



PUC Minas

Funcionamento HDFS

Diego Roberto Gonçalves de Pontes

HDFS (Hadoop Distributed File System)

- O HDFS, ou Hadoop Distributed File System, é um sistema de arquivos distribuído projetado para rodar em hardware comum. Ele é uma parte essencial do Apache Hadoop, uma plataforma de código aberto usada para o processamento distribuído de grandes conjuntos de dados. O HDFS é altamente escalável e capaz de gerenciar grandes volumes de dados, proporcionando armazenamento confiável e rápido acesso aos dados em clusters de computadores.

Principais Características do HDFS

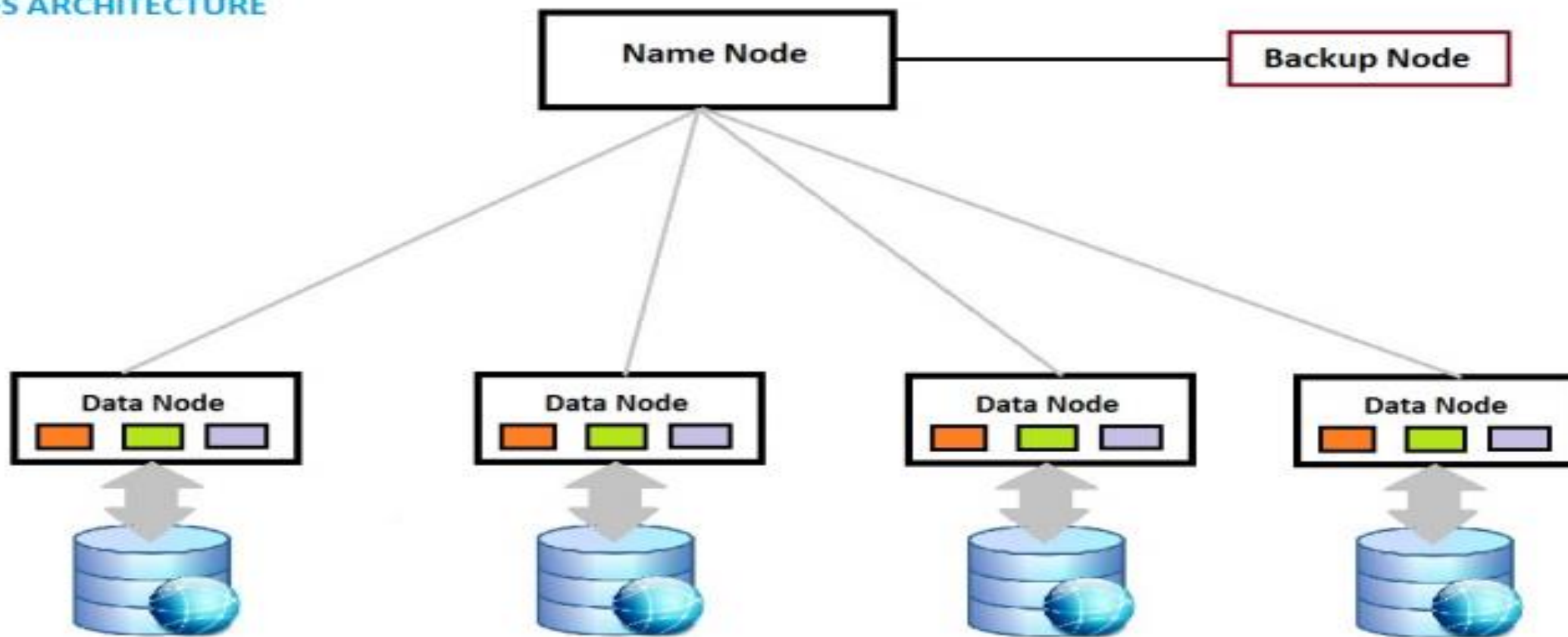
- **Arquitetura Mestre-Escravo:** O HDFS segue uma arquitetura mestre-escravo onde o NameNode atua como o servidor mestre que gerencia o namespace do sistema de arquivos e regula o acesso aos arquivos por clientes. Os DataNodes são os servidores escravos que armazenam efetivamente os dados.
- **Alta Tolerância a Falhas:** O HDFS foi projetado para ser altamente tolerante a falhas. Ele replica dados em múltiplos DataNodes, garantindo que, mesmo se alguns nós falharem, os dados ainda estarão disponíveis.
- **Manipulação de Grandes Arquivos:** O HDFS é otimizado para manipular arquivos muito grandes, frequentemente de gigabytes a terabytes de tamanho. Ele divide cada arquivo em blocos e os distribui através de diferentes nós do cluster.

Principais Características do HDFS

- Desempenho: Ao distribuir e replicar dados entre múltiplos DataNodes, o HDFS garante alta taxa de transferência de dados e pode ser dimensionado horizontalmente para adicionar mais capacidade de armazenamento e processamento.
- Escalabilidade: O HDFS pode ser facilmente escalado adicionando mais nós ao cluster, sem a necessidade de reconfigurar o sistema ou parar o serviço.

Funcionamento

HDFS ARCHITECTURE



Fonte: <https://rogeraoaraujo.com.br/2020/11/12/big-data-ferramentas-e-tecnologias-parte-2/>



PUC Minas