

Programa del curso

Nombre del curso:	Sistemas intensivos de datos
Créditos:	4
Versión PDF	Click Aquí

Descripción

Este curso se enfoca en la arquitectura de datos y las habilidades de ingeniería de datos necesarias para el procesamiento distribuido a gran escala. Los estudiantes utilizarán herramientas del ecosistema de Apache Spark para construir, optimizar y gestionar pipelines de datos. A lo largo del curso, se trabajará con conjuntos de datos representativos de escenarios reales y se aplicarán mejores prácticas de la industria, con énfasis en gobernanza, formatos de almacenamiento y generación de características (feature engineering). Al finalizar el curso, los estudiantes serán capaces de:

1. Procesar datos a gran escala utilizando Apache Spark.
2. Diseñar arquitecturas de datos para soluciones del mundo real.
3. Implementar pipelines de datos con Delta Lake, MLFlow y Unity Catalog.
4. Gestionar la ingeniería de características y el ciclo de vida de modelos de machine learning.

Este curso es una evolución del curso de análisis con BigData. Los que vieron ese curso no podrán ver el nuevo curso.