



Curso: Big Data. Arquitectura y Desarrollo

Temario

Tema 1: Fundamentos de Big Data

- Qué es big data. Características.
- Conceptos asociados a big data
- Diferentes tipos de soluciones e infraestructuras: Cloud vs On-premises

Tema 2: Puesta en marcha de un entorno big data

- Hadoop y su ecosistema
- Requisitos, instalación y puesta en marcha de Apache Hadoop
- HDFS. El Sistema de Archivos Distribuidos de Hadoop
- Paradigmas de procesamiento. MapReduce y YARN

Tema 3: Procesamiento Batch y Streaming

- Hive. Data Warehouse sobre Hadoop
- Sqoop. Transferencia de datos entre Hadoop y sistemas de bases de datos relacionales
- Flume. Recolectar y mover grandes cantidades de datos log
- Spark. Procesamiento en memoria de datos distribuidos
- Flink. Procesamiento de flujos de datos en tiempo real
- Kafka. Plataforma de streaming distribuido
- Storm. Procesamiento distribuido en tiempo real
- Samza. Framework de procesamiento distribuido de streams
- SQL sobre Hadoop: Impala, Presto, Drill

Tema 4: Bases de datos NoSQL

- NoSQL vs Bases de datos relacionales
- Teorema CAP
- ACID
- Modelos de datos en NoSQL
- HBase
- MongoDB
- Cassandra
- Neo4J

Tema 5: Contenedores y virtualización

- Conceptos sobre contenedores y virtualización
- Docker. Creación y despliegue de contenedores
- Kubernetes. Gestión de aplicaciones en contenedores



Tema 6: Infraestructura en la nube. Cloud Computing

- Por qué las empresas están migrando a la nube
- Consideraciones de diseño de aplicaciones para una infraestructura cloud
- Cloud vs On-premises
- Decisiones basadas en costes

Tema 7: Seguridad y monitorización en sistemas big data

- Conceptos sobre seguridad en entornos big data
- Herramientas de planificación y optimización de clústers
- Netflow
- Nagios

Tema 8: Arquitectura Frontend

- ReactJS
- Redux

Tema 9: Arquitectura Backend

- RESTful API
- Redis

Tema 10: Machine Learning e Inteligencia Artificial

- Tipos de algoritmos
- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje de refuerzo
- Inteligencia Artificial
- Machine Learning con Spark MLlib

Tema 11: Elasticsearch

- ¿Qué es Elastic Stack?
- Instalación de Elasticsearch
- Instalación de Kibana
- Herramientas de seguridad
- Consultas en Elasticsearch



Tema 12: Visualización de datos

- Matplotlib
- D3.js
- Tableau

Tema 13: Introducción a la ciencia de datos

- Fundamentos de ciencia de datos
- Análisis de datos con Python

Tema 14: Algorithmic Trading. Comercio basado en algoritmos

- ¿Qué es el comercio basado en algoritmos?
- ¿Cómo se construye un asesor experto?

Tema 15: Proyecto práctico. Construcción de un sistema de recomendación