



The background of the slide features a dark, atmospheric photograph of server racks in a data center, with blue light reflecting off the equipment. Overlaid on this is a faint, light-gray network diagram. The diagram consists of several circular nodes connected by thin lines. The nodes contain various icons: a cloud, a smartphone, a laptop, a server tower, a database cylinder, and a camera lens. The central part of the diagram is the largest and most prominent.

HDFS

Hadoop Distributed File System

O que é um sistema de Arquivos?

- ◆ Faz o gerenciamento de arquivos em disco:
 - ◆ Mantém integridade
 - ◆ Segurança
 - ◆ Privacidade
 - ◆ Metadados



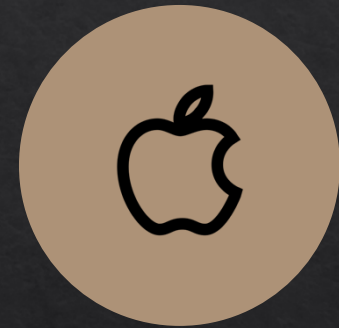
Sistemas de Arquivos



WINDOWS: FAT,
NTFS



LINUX: EXT2, EXT3,
EXT4, XFS, JFS



MACOS: APFS, HFS
PLUS

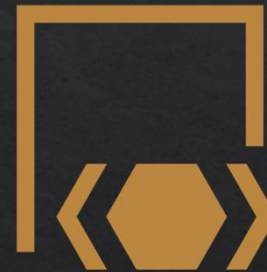
Porque não usar NTFS ou Ext3?

- ◇ Porque o Hadoop precisa de um gerenciamento com características diferenciadas:
 - ◇ Arquivo separado em blocos
 - ◇ Distribuídos em nós de redes
 - ◇ Cópias replicadas

HDFS



Armazena dados em blocos



Replicação transparente
(default 3 nós)

HDFS

◈ Hadoop Distributed File System: Sistema de Arquivos Distribuídos do Hadoop



Tipos de Arquivos

↔ Texto:

Padrão em ferramentas como Hive

📁 Sequence File:

Chave-valor binário
Podem ser divididos ou unidos facilmente

📄 AVRO

Formato binário para serialização
Ótimo para troca de dados

📁 ORC

Colunar otimizado para consultas de linhas
Formato "favorito" do ecossistema Hadoop

📄 RC

Orientado a coluna, chave-valor
Alta taxa de compressão em linha

📄 Parquet

Orientado a colunas
Binário