What is Databricks

Databricks

Multi-cloud Lakehouse Platform based on Apache Spark

Lakehouse

Data Lake

- Open
- Flexible
- ML support

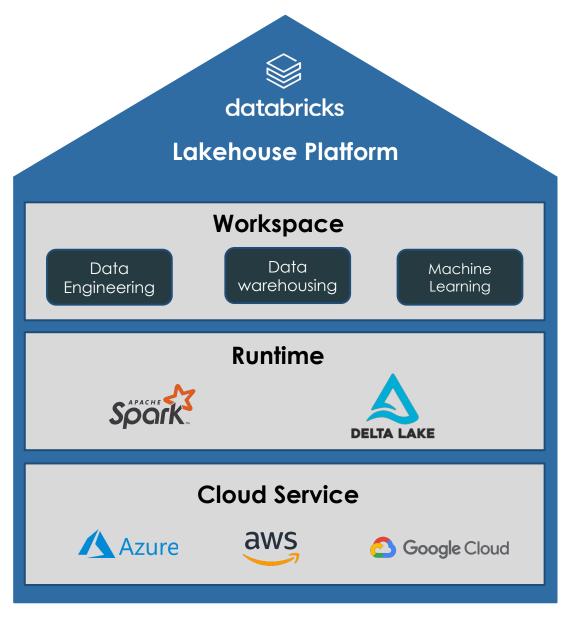
Lakehouse

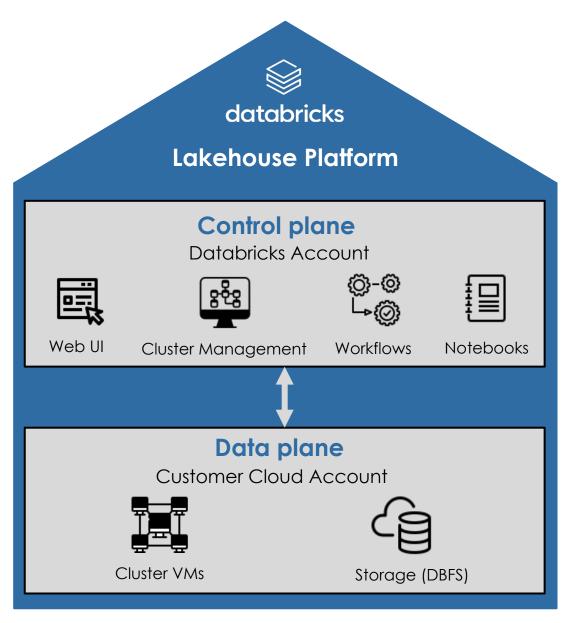
One platform that unify all of your data engineering, analytics, and Al workloads

Data warehouse

- Reliable
- Strong governance
- Performance

Derar Alhussein © Udemy | Databricks Certified Data Engineer Associate - Preparation





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 underlaying cloud storage

