

Desvendando Data Science em Sistemas Distribuídos com Apache Spark

Alexandre Castro¹, Letícia Moura², Matheus Serpa³

¹Instituto de Engenharia de Dados – Faculdade Infnet
20.771-260 – Rio de Janeiro – RJ – Brazil

²Centro Universitário Serra dos Órgãos (UNIFESO)
25.964-004 – Teresópolis – RJ – Brazil

³Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

alexandre.dcastro@al.infnet.edu.br
leticiamwork@gmail.com
msserpa@inf.ufrgs.br

Resumo. Na era atual da tecnologia, onde dados são gerados e consumidos em uma escala sem precedentes, a Ciência de Dados emergiu como um campo crucial para transformar esses dados em insights valiosos. Com a crescente complexidade e o aumento do volume de dados, torna-se essencial empregar técnicas eficientes em sistemas distribuídos para lidar com esses desafios. O Apache Spark, uma ferramenta poderosa para o processamento de dados em larga escala, desempenha um papel vital nesse cenário, permitindo a análise e o processamento de dados de maneira rápida e eficiente. Neste minicurso, os participantes serão imersos no mundo da Ciência de Dados aplicada a sistemas distribuídos. O curso abordará desde conceitos fundamentais de Data Science e suas aplicações práticas até o uso do Apache Spark. Será dada ênfase especial à capacidade do Spark de realizar processamento de dados em ambientes de alto desempenho. Os alunos aprenderão como o Apache Spark facilita o trabalho com grandes volumes de dados, além de explorar suas funcionalidades em análises de dados e machine learning. Um dos destaques do curso será o desenvolvimento de um projeto de Data Science end-to-end, que envolverá desde a extração dos dados e transformação até o treinamento de um modelo de Machine Learning, deployment e integração via frontend. Este curso oferece uma oportunidade única para profissionais e entusiastas da área entenderem como maximizar o potencial da Ciência de Dados em sistemas distribuídos, utilizando uma das ferramentas mais influentes da atualidade.