

1. DATOS BÁSICOS

Asignatura	9988001401 GRANDES VOLUMENES DE DATOS
Titulación	GRADO EN INGENIERIA INFORMATICA
Escuela/ Facultad	ESCUELA DE ARQUITECTURA , INGENIERIA Y DISEÑO (STEAM SCHOOL)
Curso	4º
ECTS	6
Carácter	OBLIGATORIA
Idioma/s	ESPAÑOL
Modalidad	PRESENCIAL
Semestre	PRIMER SEMESTRE
Curso académico	CURSO 2025-26
Docente coordinador	DRA. LAURA GARCÍA CUENCA
Docente	DRA. LAURA GARCÍA CUENCA

2. PRESENTACIÓN

La asignatura “Grandes Volúmenes de Datos” es una materia obligatoria del Grado en Ingeniería Informática, y del Doble Grado en Ingeniería Informática y Matemáticas, con una carga lectiva de 6 créditos ECTS, y se enmarca en el módulo de Computación. Su objetivo principal es dotar al alumnado de los conocimientos teóricos y prácticos necesarios para abordar los desafíos que surgen al trabajar con grandes cantidades de datos heterogéneos, generados a gran velocidad y en múltiples formatos.

Esta materia se concibe como una continuación natural de asignaturas previas como Inteligencia Artificial, Sistemas Inteligentes y Bases de Datos, ampliando y profundizando en las soluciones tecnológicas que permiten escalar los procesos de almacenamiento, análisis y extracción de conocimiento en entornos de datos masivos.

Durante el desarrollo de la asignatura, el estudiante se familiarizará con conceptos avanzados de procesamiento distribuido, almacenamiento escalable y técnicas de análisis de datos a gran escala. Asimismo, se introducirán herramientas y frameworks de uso industrial, como Hadoop, Apache Spark, Kafka, NoSQL, entre otros, que forman parte del ecosistema Big Data moderno.

Además de la infraestructura, la asignatura aborda los algoritmos de aprendizaje automático adaptados a grandes volúmenes de datos y técnicas de visualización orientadas a interpretar y comunicar eficazmente la información extraída, poniendo especial énfasis en la aplicación práctica mediante el desarrollo de proyectos y casos de estudio reales.

En definitiva, se persigue que el alumnado adquiera una visión integral del ciclo de vida del dato en entornos Big Data, desde su adquisición e ingesta hasta su análisis e interpretación, capacitando así al futuro ingeniero/a informático/a para integrarse en equipos multidisciplinares y aportar valor en organizaciones donde los datos son un activo estratégico.

3. RESULTADOS DE APRENDIZAJE

CONOCIMIENTOS

CON20. Conocimiento profundo de los principios fundamentales y modelos de la computación y saberlos aplicar para interpretar, seleccionar, valorar, modelar, y crear nuevos conceptos, teorías, usos y desarrollos tecnológicos relacionados con la informática

HABILIDADES

HAB11. Capacidad para desarrollar y evaluar sistemas interactivos y de presentación de información compleja y su aplicación a la resolución de problemas de diseño de interacción persona-computadora

HAB 12. Capacidad para aplicar 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

COMPETENCIAS

CP04. Capacidad para definir, evaluar y seleccionar plataformas hardware y software para el desarrollo y la ejecución de sistemas, servicios y aplicaciones informáticas, de acuerdo con los conocimientos adquiridos según lo establecido en las competencias específicas del título

4. RESULTADOS DE APRENDIZAJE ESPECÍFICOS Y CONTENIDOS

Resultados de aprendizaje específicos de la materia

Conocimientos específicos de la materia

- Describir las ventajas del procesamiento distribuido para la escalabilidad de los sistemas.
- Evaluar plataformas de almacenamiento y tratamiento de grandes volúmenes de información.

Habilidades específicas de la materia

- Diseñar visualizaciones de grandes volúmenes de datos.
- Aplicar algoritmos de procesamiento paralelo a grandes cantidades de datos.
- Realizar una aplicación que requiera procesamiento de datos en streaming.
- Utilizar una plataforma de computación en la nube para almacenar, tratar, procesar y visualizar grandes volúmenes de datos analizando de forma crítica las diferentes herramientas que se pueden utilizar y seleccionando la más adecuada con argumentos razonados.

Temario de la asignatura: Grandes Volúmenes de Datos (4º GII)

Unidad 0. Fundamentos de Programación para GVD

- **Objetivo:** Adquirir competencias básicas en programación con Python y en el uso de librerías para el análisis y modelado de datos, como base para el aprendizaje automático y el procesamiento de datos a gran escala.
- **Contenido:** Sintaxis básica y estructuras de control para manipulación de datos y Visualización. Introducción a scikit-learn, Evaluación básica de modelos: accuracy, overfitting, validación cruzada. Utilización de herramientas code y no code (rapidminer)
- **Contenido Opcional :**
Spanish_Python_Essentials_1_v1_0_Scope_and_Sequence . Cisco
Networking Academy & Python Institute PI

Unidad 1. Introducción a los Sistemas Big Data

- **Objetivo:** Comprender qué es Big Data, sus características distintivas y su importancia en entornos reales, tanto desde el punto de vista tecnológico como estratégico.
- **Contenido:** Definición y evolución del concepto Big Data. Las 5Vs: Volumen, Velocidad, Variedad, Veracidad, Valor. Retos y oportunidades del Big Data. Ejemplos de aplicaciones reales en diferentes sectores.

Unidad 2. Procesamiento Distribuido y Escalabilidad

- **Objetivo:** Entender los fundamentos del procesamiento distribuido de datos y cómo este permite escalar horizontalmente para gestionar grandes volúmenes de información con tolerancia a fallos.
- **Contenido:** Computación distribuida: definición, ventajas, desafíos. Modelos de paralelismo: por datos y por tareas. Escalabilidad horizontal vs vertical. Sistemas tolerantes a fallos. Introducción a MapReduce y Apache Spark.

Unidad 3. Sistemas de Almacenamiento y BBDD NoSQL

- **Objetivo:** Conocer los diferentes tipos de sistemas de almacenamiento y bases de datos NoSQL utilizados para manejar grandes volúmenes de datos, y saber cuándo utilizar cada uno.
- **Contenido:** Limitaciones de bases relacionales en entornos Big Data. Tipos de bases de datos NoSQL: Clave-valor (Redis, DynamoDB), Columnar/tabular (HBase, Cassandra), Documentales (MongoDB), De grafos (Neo4j). Casos de uso y comparación entre modelos.

Unidad 4. Plataformas y Herramientas para el Procesamiento y Almacenamiento

- **Objetivo:** Familiarizarse con las plataformas y herramientas más utilizadas para el almacenamiento distribuido y el procesamiento de grandes volúmenes de datos, tanto on-premise como en la nube.
- **Contenido:** Hadoop Distributed File System (HDFS). Herramientas de análisis: Hive, Pig, Presto, Spark SQL. Arquitecturas distribuidas y orquestación (Hadoop, YARN, Kubernetes). Introducción a plataformas cloud (AWS, GCP, Azure): S3, BigQuery, Redshift.

Unidad 5. Procesamiento de Datos en Streaming

- **Objetivo:** Comprender las particularidades del procesamiento de datos en tiempo real y conocer las herramientas más relevantes para implementarlo.
- **Contenido:** Diferencias entre procesamiento batch y en streaming. Conceptos clave: eventos, micro-batching, ventanas de tiempo. Frameworks y herramientas: Apache Kafka (mensajería), Spark Streaming Apache Flink. Casos de uso: detección de eventos, IoT, sistemas en tiempo real.
- **Proyecto Práctico Integrador**

Notas adicionales:

- Todas las unidades incorporan actividades prácticas, trabajos de investigación y/o talleres
- El **Proyecto Final** puede estar enfocado en un caso real aplicado al Big Data con visualización de resultados.

5. METODOLOGÍAS DE ENSEÑANZA-APRENDIZAJE

- Clase magistral
- Aprendizaje basado en problemas
- Aprendizaje basado en proyectos
- Aprendizaje basado en enseñanzas de taller
- Estudio del Caso

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	Tiempo Total	Tiempo en Clase	Uso de IA
Clases magistrales	12	12	Sugerido
Clases magistrales de aplicación práctica	18	18	Sugerido
Resolución de problemas	30	9	Sugerido
Investigaciones y proyectos	25	0	Sugerido

Actividades en talleres y/o laboratorios	15	15	No Permitido
Trabajo autónomo	44	0	Sugerido
Debates y coloquios	4	4	No permitido
Pruebas de evaluación presenciales	2	2	No permitido
TOTAL	150 horas	60 horas (40%)	

7. EVALUACIÓN

Modalidad presencial:

Sistema de evaluación	Peso
Pruebas de evaluación presenciales	50
Caso/problema/Actividad	10
Evaluación del desempeño	10
Investigaciones y Proyectos	20
Cuaderno de prácticas de laboratorio/taller	10

Es necesario un mínimo del 50% de asistencia en clase

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 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 5,0 en todas las pruebas de evaluación, para que la misma pueda hacer media con el resto de las actividades. En caso de no llegar al 5.0 en alguna, la nota no podrá superar el 4.0 en el sistema de actas oficial.

El profesor se reserva el derecho de poder solicitar una prueba adicional a cualquiera de las pruebas de evaluación, en el caso de tener dudas sobre la autoría del estudiante.

7.2. Convocatoria extraordinaria

Para superar la asignatura en convocatoria extraordinaria 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 5,0 en todas las pruebas de evaluación, para que la misma pueda hacer media con el resto de las actividades. En caso de no llegar al 5.0 en alguna, la nota no podrá superar el 4.0 en el sistema de actas oficial

El profesor se reserva el derecho de poder solicitar una prueba adicional a cualquiera de las pruebas de evaluación, en el caso de tener dudas sobre la autoría del estudiante.

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.

8. CRONOGRAMA

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

Actividades evaluables	Fecha
Prueba de evaluación presencial	Semana 18
Proyecto Final	Semanas 12, 13 y 14
Presentación oral	Semana 15
Ejercicios de clase (portfolio)	Semanas 2,3,5,7,10,12

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

9. BIBLIOGRAFÍA

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

10. UNIDAD DE ORIENTACIÓN EDUCATIVA Y DIVERSIDAD

Desde la Unidad de Orientación Educativa y Diversidad (ODI) ofrecemos acompañamiento a nuestros estudiantes a lo largo de su vida universitaria para ayudarles a alcanzar sus logros académicos. Otros de los pilares de nuestra actuación son la inclusión del estudiante con necesidades específicas de apoyo educativo, la accesibilidad universal en los distintos campus de la universidad y la equiparación de oportunidades.

Desde esta Unidad se ofrece a los estudiantes:

1. Acompañamiento y seguimiento mediante la realización de asesorías y planes personalizados a estudiantes que necesitan mejorar su rendimiento académico.
2. En materia de atención a la diversidad, se realizan ajustes curriculares no significativos, es decir, a nivel de metodología y evaluación, en aquellos alumnos con necesidades específicas de apoyo educativo persiguiendo con ello una equidad de oportunidades para todos los estudiantes.

3. Ofrecemos a los estudiantes diferentes recursos formativos extracurriculares para desarrollar diversas competencias que les enriquecerán en su desarrollo personal y profesional.
4. Orientación vocacional mediante la dotación de herramientas y asesorías a estudiantes con dudas vocacionales o que creen que se han equivocado en la elección de la titulación.

Los estudiantes que necesiten apoyo educativo pueden escribirnos a:
orientacioneducativa@universidadeuropea.es

11. ENCUESTAS DE SATISFACCIÓN

¡Tu 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.

PLAN DE TRABAJO DE LA ASIGNATURA

CÓMO COMUNICARTE CON TU DOCENTE

Cuando tengas una duda sobre los contenidos o actividades, no olvides escribirla en los foros de tu asignatura para que todos tus compañeros y compañeras puedan leerla.

¡Es posible que alguien tenga tu misma duda!

Si tienes alguna consulta exclusivamente dirigida al docente puedes enviarle un mensaje privado desde el Campus Virtual. Además, en caso de que necesites profundizar en algún tema, puedes acordar una tutoría.

Es conveniente que leas con regularidad los mensajes enviados por estudiantes y docentes, pues constituyen una vía más de aprendizaje.

CRONOGRAMA DE ACTIVIDADES

En este apartado se indica el cronograma de actividades formativas, así como las fechas de entrega de las actividades evaluables de la asignatura:

Programación Semanal - Grandes Volúmenes de Datos (Sept 2025 – Ene 2026)				
Semana	Fechas	Contenido	Actividad	Unidad Aprendizaje
1	8 – 12 sept	Presentación Asignatura, Normativas, Evaluación, Funcionamiento Básico Perfilado Datos Herramientas Code: Python para GVD	Taller Práctico Herramientas Code	Unidad 0. Fundamentos de Programación para GVD
2	15 – 19 sept	Herramientas Code: Python para GVD Herramientas No Code: RapidMiner (Visualización)	Taller Práctico Herramientas Code & No Code 📖 Cuaderno de prácticas Taller UA0	Unidad 0. Fundamentos de Programación para GVD
3	22 – 26 sept	Introducción a los sistemas Big Data	📖 Investigación UA1	UA1: Introducción a los sistemas Big Data
4	29 sept – 3 oct	Sistemas Altamente Distribuidos		UA2: Procesamiento Distribuido y Escalabilidad
5	6 – 10 oct	Virtualización	📖 Actividad UA2	UA2: Procesamiento Distribuido y Escalabilidad
6	13 – 17 oct	Teorema del CAP, Bases de datos NoSQL		UA3: Ssistemas de Almacenamiento y BBDD NoSQL
7	20 – 24 oct	Teorema del CAP, Bases de datos NoSQL	📖 Cuaderno de prácticas UA3	UA3: Ssistemas de Almacenamiento y BBDD NoSQL
8	27 – 31 oct	Ecosistema Hadoop, MapReduce, HDFS, YARN SPARK	Taller practico HDFS, MapReduce (Local)	UA4: Plataformas y Herramientas para el procesamiento y almacenamiento
9	3 – 7 nov	Ecosistema Hadoop, MapReduce, HDFS, YARN SPARK	Taller practico HDFS, MapReduce (Infraestructura LORCA)	UA4: Plataformas y Herramientas para el procesamiento y almacenamiento
10	10 – 14 nov	Ecosistema Hadoop, MapReduce, HDFS, YARN SPARK	📖 Cuaderno de prácticas UA4	UA4: Plataformas y Herramientas para el procesamiento y almacenamiento
11	17 – 21 nov	Apache Kafka, Flink	Taller Practico Kafka, Flink	UA5. Procesamiento de Datos en Streaming
12	24 – 28 nov	Apache Kafka, Flink	📖 Cuaderno de prácticas UA5 Proyecto Práctico Integrador - Bases	UA5. Procesamiento de Datos en Streaming
13	1 – 5 dic		Proyecto Práctico Integrador- Trabajo Estudiantes	
14	8 – 12 dic		Proyecto Práctico Integrador- Trabajo Estudiantes	
15	15 – 19 dic		Presentación Proyecto Integrador	
	22 dic – 6 ene	✗ Vacaciones de Navidad – No lectivo		
16	5 – 9 ene			
17	12 – 16 ene			
18	19-23 ene			SEMANA DE EXAMENES

Este cronograma podrá sufrir modificaciones que serán notificadas al estudiante en tiempo y forma.

REGLAMENTO PLAGIO

Atendiendo al Reglamento disciplinario de los estudiantes de la Universidad Europea:

- El plagio, en todo o en parte, de obras intelectuales de cualquier tipo se considera falta muy grave.
- Las faltas muy graves relativas a plagios y al uso de medios fraudulentos para superar las pruebas de evaluación, tendrán como consecuencia la pérdida de la convocatoria correspondiente, así como el reflejo de la falta y su motivo, en el expediente académico.