



PROGRAMA FORMATIVO:

ANALISTA DE DATOS BIG DATA CLOUDERA

Julio 2017

DATOS GENERALES DE LA ESPECIALIDAD

1. **Familia Profesional:** INFORMÁTICA Y COMUNICACIONES
Área Profesional: SISTEMAS Y TELEMÁTICA
2. **Denominación:** ANALISTA DE DATOS BIG DATA CLOUDERA
3. **Código:** IFCT35
4. **Nivel de cualificación:** 3

5. **Objetivo general:**

Adquirir los conocimientos y la solvencia necesaria para capturar, manejar, almacenar, transformar, buscar y visualizar un conjunto de datos de gran volumen y complejidad y relacionarlos entre sí, para extraer información de valor a partir de ellos, mediante una solución Big Data de código abierto como Cloudera.

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 de SQL
- Módulo 2: Fundamentos de Cloudera Apache Hadoop
- Módulo 3: Analista de Datos Cloudera

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 – Data Analyst Cloudera

MÓDULOS FORMATIVOS

Módulo 1

Denominación:

FUNDAMENTOS DE SQL

Objetivo:

Adquirir los conocimientos necesarios para entender las características y beneficios de las bases de datos relacionales, el lenguaje, los datos y tipos de columna, aprender a crear un diseño de base de datos con una estructura eficiente y ser capaz de extraer información de bases de datos de base con el comando SELECT, solucionar problemas de sintaxis y entender las advertencias y los errores típicos. Ser capaz de agrupar datos de consulta con la agregación, conectar datos de múltiples filas de la tabla con JOIN, realizar anidados de subconsultas y utilizar funciones simples.

Duración: 50 horas

Contenidos teórico prácticos:

- 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
- Conclusiones

Módulo 2

Denominación:

FUNDAMENTOS DE CLOUDERA APACHE HADOOP

Objetivo:

Adquirir los conocimientos fundamentales y las principales funcionalidades acerca del framework Apache Hadoop. Saber identificar y seleccionar los casos y situaciones más apropiadas para su aplicación. Aprender a implementar Hadoop y su ecosistema en cualquier entorno y a estimar los recursos necesarios para desplegar Hadoop.

Duración: 25 horas

Contenidos teórico-prácticos:

- Acerca de Apache Hadoop
- Acerca de Cloudera
- La motivación para Hadoop
- Problemas con los sistemas tradicionales de gran escala
- Cómo Hadoop aborda estos desafíos
- HDFS: El sistema de ficheros distribuidos de Hadoop
- El funcionamiento de MapReduce
- Anatomía de un cluster Hadoop
- Aplicaciones comunes y usos especiales de la solución Hadoop
- El ecosistema Hadoop
 - Almacenamiento de datos: Hbase
 - Integración de datos: Flume y Sqoop
 - Procesamiento de datos: Spark
 - Análisis de los datos: Hive, Pig e Impala
 - Exploración de datos: Cloudera Search
- Integración de Hadoop en el CPD
- Herramientas para gestionar Hadoop
- Planificación del proyecto

Módulo 3

Denominación:

ANALISTA DE DATOS CLOUDERA

Objetivo:

Conocer las características de Pig, Hive, e Impala para la adquisición, almacenamiento y análisis de datos, unir conjuntos de datos para obtener mayor información y realizar consultas complejas sobre conjuntos de datos en tiempo real.

Duración: 75 horas

Contenidos teórico – prácticos:

- Introducción
- Fundamentos Hadoop
- Introducción a Pig
- Análisis de datos básico con Pig
- Procesado de datos complejos con Pig
- Operaciones con multiconjuntos de datos con Pig
- Troubleshooting y optimización de Pig
- Introducción a Hive e Impala
- Consultas con Hive e Impala
- Administración de datos
- Almacenamiento y datos de rendimiento
- Análisis de datos relacional con Hive e Impala
- Trabajar con Impala
- Analizando texto y datos complejos con Hive
- Optimización Hive
- Extensión de Hive
- Elección de la mejor opción
- Conclusión