

02.Apache Hadoop

quarta-feira, 19 de julho de 2023

15:02

- É um software para gerenciar como serão armazenados os arquivos no seu servidor/Cluster
- O sistema de arquivos de um computador não foi desenvolvido para armazenamento distribuído.
- Hadoop HDFS (Hadoop Distributed File System) - Serve para indicar o gerenciamento dos arquivos, podendo armazenar os arquivos em mais de 1 máquina no cluster.
- Existem outros sistemas de arquivos além do HDFS
- É Open source
- Hardware de baixo custo
- Define como serão as distribuições dos arquivos para o cluster
- Com o HDFS podemos construir um Data Lake para rodar no cluster de computadores permitindo o armazenamento de grandes volumes de dados com hardware de baixo custo (Commodity)
- Permitindo que o Big Data funcionasse em larga escala