

- Databricks hosted on cloud platform (azure)
- Databricks hosted on opensource -- Databricks.com.
- PySpark - Databricks (opensource)
- Azure + databricks → cloud + Databricks → cloud databricks.
- Databricks hosted on cloud platform Azure.

### Cloud Services:-

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

### Core Cloud Services

Azure compute option:- (Virtual machines, Containers, Azure App Service, Serverless Computing)

Azure data storage options:- (Azure SQL Database, Azure Cosmos DB, Azure Blob storage, Azure Data Lake storage, Azure Files, Azure Queue, Disk storage)

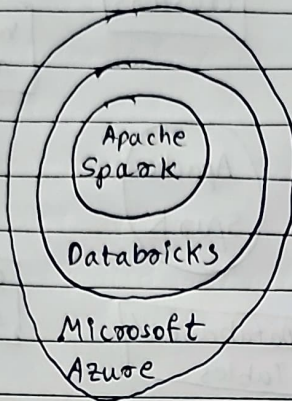
Azure networking options.

Azure IoT.

### What is Microsoft Azure?

- Azure is Microsoft's cloud computing platform
- Azure is a continually expanding set of cloud services that help your organization meet your current and future business challenges.
- Azure give you the freedom to build, manage, and deploy applications on a massive global network using your favourite tools and frameworks

### Azure Databricks:-

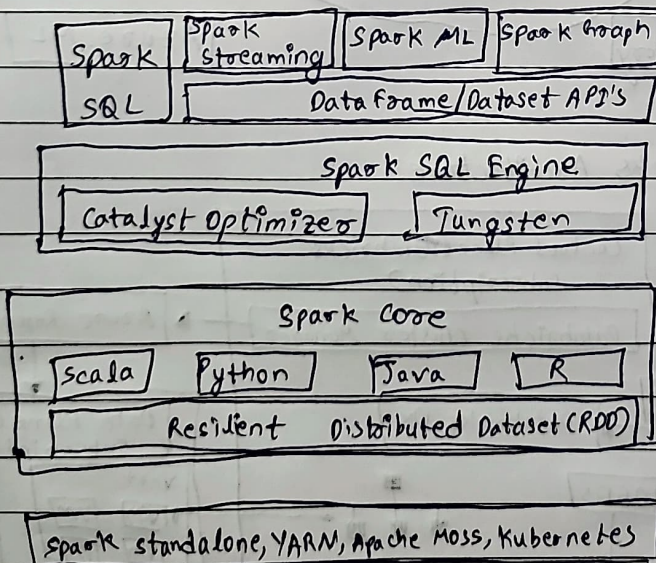


### Apache Spark:-

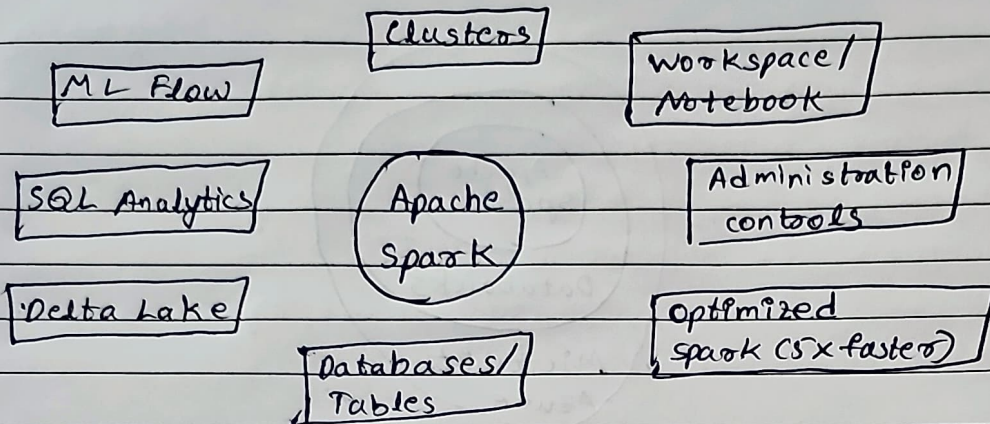
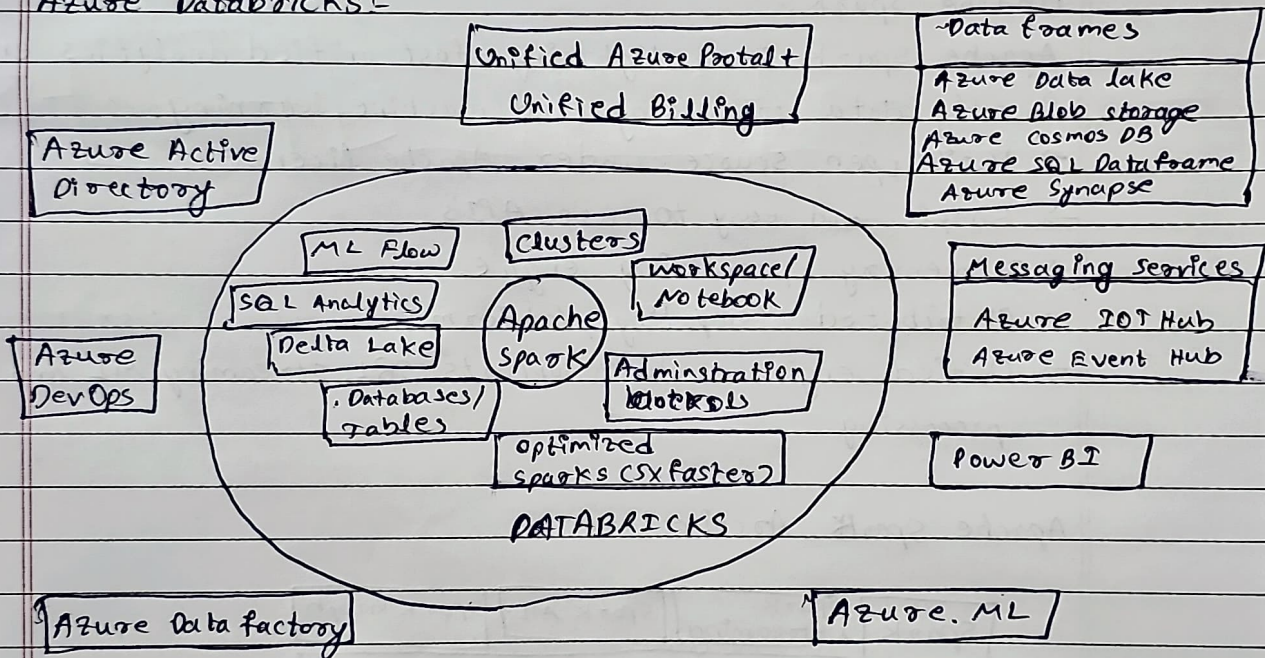
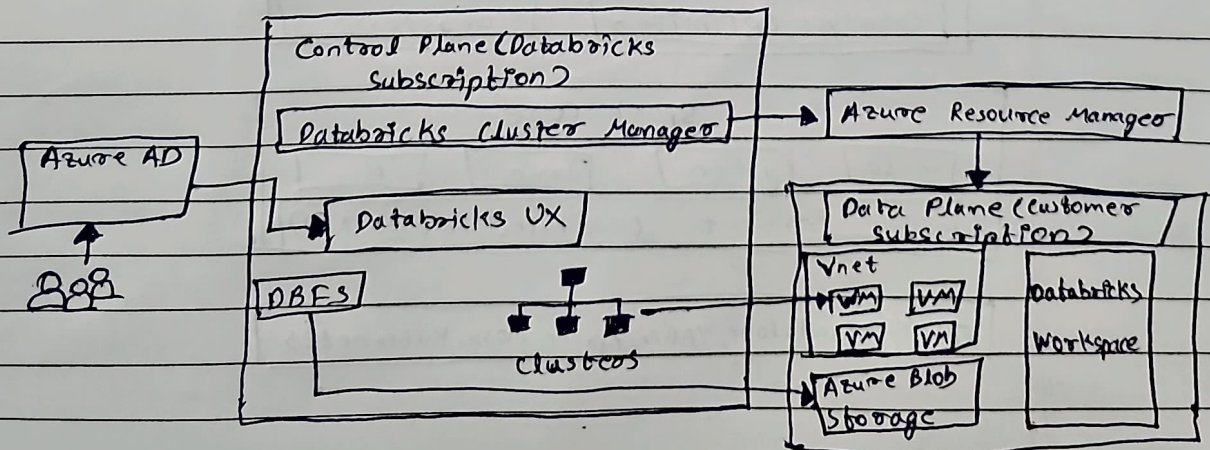
Apache Spark is a lightning-fast unified analytics engine for big data processing and machine learning.

- 100% open source under Apache license
- Simple and easy to use API's.
- In-memory processing engine.
- Distributed computing platform.
- Unified engine which supports SQL, streaming, ML and graph processing.

### Apache Spark Architecture:-





Databricks:-Azure Databricks:-Azure Databricks Architecture:-

Databricks Workspace Components.

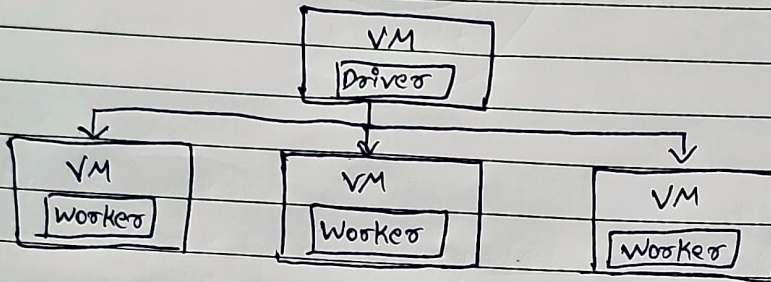
Notebooks

Data

Clusters

Jobs

Models

Databricks ClusterCluster Types.

All purpose	Job Cluster
Created manually	Created by Jobs
Persistent	Terminated at the end of the job
Suitable for interactive workloads	Suitable for automated workloads
Shared among many users.	Isolated just for the job
Expensive to run	Cheaper to run