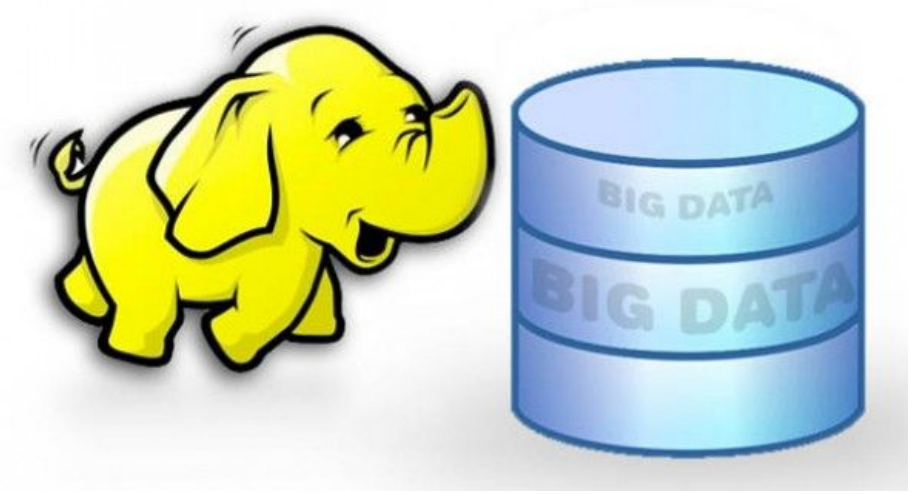
COGITI
Consejo General de la Ingeniería Técnica Industrial

COGITI
Formación e-learning

Big Data para Ingenierías

JUSTIFICACIÓN

Este curso proporciona los fundamentos necesarios para entender los principios del Big Data, así como las arquitecturas de almacenamiento y computación distribuida que lo hacen posible. El dominio de herramientas como Hadoop y Spark capacita a los ingenieros para desarrollar soluciones escalables y eficientes, mientras que el uso de tecnologías de consulta y análisis distribuido, como Hive, les permite transformar grandes volúmenes de datos en información útil. Asimismo, el conocimiento de frameworks y plataformas ampliamente utilizadas en la industria facilita la integración de los egresados en entornos profesionales altamente demandantes y tecnológicamente avanzados.



OBJETIVOS

Objetivo general:

Conocer las tecnologías disponibles para realizar estrategias de Big Data para Ingenierías, realizar un desarrollo con Spark y Hadoop y analizar datos con Hive y tecnologías distribuidas.

Objetivos específicos:

O1.Identificar los fundamentos del Big Data, comprendiendo las 5 V's y la arquitectura básica de un ecosistema distribuido.
O2.Comprender el funcionamiento del sistema de archivos distribuido de Hadoop, incluyendo la gestión de bloques, la replicación de datos y comandos de consola.
O3.Analizar el modelo de procesamiento distribuido MapReduce y el gestor de recursos YARN, evaluando la asignación eficiente de recursos en un clúster.
O4.Adquirir competencias en la configuración y mantenimiento de un clúster de Hadoop, incluyendo el monitoreo y resolución de problemas comunes.
O5.Dominar el procesamiento de datos en memoria utilizando PySpark, evaluando el uso de RDDs, DataFrames y transformaciones de datos a gran escala.
O6.Aplicar técnicas de aprendizaje automático a gran escala, validando la creación de modelos predictivos y analíticos sobre conjuntos de datos masivos.
O7.Implementar estrategias de optimización en Spark, como el manejo de skewness, caching y ajuste de parámetros para mejorar el rendimiento de los trabajos distribuidos.
O8.Evaluar la capacidad de procesar flujos de datos en tiempo real, integrando Kafka para la ingesta y Spark para el procesamiento analítico inmediato.
O9.Dominar el uso de Trino como motor de consultas SQL distribuidas para realizar análisis de datos federados, conectando múltiples fuentes de datos de forma eficiente sin necesidad de mover la información.

CONTENIDOS

1. INTRODUCCIÓN
2. BATCH PROCESSING
3. DESARROLLO SPARK Y HADOOP
4. CIENCIA DE DATOS
5. ANÁLISIS DE DATOS

80 horas /
8 semanas

Nivel de profundidad:
-*

Modalidad:
e-learning

Ampliar información:
web: www.ingenierosformacion.com
e-mail: secretaria@ingenierosformacion.com
Tlf: 985 73 28 91

* Partiendo de la base de que los cursos están dirigidos a un perfil mínimo de Ingeniero

Modalidad

Modalidad e-learning.

El curso se impartirá integralmente vía Internet en la Plataforma de Formación (https://www.ingenierosformacion.com).

Carga lectiva

80 horas

Duración

8 semanas

Precio

Curso **gratuito** por subvebción pública para la ejecución de programas de formación en el ámbito estatal, dirigidos prioritariamente a las personas ocupadas, regulada por resolución del director general del Servicio Público de Empleo Estatal.

Mínimo de alumnos

Para que la acción formativa pueda llevarse a cabo se necesitará un número mínimo de **1** alumnos.

La matrícula se cerrará cuando se hayan alcanzado un número de **80** alumnos.

Nivel de profundidad

Nivel de profundidad 0

(Partiendo de la base de que todos los cursos están dirigidos a un perfil mínimo de Ingeniero, se valorará el curso que presenta con niveles de 1 a 3 de forma que el 1 significará que el curso es de carácter básico, 2 el curso es de carácter medio y 3 el curso es de carácter avanzado.)

Perfil de Destinatarios

No es necesario tener ninguna formación específica para acceder al curso, pero las acciones formativas que componen nuestra plataforma están orientadas a la formación continua de los Ingenieros Técnicos Industriales o Graduados en Ingeniería Rama Industrial o en general cualquier ingeniero por lo que es recomendable poseer cualquiera de estas titulaciones para completar con éxito el curso.

Justificación

En la actualidad, las ingenierías se desarrollan en un contexto profundamente influenciado por la generación masiva de datos provenientes de sistemas industriales, sensores, plataformas digitales, infraestructuras críticas, procesos productivos y servicios inteligentes. Este crecimiento exponencial del volumen, variedad y velocidad de los datos ha superado las capacidades de las tecnologías tradicionales de almacenamiento y análisis, haciendo imprescindible la adopción de enfoques y arquitecturas de Big Data.

La formación en Big Data permite a los ingenieros comprender y abordar problemas complejos donde los datos son un activo estratégico. Sectores como la ingeniería industrial, informática, telecomunicaciones, eléctrica, civil, mecánica y biomédica dependen cada vez más de la analítica avanzada para optimizar procesos, mejorar la toma de decisiones, predecir fallos, diseñar sistemas inteligentes y aumentar la eficiencia operativa. Sin conocimientos en tecnologías distribuidas, los ingenieros quedan limitados frente a los desafíos reales que plantea la industria y la investigación moderna.

Este curso proporciona los fundamentos necesarios para entender los principios del Big Data, así como las arquitecturas de almacenamiento y computación distribuida que lo hacen posible. El dominio de herramientas como Hadoop y Spark capacita a los ingenieros para desarrollar soluciones escalables y eficientes, mientras que el uso de tecnologías de consulta y análisis distribuido, como Hive, les permite transformar grandes volúmenes de datos en información útil. Asimismo, el conocimiento de frameworks y plataformas ampliamente utilizadas en la industria facilita la integración de los egresados en entornos profesionales altamente demandantes y tecnológicamente avanzados.

Desde una perspectiva formativa, el curso no solo introduce conceptos teóricos, sino que promueve competencias prácticas alineadas con las necesidades actuales del mercado laboral y de la ingeniería aplicada. Al finalizar, los estudiantes estarán preparados para diseñar, implementar y evaluar estrategias de Big Data, contribuyendo de manera efectiva a proyectos de innovación, optimización y transformación digital.

En definitiva, la formación en Big Data se ha convertido en un componente esencial del perfil del ingeniero moderno. Este curso responde a esa necesidad, proporcionando una base sólida que permite a los futuros profesionales adaptarse a un entorno tecnológico en constante evolución y aportar valor real en contextos industriales, científicos y empresariales.

Objetivos

Objetivo general:

Conocer las tecnologías disponibles para realizar estrategias de Big Data para Ingenierías, realizar un desarrollo con Spark y Hadoop y analizar datos con Hive y tecnologías distribuidas.

Objetivos específicos:

O1.Identificar los fundamentos del Big Data, comprendiendo las 5 V's y la arquitectura básica de un ecosistema distribuido.
O2.Comprender el funcionamiento del sistema de archivos distribuido de Hadoop, incluyendo la gestión de bloques, la replicación de datos y comandos de consola.
O3.Analizar el modelo de procesamiento distribuido MapReduce y el gestor de recursos YARN, evaluando la asignación eficiente de recursos en un clúster.
O4.Adquirir competencias en la configuración y mantenimiento de un clúster de Hadoop, incluyendo el monitoreo y resolución de problemas comunes.
O5.Dominar el procesamiento de datos en memoria utilizando PySpark, evaluando el uso de RDDs, DataFrames y transformaciones de datos a gran escala.
O6.Aplicar técnicas de aprendizaje automático a gran escala, validando la creación de modelos predictivos y analíticos sobre conjuntos de datos masivos.
O7.Implementar estrategias de optimización en Spark, como el manejo de skewness, caching y ajuste de parámetros para mejorar el rendimiento de los trabajos distribuidos.
O8.Evaluar la capacidad de procesar flujos de datos en tiempo real, integrando Kafka para la ingesta y Spark para el procesamiento analítico inmediato.
O9.Dominar el uso de Trino como motor de consultas SQL distribuidas para realizar análisis de datos federados, conectando múltiples fuentes de datos de forma eficiente sin necesidad de mover la información.

Docente

Mario Renau Arce

Senior Data Engineer y World-Class Engineer (WCE) Senior Expert en Alstom, donde lidero el diseño y la construcción de arquitecturas de datos modernas y escalables.

Con más de diez años de experiencia especializada en ingeniería de datos, combina la excelencia técnica con una visión estratégica para transformar procesos de negocio a través del dato.

Actualmente es profesor asociado en el IE School of Human Sciences and Technology y ponente en la Escuela de Organización Industrial (EOI), impartiendo materias sobre Arquitectura de Datos, Spark y Gobernanza.

Ha formado parte de equipos estratégicos en compañías líderes como Oracle, Stratio y GFT, donde ha liderado transiciones críticas hacia arquitecturas cloud-native y sistemas en tiempo real.

Especialización Técnica: Spark, Scala, Akka, Kafka y Rust, aplicadas al procesamiento de datos a gran escala y arquitecturas reactivas.

Contenido

0. PRE-REQUISITOS
1. INTRODUCCIÓN
2. BATCH PROCESSING
3. DESARROLLO SPARK Y HADOOP
4. CIENCIA DE DATOS
5. ANÁLISIS DE DATOS

Desarrollo

El curso se desarrollará en el campus virtual de la plataforma de formación e-learning. (https://www.ingenierosformacion.com/campus/)

El día de inicio del curso los alumnos que hayan formalizado la prematrícula en la plataforma (www.ingenierosformacion.com) y hayan hecho efectivo el pago de la misma (bien por pasarela de pago, con tarjeta, directamente en el momento de la matriculación o bien por transferencia o ingreso bancario en el número de cuenta que se indica en la misma), podrán acceder al curso por medio de la plataforma, con las claves que utilizaron para registrarse como usuarios. Desde su perfil en “Mis Matrículas” podrán ver el enlace de acceso al curso.

Al ser la formación e-learning, los alumnos seguirán los distintos temas que se proponen en el curso al ritmo que ellos puedan, y en las horas que mejor se adapten a su horario.

NO se exigirá a los alumnos que estén las horas lectivas propuestas para el curso, aunque el número de horas lectivas indicado en cada curso es el recomendable para alcanzar los objetivos del curso y la adquisición de los conocimientos previstos, cada alumno va siguiendo a su ritmo los contenidos, de igual forma NO se cortará el acceso a la plataforma a aquellos alumnos que superen las horas propuestas para el curso. Sí se tendrá en cuenta que el alumno haya visto todos los contenidos o al menos la gran mayoría (más del 75 %) de los mismos durante el periodo que dura el curso, así como realizado con éxito las tareas o ejercicios, trabajos que se le vayan proponiendo durante el curso.

El alumno, además de ir estudiando los contenidos de los distintos temas, podrá participar en el foro del curso dejando sus dudas o sugerencias o intercambiando opiniones técnicas con otros alumnos, así como respondiendo aquellas que hayan dejado otros compañeros. Asimismo podrá hacer las consultas que estime oportunas al tutor del curso para que se las responda a través de la herramienta de mensajería que posee la plataforma y preferentemente en el mismo foro. Recomendamos encarecidamente el uso del foro por parte de todos los alumnos.

Para la obtención del certificado de aprovechamiento del curso el alumno tendrá que superar los objetivos mínimos marcados por el docente (superación de cuestionarios de evaluación, casos prácticos, participación, etc…).

De igual forma, los alumnos, deberán realizar la encuesta de satisfacción que nos ayudará en la mejora de la calidad de las acciones formativas que proponemos en la plataforma de formación. La encuesta estará accesible en el apartado “Mis matrículas” en la plataforma, a partir de la finalización del curso.

Matrícula

Para ampliar información mandar mail a secretaria@ingenierosformacion.com o llamando por teléfono al número 985 73 28 91.

Formación Bonificada

Este curso no es bonificable.