

Big Data for Developers

Aprenda a acelerar a Integração de Big Data por meio da ingestão em massa, transformações e processamento de arquivos complexos. Otimize o desempenho do sistema Big Data por meio de monitoramento, solução de problemas e práticas recomendadas. Este curso é aplicável a usuários do Big Data versão 10.2.1 e mais recentes.



DURAÇÃO

3 dias
(24 horas)



PÚBLICO ALVO

Desenvolvedores, Arquitetos, Administradores, Gestores e Analistas envolvidos e/ou responsáveis por projetos de Big Data.



PRÉ-REQUISITOS

Treinamento 'Informatica Developer Tool for Big Data Developers'

AGENDA

Módulo 0: Introdução ao curso Big Data

- Visão Geral do curso

Módulo 1: Big Data Management Basics

- BDM - Conceitos
- BDM - Recursos e Benefícios
- BDM - Arquitetura
- BDM - Tarefas do Desenvolvedor

Módulo 2: Ingestão e Extração

- Serviços de Aplicação do BDM 10.2.1
- Serviço de Acesso a Metadados
- Serviço de Ingestão em Massa
- Sistemas de arquivos do Hadoop
- Arquitetura de Ingestão em Massa
- Processo de Ingestão em Massa
- Interface do usuário da ferramenta de Ingestão em Massa
- Ingestão em massa para o HDFS
- Ingestão em massa para o Hive
- Integrando com bancos de dados relacionais usando o SQOOP
- SQOOP - Arquitetura
- SQOOP - Otimizações
- SQOOP para Teradata

Módulo 3: Estratégia do mecanismo do Big Data

- Descrever a estratégia do mecanismo BDM
- Hive - Discussão
- Hive - Arquitetura do mecanismo
- MapReduce
- Tez
- Spark - Arquitetura
- Spark - Suporte na versão 10.2.1
- Blaze - Arquitetura
- Passos para escolher o motor certo
- Transformações no Hadoop
- Transformação - Expression
- Transformação - Filter
- Transformação - Lookup
- Transformação - Python
- Transformação - Router
- Transformação - Union
- Transformação - Update Strategy

Módulo 4: O processo de desenvolvimento no Big Data

- Carga inicial
- Cargas Incrementais
- Ingestão de Mapeamento Dinâmico
- Computação Stateful e windowing
- Integração da ciência de dados usando Python

Módulo 5: Processamento de Arquivos Complexos

- Formatos de arquivo Big Data
- Objetos de dados de arquivos complexos
- Processando tipos de arquivos complexos
- Matrizes
- Estruturas
- Mapas
- Descoberta de estrutura inteligente

Módulo 6: Monitoramento e Solução de Problemas

- Spark - Monitoramento
- Blaze - Monitoramento
- Visualizar logs
- Solução de problemas

Módulo 7: Performance & Tuning e Melhores Práticas

- Diferenciar modo de execução Native Vs Hadoop
- Ajuste de desempenho de trabalhos do Spark
- Ajustando o desempenho de jobs do Blaze
- Melhores práticas para diferentes mecanismos