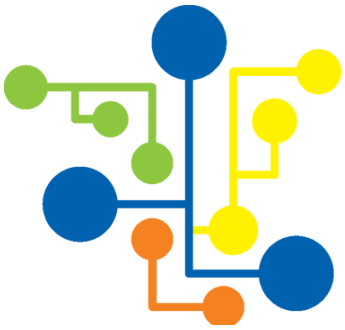

Data Science Academy

www.datascienceacademy.com.br

Data Lake – Design, Projeto e Integração

Lab 1
Implementando a Camada de Processamento no Data Lake com Apache Spark

Neste Lab você vai praticar o que aprendeu até aqui e implementar um cluster com Apache Spark.

Enquanto o Apache Hadoop HDFS é usado para o armazenamento distribuído, o Apache Spark pode ser usado para o processamento distribuído no cluster. O processo de configuração é muito similar ao que você fez com o HDFS.

No arquivo em anexo, você encontra os passos necessários para instalar, configurar e testar o cluster Spark. Reproduza os passos cuidadosamente no seus cluster criado nas aulas anteriores. O Apache Spark é executado sobre o Apache Hadoop HDFS.

Caso tenha com dúvida ou dificuldade, poste sua mensagem no fórum ou envie mensagem para suporte@datascienceacademy.com.br.

Let's do it!