26-27

GRADO EN INGENIERÍA EN INTELIGENCIA ARTIFICIAL
SEGUNDO CURSO

GUÍA DE ESTUDIO PÚBLICA

INFRAESTRUCTURAS PARA EL PROCESAMIENTO MASIVO DE DATOS Y COMPUTACIÓN EN LA NUBE

CÓDIGO 71032067

UNED

26-27

INFRAESTRUCTURAS PARA EL PROCESAMIENTO MASIVO DE DATOS Y COMPUTACIÓN EN LA NUBE

CÓDIGO 71032067

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR LA ASIGNATURA
EQUIPO DOCENTE
HORARIO DE ATENCIÓN AL ESTUDIANTE
TUTORIZACIÓN EN CENTROS ASOCIADOS
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE
RESULTADOS DE APRENDIZAJE
CONTENIDOS
METODOLOGÍA
SISTEMA DE EVALUACIÓN
BIBLIOGRAFÍA BÁSICA
BIBLIOGRAFÍA COMPLEMENTARIA
RECURSOS DE APOYO Y WEBGRAFÍA
IGUALDAD DE GÉNERO

NOMBRE DE LA ASIGNATURA	INFRAESTRUCTURAS PARA EL PROCESAMIENTO MASIVO DE DATOS Y COMPUTACIÓN EN LA NUBE
CÓDIGO	71032067
CURSO ACADÉMICO	2026/2027
DEPARTAMENTO	SISTEMAS DE COMUNICACIÓN Y CONTROL
TÍTULO EN QUE SE IMPARTE	GRADO EN INGENIERÍA EN INTELIGENCIA ARTIFICIAL
CURSO	SEGUNDO CURSO
PERIODO	SEMESTRE 2
Nº ETCS	6
HORAS	150.0
IDIOMAS EN QUE SE IMPARTE	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

El trabajo con datos masivos exige la utilización de infraestructuras computacionales específicamente diseñadas para ellos. Estas infraestructuras difieren de las infraestructuras tradicionales en varios aspectos. Para empezar, es necesario combinar la potencia de cómputo de muchos ordenadores, construyendo lo que se conoce como un cluster de ordenadores. Por otro lado, es necesario utilizar paradigmas de programación que puedan aprovechar la potencia de cómputo del cluster pero de una forma sencilla para el desarrollador encargado de implementar los programas para el análisis de datos masivos. Ambos aspectos pueden desarrollarse utilizando servicios de proveedores en la nube. En esta asignatura se muestran algunas de las tecnologías más importantes que permiten desplegar infraestructuras para el procesamiento de datos masivos, lo cual constituye el objetivo principal de la asignatura, y además una aportación muy importante para el perfil profesional del estudiantado.

Dentro de este Grado es importante adquirir una visión sólida de las herramientas más utilizadas en ese contexto, dado que son esenciales para mover y tratar datos masivos, tanto estructurados como no estructurados.

La asignatura se imparte en el segundo semestre del segundo curso del Grado en Inteligencia Artificial, se integra en la materia *TRATAMIENTO Y GESTIÓN DE DATOS* y tiene un carácter obligatorio. Para superar esta asignatura es necesario haber superado las siguientes asignaturas de este grado:

•Adquisición, procesado y tratamiento de la información.

•Modelado de la información y bases de datos.

•Fundamentos de programación.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR LA ASIGNATURA

Haber superado las siguientes asignaturas de este grado:

•Adquisición, procesado y tratamiento de la información.

•Modelado de la información y bases de datos.

•Fundamentos de programación.

Es necesario el conocimiento de la lengua extranjera inglés, para poder acceder a un mayor número de fuentes de recursos (manuales técnicos) que no siempre están traducidos al castellano, o se traducen muy posteriormente a su publicación.

Además, debido a la naturaleza multidisciplinar de la materia, se necesita que los estudiantes tengan conocimientos sobre redes, protocolos y arquitecturas de red, así como, conocimientos sobre programación.

Para realizar las prácticas, es muy recomendable que es estudiante disponga de un ordenador actual, con al menos 16G de memoria.

EQUIPO DOCENTE

Nombre y Apellidos	AGUSTIN CARLOS CAMINERO HERRAEZ (Coordinador/a de asignatura)
Correo Electrónico	accaminero@scc.uned.es
Teléfono	91398-9468
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos	RAFAEL PASTOR VARGAS
Correo Electrónico	rpastor@dia.uned.es
Teléfono	91398-8383
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos	RAFAEL PASTOR VARGAS
Correo Electrónico	rpastor@scc.uned.es
Teléfono	91398-8383
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos	ANTONIO ROBLES GOMEZ
Correo Electrónico	arobles@scc.uned.es
Teléfono	91398-8480
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL

HORARIO DE ATENCIÓN AL ESTUDIANTE

Las/os estudiantes pueden contactar con las/os profesoras/es para resolver dudas sobre la asignatura en primer lugar a través del foro de la asignatura en el campus virtual correspondiente, a través del correo electrónico o por teléfono en el horario que se indica. Si se desea una entrevista personal, debe concertarse previamente. En todo tipo de comunicación con el profesorado se deberá indicar la asignatura a la que se refiere y utilizar el correo de la UNED.

Dr. Agustín C. Caminero Herráez. Lunes de 16:00 a 20:00 horas

Dr. Rafael Pastor Vargas. Lunes de 16:00 a 20:00 horas

Dr. Antonio Robles Gómez. Martes de 10:00 a 14:00 horas

Dirección postal:

Escuela Técnica Superior de Ingeniería Informática, UNED

C. Juan del Rosal, 16. 5.ª planta.

28040 MADRID

TUTORIZACIÓN EN CENTROS ASOCIADOS

Las sesiones de tutoría se imparten presencialmente y en línea desde el Centro Asociado de referencia o, en su defecto, desde otro Centro Asociado. A continuación se muestran los centros asociados y extensiones en las que se imparten tutorías de la asignatura.

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

Ver sección de Resultados de Aprendizaje

RESULTADOS DE APRENDIZAJE

Al superar la asignatura, el estudiantado habrá adquirido los siguientes resultados de aprendizaje:

RA-CD01 - Al finalizar el aprendizaje, el o la estudiante será capaz de determinar la mejor forma de representar el conocimiento, usando formalismos basados en la lógica, aplicando los fundamentos de la gestión y procesamiento de datos, incluyendo el almacenamiento y procesamiento, la gestión de los metadatos, la gestión eficiente de flujos continuos y la gobernanza de sistemas complejos de datos. TIPO: Conocimientos o contenidos. TIPO: Conocimientos o contenidos

RA-CD02 - Al finalizar el aprendizaje, el o la estudiante será capaz de desarrollar, incluyendo la puesta en producción, aplicaciones que puedan extraer información de datos estructurados, semi estructurados o no estructurados, incluyendo texto, imagen, vídeo y audio, mediante técnicas de identificación y adquisición de datos relevantes, reducción, compresión, integración, transformación, limpieza y evaluación de su calidad, incluyendo interfaces persona computador que visualicen estos datos de manera efectiva y centrada en el usuario. TIPO: Habilidades o destrezas

RA-IA01 - Al finalizar el aprendizaje, el o la estudiante será capaz de determinar las técnicas

más adecuadas para la resolución de problemas que involucren modelos de razonamiento en entornos centralizados y distribuidos, técnicas de aprendizaje automático y estadística avanzada, técnicas de percepción y robótica cognitiva, entidades y sistemas inteligentes que permitan la adquisición y representación del conocimiento, la transformación de los datos en conocimiento y la manipulación del entorno, en problemas que requieran el uso de infraestructuras, entornos y técnicas de la inteligencia artificial. TIPO: Conocimientos o contenidos

RA06 - Al finalizar el aprendizaje, el o la estudiante será capaz de analizar las necesidades en algorítmica, complejidad computacional, programación, sistemas operativos, bases de datos, estructura, interconexión de sistemas informáticos necesarios para la resolución de problemas de ciencias e ingeniería, de acuerdo con los principios de calidad, fiabilidad y seguridad necesarios. TIPO: Habilidades o destrezas

RA08 - Al finalizar el aprendizaje, el o la estudiante será capaz de demostrar conocimiento de los procedimientos algorítmicos básicos y los tipos de datos de las tecnologías informáticas necesarios para la resolución de los problemas de inteligencia artificial. TIPO: Conocimientos o contenidos

RA10 - Al finalizar el aprendizaje, el o la estudiante será capaz de demostrar conocimiento básico de estructura y arquitectura de computadores, sistemas operativos, sistemas distribuidos redes, de computadores internet y sistemas de almacenamiento, procesamiento y acceso a datos necesarios para el análisis e implementación de aplicaciones de inteligencia artificial basadas en ellos. TIPO: Conocimientos o contenidos

RA12 - Al finalizar el aprendizaje, el o la estudiante será capaz de demostrar conocimiento básico de técnicas de programación paralela, concurrente, distribuida y en tiempo real necesarias para la implementación de aplicaciones que los requieran. TIPO: Conocimientos o contenidos

RA-IA06 - Al finalizar el aprendizaje, el o la estudiante será capaz de adaptar soluciones basadas en inteligencia artificial a entornos de aplicación específicos, considerando su integración eficiente con infraestructuras existentes. TIPO: Competencias

CONTENIDOS

MÓDULO 1: Introducción a los Datos Masivos

Este módulo presenta una introducción al Big Data, se muestran algunos ejemplos de uso de este paradigma, así como algunas tecnologías de interés como las bases de datos no relacionales, o formatos de archivos.

MÓDULO 2: Paradigmas Computacionales para Entornos Masivos

Este módulo consiste en ver tecnologías para programar en sistemas Big Data. Primero veremos el paradigma Mapreduce, pionero del Big Data, para luego pasar a estudiar su evolución, el procesamiento en memoria con Apache Spark.

MÓDULO 3: Algoritmos de IA en Plataformas de Procesamiento Masivo

Una vez tenemos una base de herramientas de programación paralela distribuida, vamos a ver de qué forma podemos aprovecharlas para entrenar modelos de inteligencia artificial.

MÓDULO 4. Servicios Gestionados en la Nube para Big Data e IA

En este último módulo veremos cómo prodemos aprovechar recursos de proveedores cloud públicos para desplegar nuestras aplicaciones y servicios. De esta forma, evitamos el coste que supone la adquisición y mantenimiento de unas infraestructuras propietarias.

METODOLOGÍA

Las diferentes asignaturas que integran este grado, se impartirán todas ellas conforme a la metodología no presencial que caracteriza a la UNED, en la cual prima el autoaprendizaje del alumno, asistido por el profesor mediante diversos sistemas de comunicación docente-discente mayoritariamente asíncronos y por las tutorías que pueden ser presenciales, impartidas en los Centros Asociados, o semipresenciales permitiendo la conexión del alumnado a las sesiones con independencia de su localización geográfica. Además, la docencia se imparte con apoyo de una plataforma virtual interactiva en la que se combinan distintos recursos, los medios impresos y los audiovisuales y virtuales. Más en concreto, la plataforma virtual contendrá: contenidos, foros de discusión, glosario de términos, tareas y enlaces de interés.

De esta forma, la metodología de esta asignatura está basada en los siguientes pilares:

1. Autoaprendizaje asistido empleando la metodología a distancia. El enfoque principal de esta metodología es el autoaprendizaje del estudiante. El profesor/a asiste al estudiante, guiándole y apoyándole durante todo el proceso de aprendizaje.
2. Comunicación docente-discente empleando la metodología a distancia. El profesor/a y el/la estudiante se comunican utilizando medios principalmente asíncronos, como correo electrónico y foros. Esta metodología también incluye comunicación síncrona, realizada a través de medios como el teléfono o plataformas de videollamadas.
3. Materiales de estudio empleando la metodología a distancia. El docente proporciona material de estudio básico y complementario para el aprendizaje autónomo del estudiante. Estos materiales incluyen problemas, pruebas de evaluación, vídeos, exámenes resueltos, entre otros.
4. Plataforma virtual interactiva empleando la metodología a distancia. El docente usa una plataforma docente que proporciona acceso a recursos diversos para el aprendizaje. Contiene materiales didácticos, tanto básicos como complementarios.
5. Tutorías mediante la metodología a distancia. Se realizan tutorías a cargo de los equipos docentes o tutores/as en los Centros Asociados para proporcionar apoyo personalizado a

los/las estudiantes.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen	Examen tipo test
Preguntas test	20
Duración del examen	120 (minutos)

Material permitido en el examen

Ninguno.

Criterios de evaluación

La prueba presencial consistirá en un test de 20 preguntas a realizar en un tiempo máximo de 2 horas. Para cada pregunta del test se propondrán 3 ó4 respuestas de las que sólo una será correcta. Únicamente puntuarán las respuestas contestadas. Si la respuesta es correcta la puntuación será de 0.5 puntos y si es incorrecta restará 0.2 puntos. Durante la realización de la prueba no se podrá utilizar ningún tipo de material. La prueba presencial se realizará en el Centro Asociado que corresponda a cada estudiante, en las fechas y horarios establecidos por la UNED.

Es obligatorio la alcanzar al menos el 50% de la puntuación del examen presencial para superar la asignatura.

% del examen sobre la nota final	60
Nota del examen para aprobar sin PEC	5
Nota máxima que aporta el examen a la calificación final sin PEC	6
Nota mínima en el examen para sumar la PEC	5

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC? Si

Descripción

Pruebas de Evaluación Continua (PEC). El objetivo de estas pruebas es evaluar los conocimientos, destrezas y habilidades adquiridos en los ejercicios de autoevaluación, por lo que las cuestiones que se planteen tendrá una relación directa con dichos ejercicios.

En la primera PEC se evaluarán los conocimientos de los Bloques 1 y 2 del temario, y vale un 5% de la nota final.

En la segunda PEC se evaluarán los conocimientos de los Bloques 3 y 4 del temario, y vale un 5% de la nota final.

Ambas PEC consistirán en cuestionarios tipo test con 3 o 4 posibles respuestas de las que sólo una será correcta. Las respuestas incorrectas o en blanco no restan puntos.

Criterios de evaluación

Ponderación de la PEC en la nota final	PEC1 = 0,05 % PEC2 = 0,05 %
Fecha aproximada de entrega	PEC1 semana 6 - PEC2 semana 12
Comentarios y observaciones	

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s?	Si

Descripción

Práctica de laboratorio. El objetivo de esta prueba de evaluación es el diseño e implementación de un caso práctico de manejo de datos masivos, o Big Data. Esta práctica de laboratorio vale hasta el 30% de la nota final de la asignatura.

Es obligatorio la alcanzar al menos el 50% de la puntuación de la práctica de laboratorio para superar la asignatura.

Criterios de evaluación

Se valorará que los desarrollos consigan los resultados que se indiquen en el enunciado de la práctica. Además también se valorará la corrección del diseño, la estructura y calidad de la memoria, y la presencia de un apartado de conclusiones y valoraciones de la práctica.

Ponderación en la nota final	30%
Fecha aproximada de entrega	Semana 12
Comentarios y observaciones	

¿CÓMO SE OBTIENE LA NOTA FINAL?

Siendo:

NF= nota final de la asignatura

EX = nota del examen presencial

PEC1 y PEC2= nota de las PEC1 y PEC2

PL = nota de la práctica de laboratorio

La nota final de la asignatura se calcula como sigue:

Si (EX >=5 y PL >=5) ENTONCES

NF = EX * 0,6 + PEC1 * 0,05 + PEC2 * 0,05 + PL * 0,3

SI NO

NF = 4 (SUSPENSO)

De cara a la convocatoria extraordinaria:

Se reabrirá la PL para su entrega.

Se guardarán las notas de las PEC y PL realizadas en la convocatoria ordinaria.

BIBLIOGRAFÍA BÁSICA

La bibliografía básica será proporcionada al estudiante dentro del curso virtual, estará compuesta por materiales teórico-prácticos propuestos por el equipo docente.
Gran parte de la bibliografía, así como los recursos proporcionados al estudiante en el curso virtual pueden estar únicamente en inglés, debido a la novedad de algunos de los contenidos propuestos para la asignatura.

BIBLIOGRAFÍA COMPLEMENTARIA

Título: Designing Data-Intensive Applications
Autor: Martin Kleppmann
Editorial: O'Reilly Media, Inc.
Año: 2017
ISBN-13 en papel: 978-1-4493-7332-0
URL (solamente funciona tras autenticarse en UNED.es):
https://learning.oreilly.com/library/view/designing-data-intensive-applications/9781491903063/

Título: Apache Hive Cookbook
Autores: Hanish Bansal; Saurabh Chauhan; Shrey Mehrotra
Editorial: Packt Publishing
Año: 2016.
ISBN-13 en papel: 978-1-78216-108-0
ISBN-13 web: 978-1-78216-109-7
URL (solamente funciona tras autenticarse en UNED.es):
https://learning.oreilly.com/library/view/apache-hive-cookbook/9781782161080/

Título: Data Analytics with Hadoop
Autores: Benjamin Bengfort; Jenny Kim
Editorial: O'Reilly Media, Inc.
Año: 2016
ISBN-13 en papel: 978-1-4919-1370-3
URL (solamente funciona tras autenticarse en UNED.es):
https://learning.oreilly.com/library/view/data-analytics-with/9781491913734/

Título: Hadoop Real-World Solutions Cookbook - Second Edition
Autor: Tanmay Deshpande

Editorial: Packt Publishing
Año: 2016
ISBN-13 en papel: 978-1-78439-550-6
ISBN-13 web: 978-1-78439-800-2
URL (solamente funciona tras autenticarse en UNED.es):
https://learning.oreilly.com/library/view/hadoop-real-world-solutions/9781784395506/

Título: Practical Hive: A Guide to Hadoop's Data Warehouse System
Autores: Scott Shaw; Andreas François Vermeulen; Ankur Gupta; David Kjerrumgaard
Editorial: Apress
Año: 2016
ISBN-13: 978-1-4842-0272-2
URL (solamente funciona tras autenticarse en UNED.es):
https://learning.oreilly.com/library/view/practical-hive-a/9781484202715/

Título: Big Data for Chimps
Autores: Philip Kromer; Russell Jurney
Editorial: O'Reilly Media, Inc.
Año: 2015
ISBN-13 en papel: 978-1-4919-2394-8
URL (solamente funciona tras autenticarse en UNED.es):
https://learning.oreilly.com/library/view/big-data-for/9781491923931/

Título: MapReduce Design Patterns
Autores: Donald Miner; Adam Shook
Editorial: O'Reilly Media, Inc.
Año, 2012
ISBN-13 en papel: 978-1-4493-2717-0
URL (solamente funciona tras autenticarse en UNED.es):
https://learning.oreilly.com/library/view/mapreduce-design-patterns/9781449341954/

RECURSOS DE APOYO Y WEBGRAFÍA

La asignatura se encontrará virtualizada en la plataforma que para ello disponga la Universidad. En ella el estudiante encontrará las últimas novedades relativas a la asignatura y los materiales necesarios para su estudio, así como una serie de foros para comunicarse con el equipo docente y con sus compañeros.

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.