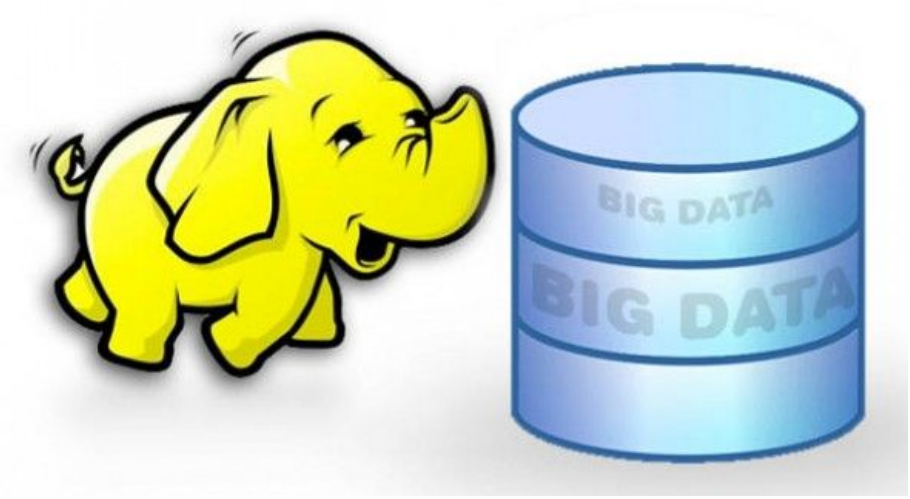
COGITI
Consejo General de la Ingeniería Técnica Industrial

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