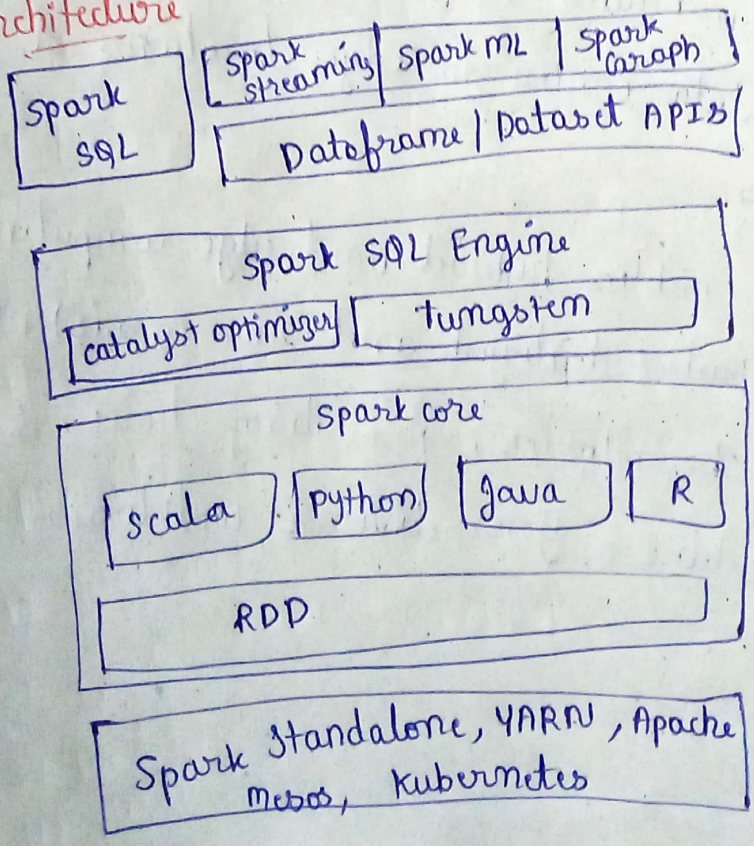
Apache Spark

Apache Spark is a lightning-fast unified analytics engine for big data processing and machine learning

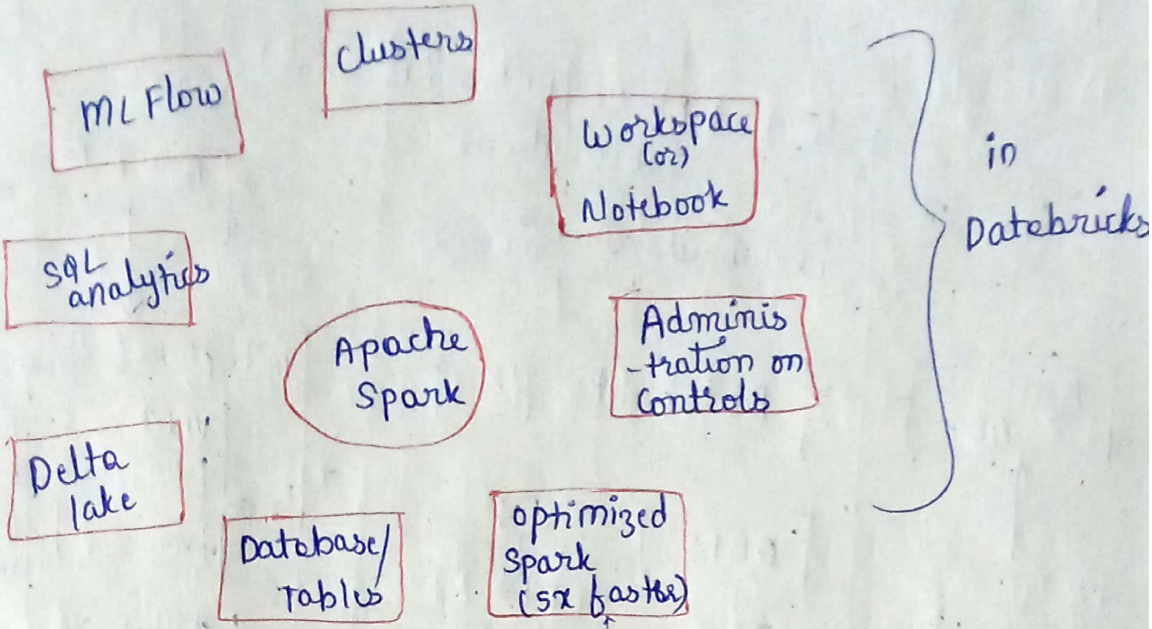
Features :-

1) 100% open source under Apache License
2) simple & easy to use APIs
3) In-memory processing engine
4) Distributed computing platform
5) unified engine which supports SQL, streaming, ML

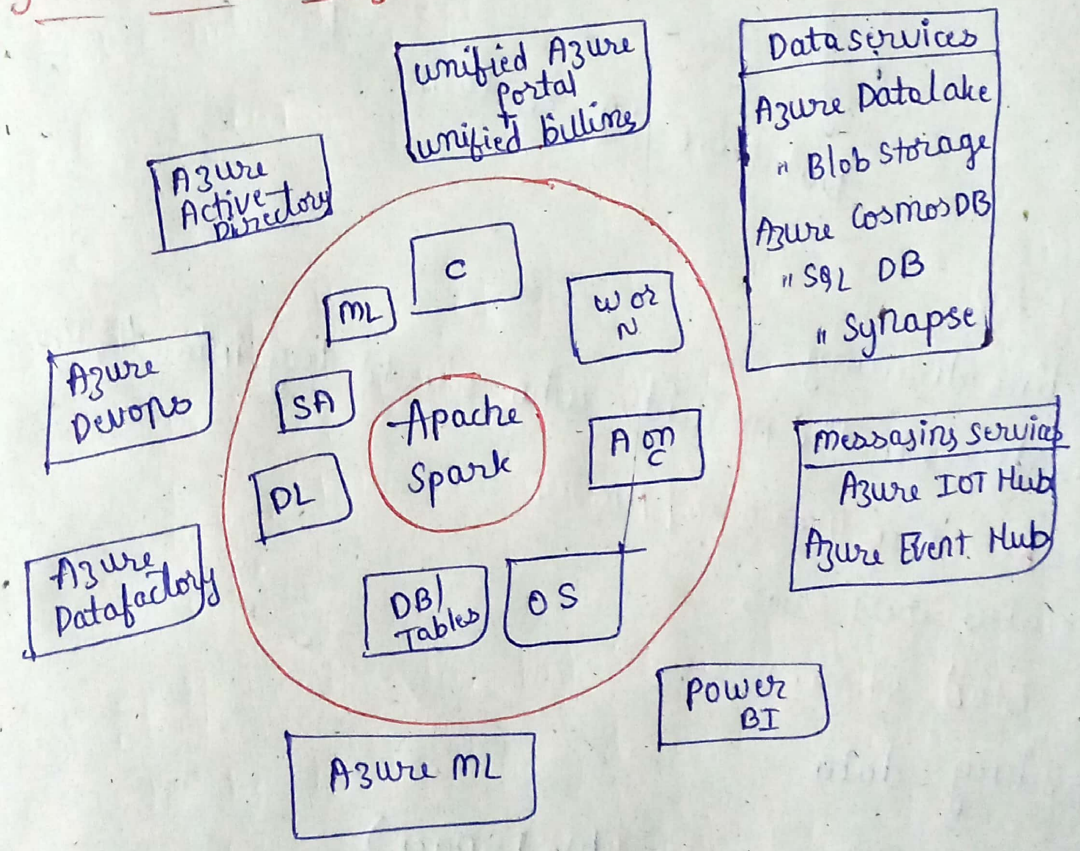
Architecture



Databricks

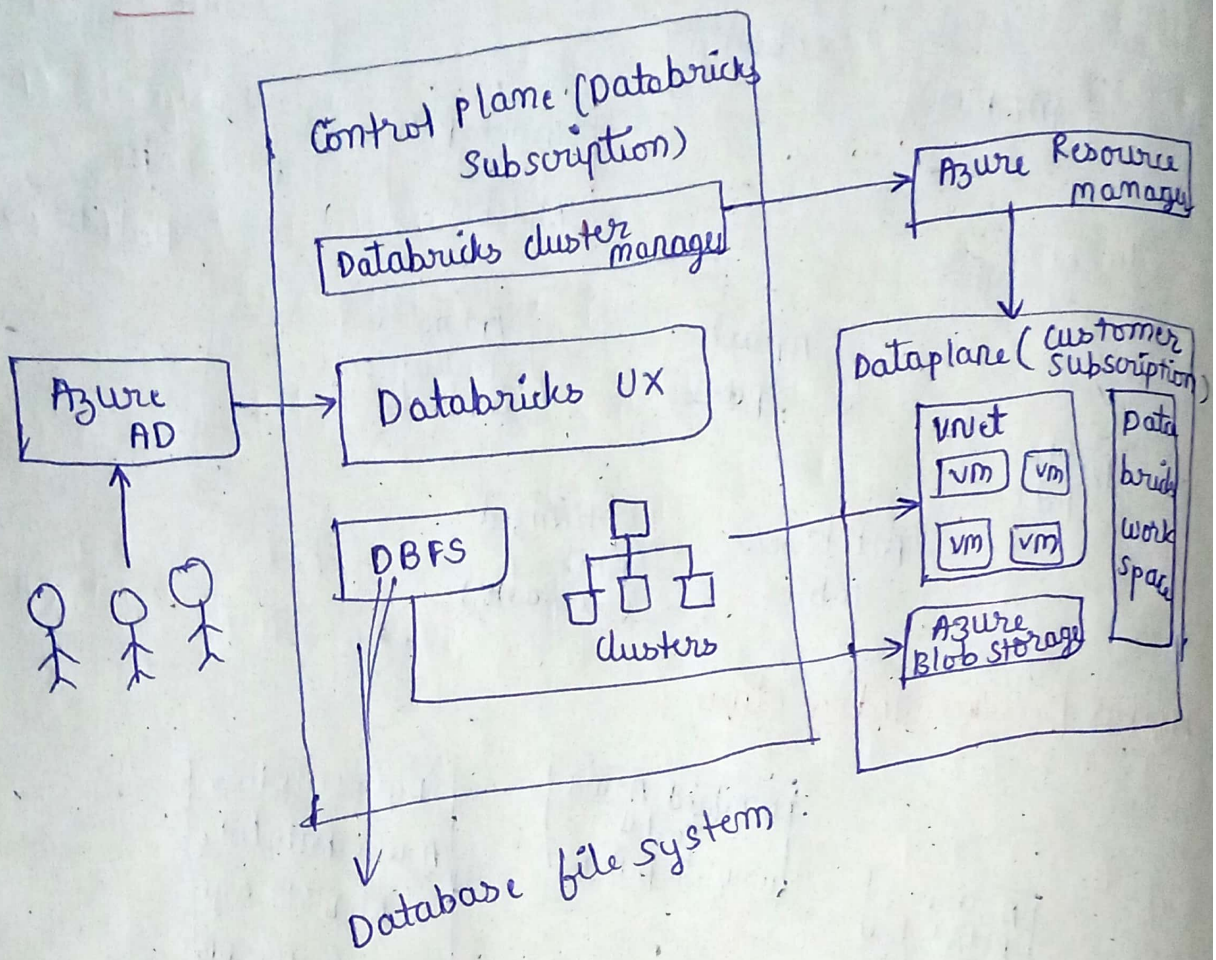


Azure Bricks Integration :-



We are going to use, Azure Active Directory, Azure Devops, Data Factory, Unified Azure portal + unified billing etc.,

Azure Databricks Architecture



Azure AD

* Active directory
* creates all the mail ids who should log into Azure.

Azure Resource manager / Mover

* allocates resources

Azure blob storage

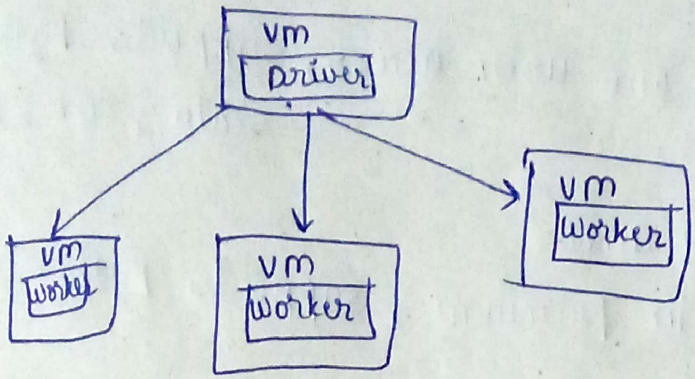
* storing data

Components of cluster (Workspace)

1) Notebooks
2) Jobs
3) Models
4) Data

5) cluster

Databricks Cluster



1) Databricks cluster
2) cluster types
3) " Configuration
4) creating cluster
5) Pricing
6) Cost control
7) cluster pools
8) cluster policy

Types of cluster

All purpose	Job cluster
1) created manually	1) Created by jobs
2) persistent	2) Terminated at end of job
3) suitable for interactive workloads	3) suitable for automated workloads
4) shared among many users	4) Isolated just for job
5) Expensive to run	5) cheaper to run

* cluster ⇒ group of virtual machines working together
* If capacity of vm is high, then time taken will be less.
* Configuring cluster :- giving components to create a cluster
* Azure databricks has many operations as compared to open source databricks.

Cluster Configuration

- single node
- multi node

~~cluster~~

Access mode

1) Single user - only one user access, supports Pyth
Scala, SQL, R

2) Shared
- Multi user access
- available in Premium, supports Python,

3) No Isolation shared
- multiuser access
- supports Python, SQL

4) Custom → Legacy configuration