Databricks

- → Databricks hosted on cloud Platform (Azure)
- → Databricks hosted on open source → Databricks.com
- → Pyspark - Databricks (open source)
- → 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)

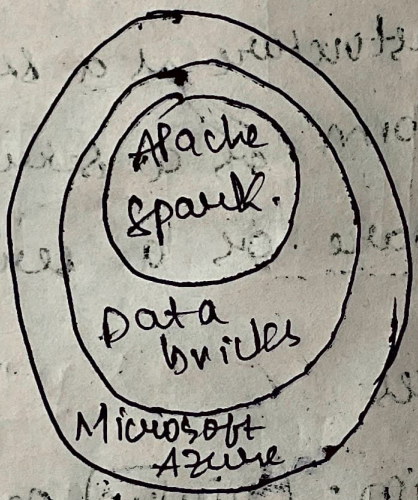
Azure data storage options:-

(Azure SQL Database, Azure cosmo DB, Azure Blob storage, Azure Data Lake storage, Azure files, Azure queue)

Microsoft Azure:-

→ It is cloud computing platform
→ It is continually expanding set of cloud services that helps your organization
→ It gives you the freedom to build, manage and deploy applications.

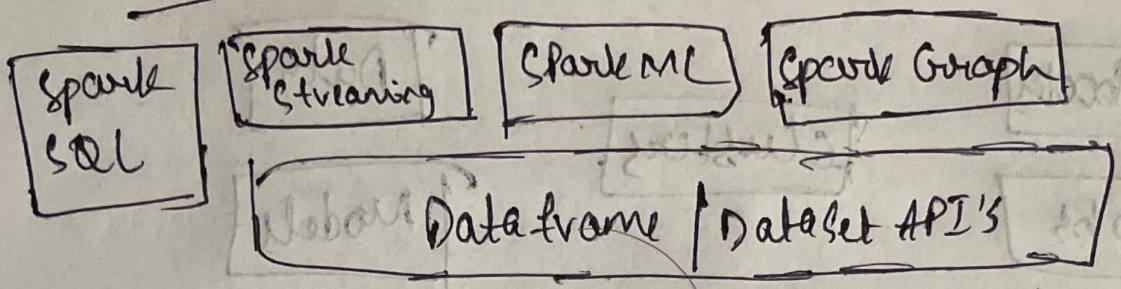
~~Apache~~ Azure Databricks:-

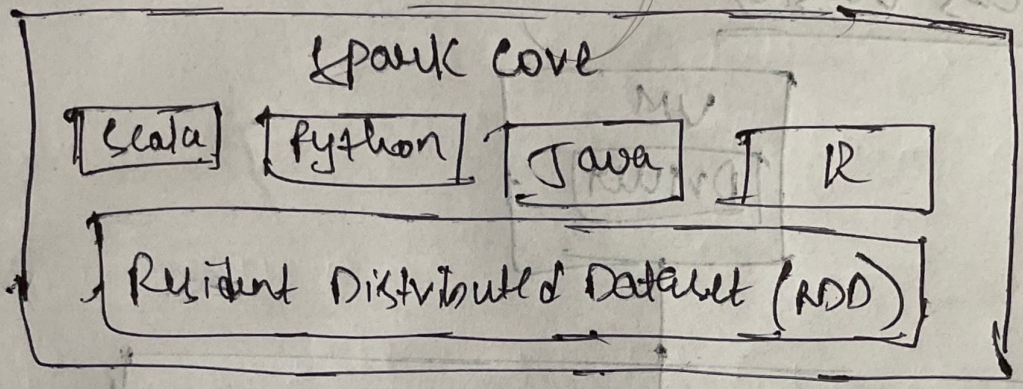


~~Apache~~ Spark:-

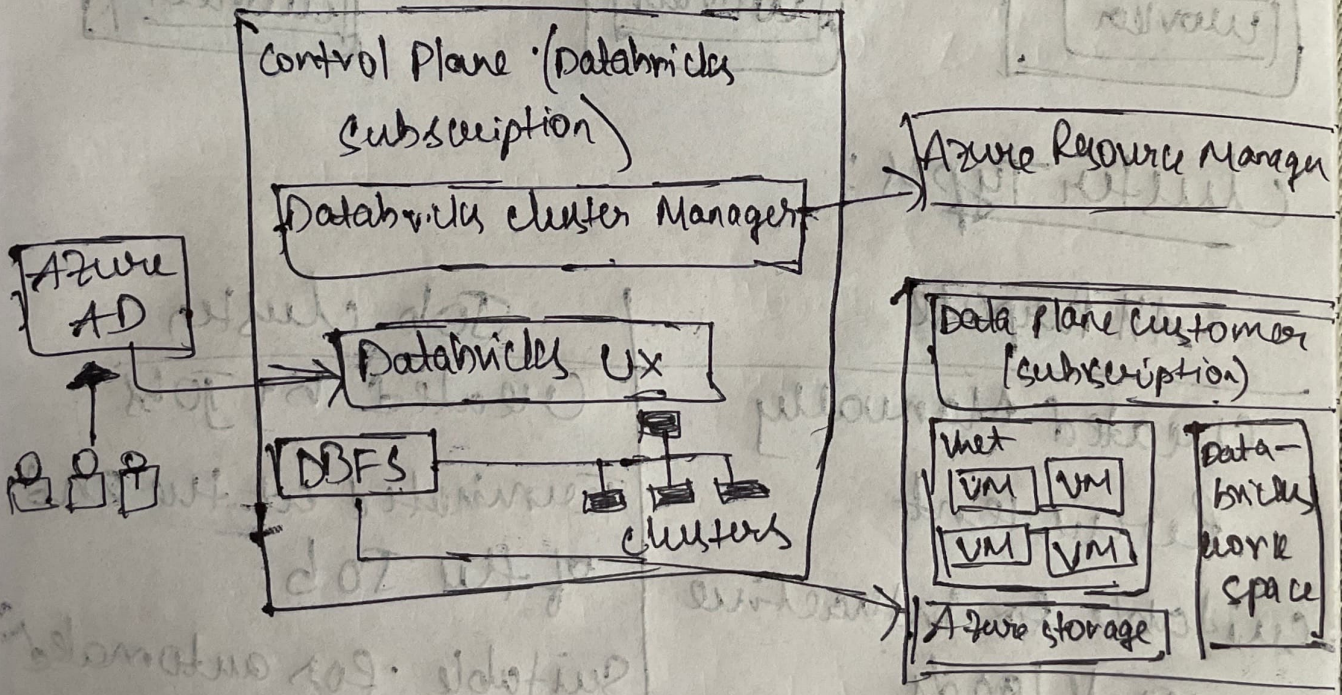
→ It is a lightning-fast unified analytics engine for big data.
→ It is open source under Apache License
→ Simple & easy to use API's
→ In-memory Processing
→ Distributed computing platform

Apache Spark Architecture:

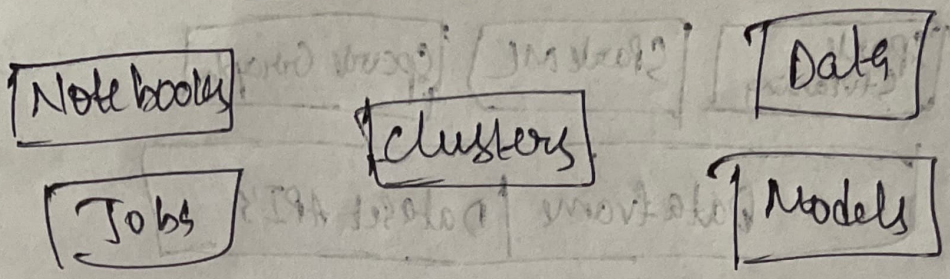




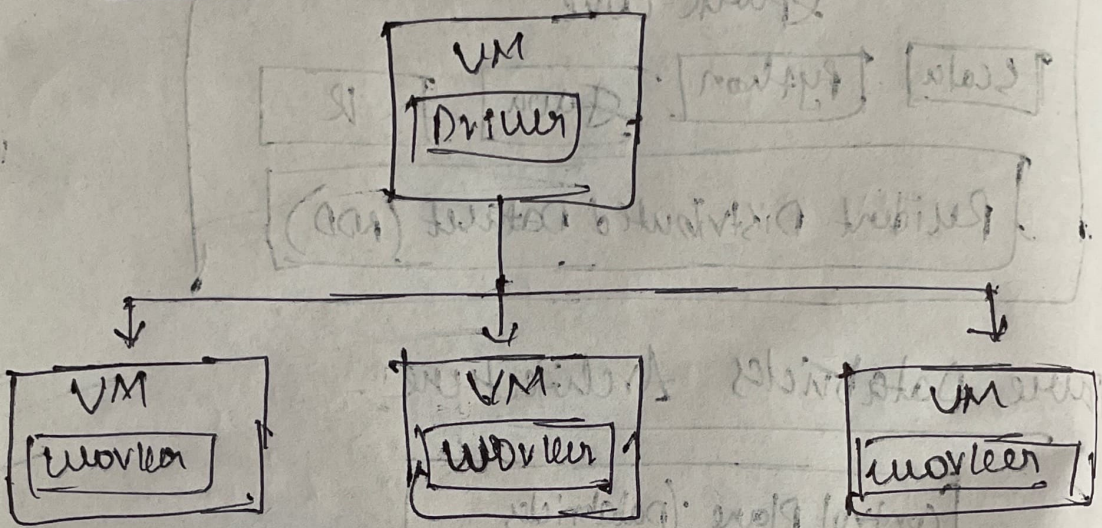
Azure Databricks Architecture:-



Databricks workspace Components



Databricks cluster



Cluster 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	Alloted just for the Job
Expensive to run	Cheaper to run