

Sistema de Arquivos

- Um sistema de arquivos é uma maneira de organizar, armazenar e nomear dados em um dispositivo de armazenamento, como um disco rígido, SSD, cartão de memória ou pen drive. Ele controla como os arquivos são colocados no armazenamento e como são recuperados.



Sistemas de Arquivos Distribuidos

- Sistema de arquivos que permite que o acesso a arquivos e dados seja transparente, independentemente de onde os dados estejam fisicamente armazenados.
- Projetado para **armazenar e analisar grandes volumes de dados de maneira distribuída e paralela em um cluster de computadores**. Ele divide os arquivos em blocos, distribui-os em todo o cluster e mantém cópias redundantes para garantir a confiabilidade e a disponibilidade dos dados.
- Exemplos:
 - S3
 - Google File System
 - GlusterFS
 - HDFS



DBFS (Databricks File System)

- Abstração sobre o armazenamento de objetos na nuvem, como o Amazon S3 ou o Azure Blob Storage
- **Permite que você armazene dados como se estivesse usando um sistema de arquivos distribuídos local**, mas com a durabilidade e escalabilidade da infraestrutura de armazenamento em nuvem
- Para **tarefas de processamento de dados**, o Databricks utiliza o Apache Spark
- Também fornece **uma camada de armazenamento transacional ACID para aprimorar a confiabilidade e o desempenho do sistema de arquivos (Delta Lake)**

