

Desenvolvemento Big Data Cloudera para Spark e Hadoop - Virtual

Descrición

Este curso ofrece os conceptos clave e a experiencia que os participantes necesitan para introducir e procesar datos nun clúster Hadoop utilizando as ferramentas e técnicas máis actuais tales como Spark 2 (incluíndo Spark Streaming e Spark SQL), Flume, Kafka e Sqoop.

Este curso proporciona a mellor preparación para os desafíos do mundo real que afrontan os desenvolvedores de Hadoop. Con Spark 2, os desenvolvedores poden escribir sofisticadas aplicacións en paralelo para tomar mellores decisións, máis rápidas e accións interactivas, aplicadas a unha ampla variedade de casos de uso, arquitecturas e industrias.

Obxectivos

Á finalización da formación, o participante coñecerá:

- Uso do Spark Shell para análise interactiva de datos
- As características dos conxuntos de datos distribuídos resilientes de Spark
- Como executa Spark nun clúster
- Programación paralela con Spark
- Creación de aplicacións Spark
- Procesamento de datos en streaming con Spark

Exame de certificación incluído:

CCA175 - CCA Spark and Hadoop Developer Exam

Dirixido a

O curso está especialmente ideado para desenvolvedores/as e enxeñeiros/as con experiencia en programación.

Os exemplos e exercicios preséntanse en Python e Scala, por iso requírese o coñecemento dun destas linguaxes de programación.

Tamén é necesario ter coñecementos básicos de Linux. Non se require coñecemento previo de Hadoop.

É necesario ter a capacidade de ler textos técnicos en inglés.

Perfil do docente

O noso equipo de formación son persoas con máis de 5 anos de experiencia en áreas de alta especialización técnica nos ámbitos de aplicación. Dispoñen das certificacións oficiais do fabricante (neste caso Cloudera) para impartir estes cursos.

DURACIÓN	48 horas
PROGRAMA	Programación 2018/19
MATRÍCULA	Gratuíta
METODOLOXÍA	Virtual
TIPO	CURSO
CERTIFICACIÓN OFICIAL	Sí
EXAME CERTIFICACIÓN	CCA175 - CCA Spark and Hadoop Developer Exam
BENEFICIOS	
HORARIO	De luns a venres de 09:30 a 13:30 horas.
PERIODO INSCRICIÓN	09/05/2019 - 23/05/2019

PROBA DE SELECCIÓN	28/05/2019, 10:00
PERIODO DOCENCIA	20/06/2019 - 05/07/2019
LUGAR DE DOCENCIA	Edificio localizado na r/Airas Nunes s/n, barrio de Conxo, en
Nº PRAZAS	20 (Mínimo 10)

Temario

Introdución ao Apache Hadoop e ao ecosistema de Hadoop
Arquivos de almacenamento de Apache Hadoop
Procesamento de datos nun clúster de Apache Hadoop
Fundamentos de Apache Spark
Traballar con Dataframes e Schemas
Análise de datos con consultas DataFrame
RDDs: Resumo
Transformación de datos con RDD
Agregación de datos con par RDD
Consulta e ver táboas con Spark SQL
Creación, configuración e execución de aplicacións Spark
Procesamento distribuído
Persistencia de datos distribuídos
Patróns comúns ao procesar datos con Spark
Spark Streaming: Introducción a DStreams
Spark Streaming: procesando varios lotes
Streaming de Apache Spark: fontes de datos
Conclusión