

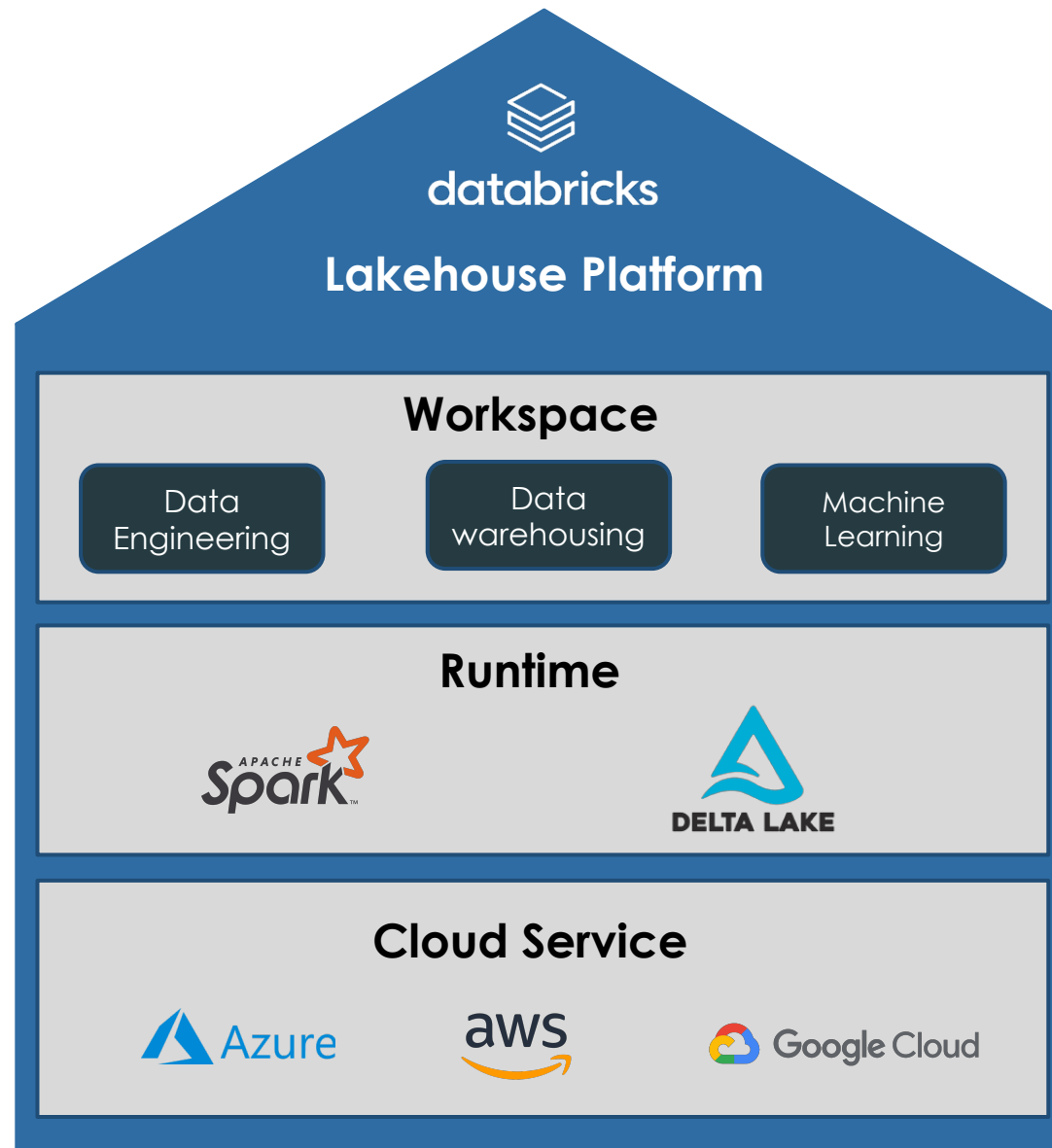
What is Databricks

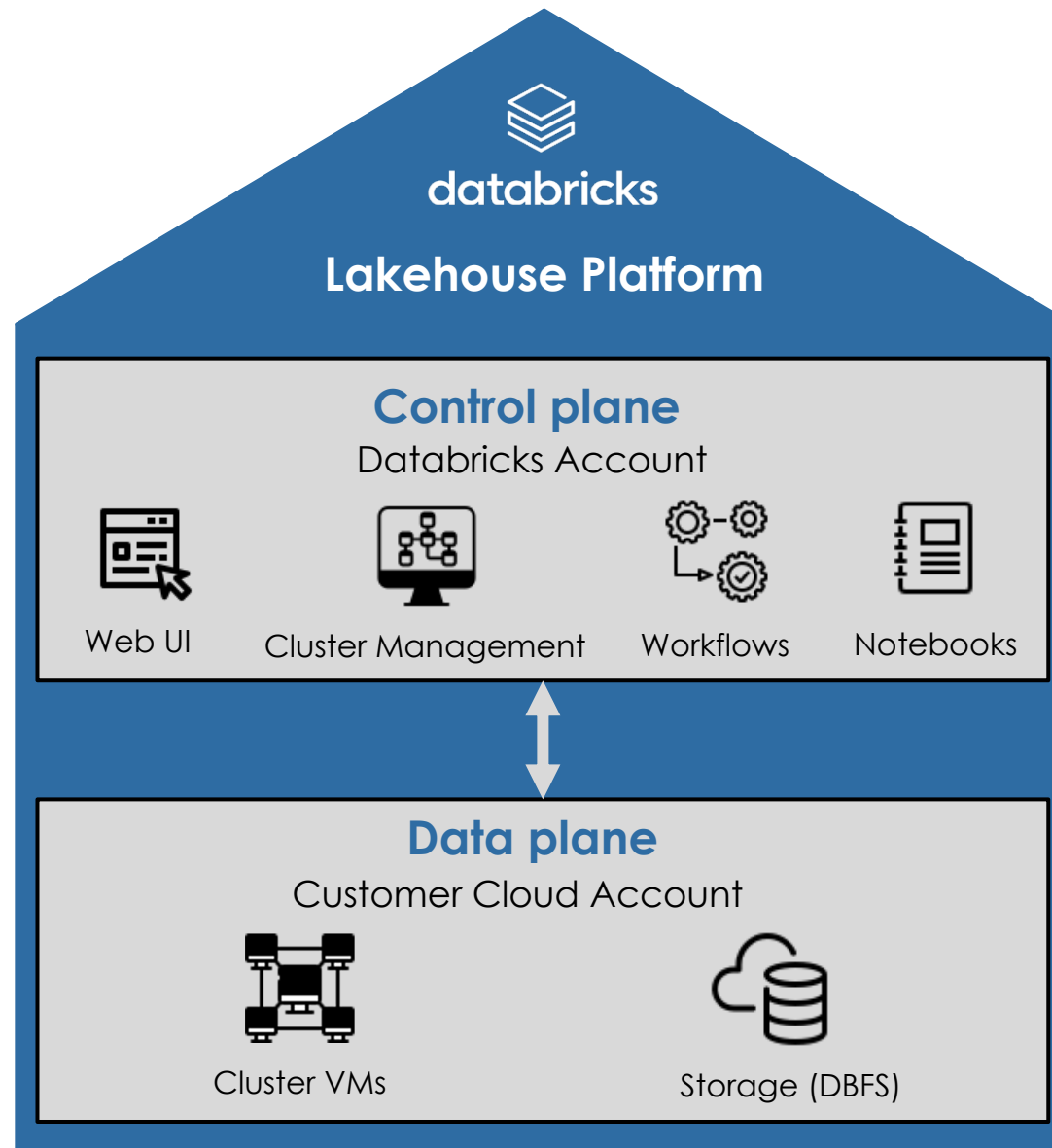
Databricks

► Multi-cloud Lakehouse Platform
based on Apache Spark

Lakehouse







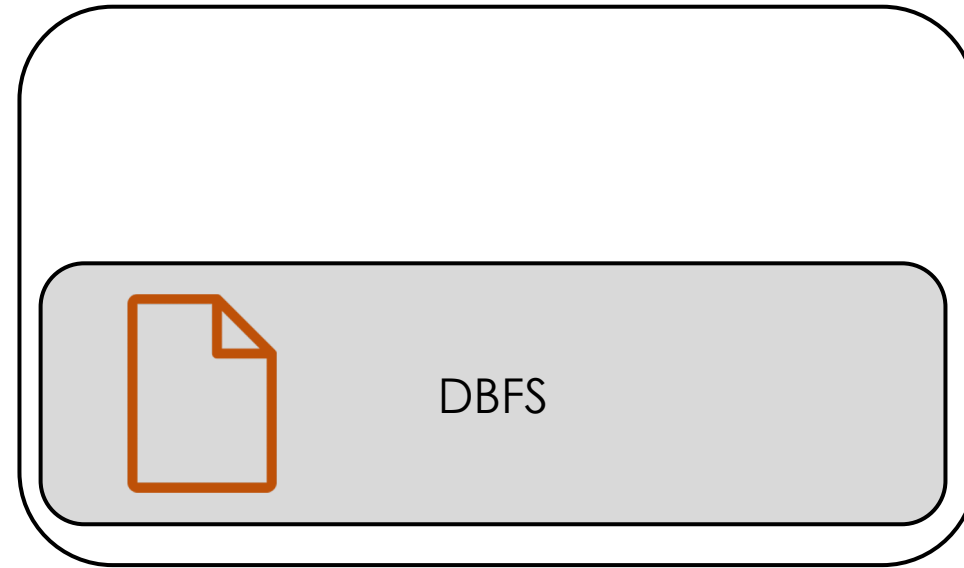
Spark on Databricks

- ▶ In-memory, distributed data processing
- ▶ All languages support
 - ▶ Scala, Python, SQL, R, & Java
- ▶ Batch processing & stream processing
- ▶ Structured, semi structured, & unstructured data

Databricks File System (DBFS)

- ▶ Distributed file system
- ▶ Preinstalled in Databricks clusters
- ▶ Abstraction layer
 - ▶ data persisted to the underlying cloud storage

Cluster



**Cloud
Storage**

