

CURSO GRATUITO

Introducción al Big Data con Apache Spark

impartido por Dr. Roberto Tardío Olmos

(Responsable de Big Data en StrateBI)

Destinado a estudiantes de la UPM de áreas TIC, y en especial, a alumnos de la asignatura Bases de Datos Avanzadas. Para estudiantes de Grado de la ETSISI podrá tener el reconocimiento de 1 crédito. Los interesados deberán enviar un email con asunto "Curso BigData" adjuntando carta o texto de motivación e indicando las asignaturas del área de bases de datos cursadas, al profesor Pedro Pablo Alarcón (pedropablo.alarcon@upm.es) antes del día 3 de julio de 2023.

Lugar impartición: Laboratorio 1301 de la ETSISI (UPM)

Fechas: 10 al 13 de julio de 2023 en horario de 16:00 a 20:00

Descripción curso: Las aplicaciones de Inteligencia de Negocio (BI), como los cuadros de mando o los modelos predictivos en Machine Learning (ML), son cada vez más demandadas por las empresas para explotar uno de sus principales activos, los datos. Además, cada vez es más frecuente encontrarse con escenarios de Big Data, donde el volumen datos, su naturaleza diversa y compleja, y la necesidad de procesarlos de manera muy eficiente, incluso en tiempo real, suponen un reto para las tecnologías y métodos existentes. Apache Spark es una de las tecnologías más potentes y estandarizadas para afrontar con éxito este tipo de escenarios. El objetivo de este curso es introducir al desarrollo en Spark para escenarios Big Data y ML, definiendo previamente estos conceptos para aquellos alumnos de ingeniería sin conocimientos previos en la materia.

Impartido por: Dr. Roberto Tardío Olmos (Responsable de Big Data en StrateBI)

Biografía: Roberto Tardío es Doctor en Ingeniería Informática y trabaja como Responsable de Big Data en StrateBI. Esta empresa, con sedes en Madrid y Alicante, cuenta con más de 15 años de experiencia en el desarrollo de soluciones Business Intelligence, Big Data y Machine Learning para grandes multinacionales como Repsol, Toyota Financial Services, Burger King o Abanca. Para más información visita <https://www.stratebi.com/> y consulta las ofertas de prácticas de la UPM.



Contenido del curso:

- Día 1
 - Apertura, presentación del curso y sobre StrateBI
 - Conceptos básicos de sistemas BI para Big Data. ¿Qué es el Big Data?
 - Tecnologías actuales y metodología para el diseño de arquitecturas BI/Big Data
 - Casos de uso reales
 - Introducción a Apache Spark: Arquitectura, características principales y lenguajes de programación soportados (Python, Scala, SQL,...)
- Día 2
 - Introducción a Jupyter Notebooks
 - Introducción a Python: Conceptos básicos, tipos de datos, Colecciones, estructura de control y clases
 - Prácticas:
 - Instalación de una máquina virtual (Virtual Box) con Spark y Jupyter Notebooks
 - Ejercicios de Python con Jupyter Notebooks.
- Día 3
 - PySpark SQL: Desarrollo en Spark usando Python y Data Frames
 - Lectura/Escritura de datos (CSV, JSON, Parquet, JDBC, NoSQL...)
 - Operaciones fundamentales: *select* , *filter/where*, *group by*, *distinct*, *group by*, *join*, *union*...
 - Prácticas:
 - Ejercicios Spark con Jupyter Notebooks.
- Día 4
 - Introducción al Machine Learning con Spark
 - Definición e introducción al Machine Learning
 - Creación y ejecución de Pipelines de ML en Spark: Dataframes, Estimators, Transformers, Pipelines
 - Desarrollo de modelos de clasificación y regresión con Spark
 - Prácticas: Ejemplo práctico guiado de Clasificación con Spark.
 - Introducción a Databricks: Versión comercial de Spark en la nube de Azure, AWS y Google. Demo.

Para el reconocimiento de 1 crédito, además de asistir a todas las sesiones, se deberá entregar un trabajo práctico adicional estimado en unas cuatro horas de realización.