

## **PROGRAMA FORMATIVO**

# Administrador Big Data Cloudera Apache Hadoop

## DATOS GENERALES DE LA ESPECIALIDAD

1. **Familia Profesional:** INFORMÁTICA Y COMUNICACIONES (IFC)  
**Área Profesional:** SISTEMAS Y TELEMÁTICA
2. **Denominación:** ADMINISTRACIÓN BIG DATA CLOUDERA APACHE HADOOP
3. **Código:** IFCT23
4. **Nivel de cualificación:** 3

### 5. **Objetivo general:**

Adquirir las competencias y el expertise de un profesional para la Administración Big Data Cloudera de Apache Hadoop.

Alcanzar los conocimientos y habilidades necesarias en UNIX y Linux para ser capaz de interactuar con el sistema operativo y realizar las principales tareas de Administración.

Conocer los componentes principales de Hadoop, HDFS, MapReduce y el amplio ecosistema existente alrededor de Hadoop como Hive, Pig, Sqoop, Flume e Impala, para ser capaces de operar y mantener un clúster Hadoop desde la instalación y la configuración, incluyendo el balanceo de carga y tuning.

### 6. **Prescripción de los formadores:**

#### 6.1. Titulación requerida:

El formador deberá estar homologado como instructor en la correspondiente especialidad de la tecnología específica del fabricante y contar con la certificación vigente " Cloudera Certified Administrator for Apache Hadoop (CCAHA)".

#### 6.2. Experiencia profesional requerida:

Tener experiencia acreditable en ocupaciones relacionadas con la especialidad a impartir de al menos un año, excluyendo la experiencia docente.

#### 6.3. Competencia docente

Los formadores deberán contar con formación metodológica, o experiencia docente contrastada superior a 350 horas relacionadas con la familia de Informática y Comunicaciones.

### 7. **Criterios de acceso del alumnado:**

#### 7.1. Nivel académico o de conocimientos generales:

- Título de Bachiller o equivalente. Asimismo, podrán acceder quienes posean un título de Formación Profesional de grado superior, un certificado de profesionalidad de nivel 3, un certificado de profesionalidad de la familia profesional de Informática y comunicaciones de nivel 2, cumplir el requisito académico de acceso al ciclo formativo de grado superior, haber superado las correspondientes pruebas de acceso reguladas por las administraciones educativas o tener superada la prueba de acceso a la universidad para mayores de 25 años y/o de 45 años.

- Se requieren conocimientos de sistemas operativos e informática.
- Cuando el aspirante al curso no posea el nivel académico indicado demostrará conocimientos suficientes a través de una prueba de acceso que evalúe los conocimientos específicos iniciales del alumno respecto a la unidad de competencia.

## 8. Número de participantes:

Máximo 25 participantes.

## 9. Relación secuencial de módulos formativos:

- Módulo 1: Fundamentos Unix y Linux
- Módulo 2: Administración de Sistemas Linux
- Módulo 3: Administración Cloudera para Apache Hadoop

## 10. Duración:

Horas totales: 110 h

## 11. Requisitos mínimos de espacios, instalaciones y equipamiento

### 11.1. Espacio formativo:

- Aula de Informática: Superficie: 45 m<sup>2</sup> para grupos de 15 alumnos (3 m<sup>2</sup> por alumno).
- Cada espacio estará equipado con mobiliario docente adecuado al número de alumnos, así mismo constará de las instalaciones y equipos de trabajo suficientes para el desarrollo del curso.

### 11.2. Equipamiento:

Los equipos tendrán unas características equivalentes a las enumeradas a continuación, consideradas siempre como mínimas:

- 1 ordenadores por alumno y 1 para el profesor con las siguientes características mínimas:
  - Hardware:
    - Procesador mínimo 3,2 GHz
    - Memoria RAM mínima 4 GB
    - Disco duro mínimo 160 GB
    - Tarjeta de red 10/100/1000 Mbps
    - Tarjeta gráfica 256 Mb. PCIe
    - Tarjeta de sonido
    - Lector grabador de DVD
    - Periféricos: Teclado, Ratón y Monitor color 17"
  - Software:
    - Licencias de Sistema Operativo.
    - Licencias del software ofimático necesario para la impartición del curso
  - Conectividad a Internet para acceso a los sistemas de laboratorios Oracle
    - Navegadores soportados: Internet Explorer 7.0, Mozilla 1.5+, Firefox 2.0 o superiores
    - Javascript y cookies habilitadas
    - Ancho de banda de 250K por estación
- Pantalla y cañón de proyección.

A los alumnos se les proporcionará la documentación oficial de Cloudera necesaria para la impartición del curso.

Las instalaciones y equipamientos deberán cumplir con la normativa industrial e higiénico - sanitaria

correspondiente y responderán a medidas de accesibilidad universal y seguridad de los participantes. En el caso de que la formación se dirija a personas con discapacidad se realizarán las adaptaciones y los ajustes razonables para asegurar su participación en condiciones de igualdad.

## 12. Requisitos oficiales de los centros

Para la impartición de esta especialidad formativa el Centro ha de estar homologado como:

- Centro de formación oficial del fabricante bajo la figura de Cloudera Authorized Training Center de Cloudera.

## 13. Evaluación del aprendizaje

Se llevará a cabo una evaluación continua y sistemática durante el proceso de aprendizaje y al final del mismo para comprobar si los alumnos han alcanzado los objetivos establecidos en cada módulo y, por consiguiente, han realizado el curso con el aprovechamiento requerido.

## 14. Certificación oficial del fabricante

La ejecución y financiación del programa formativo incluye la presentación de los alumnos que han realizado el curso con aprovechamiento, al siguiente examen para obtener la certificación oficial del fabricante, que gestionará el centro y que en ningún caso supondrá coste alguno para el alumno.

En concreto, para esta acción formativa está incluido el siguiente examen de certificación oficial de Cloudera (1 convocatoria a examen por alumno):

- CCA-500 - Cloudera Certified Administrator for Apache Hadoop (CCAH)

## MÓDULOS FORMATIVOS

### Módulo nº 1

#### Denominación:

FUNDAMENTOS UNIX y LINUX

#### Objetivo:

Usar adecuadamente comandos y recursos; Crear, modificar, tramitar permisos y gestionar archivos y directorios; Controlar el entorno de trabajo; Realizar la transferencia de archivos a distancia y tareas de proceso y control de trabajos.

**Duración:** 20 horas

#### Contenidos teórico- prácticos:

- Visualización de directorios y archivos
- Trabajo con archivos y directorios de Solaris
- Uso del editor vi
- Uso de comandos en el Shell
- Uso de los permisos de archivos básicos
- Configuración las listas de control de acceso (ACL)
- Búsqueda de archivos y directorios
- Control básico de los procesos
- Funcionalidad avanzada de shell
- Lectura de secuencias de comandos de shell
- Creación de Archivos

- Compresión, visualización y descompresión de archivos
- Conexiones remotas y transferencias de archivos

## **Módulo nº 2**

### **Denominación:**

ADMINISTRACIÓN DE SISTEMAS LINUX

### **Objetivo:**

Instalar y configurar terminales virtuales en entornos Linux, configurar Sistemas Linux para conectar a la red de la empresa o Internet, implementar cualquier tipo de periférico, gestionar usuarios y ficheros, controlar el almacenamiento, administrar la seguridad, supervisar un Sistema Linux y solucionar problemas.

**Duración:** 30 horas

### **Contenidos teórico - prácticos:**

- Introducción a Linux
- Instalación de Linux
- Linux Boot Process
- Configuración del Sistema
- Gestión de paquetes
- Automatización de tareas
- Logging y acceso al sistema
- Configuración del módulo del kernel
- Administración de usuarios y grupos
- Sistemas de archivos
- Administración de Almacenamiento
- Configuración de la red
- Direccionamiento de red y servicios de nombres
- Intercambio de archivos
- Servicio de OpenSSH
- Autenticación y Directorio de Servicios
- Módulos de autenticación conectables (PAM)
- Administración de la Seguridad
- Web y servicios de correo electrónico
- Supervisión del sistema y solución de problemas

## **Módulo nº 3**

### **Denominación:**

ADMINISTRACIÓN CLOUDERA PARA APACHE HADOOP

### **Objetivo:**

Conocer los detalles internos de YARN, MapReduce y HDFS; Identificar y aprender a seleccionar las configuraciones adecuadas de hardware e infraestructura para el funcionamiento óptimo del clúster; Configurar, desplegar e integrar el clúster como centro de datos; Automatizar la carga utilizando canal de flujo RDBMS y Sqoop; Configurar el FairScheduler; Proporcionar acuerdos de nivel de servicio para múltiples usuarios de un grupo; Aprender a instalar e implementar la seguridad basada en Kerberos. Mantener y monitorizar el clúster, solucionar problemas que pueden surgir al trabajar con otras plataformas y frameworks como Hive, Pig, y HBase y en definitiva, ser capaces de configurar e implementar las mejores prácticas para obtener óptimos resultados con Apache Hadoop.

**Duración:** 60 horas

**Contenidos teórico – prácticos:**

- Introducción
- Historia de Apache Hadoop
- Características de HDFS
- Obtención de datos en HDFS
- MapReduce
- Planificación de un clúster Hadoop
- Consideraciones generales de planificación
- Instalación de Hadoop
- Instalación y configuración de Hive, Pig e Impala
- Clientes Hadoop
- Cloudera Manager
- Configuración avanzada de un cluster
- Seguridad Hadoop
- Gestión y planificación de tareas o trabajos
- Mantenimiento de un clúster
- Solución de problemas habituales en el clúster de Hadoop
- monitorización de un clúster