



# Data Science Academy

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

## Data Lake – Design, Projeto e Integração

### Vantagens e Desvantagens do Apache Hadoop



Vejamos algumas vantagens e desvantagens do Apache Hadoop.

## Vantagens

- Escalabilidade – O Hadoop pode tratar grandes volumes e variedades de dados e podemos adicionar mais nós ao cluster, garantindo assim escalabilidade horizontal e em tempo real.
- Flexibilidade – Não há esquema definido e não precisamos limpar ou organizar os dados antes do armazenamento.
- Performance – Mesmo com grandes volumes de dados a performance do Hadoop é um dos seus diferenciais.
- Baixo-custo – O Hadoop é gratuito e pode ser executado em máquinas de baixo custo, tornando a implementação um projeto de baixo custo. A maior dificuldade está em encontrar profissionais qualificados.
- Análises Avançadas – Como o Hadoop oferece alta capacidade computacional devido a execução paralela em clusters, pode ser usado para análises avançadas e complexas com muitas operações matemáticas.
- Ecossistema – O Hadoop possui um grande ecossistema de produtos que permitem construir uma arquitetura única para análise de Big Data.

## Desvantagens

Embora as vantagens superem as desvantagens, o Hadoop também possui problemas. Com os dados são armazenados em blocos, tarefas de consulta terão que buscar todos os blocos, para montar o arquivo, tornando o acesso aos dados problemático.

Alguns algoritmos analíticos também podem não executar de forma apropriada no Hadoop, uma vez que podem requerer uso de instruções específicas da CPU.



A segurança é outro ponto crítico no Hadoop, que vem ganhando cada vez mais atenção dos desenvolvedores do framework.