escola britânica de artes criativas & tecnologia

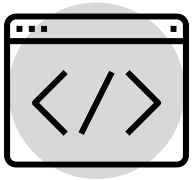
Profissão: Analista de dados

escola britânica de artes criativas & tecnologia

DE OLHO NO CÓDIGO

Big Data I Processamento

- **Entenda o Apache Spark**



Confira boas práticas sobre Big Data por assunto relacionado às aulas.

Entenda o Apache Spark

- **Compreenda a arquitetura de processamento distribuído do Spark**: Para obter o máximo desempenho do Spark, é importante entender como ele processa dados em um cluster de máquinas. O Spark executa operações em paralelo em várias máquinas, usando memória compartilhada para reduzir o tempo de leitura e gravação em disco.

- **Use o Spark em conjunto com outras ferramentas de big data**: O Spark é uma ferramenta poderosa, mas não é a única disponível para trabalhar com big data. Use-o em conjunto com outras ferramentas, como Hadoop, Hive e HBase, para obter os melhores resultados possíveis.

Entenda o Apache Spark

- **Otimize suas operações**: O Spark oferece muitas maneiras de otimizar suas operações. Use técnicas como particionamento de dados, cache e uso de algoritmos de processamento específicos para obter os melhores resultados.

- **Use a linguagem Scala para desenvolver suas aplicações**: O Spark foi desenvolvido originalmente em Scala e é a linguagem de programação mais usada pelos desenvolvedores do Spark. Dominar Scala é uma vantagem para trabalhar com Spark, pois permite escrever código mais eficiente e eficaz.

Entenda o Apache Spark

- **Use notebooks interativos para explorar e prototipar**: Notebooks interativos, como Jupyter e Zeppelin, são uma excelente ferramenta para explorar e prototipar suas soluções de big data com Spark. Eles permitem visualizar os dados e os resultados dos algoritmos, além de poder compartilhar seu trabalho com outras pessoas.

- **Utilize o Spark SQL**: O Spark SQL é uma extensão do Spark que permite consultar dados estruturados usando a linguagem SQL. Use-o para executar consultas SQL em dados distribuídos e para criar pipelines de processamento de dados mais complexos.

Entenda o Apache Spark

- **Aprenda a lidar com dados não estruturados**: O Spark é uma ótima ferramenta para processar dados estruturados, mas também pode ser usado para trabalhar com dados não estruturados, como arquivos de texto e JSON. Aprenda a usar o Spark para processar esses tipos de dados e aproveite ao máximo suas capacidades.
- **Use bibliotecas e frameworks complementares**: Existem muitas bibliotecas e frameworks complementares disponíveis para trabalhar com Spark, como o MLlib para machine learning e o GraphX para processamento de gráficos. Use-os para estender as capacidades do Spark e criar soluções de big data mais avançadas.

Entenda o Apache Spark

- **Faça ajustes finos nas configurações do Spark**: Para obter o máximo desempenho do Spark, é importante fazer ajustes finos nas configurações de cluster e em outras opções de configuração. Isso pode ser feito por meio de arquivos de configuração ou por meio de variáveis de ambiente.

- **Utilize práticas de segurança**: O Spark pode manipular grandes volumes de dados, alguns dos quais podem ser sensíveis ou confidenciais. Utilize práticas de segurança, como criptografia de dados em repouso e em trânsito, controle de acesso e autenticação, para proteger seus dados.

escola britânica de artes criativas & tecnologia

Bons estudos!