FICHA IDENTIFICATIVA

Datos de la Asignatura

Código	44658
Nombre	Big Data
Ciclo	Máster
Créditos ECTS	6.0
Curso académico	2019 - 2020

Titulación(es)

Titulación	Centro	Curso	Periodo
2221 - Máster Universitario en Ciencia de Datos	Escuela Técnica Superior de Ingeniería	1	Segundo cuatrimestre

Materias

Titulación	Materia	Carácter
2221 - Máster Universitario en Ciencia de Datos	10 - Big data	Obligatoria

Coordinación

Nombre	Departamento
LOZANO IBAÑEZ, MIGUEL	240 - Informática
MUÑOZ MARI, JORDI	242 - Ingeniería Electrónica

RESUMEN

Esta asignatura se divide en tres bloques temáticos. En el primero, se presentan los principales problemas asociados al procesamiento de grandes volúmenes de datos así como las plataformas computacionales capaces de manejar estos datos eficientemente. El primer bloque se cierra revisando casos reales de éxito derivados de desarrollos empresariales en este contexto.

La segunda parte se centra en la presentación del modelo MapReduce para la explotación de bases de datos sobre plataformas de Big Data.

El tercer bloque se orienta al análisis y desarrollo de aplicaciones basadas en aprendizaje máquina sobre plataformas de Big Data. En esta parte, se revisan los lenguajes ofrecidos por las principales plataformas Big Data para la implementación de aplicaciones, prestando especial atención a Python y R.

CONOCIMIENTOS PREVIOS

Relación con otras asignaturas de la misma titulación

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

Otros tipos de requisitos

Análisis de señales
Análisis exploratorio de datos
Aprendizaje máquina (I)
Aprendizaje máquina (II)
Visualización avanzada de datos

COMPETENCIAS (RD 1393/2007) // RESULTADOS DEL APRENDIZAJE (RD 822/2021)

2221 - Máster Universitario en Ciencia de Datos

- Que los/las estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- Que los/las estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo
- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- Ser capaces de valorar la necesidad de completar su formación técnica, científica, en lenguas, en informática, en literatura, en ética, social y humana en general, y de organizar su propio autoaprendizaje con un alto grado de autonomía
- Capacidad de acceso y gestión de la información en diferentes formatos para su posterior análisis con el fin de obtener conocimiento a partir de datos.
- Ser capaces de acceder a herramientas de información (bibliográficas y de empleo) y utilizarlas apropiadamente.
- Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio, aplicando los conocimientos adquiridos en la identificación de salidas profesionales y yacimientos de empleo.
- Conocer y utilizar los distintos modelos de almacenamiento de datos y los sistemas de gestión de las bases de datos utilizando lenguajes de programación de definición, consulta y manipulación de los mismos.
- Seleccionar, atendiendo a criterios de eficiencia, escalabilidad, tolerancia a fallos y adecuación al entorno de producción el paradigma de datos óptimo en soluciones Big Data. Entender como las técnicas Big Data se utilizan para soportar y realizar la toma
  de decisiones basadas en datos.

RESULTADOS DE APRENDIZAJE (RD 1393/2007) // SIN CONTENIDO (RD 822/2021)

Conocer el contexto actual y casos de éxito en Big Data.
Conocer e implementar las principales fases empleadas en la resolución de problemas de análisis de grandes volúmenes de datos.
Conocer y manejar las principales plataformas y herramientas para procesar grandes volúmenes de información.
Procesar y analizar datos numéricos (señales) y alfanuméricos (texto) con plataformas de Big Data.
Conocer e implementar el modelo de MapReduce para análisis de grandes volúmenes de datos (bases de datos).
Aplicar los modelos de aprendizaje estudiados a problemas de BigData

DESCRIPCIÓN DE CONTENIDOS

1. Introducción al Big Data

Contexto, problemas y soluciones de BigData

2. Plataformas de BigData

Revisión de las plataformas y herramientas actuales para el análisis de grandes volumenes de datos

3. Casos de éxito: BigData en la empresa

Casos de éxito del BigData en el sector empresarial: modelos de negocio

4. Plataformas BigData: Hadoop. MapREduce

1 Entorno de programación y ejemplos básicos de aplicación de MapReduce: desarrollo y puesta en práctica de aplicaciones.
2. Arquitectura, componentes y entorno software de Hadoop.
3. Hadoop y Python

5. Proyectos relacionados con Hadoop. Kafka

1.Kafka: Sistemas basados en paso de mensajes
2.Arquitectura de Kafka (Streaming platform).
3.Desarrollo de aplicaciones.

6. Spark

1. DataFrames, Datasets y SQL.
2. API Spark en Python.
3. SparkR: API Spark en R.
4. Machine Learning Library: Mlib.
5. Spark Streaming.

7. Casos prácticos

Google Earth Engine.

VOLUMEN DE TRABAJO

ACTIVIDAD	Horas	% Presencial
Clases teórico-prácticas	60,00	100
Elaboración de trabajos individuales	20,00	0
Estudio y trabajo autónomo	12,00	0
Lecturas de material complementario	3,00	0
Preparación de actividades de evaluación	12,00	0
Preparación de clases de teoría	20,00	0
Preparación de clases prácticas y de problemas	13,00	0
Resolución de casos prácticos	10,00	0
TOTAL	**150,00**	

METODOLOGÍA DOCENTE

Las clases combinarán la parte teórica con la práctica sin distinción entre sesiones dedicadas a teoría y práctica. Todas las sesiones se impartirán en aula de informática.

En las sesiones teóricas se realizará un desarrollo expositivo de la materia con la participación del estudiante en la resolución de cuestiones puntuales. Las sesiones prácticas irán sincronizadas con las teóricas, y en ellas el aprendizaje se realizará mediante la resolución de problemas, ejercicios y casos de estudio a través de los cuales se adquieren competencias sobre los diferentes aspectos de la materia.

EVALUACIÓN

La evaluación del aprendizaje de los conocimientos y competencias conseguidas por los estudiantes se hará de forma continuada a lo largo del curso, y constará de los siguientes bloques de evaluación:

1. Ejercicios y trabajos entregados durante el curso y/o exámenes parciales: 80% de la nota final**.**
2. Examen final: 20% de la nota final.

Las calificaciones obtenidas en el apartado 1 se conservarán en las dos convocatorias del curso académico en que hayan sido realizadas, dado que su evaluación sólo es posible en el periodo de docencia.

REFERENCIAS

Básicas

- Bowles, M. (2015). Machine Learning in Python: Essential Techniques for Predictive Analysis. Ed. Wiley
- Wes McKinney, W (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython. Ed. O'Reilly Media
- Scalable Big Data Architecture. Bahaaldine Azarmi. Apress, (2016)
- Amit Nandi. (2015).Spark for Python Developers. Ed. Packt Publishing
- Matei Zaharia, Bill Chambers (2017). Spark: The Definitive Guide. Big data processing made simple. Ed. O'Reilly Media.
- Matei Zaharia, Holden Karau, Andy Konwinski, Patrick Wendell (2015). Learning Spark Lightning-Fast Big Data Analysis. Ed. OReilly Media.
- Tom White. (2015) Hadoop: The Definitive Guide. 4th Edition. Storage and Análysis at Internate Scale. O'reilly.
- Zachary Radtka and Donald Miner. (2016) Hadoop with Python. O'reilly
- Kafka: The Definitive Guide: Real-time data and stream processing at scale Neha Narkhede , OReally 2017

ADENDA COVID-19

Esta adenda solo se activará si la situación sanitaria lo requiere y previo acuerdo del Consejo de Gobierno

1. **Contenidos.**

Se mantienen todos los contenidos inicialmente programados en la guía docente para las sesiones teóricas.

1. **2. Volumen de trabajo y planificación temporal de la docencia.**

Mantenimiento de la planificación temporal docente tanto en días como en horario.

1. **3. Metodología docente.**

Se mantiene la metodología docente.

Sistema de tutorías.

Se mantiene el programa de tutorías virtuales (atención en 48 horas laborables máximo por correo electrónico) y en el horario de tutorías presenciales disponibilidad del profesor en el canal de la asignatura de Microsoft Teams.

4. Evaluación.

Se mantiene la ponderación inicial.

El examen final se realizará mediante un cuestionario en el aula virtual.

5. Bibliografía.

La bibliografía recomendada se mantiene.