



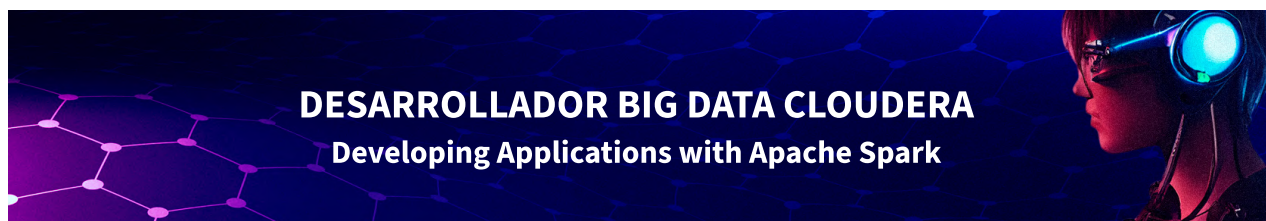
PRIMERA EDICIÓN

#DATA  
SENTINEL

# Desarrollador Big Data Cloudera

Developing Applications  
with Apache Spark





## ACERCA DE ESTE CURSO

El presente itinerario, oficial del Linux Professional Institute y de Cloudera, proporciona al candidato los conceptos clave y el conocimiento que los desarrolladores necesitan para utilizar Apache Spark para desarrollar aplicaciones paralelas de alto rendimiento en Cloudera Data Platform (CDP).

Los ejercicios que contiene esta formación permitirán a los alumnos practicar la escritura de aplicaciones Spark que se integran con los componentes principales de CDP, como Hive y Kafka.

Los participantes aprenderán cómo usar Spark SQL para consultar datos estructurados, cómo usar Spark Streaming para procesar datos de transmisión en tiempo real y cómo trabajar con "big data" almacenado en un sistema de archivos distribuido.

Una vez finalizado el curso, los candidatos serán capaces de crear aplicaciones para tomar decisiones mejores y más rápidas, además de saber ejecutar un análisis interactivo aplicado a una amplia variedad de casos de uso, arquitecturas e industrias.

PUE es [Platinum Training Partner de LPI](#), autorizado oficialmente por esta organización para impartir formación oficial en tecnologías Linux.

PUE es [Platinum Partner de Cloudera](#), autorizado oficialmente por esta multinacional para impartir formación y certificación oficial en tecnologías Cloudera.

## A QUIÉN VA DIRIGIDO

Este curso está diseñado para alumnos interesados en formarse profesionalmente como desarrolladores e ingenieros de datos.

Aunque es objeto de la formación del curso, se valorará en la selección el poseer formación reglada y/o experiencia profesional previa en el ámbito de la administración de sistemas operativos Linux y también en el desarrollo de aplicaciones.

Ciclos formativos de Grado Superior o Ingenierías serán de gran utilidad para maximizar la correcta asimilación de los contenidos.

## OBJETIVOS DEL CURSO

Una vez finalizado el curso el alumno habrá adquirido los siguientes conocimientos y habilidades:

Este curso enseña a los participantes las siguientes habilidades:

- Distribuir, almacenar y procesar datos en un clúster CDP
- Escribir, configurar e implementar aplicaciones Apache Spark
- Usar los intérpretes de Spark y las aplicaciones de Spark para explorar, procesar y analizar datos distribuidos
- Consultar datos con tablas Spark SQL, DataFrames y Hive
- Usar Spark Streaming junto con Kafka para procesar un flujo de datos

## CERTIFICACIÓN INCLUIDA

Este curso oficial es el recomendado por LPI para la preparación de la certificación oficial:

- [101-500: LPIC-1 General Linux Part 1](#)
- [102-500: LPIC-1 General Linux Part 2](#)

La superación de ambos exámenes es un requisito imprescindible para obtener la certificación [LPIC-1: System Administrator](#).

Este curso oficial es el recomendado por Cloudera para la preparación del siguiente examen de certificación:

- [CDP-3001: CDP Data Developer](#)

La superación de este examen es un requisito imprescindible para obtener la certificación **CDP Data Developer**.

Los alumnos que finalicen la presente formación con aprovechamiento tendrán la oportunidad de obtener, sin coste alguno, las certificaciones oficiales de LPI y Cloudera.

## CONTENIDOS

### Módulo 1. LPI 101-General Linux Part 1

#### 1. Arquitectura del sistema

- 1.1 Selección y configuración del hardware
- 1.2 Arranque del sistema
- 1.3 Cambiar de nivel de ejecución / objetivo de arranque y apagar o reiniciar el sistema

#### 2. Instalación de Linux y Gestión de paquetes

- 2.1 Diseño del esquema de particionado del disco duro
- 2.2 Instalación de un gestor de arranque
- 2.3 Gestión de librerías compartidas
- 2.4 Gestión de paquetes Debian
- 2.5 Gestión de paquetes RPM y YUM
- 2.6 Virtualización en Linux

#### 3. Comandos GNU y UNIX

- 3.1 Uso de la línea de comandos
- 3.2 Procesamiento de secuencias de texto usando filtros
- 3.3 Gestión básica de ficheros
- 3.4 Uso de Streams, pipes y redirecciones
- 3.5 Creación, monitorización y eliminación de procesos
- 3.6 Modificación de las prioridades de ejecución de procesos
- 3.7 Búsqueda de ficheros de texto mediante el uso de expresiones regulares
- 3.8 Operaciones de edición básica de archivos

#### 4. Dispositivos, Sistemas de Archivos de Linux y Jerarquía Estándar del Sistema de ficheros

- 4.1 Creación de particiones y sistemas de archivos
- 4.2 Mantenimiento de la integridad del sistema de archivos
- 4.3 Montaje y desmontaje de sistemas de archivos
- 4.4 Gestión de permisos y propiedad de archivos
- 4.5 Creación y modificación de enlaces simbólicos y duros
- 4.6 Búsqueda de ficheros del sistema y colocación en su correcta localización

### Módulo 2. LPI 102-General Linux Part 2

#### 1. Shells y Shell Scripting

- 1.1 Personalización y uso del intérprete de comandos
- 1.2 Personalización y creación de scripts simples

## **2. Interfaces de Usuario y Escritorios**

- 2.1 Instalación y configuración de X11
- 2.2 Interfaz gráfica
- 2.3 Accesibilidad

## **3. Tareas Administrativas**

- 3.1 Gestión de cuentas de usuarios y grupos y archivos relacionados
- 3.2 Automatización las tareas de administración del sistema mediante la planificación de trabajos
- 3.3 Localización e internacionalización

## **4. Servicios esenciales del sistema**

- 4.1 Mantenimiento de la hora del sistema
- 4.2 Logs del sistema
- 4.3 Fundamentos del agente de transferencia de correo (MTA)
- 4.4 Impresión y gestión de impresoras

## **5. Fundamentos de redes**

- 5.1 Conocimientos básicos de los protocolos de internet
- 5.2 Configuración básica de red
- 5.3 Detección de problemas de red
- 5.4 Configuración de un cliente DNS

## **6. Seguridad**

- 6.1 Realización de tareas de gestión de seguridad del sistema
- 6.2 Configuración de la seguridad de un host
- 6.3 Protección de datos mediante encriptación

## **Módulo 3: Introducción a SQL**

- Introducción
- Base de datos básica
- Estructura de base de datos
- El uso de SQL
- Creación de una base de datos
- Las consultas básicas
- Mantenimiento de la Base de datos
- Copia de tablas y modificaciones de columnas
- Índices y restricciones

- Eliminar o modificar filas de la tabla de datos
- Funciones: uso y tipos
- Funciones específicas
- Joins
- Exportación de datos, consultas y utilidades
- Importación de datos y de archivos de datos
- Importación con sentencias y utilidades
- Subconsultas generales y básicas
- Subconsulta no correlacionada y correlacionada
- Tipos de subconsultas
- Modificación de la tabla con subconsultas
- Motores de almacenamiento
- Creación de Vistas
- Transacciones
- Recuperación de Metadatos

#### **Módulo 4: Programación con Python**

- Introducción a Python
- Variables en Python
- Colecciones en Python
- Flow Control en Python
- Estructura de Programación en Python
- Trabajo con librerías en Python

#### **Módulo 5: Programación con Scala**

- Introducción
- Scala Overview
- Introducción a Scala
- Scala's Role in Distributed
- Procesamiento de Datos
- La motivación de Scala
- Básicos Scala
- Trabajo con tipos de datos
- Agrupación de datos
- Flow Control en Scala
- Uso y creación de librerías

## **Módulo 6: Introducción a Hadoop y Big Data**

- Introducción a Hadoop.
- Ecosistema y gestión de las soluciones de Hadoop.
- MapReduce.
- Clusters de Hadoop.
- Herramientas de Hadoop.
- Pig.
- Hive.
- Impala.

## **Módulo 7: Cloudera Data Engineering: Developing Applications with Apache Spark**

- Introducción a Zeppelin
- Introducción a HDFS
- Introducción a YARN
- Historial de procesamiento distribuido
- Trabajar con RDDs
- Trabajar con DataFrames
- Introducción a Apache Hive
- Integración de Hive y Spark
- Visualización de datos con Zeppelin
- Desafíos del procesamiento distribuido
- Procesamiento distribuido Spark
- Persistencia distribuida de Spark
- Escribir, configurar y ejecutar aplicaciones Spark
- Introducción a Structured Streaming
- Procesamiento de mensajes con Apache Kafka
- Structured Streaming con Apache Kafka
- Agregar y unir Streaming DataFrames

Plazas limitadas.

### **¿TIENES DUDAS?**

Para cualquier consulta no dudes en contactarnos a través del siguiente correo electrónico:  
[puetalent@pue.es](mailto:puetalent@pue.es)

PRIMERA EDICIÓN

#DATA  
SENTINEL

