Catálogo de Especialidades Formativas

PROGRAMA FORMATIVO

CLOUDERA - DESARROLLO CON SPARK Y HADOOP

Mayo 2023

GOBIERNO DE ESPAÑA
MINISTERIO DE TRABAJO Y ECONOMÍA SOCIAL
SERVICIO PÚBLICO DE EMPLEO ESTATAL SEPE
SISTEMA NACIONAL DE EMPLEO

IDENTIFICACIÓN DE LA ESPECIALIDAD Y PARÁMETROS DEL CONTEXTO FORMATIVO

Denominación de la especialidad:	CLOUDERA - DESARROLLO CON SPARK Y HADOOP
Familia Profesional:	INFORMÁTICA Y COMUNICACIONES
Área Profesional:	DESARROLLO
Código:	IFCD73
Nivel de cualificación profesional:	3

Objetivo general

Conocer los sistemas Apache Hadoop y Apache Spark y sus usos más comunes respecto al procesamiento de datos.

Relación de módulos de formación

Módulo 1	FUNDAMENTOS DE HADOOP Y SPARK	10 horas
Módulo 2	RDDs	10 horas
Módulo 3	APLICACIONES CON SPARK Y ESTRUCTURAS STREAMING CON KAFKA	10 horas

Modalidad de impartición

Presencial

Teleformación

Duración de la formación

Duración total en cualquier modalidad de impartición	30 horas
Teleformación	Duración total de las tutorías presenciales: 0 horas

Requisitos de acceso del alumnado

Acreditaciones / titulaciones	Cumplir como mínimo alguno de los siguientes requisitos: Haber superado la prueba de acceso a Ciclos Formativos de Grado Superior Título de Postgrado (Máster) o equivalente Título de Grado o equivalente Haber superado cualquier prueba oficial de acceso a la universidad Certificado de profesionalidad de nivel 3 Título de Técnico Superior (FP Grado superior) o equivalente
Experiencia profesional	No se requiere
Otros	Cuando el aspirante no disponga del nivel académico demostrará conocimientos y competencias suficientes para participar en el curso con aprovechamiento mediante una prueba de acceso

Otros	Además de lo indicado anteriormente, los participantes han de tener las destrezas suficientes para ser usuarios de la plataforma virtual en la que se apoya la acción formativa.
Modalidad de teleformación	Además de lo indicado anteriormente, los participantes han de tener las destrezas suficientes para ser usuarios de la plataforma virtual en la que se apoya la acción formativa.

Prescripciones de formadores y tutores

Acreditación requerida	Cumplir como mínimo alguno de los siguientes requisitos: • Licenciado, Ingeniero, Arquitecto o Título de Grado correspondiente u otros títulos equivalentes. • Diplomado, Ingeniero Técnico, Arquitecto Técnico o el Título de Grado correspondiente u otros títulos equivalentes.
Experiencia profesional mínima requerida	- 1 año en el ámbito de Informática y comunicaciones, excluyendo la docencia.
Competencia docente	Cumplir como mínimo alguno de los siguientes requisitos: • Será necesario tener formación metodológica o experiencia docente. • Certificado de Profesionalidad de Docencia de la Formación Profesional para la Ocupación. • Máster Universitario de Formador de Formadores u otras acreditaciones oficiales equivalentes.
Otros	Además de cumplir con las prescripciones establecidas anteriormente, los tutores-formadores deben acreditar una formación, de al menos 30 horas, o experiencia, de al menos 60 horas, en esta modalidad y en la utilización de las tecnologías de la información y comunicación.
Modalidad de teleformación	Además de cumplir con las prescripciones establecidas anteriormente, los tutores-formadores deben acreditar una formación, de al menos 30 horas, o experiencia, de al menos 60 horas, en esta modalidad y en la utilización de las tecnologías de la información y comunicación.

Requisitos mínimos de espacios, instalaciones y equipamientos

Espacios formativos	Superficie m² para 15 participantes	Incremento Superficie/ participante (Máximo 30 participantes)
Aula de gestión	45.0 m²	2.4 m² / participante

Espacio formativo	Equipamiento
Aula de gestión	- Mesa y silla para el formador - Mesas y sillas para el alumnado - Material de aula - Pizarra - PC instalado en red con posibilidad de impresión de documentos, cañón con proyección e Internet para el formador - PCs instalados en red e Internet con posibilidad de impresión para los participantes - Software específico para el aprendizaje de cada acción formativa.

La superficie de los espacios e instalaciones estarán en función de su tipología y del número de participantes. Tendrán como mínimo los metros cuadrados que se indican para 15 participantes y el equipamiento suficiente para los mismos.

En el caso de que aumente el número de participantes, hasta un máximo de 30, la superficie de las aulas se incrementará proporcionalmente (según se indica en la tabla en lo relativo a m²/participante) y el equipamiento estará en consonancia con dicho aumento.

No debe interpretarse que los diversos espacios formativos identificados deban diferenciarse necesariamente mediante cerramientos.

Las instalaciones y equipamientos deberán cumplir con la normativa industrial e higiénico-sanitaria correspondiente y responderán a medidas de accesibilidad 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.

Aula virtual

Características
• La impartición de la formación mediante aula virtual se ha de estructurar y organizar de forma que se garantice en todo momento que exista conectividad sincronizada entre las personas formadoras y el alumnado participante así como bidireccionalidad en las comunicaciones. • Se deberá contar con un registro de conexiones generado por la aplicación del aula virtual en que se identifique, para cada acción formativa desarrollada a través de este medio, las personas participantes en el aula, así como sus fechas y tiempos de conexión.

Si la especialidad se imparte en **modalidad de teleformación**, cuando haya tutorías presenciales, se utilizarán los espacios formativos y equipamientos necesarios indicados anteriormente.

Para impartir la formación en **modalidad de teleformación**, se ha de disponer del siguiente equipamiento.

Plataforma de teleformación

La plataforma de teleformación que se utilice para impartir acciones formativas deberá alojar el material virtual de aprendizaje correspondiente, poseer capacidad suficiente para desarrollar el proceso de aprendizaje y gestionar y garantizar la formación del alumnado, permitiendo la interactividad y el trabajo cooperativo, y reunir los siguientes requisitos técnicos de infraestructura, software y servicios:

- **Infraestructura:**

- Tener un rendimiento, entendido como número de alumnos que soporte la plataforma, velocidad de respuesta del servidor a los usuarios, y tiempo de carga de las páginas Web o de descarga de archivos, que permita:

  a) Soportar un número de alumnos equivalente al número total de participantes en las acciones formativas de formación profesional para el empleo que esté impartiendo el centro o entidad de formación, garantizando un hospedaje mínimo igual al total del alumnado de dichas acciones, considerando que el número máximo de alumnos por tutor es de 80 y un número de usuarios concurrentes del 40% de ese alumnado.

  b) Disponer de la capacidad de transferencia necesaria para que no se produzca efecto retardo en la comunicación audiovisual en tiempo real, debiendo tener el servidor en el que se aloja la plataforma un ancho de banda mínimo de 300 Mbs, suficiente en bajada y subida.

- Estar en funcionamiento 24 horas al día, los 7 días de la semana.

- **Software:**

- Compatibilidad con el estándar SCORM y paquetes de contenidos IMS.

- Niveles de accesibilidad e interactividad de los contenidos disponibles mediante tecnologías web que como mínimo cumplan las prioridades 1 y 2 de la Norma UNE 139803:2012 o posteriores actualizaciones, según lo estipulado en el capítulo III del Real Decreto 1494/2007, de 12 de noviembre.

- El servidor de la plataforma de teleformación ha de cumplir con los requisitos establecidos en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, por lo que el responsable de dicha plataforma ha de identificar la localización física del servidor y el cumplimento de lo establecido sobre transferencias internacionales de datos en los artículos 40 a 43 de la citada Ley Orgánica 3/2018, de 5 de diciembre, así como, en lo que resulte de aplicación, en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas respecto del tratamiento de datos personales y la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE.

- Compatibilidad tecnológica y posibilidades de integración con cualquier sistema operativo, base de datos, navegador de Internet de los más usuales o servidor web, debiendo ser posible utilizar las funciones de la plataforma con complementos (plug-in) y visualizadores compatibles. Si se requiriese la instalación adicional de algún soporte para funcionalidades avanzadas, la plataforma debe facilitar el acceso al mismo sin coste.

- Disponibilidad del servicio web de seguimiento (operativo y en funcionamiento) de las acciones formativas impartidas, conforme al modelo de datos y protocolo de transmisión establecidos en el anexo V de la Orden/TMS/369/2019, de 28 de marzo.

- **Servicios y soporte:**

- Sustentar el material virtual de aprendizaje de la especialidad formativa que a través de ella se imparta.

- Disponibilidad de un servicio de atención a usuarios que de soporte técnico y mantenga la infraestructura tecnológica y que, de forma estructurada y centralizada, atienda y resuelva las consultas e incidencias técnicas del alumnado. Las formas de establecer contacto con este servicio, que serán mediante teléfono y mensajería electrónica, tienen que estar disponibles para el alumnado desde el inicio hasta la finalización de la acción formativa, manteniendo un horario de funcionamiento de mañana y de tarde y un tiempo de demora en la respuesta no superior a 48 horas laborables.

- Personalización con la imagen institucional de la administración laboral correspondiente, con las pautas de imagen corporativa que se establezcan.
  Con el objeto de gestionar, administrar, organizar, diseñar, impartir y evaluar acciones formativas a través de Internet, la plataforma de teleformación integrará las herramientas y recursos necesarios a tal fin, disponiendo, específicamente, de herramientas de:

  - Comunicación, que permitan que cada alumno pueda interaccionar a través del navegador con el tutor-formador, el sistema y con los demás alumnos. Esta comunicación electrónica ha de llevarse a cabo mediante herramientas de comunicación síncronas (aula virtual, chat, pizarra electrónica) y asíncronas (correo electrónico, foro, calendario, tablón de anuncios, avisos). Será obligatorio que cada acción formativa en modalidad de teleformación disponga, como mínimo, de un servicio de mensajería, un foro y un chat.

  - Colaboración, que permitan tanto el trabajo cooperativo entre los miembros de un grupo, como la gestión de grupos. Mediante tales herramientas ha de ser posible realizar operaciones de alta, modificación o borrado de grupos de alumnos, así como creación de «escenarios virtuales» para el trabajo cooperativo de los miembros de un grupo (directorios o «carpetas» para el intercambio de archivos, herramientas para la publicación de los contenidos, y foros o chats privados para los miembros de cada grupo).

- Administración, que permitan la gestión de usuarios (altas, modificaciones, borrado, gestión de la lista de clase, definición, asignación y gestión de permisos, perfiles y roles, autenticación y asignación de niveles de seguridad) y la gestión de acciones

- Gestión de contenidos, que posibiliten el almacenamiento y la gestión de archivos (visualizar archivos, organizarlos en carpetas –directorios- y subcarpetas, copiar, pegar, eliminar, comprimir, descargar o cargar archivos), la publicación organizada y selectiva de los contenidos de dichos archivos, y la creación de contenidos.

- Evaluación y control del progreso del alumnado, que permitan la creación, edición y realización de pruebas de evaluación y autoevaluación y de actividades y trabajos evaluables, su autocorrección o su corrección (con retroalimentación), su calificación, la asignación de puntuaciones y la ponderación de las mismas, el registro personalizado y la publicación de calificaciones, la visualización de información estadística sobre los resultados y el progreso de cada alumno y la obtención de informes de seguimiento.

Material virtual de aprendizaje:

El material virtual de aprendizaje para el alumnado mediante el que se imparta la formación se concretará en el curso completo en formato multimedia (que mantenga una estructura y funcionalidad homogénea), debiendo ajustarse a todos los elementos de la programación (objetivos y resultados de aprendizaje) de este programa formativo que figura en el Catálogo de Especialidades Formativas y cuyo contenido cumpla estos requisitos:

- Como mínimo, ser el establecido en el citado programa formativo del Catálogo de Especialidades Formativas.

- Estar referido tanto a los objetivos como a los conocimientos/ capacidades cognitivas y prácticas, y habilidades de gestión, personales y sociales, de manera que en su conjunto permitan conseguir los resultados de aprendizaje previstos.

- Organizarse a través de índices, mapas, tablas de contenido, esquemas, epígrafes o titulares de fácil discriminación y secuenciase pedagógicamente de tal manera que permiten su comprensión y retención.

- No ser meramente informativos, promoviendo su aplicación práctica a través de actividades de aprendizaje (autoevaluables o valoradas por el tutor-formador) relevantes para la adquisición de competencias, que sirvan para verificar el progreso del aprendizaje del alumnado, hacer un seguimiento de sus dificultades de aprendizaje y prestarle el apoyo adecuado.

- No ser exclusivamente textuales, incluyendo variados recursos (necesarios y relevantes), tanto estáticos como interactivos (imágenes, gráficos, audio, video, animaciones, enlaces, simulaciones, artículos, foro, chat, etc.). de forma periódica.

- Poder ser ampliados o complementados mediante diferentes recursos adicionales a los que el alumnado pueda acceder y consultar a voluntad.

- Dar lugar a resúmenes o síntesis y a glosarios que identifiquen y definan los términos o vocablos básicos, relevantes o claves para la comprensión de los aprendizajes.

- Evaluar su adquisición durante y a la finalización de la acción formativa a través de actividades de evaluación (ejercicios, preguntas, trabajos, problemas, casos, pruebas, etc.), que permitan medir el rendimiento o desempeño del alumnado.

Ocupaciones y puestos de trabajo relacionados

27211027 ADMINISTRADORES DE BASES DE DATOS

2721 DISEÑADORES Y ADMINISTRADORES DE BASES DE DATOS.

Requisitos oficiales de las entidades o centros de formación

Estar inscrito en el Registro de entidades de formación (Servicios Públicos de Empleo).

Estar inscrito en el Registro de entidades de formación (Servicios Públicos de Empleo)

MÓDULO DE FORMACIÓN 1: FUNDAMENTOS DE HADOOP Y SPARK

OBJETIVO

Conocer los fundamentos de las herramientas Apache Hadoop y Apache Spark y sus utilidades en el proceso de datos.

DURACIÓN TOTAL EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 10 horas

Teleformación: Duración de las tutorías presenciales: 0 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- **Introducción a Apache Hadoop y el ecosistema Hadoop**
- Aspectos generales de Apache Hadoop
- Proceso de datos
- Ejercicios prácticos
- **Archivos de almacenamiento de Apache Hadoop**
- Componentes del Clúster Apache Hadoop
- Arquitectura HDFS
- Uso de HDFS
- **Proceso de datos en un clúster de Apache Hadoop**
- Arquitectura YARN
- Trabajando con YARN
- **Fundamentos de Apache Spark**
- ¿Qué es Apache Spark?
- Arranque del Spark Shell
- Uso de Spark Shell
- Primeros pasos con Datasets y Dataframes
- Operaciones en Dataframe
- **Trabajo con Dataframes y Schemas**
- Crear Dataframes a partir de Data Sources
- Guardar DataFrames en Data Sources
- DataFrame Schemas
- Primeros pasos con Datasets y Dataframes
- Rapidez y lentitud de ejecución
- **Análisis de datos con consultas de DataFrame**
- Consultar DataFrames con el empleo de expresiones de columna
- Agrupación y agregación de consultas
- Unión de DataFrames

Habilidades de gestión, personales y sociales

- Concienciación de la importancia del conocimiento de estas dos herramientas Apache, para poder incluirlas en el trabajo diario y su importancia y consecuencias en los objetivos de la empresa.

- Conocimiento del papel personal a jugar en la estrategia general de la empresa, gracias al conocimiento de las herramientas que se tratan en el módulo y sus consecuencias a medio y largo plazo.

MÓDULO DE FORMACIÓN 2: RDDs

OBJETIVO

Comprender el papel de los RDD dentro de Apache Spark y su uso con datos.

DURACIÓN TOTAL EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 10 horas

Teleformación: Duración de las tutorías presenciales: 0 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- **RDDs: Sumario**
- Vistazo general sobre RDD
- Operaciones con RDDs
- Creando y guardando RDDs

- RDD Data Sources

- **Transformación de datos con RDDs**
- Escritura y paso de funciones de transformación
- Ejecuciones de transformación
- Conversión entre RDDs y DataFrames
- **Agregación de datos con Pair RDDs**
- Key-Value Pair RDDs
- Map-Reduce
- Otras operaciones Pair RDD
- **Consulta y vistas de tablas mediante SQL**
- Datasets y DataFrames
- Creación de Datasets
- Ejecución y guardado de Datasets
- Operaciones de Dataset

Habilidades de gestión, personales y sociales

- Concienciación del uso de datos con RDDs y su influencia en la consecución de los objetivos de la empresa
- Conocimiento del papel personal en la buena gestión de los conocimientos de esta temática para integrarlos dentro de los objetivos integrales de la empresa

MÓDULO DE FORMACIÓN 3:	APLICACIONES CON SPARK Y ESTRUCTURAS STREAMING CON KAFKA

OBJETIVO

Conocer el funcionamiento de Spark relativo a las aplicaciones, así como el de las estructuras de streaming con Kafka y sus usos más comunes

DURACIÓN TOTAL EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 10 horas

Teleformación: Duración de las tutorías presenciales: 0 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

Creación, configuración y ejecución de aplicaciones Spark

- Creación de una aplicación Spark
- Compilar y ejecutar la aplicación
- Application Deployment Mode
- La interfaz Spark Application Web UI
- Configuración de las propiedades de la aplicación

Procesamiento distribuido Spark

-
-
-
-
-
-
  - Apache Spark en un Clúster
  - Particiones RDD
  - Ejemplo: Particionamiento en consultas
  - Etapas y Tareas
  - Planificación de tareas de ejecución
  - Ejemplo: Programar la ejecución de Catalyst
  - Ejemplo: Programar la ejecución de un RDD

Persistencia de datos distribuidos

-
-
  - Persistencia en Datasets y DataFrames
  - Persistencia en niveles de almacenamiento
  - Visualización de RDDs persistentes

Patrones comunes al procesar datos con Spark

- Casos comunes de uso de Spark

- Algoritmos de iteración en Apache Spark
- Machine Learning

- Caso práctico

Introducción a las Estructuras Streaming

-
-
-
  - Visión general de Apache Spark Streaming
  - Creación de Streaming DataFrames
  - Transformación de Dataframes
  - Ejecución de consultas Streaming

Estructuras Streaming con Apache Kafka

- Vistazo general
- Recepción de mensajes Kafka
- Envío de mensajes Kafka

Agregación y unión de Streaming Dataframes

- o Agregación Streaming
- o Unión de Streaming Dataframes

Conclusión

Procesamiento de mensajes con Apache Kafka

- ¿Qué es Apache Kafka?
- Visión general de Apache Kafka
- Escalado de Apache Kafka
- Arquitectura de un Clúster Apache Kafka
- Herramientas Apache Kafka de la linea de comandos

Habilidades de gestión, personales y sociales

- Concienciación del uso de aplicaciones con Spark y los usos de Kafka, así como dominio de las herramientas para su aportación a los objetivos de la empresa

- Conocimiento del papel personal en la buena gestión de los conocimientos de esta temática para integrarlos dentro de los objetivos integrales de la empresa

EVALUACIÓN DEL APRENDIZAJE EN LA ACCIÓN FORMATIVA

• La evaluación tendrá un carácter teórico-práctico y se realizará de forma sistemática y continua, durante el desarrollo de cada módulo y al final del curso.
• Puede incluir una evaluación inicial de carácter diagnóstico para detectar el nivel de partida del alumnado.
• La evaluación se llevará a cabo mediante los métodos e instrumentos más adecuados para comprobar los distintos resultados de aprendizaje, y que garanticen la fiabilidad y validez de la misma.

• Cada instrumento de evaluación se acompañará de su correspondiente sistema de corrección y puntuación en el que se explicite, de forma clara e inequívoca, los criterios de medida para evaluar los resultados alcanzados por los participantes.
• La puntuación final alcanzada se expresará en términos de Apto/ No Apto.