



PROGRAMA FORMATIVO:

BIG DATA DEVELOPER CON CLOUDERA
APACHE HADOOP

Noviembre 2015

DATOS GENERALES DEL CURSO

1. **Familia Profesional:**
INFORMÁTICA Y COMUNICACIONES (IFC)

Área Profesional:
DESARROLLO

2. **Denominación:**
BIG DATA DEVELOPER CON CLOUDERA APACHE HADOOP

3. **Código:** IFCD23

4. **Nivel de cualificación:** 3

5. **Objetivo general:**

El objetivo de la presente formación es preparar al alumno para analizar y resolver problemas del mundo real utilizando Apache Hadoop y herramientas asociadas al centro de datos empresariales. Se trabajara a través de todo el proceso de diseño y construcción de soluciones, incluyendo los datos de ingestión, determinar el formato de archivo adecuado para el almacenamiento, el procesamiento de los datos almacenados, y presentar los resultados al usuario final en una forma fácil de interpretar. Irá más allá de MapReduce para utilizar elementos adicionales del centro de datos empresarial y desarrollar aplicaciones de convergencia que son altamente relevantes para el negocio.

Los alumnos adquirirán los conocimientos básicos del lenguaje Java y los conocimientos necesarios para, una vez familiarizados con los componentes principales de Hadoop, HDFS, MapReduce y el amplio ecosistema alrededor de Hadoop, elaborar código sobre la API de Hadoop, así como depurar programas Mapreduce y optimizar su funcionamiento. A lo largo del curso también se introduce a los asistentes en otros elementos del ecosistema Hadoop como son Hive, Pig, HBase, Flume y Oozie.

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.

- 6.3. Competencia docente:

Será necesario tener experiencia metodológica o experiencia docente contrastada de 350 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 básicos de programación.

8. Número de alumnos:

Máximo 25 participantes para cursos presenciales.

9. Relación secuencial de módulos:

- Módulo 1. DESARROLLO DE APLICACIONES PYTHON
- Módulo 2. DESARROLLADOR CLOUDERA PARA APACHE SPARK I
- Módulo 3. DESARROLLADOR CLOUDERA PARA APACHE SPARK II

10. Duración:

Horas totales: 175

Distribución horas:

- Presencial: 175
- Teleformación: 0
- A distancia convencional: 0

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:

- 16 ordenadores (15 alumnos y 1 profesor) con las siguientes características mínimas:
 - CPU: procesador Intel D, a 3,2 GHz, 2 MB de cache de nivel 2 y 8 GB de RAM
 - Disco duro de 160 GB
 - Disquetera interna de 3,5 pulgadas de 1,44 MB.

- Lector-grabador DVD+/- RW
- Tarjeta de red 10/100/1000 Mbps
- Tarjeta gráfica 128 Mb.
- Tarjeta de sonido de última generación
- Teclado multimedia USB
- Ratón sensor óptico USB de 2 botones y rueda de desplazamiento.
- Monitor color de 17" TFT
- Sistema operativo WINDOWS 7 y antivirus.
- Software del fabricante
- Conectividad a Internet.
- Impresora láser con conexión a red.
- Servidor con sistema operativo
- Software antivirus.
- 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. 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.

13. 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.

14. Requisitos oficiales de los centros

Acreditación de encontrarse la entidad licitante autorizada por el fabricante del lote al que opta, para poder impartir adecuadamente cursos de formación con certificación oficial del mismo, que es el objeto del contrato. 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".

MÓDULOS FORMATIVOS

Módulo nº 1

Denominación:

DESARROLLO DE APLICACIONES PYTHON

Objetivo:

Al finalizar el módulo el alumno será capaz de desarrollar aplicaciones en Python.

Duración:

75 horas

Contenidos teórico-prácticos:

- Introducción a Python
- Cómo ejecuta Python los programas
- Ejecutar programas en Python
- Introducción a los tipos de objetos Python
- Tipos Numéricos
- Tipado Dinámico
- Fundamentos de Cadenas
- Listas y Diccionarios
- Tuplas, Ficheros y todo lo demás
- Introducción a las Sentencias Python
- Asignaciones, Expresiones y Prints
- if Tests y Reglas de Sintaxis
- Los Bucles While y For
- Iteraciones y Comprehensions
- La Documentación
- Fundamentos de las Funciones
- Scopes
- Argumentos
- Funciones Avanzadas
- Comprehensions y Generations
- Benchmarking
- Módulos
- Programación de Módulos
- Programación Orientada a Objetos
- Fundamentos de Programación de Clases
- Sobrecarga de Operadores

Módulo nº 2

Denominación:

DESARROLLADOR CLOUDERA PARA APACHE SPARK I

Objetivo:

Al finalizar el módulo el alumno será capaz de importar datos a Apache Hadoop Cluster y procesarlos con Spark, Hive, Flume, Sqoop, Impala, y otras herramientas del ecosistema de Hadoop.

Duración:

50 horas

Contenidos teórico-prácticos:

- Introduction
- Introduction to Hadoop and the Hadoop Ecosystem
- Hadoop Architecture and HDFS
- Importing Relational Data with Apache Sqoop
- Introduction to Impala and Hive
- Modeling and Managing Data with Impala and Hive
- Data Formats
- Data Partitioning
- Capturing Data with Apache Flume
- Spark Basics
- Working with RDDs in Spark
- Writing and Deploying Spark Applications
- Parallel Programming with Spark
- Spark Caching and Persistence
- Common Patterns in Spark Data Processing
- Preview: Spark SQL
- Conclusion

Módulo nº 3

Denominación:

DESARROLLADOR CLOUDERA PARA APACHE SPARK II

Objetivo:

Al finalizar el módulo el alumno será capaz de: simplificar el desarrollo con Kite SDK, definir y usar Data Sets, importar datos relacionales con Apache Sqoop, capturar datos con Apache Flume y desarrollar componentes personalizados, manejar Workflows con Apache Oozie, procesar pipeline de datos con Apache Crunch, leer y analizar formatos de datos customizados en Apache Hive, responder queries interactivas con Impala, transformar Data Streams con Morphlines, autorizar búsqueda completa en los datos guardados con HDFS y presentar resultados a los usuarios

Duración:

50 horas

Contenidos teórico-prácticos:

- Introducción
- Arquitectura de aplicaciones
- Simplificando el desarrollo con Kite SDK
- Definiendo y usando Data Sets
- Importación de datos relacionales con Apache Sqoop
- Captura de datos con Apache Flume
- Desarrollo de componentes Flume customizados
- Manejo de Workflows con Apache Oozie
- Procesamiento de pipeline de datos con Apache Crunch
- Leer y analizar formatos de datos customizados en Apache Hive
- Responder queries interactivas con Impala
- Transformación de Data Streams con Morphlines
- Autorizar Búsqueda completa en los datos guardados con HDFS
- Presentación de resultados a los usuarios
- Trabajo de RDDs
- El Hadoop Distributed File System
- Ejecutar Spark en un Cluster
- Programación paralela con Spark
- Caching y Persistence
- Escritura de aplicaciones Spark
- Spark, Hadoop y el Enterprise Data Center
- Spark Streaming
- Algoritmos usuales en Spark
- Mejora de rendimiento Spark
- Conclusion