

What is Databricks

Databricks

- ▶ Multi-cloud Lakehouse Platform
based on Apache Spark

Lakehouse





databricks

Lakehouse Platform

Workspace

Data
Engineering

Data
warehousing

Machine
Learning

Runtime



Cloud Service





databricks

Lakehouse Platform

Control plane

Databricks Account



Web UI



Cluster Management



Workflows



Notebooks



Data plane

Customer Cloud Account



Cluster VMs



Storage (DBFS)

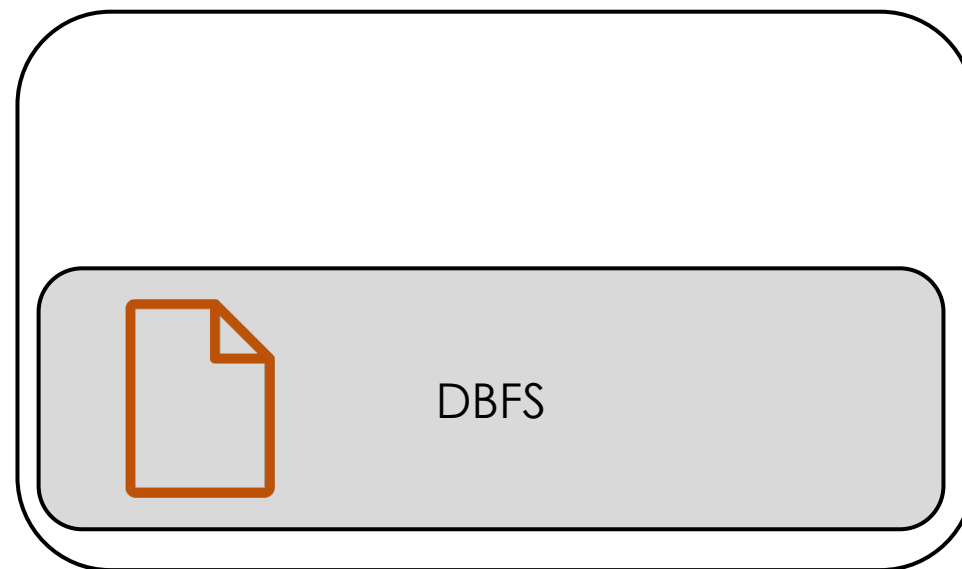
Spark on Databricks

- ▶ 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**

