

1. DATOS BÁSICOS

Asignatura	Grandes Volúmenes de Datos
Titulación	Grado en Ingeniería Informática
Escuela/ Facultad	Arquitectura, Ingeniería y Diseño
Curso	Cuarto
ECTS	6 ECTS
Carácter	Obligatorio
Idioma/s	Castellano
Modalidad	Presencial / Online
Semestre	Primer semestre
Curso académico	2022/2023
Docente coordinador	Enrique Puertas

2. PRESENTACIÓN

La materia de Grandes Volúmenes de datos es una materia obligatoria dentro del Grado en Ingeniería Informática, con una carga lectiva de 6 créditos ECTS.

Esta asignatura está dentro de la materia de Computación y se plantea como una continuación de las asignaturas de Inteligencia Artificial, Sistemas Inteligentes y Bases de Datos, para profundizar en las soluciones al reto que plantean los grandes datos en esas áreas anteriormente mencionadas. Esta asignatura introduce contenidos avanzados relacionados con la infraestructura necesaria para procesar de forma distribuida grandes cantidades de datos, y los algoritmos de Aprendizaje Automático y Visualización para extraer valor de los datos.

3. COMPETENCIAS Y RESULTADOS DE APRENDIZAJE

Competencias básicas:

- **CB4:** Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

Competencias transversales:

- **CT10:** Iniciativa y espíritu emprendedor: Capacidad para acometer con resolución acciones difíciles o azarosas. Capacidad para anticipar problemas, proponer mejoras y perseverar en su consecución. Preferencia por asumir y llevar a cabo actividades.

Competencias específicas:

- **CE31:** Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.

Resultados de aprendizaje:

- **RA8:** Definir, evaluar y seleccionar plataformas de almacenamiento y tratamiento de grandes volúmenes de información, así como configurarlas para utilizarlas para el análisis de dicha información.

En la tabla inferior se muestra la relación entre las competencias que se desarrollan en la asignatura y los resultados de aprendizaje que se persiguen:

Competencias	Resultados de aprendizaje
CB4, CT10, CE31	RA8: Definir, evaluar y seleccionar plataformas de almacenamiento y tratamiento de grandes volúmenes de información, así como configurarlas para utilizarlas para el análisis de dicha información.

4. CONTENIDOS

La materia está organizada en seis unidades de aprendizaje, las cuales, a su vez, están divididas en temas:

Tema 1. Introducción a los Sistemas Big Data.

- ¿Qué es Big Data?
- Las 5 Vs del Big Data
- Sistemas distribuidos y tolerancia a fallos

Tema 2. Ecosistema HADOOP.

- Sistema de archivos HDFS.
- Paradigma de programación MapReduce.
- Gestor de recursos YARN
- Apache Hive y apache Pig

Tema 3. Sistemas de almacenamiento y BBDD No SQL.

- Almacenamiento distribuido.
- Separación entre almacenamiento y computación.
- Bases de datos NoSQL

- BBDD Clave-Valor.
- BBDD Tabulares.
- BBDD de Grafos.
- BBDD documentales.

Tema 4. Apache Spark.

- Qué es Apache Spark.
- Resilient Distributed Datasets y Spark Dataframes
- Spark SQL.
- Spark Machine Learning.

Tema 5. Procesamiento de datos en Streaming.

- Introducción al procesamiento en Streaming.
- Mensajes y modelo Publicador/Suscriptor
- Ventanas y procesamiento en tiempo real.

Tema 6. Visualización de Grandes Volúmenes de Datos

- Visualización de datos.
- Metáforas Visuales.
- Cuadros de mando en tiempo real.

5. METODOLOGÍAS DE ENSEÑANZA-APRENDIZAJE

A continuación, se indican los tipos de metodologías de enseñanza-aprendizaje que se aplicarán:

- Clase magistral.
- Trabajo en grupo.
- Trabajo Autónomo.
- Lectura individual de temas y debate.
- Tutorías, seguimiento académico y evaluación.

6. ACTIVIDADES FORMATIVAS

A continuación, se identifican los tipos de actividades formativas que se realizarán y la dedicación en horas del estudiante a cada una de ellas:

Modalidad presencial:

Actividad formativa	Número de horas
Clases magistrales, lectura de temas principales y materiales complementarios	50 h
Actividades aplicativas de carácter grupal.	25 h
Trabajo autónomo por parte del estudiante.	50 h
Tutorías, seguimiento académico y evaluación	25 h

TOTAL	150
--------------	------------

Modalidad online:

Actividad formativa	Número de horas
Trabajo autónomo por parte del estudiante.	50 h
Actividades aplicativas de carácter grupal.	50 h
Lectura individual de temas y debate	25 h
Tutorías, seguimiento académico y evaluación	25 h
TOTAL	150

7. EVALUACIÓN

A continuación, se relacionan los sistemas de evaluación, así como su peso sobre la calificación total de la asignatura:

Modalidad presencial:

Sistema de evaluación	Peso
Exámenes y tests.	30%
Elaboración de artículos, informes o memorias de diseños.	30%
Portfolios, evaluación entre compañeros.	15%
Conferencias, visitas a empresas e instituciones.	10%
Casos prácticos, diseños, simulaciones e investigación.	15%

Modalidad online:

Sistema de evaluación	Peso
Exámenes y tests.	60%
Elaboración de artículos, informes o memorias de diseños.	10%
Portfolios, evaluación entre compañeros.	10%
Casos prácticos, diseños, simulaciones e investigación.	20%

En el Campus Virtual, cuando accedas a la asignatura, podrás consultar en detalle las actividades de evaluación que debes realizar, así como las fechas de entrega y los procedimientos de evaluación de cada una de ellas.

7.1. Convocatoria ordinaria

Para superar la asignatura en convocatoria ordinaria deberá obtenerse una media ponderada igual o superior a 5 sobre 10 en la calificación final. Además, es necesario cumplir con las siguientes condiciones:

- En las actividades prácticas de evaluación continua (actividades 1-5) se debe tener una asistencia a clase superior al 70% (solo en versión presencial).
- Obtener una calificación superior a 5 en la actividad prueba final.

7.2. Convocatoria extraordinaria

Para superar la asignatura en convocatoria ordinaria deberás obtener una calificación mayor o igual que 5,0 sobre 10,0 en la calificación final (media ponderada) de la asignatura.

En todo caso, será necesario que obtengas una calificación mayor o igual que 4,0 en la prueba final, para que la misma pueda hacer media con el resto de las actividades.

Se deben entregar las actividades no superadas en convocatoria ordinaria, tras haber recibido las correcciones correspondientes a las mismas por parte del docente, o bien aquellas que no fueron entregadas. Las actividades de la convocatoria extraordinaria se realizarán todas de forma individual y el profesor podrá introducir cambios en los enunciados con respecto a las actividades de ordinaria

8. CRONOGRAMA

En este apartado se indica el cronograma con fechas de entrega de actividades evaluables de la asignatura:

Actividades evaluables	Fecha
Actividad 1. Trabajo de investigación sobre el sector del Big Data.	Semana 2
Actividad 2. Práctica de almacenamiento.	Semana 4
Actividad 3. Programación MapReduce	Semana 7
Actividad 4. Programa con apache Spark	Semana 11-12
Actividad 5. Procesamiento de datos en streaming	Semana 15-16
Actividad 6. Desarrollo de un proyecto de análisis Big Data end-to-end.	Semana 18
Actividad 7. Prueba final	Semana 18

Este cronograma podrá sufrir modificaciones por razones logísticas de las actividades. Cualquier modificación será notificada al estudiante en tiempo y forma.

9. BIBLIOGRAFÍA

A continuación, se indica bibliografía recomendada:

- EIJKHOUT, V. (2014). Introduction to High Performance Scientific Computing. (2ª edición).
- TANENBAUM, AS., VAN STEEN, M. (2001). Distributed Systems: Principles and Paradigms. (2ª edición). Pearson Prentice Hall.
- WHITE, T. (2015). Hadoop: The Definitive Guide. O'Reilly.
- ODERSKY, M., SPOON, L. & VENNERS, B. (2008). Programming in Scala. Ed. Artima.
- ODERSKY, M. (2010). Scala by Example. Programming Method Laboratory, Switzerland: EPFL.
- WAMPLER, D. & PAYNE, A. (2009). Programming Scala. Ed. O'Reilly.
- KARAU, H., KONWINSKI, A., WENDELL, P. & ZAHARIA, M. (2015). Learning Spark. O'Reilly.
- BIHANIC, D. (2015). New Challenges for Data Design. Springer.
- BIHANIC, D. (2015). Empowering Users through Design. Springer.
- FISHER, D., MEYER, M. (2016). Making sense of data. O'Reilly.
- GRAY, J., BOUNEGRU, L., CHAMBERS, L. (2012). The Data Journalism Handbook. O'Reilly.
- ILIINSKY, N., STEELE, J. (2011). Designing Data Visualizations. O'Reilly.
- MUNZNER, T. (2014). Visualization Analysis & Design. CRC Press.
- RYAN, L. (2016). The Visual Imperative. Creating a Visual Culture of Data Discovery.
- SIMON, P. (2014). The Visual Organization. Wiley.
- TELEA, A. (2007). Data Visualization Principles and Practice.
- WARD, M., GRINSTEIN, G., KEIM, D. (2015). Interactive Data Visualisation.
- YAU, N. (2013). Data Points. Visualization That Means Something. Wiley.

10. UNIDAD DE ATENCIÓN A LA DIVERSIDAD

Estudiantes con necesidades específicas de apoyo educativo:

Las adaptaciones o ajustes curriculares para estudiantes con necesidades específicas de apoyo educativo, a fin de garantizar la equidad de oportunidades, serán pautadas por la Unidad de Atención a la Diversidad (UAD).

Será requisito imprescindible la emisión de un informe de adaptaciones/ajustes curriculares por parte de dicha Unidad, por lo que los estudiantes con necesidades específicas de apoyo educativo deberán contactar a través de: unidad.diversidad@universidadeuropea.es al comienzo de cada semestre.

11. ENCUESTAS DE SATISFACCIÓN

¡Tú opinión importa!

La Universidad Europea te anima a participar en las encuestas de satisfacción para detectar puntos fuertes y áreas de mejora sobre el profesorado, la titulación y el proceso de enseñanza-aprendizaje.

Las encuestas estarán disponibles en el espacio de encuestas de tu campus virtual o a través de tu correo electrónico.

Tu valoración es necesaria para mejorar la calidad de la titulación.

Muchas gracias por tu participación.