

# GUIA DOCENTE

CURSO 2024-25

## MÁSTER UNIVERSITARIO EN ECONOMÍA, FINANZAS Y COMPUTACIÓN

### DATOS DE LA ASIGNATURA

**Nombre:**

SISTEMA DE PROCESADO MASIVO DE DATOS Y BIG DATA

**Denominación en Inglés:**

MASSIVE DATA PROCESSING SYSTEMS AND BIG DATA

**Código:**

1210102

**Tipo Docencia:**

Presencial

**Carácter:**

Optativa

**Horas:**

**Totales**

**Presenciales**

**No Presenciales**

**Trabajo Estimado**

150

45

105

**Créditos:**

<b>Grupos Grandes</b>	<b>Grupos Reducidos</b>			
	<b>Aula estándar</b>	<b>Laboratorio</b>	<b>Prácticas de campo</b>	<b>Aula de informática</b>
4.8	1.2	0	0	0

**Departamentos:**

**Áreas de Conocimiento:**

TECNOLOGIAS DE LA INFORMACION

CIENCIA DE LA COMPUTACION E INTELIG. ARTIFICIAL

TECNOLOGIAS DE LA INFORMACION

LENGUAJES Y SISTEMA INFORMATICOS

**Curso:**

**Cuatrimestre**

1º - Primero

Anual

**DATOS DEL PROFESORADO (\*Profesorado coordinador de la asignatura)**

<b>Nombre:</b>	<b>E-mail:</b>	<b>Teléfono:</b>
* Antonio Peregrin Rubio	peregrin@dti.uhu.es	959 217 653
Francisco Alfredo Marquez Hernandez	alfredo.marquez@dti.uhu.es	959 217 641
Manuel Jesus Mana Lopez	manuel.mana@dti.uhu.es	959 217 389

**Datos adicionales del profesorado (Tutorías, Horarios, Despachos, etc... )**

Antonio Peregrín Rubio, peregrin@dti.uhu.es, Telf. 959 21 7653, Despacho: ETSI (El Carmen), ETP-156

Manuel Jesús Maña López, manuel.mana@dti.uhu.es, Telf. 959 21 7389, Despacho: ETSI (El Carmen), ETP-135

Francisco Alfredo Márquez Hernández, alfredo.marquez@dti.uhu.es, Telf. 959 21 7641, Despacho: ETSI (El Carmen), ETP-129

Horarios de tutorías: <https://guiadocente.uhu.es/tutoria/titulacion>

## DATOS ESPECÍFICOS DE LA ASIGNATURA

### 1. Descripción de Contenidos:

#### 1.1 Breve descripción (en Castellano):

Sistemas de almacenamiento masivo y arquitecturas, tecnologías de los sistemas de procesado masivo de datos, principios de la minería de texto, principios de la lingüística computacional, introducción a la ciencia de los datos y el big data, framework Spark, procesamiento con Python en entornos Spark, conectores con bases de datos, uso del aprendizaje automático en entornos de big data, y aplicaciones de la tecnología de big data con datos de economía.

#### 1.2 Breve descripción (en Inglés):

Mass storage systems and architectures, technologies of mass data processing systems, principles of text mining, principles of computational linguistics, introduction to data science and big data, Spark framework, Python processing in Spark environments, connectors with databases, use of machine learning in big data environments, and applications of Big Data technology with economics data.

### 2. Situación de la asignatura:

#### 2.1 Contexto dentro de la titulación:

Esta asignatura proporciona el marco formal para entender el proceso de almacenamiento, manipulación y extracción de conocimiento a partir de grandes volúmenes de datos estructurados y no estructurados, y de texto. Se introducirá a los alumnos en las tecnologías del Big Data, los paradigmas entorno a él, sus retos y las tecnologías adyacentes y relacionadas con el mismo como la Computación en la Nube, el Internet de las Cosas o las Bases de Datos No Estructuradas. Asimismo, se buscará la conexión entre las tecnologías del Big Data con la Analítica de Datos, la Inteligencia de Negocio, y la Minería de Datos, y las herramientas disponibles para llevarlas a cabo, las infraestructuras necesarias, y las alternativas para las empresas. Se abordarán las técnicas fundamentales de procesamiento de texto y las aplicaciones más relevantes de la minería de textos. Los conceptos estudiados en esta asignatura son de carácter transversal para las asignaturas del Máster en las que se tenga acceso a grandes volúmenes de información.

#### 2.2 Recomendaciones

Prerequisitos: Conocimientos básicos de programación en Python (cursados en a asignatura “Programación”).

Recomendaciones:

- Poseer destreza para buscar información útil en la Red
- Poseer conocimientos básicos de inglés
- Saber manejar fuentes bibliográficas

### 3. Objetivos (resultado del aprendizaje, y/o habilidades o destrezas y conocimientos):

El objetivo general es dotar al alumno de los conocimientos fundamentales necesarios para comprender y utilizar sistemas de procesado y almacenamiento masivo de datos. Este objetivo general se puede refinar en los siguientes objetivos específicos:

- Conocer las tecnologías del Big Data, el alcance del problema, y los recursos necesarios para utilizarlo.
- Comprender la relación entre el tratamiento masivo de datos, y otras tecnologías de la información emergentes directamente relacionadas.
- Conocer los recursos disponibles que integran Big Data con Analítica de Datos, Inteligencia de Negocio, etc.
- Comprender la dificultad de analizar grandes bases de datos de información textual y extraer conocimiento de las mismas
- Conocer los retos más relevantes y las áreas de aplicación de la minería de textos
- Conocer las herramientas más importantes que existen actualmente y que pueden aplicarse en algunos de los problemas de minería de textos
- Comprender las técnicas fundamentales relacionadas con el agrupamiento, clasificación, análisis de sentimientos y minería de opiniones

### 4. Competencias a adquirir por los estudiantes

#### 4.1 Competencias específicas:

**COM3:** Organizar, planificar y desarrollar trabajos y proyectos propios de su ámbito científico o profesional.

**COM4:** Gestionar adecuadamente la información adquirida expresando conocimientos avanzados y demostrando, en un contexto de investigación científica y tecnológica o altamente especializado, una comprensión detallada y fundamentada de los aspectos teóricos y prácticos y de la metodología de trabajo en el campo de estudio

**COM5:** Dominar el proyecto académico y profesional, habiendo desarrollado la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro de su ámbito temático, en contextos interdisciplinares y, en su caso, con un alto componente de transferencia del conocimiento

**COM6:** Desarrollar una actitud y una aptitud de búsqueda permanente de la excelencia en el quehacer académico y en el ejercicio profesional futuro.

#### 4.2 Competencias básicas, generales o transversales:

**CO4:** Conocer los principios de programación informática y de gestión de bases de datos

**HD05:** Manejar los Sistemas de Información en la empresa

**HD06:** Dispone de habilidades de aprendizaje para continuar aprendiendo de forma autodirigida y autónomo autónoma

## 5. Actividades Formativas y Metodologías Docentes

### 5.1 Actividades formativas:

- Clases magistrales teóricas
- Clases prácticas
- Evaluación

### 5.2 Metodologías Docentes:

- Clases magistrales participativas
- prácticas asistidas con software
- Discusión de artículos científicos
- Participación en tutorías virtuales y foros
- Trabajos dirigidos

### 5.3 Desarrollo y Justificación:

Todo el material oportuno para el seguimiento de las clases teóricas y prácticas estará disponible en la plataforma de teleformación. Para las clases teóricas, los recursos que se utilizarán son la pizarra (tradicional y en su versión electrónica), las proyecciones de presentaciones con la ayuda del ordenador y material suplementario suministrado por el profesorado (fotocopias, archivos electrónicos, etc.). En las clases prácticas se aplicarán los contenidos abordados en las clases teóricas, se hará hincapié en los mecanismos de resolución, sus limitaciones y ventajas, así como un análisis crítico de los resultados alcanzados. Estas clases prácticas serán interactivas y la participación del/la alumno/a será tenida en cuenta a la hora de valorar su adaptación al grado de aprendizaje.

## 6. Temario Desarrollado

### Módulo I: BIG DATA (BD) (Profesores Dr. A. Peregrín y Dr. F.A. Márquez)

- Tema 1.1. Introducción al Big Data Motivación y Objetivos. Tecnologías de la Información emergentes relacionadas con Big Data. La Ciencia de los Datos y el Big Data en relación con la Analítica de Datos, la Inteligencia de Negocio, el Aprendizaje Automático y la Minería de Datos.
- Tema 1.2. Minería de Datos Computacional Definiciones, proceso, algoritmos, validación, etc.
- Tema 1.3. Tecnologías para Big Data I Paradigma MapReduce y Apache Hadoop: Modelos de procesamiento y almacenamiento.
- Tema 1.4. Ecosistema Hadoop Elementos del ecosistema, Hive, Pig, Sqoop, Flume, etc.
- Tema 1.5. Tecnologías para Big Data II Apache Spark, DataFrames, Machine Learning con

Spark.

- Tema 1.6. Ecosistema Spark Esquema del ecosistema Spark, SparqSQL, GraphX, etc.
- Tema 1.7. Introducción al Deep Learning y su relación con el Big Data.

## Módulo II: PROCESAMIENTO DEL TEXTO (MT) (Profesor Dr. M.J. Maña)

- Tema 2.1. Acceso a la información textual Motivación: la era del conocimiento, el acceso a la información y la minería de textos. Características del texto. La investigación en acceso a la información textual en la UHU.
- Tema 2.2. Análisis textual Introducción. Análisis léxico. Eliminación de palabras vacías. Reducción a la raíz. Análisis morfológico. Análisis sintáctico. Reconocimiento de entidades nombradas. Índice invertido: construcción y búsqueda. Recursos léxicos
- Tema 2.3. Minería de textos Motivación: definiciones y aplicaciones de la minería de textos. Evaluación. Clasificación. Agrupación. Análisis de sentimiento y minería de opiniones.

## 7. Bibliografía

### 7.1 Bibliografía básica:

#### Módulo I

- Large-Scale Data Analytics with Python and Spark: A Hands-on Guide to Implementing Machine Learning Solutions. Isaac Triguero, Mikel Galar. Cambridge University Press, 2023.
- Big Data. La revolución de los datos masivos. Viktor Mayer-Schönberger, Kenneth Cukier, HoughtonMifflin Harcourt, 2013
- Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. IBM. Paul Zikopoulos, Chris Eaton. McGraw Hill Professional, 29/05/2015.

#### Módulo II

- R. Baeza-Yates y B. Ribeiro-Neto. Modern Information Retrieval: The Concepts and Technology behind Search. Second Edition. Pearson, 2011.
- Christopher D. Manning, Prabhakar Raghavan y Hinrich Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.

### 7.2 Bibliografía complementaria:

#### Módulo I

- A. Fernandez, S. Río, V. López, A. Bawakid, M.J. del Jesus, J.M. Benítez, F. Herrera, Big Data with Cloud Computing: An Insight on the Computing Environment, MapReduce and Programming Frameworks. WIREs Data Mining and Knowledge Discovery 4:5 (2014) 380-409.
- Hadoop, Soluciones Big Data., Boris Lublinsky, Kevin Smith, Alex Yakubovich, Anaya Multimedia 2014
- Del Cloud Computing al Big Data: Visión introductoria para jóvenes emprendedores. Jordi Torres i Vinnals. Editorial UOC – PID\_00194204. Primera edición: septiembre 2012. FUOC, 2012. v.3.0 España de Creative Commons.
- Big Data, Técnicas, Herramientas y Aplicaciones, María Pérez Marqués, 2015.

## Módulo II

- Bo Pang y Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. Journal Foundations and Trends in Information Retrieval, vol. 2: 1-2, pp. 1-135.

## 8. Sistemas y criterios de evaluación

### 8.1 Sistemas de evaluación:

- Asistencia y participación en clase
- Examen final de la asignatura con preguntas teóricas y problemas
- Trabajos e informes realizados por el alumno o grupo de trabajo

### 8.2 Criterios de evaluación relativos a cada convocatoria:

#### 8.2.1 Convocatoria I:

En la convocatoria ordinaria I (Junio), salvo que se solicite expresamente, se empleará la Evaluación Continua, la cual consiste en:

- Los conocimientos teóricos y prácticos (grado de capacidad de resolución de problemas y aplicación de los contenidos teóricos a la práctica, y grado de conocimiento y comprensión) de la materia se evaluarán mediante una prueba o examen (de tipo test en la convocatoria I, escrito de preguntas cortas en las demás convocatorias) presencial e individual en el aula de informática a través de la plataforma de apoyo a la enseñanza presencial Moodle, según la convocatoria oficial para esta materia. En dicho examen, además del ordenador del aula, no se permitirá el uso de ningún otro dispositivo electrónico, y no se empleará ningún elemento documental externo. El peso de este examen será de un 50%, y el alumno debe obtener al menos un 4 sobre 10 puntos para el cómputo de la nota final considerando esta parte superada. La duración del examen será de 1h.
- La participación activa del estudiante en las actividades propuestas (seguimiento del estudiante), es decir, su contribución a los debates, foros de ideas, respuesta a ejercicios, actividades académicas, etc., será un elemento que se valorará con un peso de un 50%, a criterio de los profesores de la asignatura. El alumno debe obtener al menos un 4 sobre 10 puntos para el cómputo de la nota final considerando esta parte superada, y mantenerla en un mismo curso académico entre convocatorias ordinarias.

Nota sobre la calificación “Matrícula de Honor”: si existieran más alumnos con una calificación que les permita aspirar a la matrícula de honor (10 sobre 10 puntos en la media ponderada), es decir, en caso de equidad, la calificación de Matrícula de Honor se asignará basándose en la mayor participación en clase e implicación del alumno en la asignatura (en todo tipo de sesiones), a juicio de los profesores de la asignatura.

#### 8.2.2 Convocatoria II:

Mismas consideraciones y descripción que en la convocatoria I (Junio) para la convocatoria II (Septiembre).

#### 8.2.3 Convocatoria III:

Mismas consideraciones y descripción que en la convocatoria I (Junio) y convocatoria II (Septiembre) para la convocatoria III (Diciembre).

#### 8.2.4 Convocatoria extraordinaria:

Sólo será posible realizar en la convocatoria extraordinaria IV (Noviembre) evaluación única, por tanto, no será necesario solicitarla.

Consiste en:

- Un examen presencial e individual (en el que se medirá el grado de capacidad de resolución de problemas y aplicación de los contenidos teóricos a la práctica, y grado de conocimiento y comprensión de la materia) escrito en papel (que se le suministrará) y se rellenará a bolígrafo (azul o negro), consistente en preguntas cortas, largas y problemas combinados, en la fecha, hora y lugar en la que se convoque oficialmente para esta materia. No se permitirá el uso de dispositivos electrónicos (teléfonos móviles, tabletas, auriculares, ordenadores, relojes inteligentes, wearables, etc.) ni cualquier otro material didáctico que no haya sido expresamente autorizado por el profesor. El peso del examen es de un 100% en la calificación. La duración de este examen será de entre 2h y 2h y 30m

#### 8.3 Evaluación única final:

##### 8.3.1 Convocatoria I:

El alumno debe solicitarlo (mediante escrito firmado y entregado a través del Registro General de la Universidad (presencial o telemático) dirigido al Departamento de Tecnologías de la Información y a la atención del profesor coordinador de la materia, Antonio Peregrín Rubio (recomendándose, enviar también copia de dicho escrito al profesor por correo electrónico (peregrin@dti.uhu.es) sólo como medida complementaria, y por agilidad administrativa).

Consiste en:

- Un examen presencial e individual (en el que se medirá el grado de capacidad de resolución de problemas y aplicación de los contenidos teóricos a la práctica, y grado de conocimiento y comprensión de la materia) escrito en papel (que se le suministrará) y se rellenará a bolígrafo (azul o negro), consistente en preguntas cortas, largas y problemas combinados, en la fecha, hora y lugar en la que se convoque oficialmente para esta materia. No se permitirá el uso de dispositivos electrónicos (teléfonos móviles, tabletas, auriculares, ordenadores, relojes inteligentes, wearables, etc.) ni cualquier otro material didáctico que no haya sido expresamente autorizado por el profesor. El peso del examen es de un 100% en la calificación. La duración de este examen será de entre 2h y 2h y 30m

##### 8.3.2 Convocatoria II:

Mismas consideraciones y descripción que para la evaluación única final de la convocatoria I.

##### 8.3.3 Convocatoria III:

Mismas consideraciones y descripción que para las evaluaciones únicas finales de la convocatoria I y II.

#### 8.3.4 Convocatoria Extraordinaria:

El alumno no debe solicitarla en este caso, pues es la única disponible para esta convocatoria.

Consistirá en el mismo tipo de ejercicio que se describe para la evaluación única final de la convocatoria I.

9. Organización docente semanal orientativa:							
Fecha	Grupos Grandes	G. Reducidos				Pruebas y/o act. evaluables	Contenido desarrollado
		Aul. Est.	Lab.	P. Camp	Aul. Inf.		
21-10-2024	0	0	0	0	0		
28-10-2024	0	0	0	0	0		
04-11-2024	0	0	0	0	0		
11-11-2024	0	0	0	0	0		
18-11-2024	0	0	0	0	0		
25-11-2024	0	0	0	0	0		
04-11-2024	0	0	0	0	0		
11-11-2024	0	0	0	0	0		
18-11-2024	0	0	0	0	0		
06-01-2025	0	0	0	0	0		
13-01-2025	0	0	0	0	0		
20-01-2025	0	0	0	0	0		
27-01-2025	0	0	0	0	0		
03-02-2025	0	0	0	0	0		
10-02-2025	0	0	0	0	0		
17-02-2025	0	0	0	0	0		
24-02-2025	0	0	0	0	0		
03-03-2025	0	0	0	0	0		
10-03-2025	0	0	0	0	0		
17-03-2025	0	0	0	0	0		
24-03-2025	0	0	0	0	0		
31-03-2025	0	0	0	0	0		
07-04-2025	0	0	0	0	0		

21-04-2025	0	6	0	0	2		BD: Tema 1.1 MT: Tema 2.1
28-04-2025	0	6	0	0	2		BD: Tema 1.2 MT: Tema 2.2
05-05-2025	0	6	0	0	2		BD: Tema 1.3 MT: Tema 2.2
12-05-2025	0	6	0	0	2	BD: Práctica I (Eco)	BD: Tema 1.4 MT: Tema 2.3
19-05-2025	0	6	0	0	2	BD: Práctica II (MLlib)	BD: Tema 1.5
26-05-2025	0	4	0	0	1	MT: Trabajos BD: P-I y P-II	BD: Tema 1.6 MT: Tema 2.3
02-06-2025	0	0	0	0	0		

**TOTAL**      **0**      **34**      **0**      **0**      **11**