



Data Science Academy

www.datascienceacademy.com.br

Data Lake – Design, Projeto e Integração

Experimentando o Data Lake



Parabéns! Implementar um Data Lake em um Cluster Hadoop não é uma tarefa simples e são poucos os profissionais habilitados a fazê-lo, uma vez que implementar um cluster é muito mais do que apenas executar comandos. É necessário conhecimento sobre sistemas operacionais, ambiente de rede, conexão entre servidores e o próprio Hadoop em si. Chegar até aqui é uma conquista, mas ainda estamos no começo da caminhada na Formação Engenheiro de Dados.

Agora é com você. Experimente o Data Lake. Estude os arquivos de configuração, carregue mais dados no cluster, analise os relatórios de uso de espaço em disco e de blocos, altere os parâmetros de configuração do Hadoop, “derrube” um dos Datanodes e teste o Data Lake. Faça um backup das suas VMs antes de experimentar seu cluster.

Voltaremos a construir um cluster completo, com mais camadas, quando construirmos o Data Lake em nuvem!