



PROGRAMA FORMATIVO CIENTÍFICO DE DATOS CLOUDERA

Octubre 2019

DATOS GENERALES DE LA ESPECIALIDAD

1. **Familia Profesional:** INFORMÁTICA Y COMUNICACIONES

Área Profesional: SISTEMAS Y TELEMÁTICA

2. **Denominación:** CIENTÍFICO DE DATOS CLOUDERA

3. **Código:** IFCT63

4. **Nivel de cualificación:** 3

5. **Objetivo general:**

Extraer información de valor, mediante el manejo, captura, almacenamiento, transformación, búsqueda, relación y visualización de datos de gran volumen y complejidad, con la solución Big Data de código abierto de Cloudera

6. **Prescripción de los formadores:**

6.1. Titulación requerida:

Titulación universitaria o Ciclo Formativo de Grado Superior, en su defecto, capacitación profesional equivalente en la especialización relacionada con el curso.

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:

Al menos doce meses de experiencia profesional en la especialidad objeto, 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.
- Dominio de inglés a nivel de lectura.
- Conocimientos y/o experiencia profesional en el campo de análisis de datos.
- Conocimiento de algún lenguaje de programación de uso general (Python, R,..).

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

8. **Número de participantes:** Máximo 25 participantes para cursos presenciales.

9. **Relación secuencial de módulos formativos:**

- Módulo 1: La ciencia de datos y el aprendizaje autónomo a escala. (30 horas)
- Módulo 2: Gestión del Dato. Apache Spark. (40 horas)
- Módulo 3: Aprendizaje autónomo (Machine Learning). Apache Spark Mlib. (80 horas)

10. Duración:

Horas totales: 150 horas.

11. Requisitos mínimos de espacios, instalaciones y equipamiento

11.1. Espacio formativo: aula de gestión de 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:

- Mesa y silla para el formador
- Mesas y sillas para el alumnado
- Material de aula
- Pizarra
- Impresora láser con conexión a red
- Pantalla y cañón de proyección
- PC instalado en red con posibilidad de impresión de documentos, cañón con proyector e internet para el formador (procesador de 64 bits Intel i5 de séptima generación o superior, o AMD similar, 8 GB de memoria R.A.M. o superior, disco duro de 256 GB o superior, preferiblemente SSD, pantalla 17' o superior con resolución mínima WXGA; teclado multimedia USB, ratón sensor óptico USB de 2 botones y rueda de desplazamiento)
- PCs instalados en red e internet con posibilidad de impresión para los alumnos (procesador de 64 bits Intel i5 de séptima generación o superior, o AMD similar, 8 GB de memoria R.A.M. o superior, disco duro de 256 GB o superior, preferiblemente SSD, pantalla 17' o superior con resolución mínima WXGA, teclado multimedia USB, ratón sensor óptico USB de 2 botones y rueda de desplazamiento)
- Licencia de sistema operativo
- Licencia de software antivirus.
- Licencia de software ofimático y herramientas necesarias para la impartición del curso (versión actualizada)
- Licencias del fabricante para la correcta impartición del curso.

En todo caso los requisitos mínimos tanto HW como SW serán los que marque el fabricante como recomendados en cada momento para las versiones actualizadas.

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 de los centros

Los centros impartidores de formación Oficial de Cloudera, deben estar homologados 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

Actualmente no existe certificación de fabricante para esta especialidad.

MÓDULOS FORMATIVOS

Módulo 1

Denominación: LA CIENCIA DE DATOS Y EL APRENDIZAJE AUTÓNOMO A ESCALA.

Objetivo: Evaluar las principales herramientas y procesos que se utilizarán en el análisis e interpretación de grandes bases de datos.

Duración: 30 horas

Contenidos teórico prácticos:

- Overview of data science and machine learning at scale.
- Overview of the Hadoop ecosystem.
- Working with HDFS data and Hive tables using Hue.
- Introduction to Cloudera Data Science Workbench.

Módulo 2

Denominación: GESTIÓN DEL DATO. APACHE SPARK.

Objetivo: Trabajar en formatos, procesos y lenguajes de programación que facilitan la gestión del dato (lectura, escritura, transformación, agrupación, limpieza, calidad,...) con Apache Spark.

Duración: 40 horas

Contenidos teórico prácticos:

- Overview of Apache Spark 2.
- Reading and writing data.
- Inspecting data quality.
- Cleansing and transforming data.
- Summarizing and grouping data.
- Combining, splitting, and reshaping data.
- Exploring data.
- Configuring, monitoring, and troubleshooting Spark applications.

Módulo 3

Denominación: APRENDIZAJE AUTÓNOMO (MACHINE LEARNING). APACHE SPARK MLIB.

Objetivo: Crear, validar e implementar modelos de aprendizaje autónomo con las funciones que proporciona MLlib.

Duración: 80 horas

Contenidos teórico prácticos:

- Overview of machine learning in Spark MLlib.
- Extracting, transforming, and selecting features.
- Building and evaluating regression models.
- Building and evaluating classification models.
- Building and evaluating clustering models.

- Cross-validating models and tuning hyperparameters.
- Building machine learning pipelines.
- Deploying machine learning models.