



PROGRAMA FORMATIVO:

ADMINISTRADOR BIG DATA CLOUDERA

Julio 2017

DATOS GENERALES DEL CURSO

1. **Familia Profesional:** INFORMÁTICA Y COMUNICACIONES
Área Profesional: SISTEMAS Y TELEMÁTICA
2. **Denominación:** ADMINISTRADOR BIG DATA CLOUDERA
3. **Código:** IFCT34
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:

Titulación universitaria u otros títulos equivalentes, o capacitación profesional equivalente acreditada por el fabricante.

El formador deberá estar homologado como instructor en la correspondiente tecnología específica del fabricante y contar con todas las certificaciones de la especialidad a impartir vigentes y actualizadas.

6.2. Experiencia profesional requerida:

Experiencia Profesional de al menos 12 meses en la ocupación relacionada con la especialidad a impartir, excluyendo la experiencia docente.

6.3. Competencia docente:

Será necesario tener experiencia metodológica o experiencia docente contrastada de 500 horas de formación en especialidades relacionadas con la especialidad a impartir.

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

7.1. Nivel académico o de conocimientos generales:

Título de FP Grado superior, bachillerato.

Cuando el aspirante al curso no posea el nivel académico indicado, demostrará conocimientos suficientes a través de una prueba de acceso.

Se requiere inglés a nivel de lectura y conocimientos de sistemas operativos e informática.

8. **Número de alumnos:**

Máximo 25 participantes para cursos presenciales.

9. Relación secuencial de módulos:

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

10. Duración:

Horas totales: 150

Distribución horas:

- Presencial: 150

11. Requisitos mínimos de espacios, instalaciones y equipamiento.

11.1. Espacio formativo:

- Aula de Informática: Superficie: 45 m² para grupos de 15 alumnos (3 m² 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:

- Un ordenador por alumno y uno para el profesor con las siguientes características mínimas:
 - CPU: procesador Intel Core i5 de 6 generación o similar.
 - 16 GB de RAM
 - Disco duro de 500 GB
 - Lector-grabador DVD
 - Tarjeta de red 10/100/1000 Mbps
 - Teclado multimedia USB
 - Ratón sensor óptico USB de 2 botones y rueda de desplazamiento.
 - Monitor color de 17" TFT
 - Licencia de sistema operativo WINDOWS 7 o similar.
 - Licencia de software antivirus.
 - Licencias del software del fabricante necesario para la impartición del curso.
- Conectividad a Internet.
- Impresora láser con conexión a red.
- Pantalla y cañón de proyección.

A los alumnos se le proporcionará la documentación oficial de CLOUDERA necesaria para el seguimiento 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 a los exámenes 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 o el que lo sustituya actualizado al momento de su impartición:

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

MÓDULOS FORMATIVOS

Módulo 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: 40 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
- Troubleshooting

Módulo 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: 40 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
- Troubleshooting

Módulo 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: 70 horas

Contenidos teórico – prácticos:

- Introducción
- Historia de Apache Hadoop
- Instalación del Clúster Hadoop
- El Hadoop Distributed File System (HDFS)
- MapReduce y Spark en YARN
- Configuración de Hadoop y registros de Daemon
- Obtención de datos en HDFS
- Planificación de un cluster Hadoop
- Instalación y configuración de Hive, Pig e Impala
- Clientes Hadoop incluidos en Hue
- Configuración avanzada de un cluster
- Seguridad Hadoop
- Gestión de recursos
- Mantenimiento de un cluster
- Solución de problemas y monitorización de un cluster
- Conclusión