Spark com Scala para Programadores

Tecnologias de Informação - Desenvolvimento / Programação

Live Training (também disponível em presencial)

- **Localidade:** Aveiro
- **Data:** 09 Oct 2023
- **Preço:** 1815 € (Os valores apresentados não incluem IVA. Oferta de IVA a particulares e estudantes.)
- **Horário:** Laboral das 09h30 - 16h30
- **Nível:**
- **Duração:** 33h

Sobre o curso

Big Data é um tema central na abordagem aos dados, processamento e tratamento da informação em larga escala, na atual realidade.

Organizações só conseguirão ser verdadeiramente competitivas e bem sucedidas se reunirem, analisarem e visualizarem a informação das múltiplas variáveis dos seus produtos, serviços e mercados onde operam por forma a responderem de forma eficaz às necessidades de cada momento.

Organizações como a Google, Amazon, NASA ou Yahoo, empregam a utilização de Clusters Spark para extrair informação de bases de dados massivas.

Spark é uma ferramenta de Big Data, uma das tecnologias mais modernas para processar grandes conjuntos de dados de forma paralela e distribuída. Embora possam ser utilizadas várias linguagens de programação como o Python, R ou Java, **Scala é a linguagem de eleição** para programação em Spark.

Numa ótica de sessões mistas de teoria e de prática, neste curso, serão abordados conhecimentos fundamentais para a programação com a linguagem Scala e o framework Spark.

Objetivos:

No final da acção de formação os participantes deverão estar aptos a:

- Desenvolver código distribuído utilizando o a linguagem Scala
- Instalar e configurar um Standalone Apache Spark 3.0 Cluster.

- Analisar BigData e extrair informação útil.
- Aplicar transformações a bases de dados e data frames
- Aplicar analises gráficas usando GraphX
- Monitorizar o funcionamento de um cluster spark

Destinatários

- Engenheiros de software
- Programadores
- Analistas de dados
- Interessados em ampliar conhecimentos de BigData

Pré-requisitos

- Conhecimentos de Inglês
- Conhecimentos de uma linguagem orientada por objectos: Java, C#, Python, MatLab, ou outra
- Conhecimentos de estatística simples
- Conhecimentos de Linux

Metodologia

Este curso é constituído por módulos de formação online síncronos (em tempo real em sala virtual) integrados numa ótica de sessões mistas de teoria e de prática. Durante todo o curso, os formandos terão acompanhamento permanente e personalizado por parte do formador.

Serão elaborados projetos práticos em grupo, garantindo uma aprendizagem mais eficaz.

Composição:

- 21 horas (formação com formador em sala virtual)
- 12 horas (desenvolvimento de projectos práticos em grupo com moderação do formador)

Programa

- Implementação de um standalone Apache Spark Server
- Instrução ao Scala
- Exemplos básicos em Spark com Scala

- Exemplos avançados
- SparkSQL, DataFrames, DataSets
- Machine Learning com MLLib
- Introdução ao GraphX

Implementação de um standalone Apache Spark Server

- Criação de virtualizações
- Instalação do Apache Spark 3
- Configuração de Master VM’s e Slaves VM’s
- Monitorização de recursos
- Diagnóstico de falhas

Instrução ao Scala

- Scala crash-course
- Controle de fluxo
- Funções em Scala
- Estruturas de dados

Exemplos básicos em Spark com Scala

- Introdução ao Spark
- Resilient Distributed Dataset – RRD
- Funcionamento interior do Spark
- Desenvolvimento de exemplos em Scala

Exemplos avançados

- Exemplos de utilização em média social
- Exemplos de utilização em dados de automação (IoT)
- Exemplos de utilização em documentos

SparkSQL, DataFrames, DataSets

- Introdução ao SparkSQL
- Utilização de DataFrames
- Utilização de DataSets

Machine Learning com MLLib

- Introdução ao MLLib
- Exemplo de regressão linear com MLLib
- Utilização de DataFrames com MLLib

Introdução ao GraphX

- Criação de gráficos do GraphX
- Exemplo de aplicação do GraphX