

IFCT0118 BIG DATA PARA INGENIERÍAS

Metodología

Teleformación

Duración

100 horas

Objetivos Didácticos

Adquirir competencias en Big Data aplicadas a ingenierías, abarcando la gestión, procesamiento y análisis de grandes volúmenes de datos mediante Hadoop, Spark y herramientas de consulta como Hive e Impala, además de aprender técnicas de machine learning y procesamiento en tiempo real.

Temario del Curso

1. Fundamentos de Big Data y Procesamiento de datos

- Conocimiento de los fundamentos de Big Data
 - Definición y conceptos básicos de Big Data
 - Características de Big Data
 - Principales paradigmas de procesamiento en Big Data
 - Introducción a los sistemas de almacenamiento distribuido.
- Procesamiento de datos
 - Introducción a Hadoop y el sistema de archivos HDFS
 - Funcionamiento de MapReduce en Hadoop
 - Gestión de recursos con YARN.
 - Ingestión de datos con Apache Flume y Sqoop.
 - Configuración de seguridad en Hadoop con Kerberos.
- Conocimiento de procesos y herramientas que favorecen la eficiencia energética en el trabajo con grandes volúmenes de datos

2. Procesamiento avanzado, ciencia de datos y aprendizaje automático

- Procesamiento avanzado con Spark
 - Introducción a Spark MLlib y algoritmos básicos de machine learning.
 - Modelado, evaluación y validación de modelos de datos.
 - Librerías y funciones de MLlib para procesamiento de datos.
 - Implementación de pipelines de Machine Learning

3. Análisis y procesamiento en tiempo real

- Análisis y almacenamiento de datos
 - Introducción a Hive e Impala para consultas SQL en Hadoop.
 - Creación y gestión de tablas y bases de datos
 - Consultas avanzadas con hiveQL y optimización de Impala
 - Análisis de datos complejos con Hive y consultas de Impala
- Procesamiento en tiempo real
 - Fundamentos de Spark Streaming y fuentes de datos en tiempo real
 - Configuración y ejecución de flujos de datos en spark
 - Integración con Apache para ingestión de datos en tiempo real
 - Análisis en streaming y actualización en tiempo real