

# Desenvolvemento Big Data Cloudera para Spark e Hadoop

## Descrición

Este curso ofrece os conceptos clave e a experiencia que o alumnado necesita para introducir e procesar datos nunha agrupación industrial Hadoop utilizando as ferramentas e técnicas máis actuais tales como Spark (incluíndo Spark Streaming e Spark SQL), Flume, Kafka e Sqoop.

Proporciona a mellor preparación para os desafíos do mundo real que afrontan os desenvolvedores/as de Hadoop. Con Spark, os desenvolvedores/as 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/a alumno/a coñecerá:

- Como se axusta o ecosistema de Apache Hadoop co ciclo de vida do procesamento de datos
- Como se distribúen, almacenan e procesan os datos nun Cluster de Hadoop
- Escribir, configurar e despregar aplicacións Apache Spark nun Cluster de Hadoop
- Utilizar a Spark shell e aplicacións Spark para explorar, procesar e analizar datos distribuídos
- Realizar consultas de datos con Spark SQL, Dataframes e Datasets
- Uso de Spark Streaming para procesar datos en tempo real.

## 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.

O coñecemento básico de SQL é útil.

Recomendable ter coñecemento nivel medio comprensión lectora de ínguas.

## Perfil do docente

Os/As nosos/as formadores/as 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.

<b>DURACIÓN</b>	48 horas
<b>PROGRAMA</b>	Programación 2019/20
<b>MATRÍCULA</b>	Gratuíta
<b>METODOLOXÍA</b>	Virtual
<b>TIPO</b>	CURSO
<b>CERTIFICACIÓN OFICIAL</b>	Sí
<b>EXAME CERTIFICACIÓN</b>	Cloudera CCA175 - CCA Spark and Hadoop Developer Exam
<b>BENEFICIOS</b>	
<b>HORARIO</b>	De luns a venres de 16:30 a 20:30 horas.
<b>PERIODO INSCRICIÓN</b>	19/12/2019 - 16/01/2020
<b>PROBA DE SELECCIÓN</b>	21/01/2020, 17:30
<b>PERIODO DOCENCIA</b>	03/02/2020 - 18/02/2020
<b>LUGAR DE DOCENCIA</b>	Edificio localizado na r/Airas Nunes s/n, barrio de Conxo, en
<b>Nº PRAZAS</b>	20 (Mínimo 10)

## Temario

- Introducción
- Introducción a Apache Hadoop e o ecosistema Hadoop
- Arquivos de almacenamento de Apache Hadoop
- Proceso de datos nunha agrupación industrial de Apache Hadoop
- Fundamentos de Apache Spark
- Traballar con Dataframes e Schemas
- Análise de datos con consultas de DataFrame
- RDDs: Sumario
- Transformación de datos con RDDs
- Agregación de datos con Pair RDDs
- Consulta e vistas de táboas con Spark SQL
- Creación, configuración e execución de aplicacións Spark
- Procesamento distribuído Spark
- Persistencia de datos distribuídos
- Patróns comúns ao procesar datos con Spark
- Introducción ás Estruturas Streaming
- Estruturas Streaming con Apache Kafka
- Agregación e unión de Streaming Dataframes
- Procesamento de mensaxes con Apache Kafka
- Conclusión