

19-20

MÁSTER UNIVERSITARIO EN
INGENIERÍA Y CIENCIA DE DATOS

GUÍA DE ESTUDIO PÚBLICA



INFRAESTRUCTURAS COMPUTACIONALES PARA PROCESAMIENTO DE DATOS MASIVOS

CÓDIGO 31110056

UNED

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN	
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA	
EQUIPO DOCENTE	
HORARIO DE ATENCIÓN AL ESTUDIANTE	
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE	
RESULTADOS DE APRENDIZAJE	
CONTENIDOS	
METODOLOGÍA	
SISTEMA DE EVALUACIÓN	
BIBLIOGRAFÍA BÁSICA	
BIBLIOGRAFÍA COMPLEMENTARIA	
RECURSOS DE APOYO Y WEBGRAFÍA	
ADENDA AL SISTEMA DE EVALUACIÓN CON MOTIVO DE LA PANDEMIA COVID 19	

Nombre de la asignatura	INFRAESTRUCTURAS COMPUTACIONALES PARA PROCESAMIENTO DE DATOS MASIVOS
Código	31110056
Curso académico	2019/2020
Título en que se imparte	MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA DE DATOS
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	SEMESTRE 1
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

PRESENTACIÓN

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.

CONTEXTUALIZACIÓN

La asignatura de Infraestructuras Computacionales para Procesamiento de Datos masivos se trata de una asignatura de 6 créditos ECTS, con carácter obligatorio, impartida en el primer semestre del Máster Universitario en Ingeniería y Ciencia de Datos. Guarda relación con las siguientes asignaturas también disponibles en el mismo Máster:

- Programación en entorno de datos
- Visualización de datos
- Gestión/almacenamiento de información no estructurada.
- Seguridad de la gestión de datos

-

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Se recomienda que los interesados en cursar el Máster tengan un nivel de lectura en inglés suficiente como para entender contenidos técnicos en dicha lengua. 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.

Dado que se verán diferentes tipos de despliegues y/o tecnologías, es necesario que los

estudiantes dispongan de sólidos conocimientos en sistemas operativos y redes, a nivel de comandos de gestión y manipulación de ficheros (especialmente, Linux)

Se fomentará el uso de software libre siempre y cuando sea posible para la realización de las actividades y las prácticas propuestas.

EQUIPO DOCENTE

Nombre y Apellidos

Correo Electrónico

Teléfono

Facultad

Departamento

RAFAEL PASTOR VARGAS

rpastor@dia.uned.es

91398-8383

ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA

SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos

Correo Electrónico

Teléfono

Facultad

Departamento

RAFAEL PASTOR VARGAS

rpastor@scc.uned.es

91398-8383

ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA

SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos

Correo Electrónico

Teléfono

Facultad

Departamento

AGUSTIN CARLOS CAMINERO HERRAEZ (Coordinador de asignatura)

accaminero@scc.uned.es

91398-9468

ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA

SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos

Correo Electrónico

Teléfono

Facultad

Departamento

SALVADOR ROS MUÑOZ

sros@dia.uned.es

7205/7564

ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA

SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos

Correo Electrónico

Teléfono

Facultad

Departamento

SALVADOR ROS MUÑOZ

sros@scc.uned.es

7205/7564

ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA

SISTEMAS DE COMUNICACIÓN Y CONTROL

HORARIO DE ATENCIÓN AL ESTUDIANTE

La tutorización de los estudiantes tendrá lugar esencialmente a través de los foros de la plataforma, aunque también podrán utilizarse ocasionalmente otros medios, tales como chats interactivos, servicios de mensajería instantánea y el correo electrónico. Adicionalmente, está también previsto, para temas personales que no afecten al resto de los estudiantes, atender consultas en persona o por teléfono.

El seguimiento del aprendizaje se realizará revisando la participación de los alumnos en los distintos foros de debate y las aportaciones de material nuevo además de la entrega en fecha de los diferentes trabajos prácticos que se han planificado durante la evolución del curso.

En caso de necesitar contactar con el Equipo Docente por medios distintos al curso virtual,

se utilizará preferentemente el correo electrónico, pudiéndose también realizar consultas telefónicas y entrevista personal en los horarios establecidos y que se muestran a continuación en la siguiente tabla.

Profesor	Horario de atención	Correo electrónico	Teléfono de contacto
Agustín Carlos Caminero Herráez	Lunes lectivos de 11h a 13h y de 15h a 17h	accaminero@scc.uned.es	91 398 9468
Andrés Duque Fernández	Lunes lectivos de 12h a 13h y de 15h a 18h	aduque@scc.uned.es	91 398 XXXX
Rafael Pastor Vargas	Lunes lectivos de 16h a 20 h	rpastor@scc.uned.es	91 398 8383
Salvador Ros Muñoz	Lunes y martes lectivos de 16h a 18h.	sros@scc.uned.es	91 398 7205

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

COMPETENCIAS BÁSICAS

CB6 - 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.

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.

CB9 - Que los 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.

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

COMPETENCIAS GENERALES

CG3 - Desarrollar sistemas de gestión/almacenamiento/procesamiento de grandes volúmenes de datos de una manera eficiente y segura, teniendo en cuenta las normativas/legislaciones existentes

CG5 - Utilizar las habilidades de científico de datos y/o ingeniero de datos en entornos de trabajo multidisciplinares y ser capaz de distinguir/organizar las diferentes actividades de los roles en dicho entorno

COMPETENCIAS TRANSVERSALES

CT1 - Ser capaz de abordar y desarrollar proyectos innovadores en entornos científicos,

tecnológicos y multidisciplinares.

CT2 - Ser capaz de tomar decisiones y formular juicios basados en criterios objetivos (datos experimentales, científicos o de simulación disponibles).

COMPETENCIAS ESPECÍFICAS

CE2 - Desarrollar aplicaciones/servicios/scripts orientados a la analítica de datos y analizar el uso de diferentes librerías para el desarrollo e implementación de métodos numéricos, algoritmos y modelos asociados a los datos

CE8 - Diseñar y utilizar modelos de gestión de datos masivos (Big Data) basados en herramientas disponibles sobre clústeres y en la nube

CE9 - Identificar y utilizar técnicas de desarrollo de algoritmos de manipulación de datos en entornos de gestión de datos masivos

RESULTADOS DE APRENDIZAJE

Los resultados que se pretenden alcanzar con el estudio de esta asignatura son los siguientes:

- Distinguir entre las principales herramientas de inyección, programación y almacenamiento de datos masivos, tanto en batch como en streaming.
- Diseñar programas para el análisis de datos masivos utilizando las herramientas adecuadas para la inyección, análisis y almacenamiento de dichos datos.
- Describir las características más importantes de las principales arquitecturas de programación de Big Data y de sus formas de despliegue tanto local como en la nube.
- Identificar y seleccionar las diferentes opciones de configuración con el objetivo de optimizar las infraestructuras de Big Data.

CONTENIDOS

Módulo 1: Ecosistema de procesamiento paralelo para datos masivos: Apache Hadoop.

Módulo 2: Procesamiento paralelo basado en memoria: Apache Spark.

Módulo 3: Gestión de datos en tiempo real (Streaming).

Módulo 4: Servicios gestionados en la nube para el procesamiento de datos masivos.

METODOLOGÍA

Esta asignatura ha sido diseñada para la enseñanza a distancia. Por tanto, el sistema de enseñanza-aprendizaje estará basado en gran parte en el estudio independiente o autónomo del estudiante. Para ello, el estudiante contará con diversos materiales que permitirán su trabajo autónomo y la Guía de Estudio de la asignatura, que incluye orientaciones para la realización de las actividades prácticas. Asimismo, mediante la plataforma virtual de la UNED existirá un contacto continuo entre el equipo docente y los/as estudiantes, así como una interrelación entre los propios estudiantes a través de los foros, importantísimo en la enseñanza no presencial.

El estudio de esta asignatura se realizará a través de los materiales que el Equipo Docente publicará en el curso virtual.

Las actividades formativas para el estudio de la asignatura son las siguientes:

- Estudios de contenidos (67 horas)
- Tutorías (13 horas)
- Actividades en la plataforma virtual (2 horas)
- Prácticas Informáticas. (58 horas)
- Trabajos (10 horas)
- Total: 150 horas

Los medios necesarios para el aprendizaje son:

1. Materiales teórico-prácticos preparados por el Equipo Docente para cubrir los conceptos básicos del temario.

2. Bibliografía complementaria. El estudiante puede encontrar en ella información adicional para completar su formación.

3. Curso Virtual de la asignatura, donde el estudiante encontrará:

- Una guía de la asignatura en la que se hace una descripción detallada del plan de trabajo propuesto.
- Un calendario con la distribución temporal de los temas propuesta por el Equipo Docente y con las fechas de entrega de las actividades teórico-prácticas que el estudiante tiene que realizar para su evaluación.
- Enunciado de las actividades teórico-prácticas propuestas y zona donde depositar los entregables asociados a dichas actividades.
- Los foros por medio de los cuales el Equipo Docente aclarará las dudas de carácter general y que se usarán también para comunicar todas aquellas novedades que surjan a lo largo del curso. Éste será el principal medio de comunicación entre los distintos participantes en la asignatura.

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 ó 4 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.

% del examen sobre la nota final	40
Nota del examen para aprobar sin PEC	0
Nota máxima que aporta el examen a la calificación final sin PEC	4
Nota mínima en el examen para sumar la PEC	5

Comentarios y observaciones

Es necesario conseguir un mínimo de 4 puntos en la prueba presencial antes de ponderarla para poder aprobar la asignatura.

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad	Si
-------------------------	----

Descripción

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 ó 4 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.

Criterios de evaluación

Ponderación de la prueba presencial y/o los trabajos en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?

Si,PEC no presencial

Descripción

PEC1: Módulo 1. 3 % de la nota.

PEC2: Módulo 2. 2 % de la nota.

PEC3: Módulo 3. 3 % de la nota.

PEC4: Módulo 4. 2 % de la nota.

Criterios de evaluación

Ponderación de la PEC en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s?

Si,no presencial

Descripción

Existen cuatro prácticas informáticas:

Trabajo práctico 1: Módulo 1. 15% de la nota.

Trabajo práctico 2: Módulo 2. 10% de la nota.

Trabajo práctico 3: Módulo 3. 15% de la nota.

Trabajo práctico 4: Módulo 4. 10% de la nota.

Criterios de evaluación

Ponderación en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

La calificación final de la asignatura se calculará de la siguiente forma:

Si en el examen presencial, los trabajos o las PEC (***de manera individual por tipo cada actividad***) no se obtiene al menos el 40% de la puntuación, entonces el/la estudiante estará suspenso.

En otro caso, se calculará la nota final sumando las diferentes pruebas de evaluación ponderadas con los porcentajes descritos arriba.

Aprobarán la asignatura los estudiantes que consigan al menos 5 puntos en la nota final.

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

El Equipo Docente propone una serie de libros disponibles de forma gratuita dentro de la biblioteca digital de la UNED. Se proporcionan enlaces a los libros que funcionan tras autenticarse en UNED.es:

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/>

Título: Hadoop: The Definitive Guide, 4th Edition

Autor: Tom White

Editorial: O'Reilly Media, Inc.

Año: 2015

ISBN-13 en papel: 978-1-4919-0163-2

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/hadoop-the-definitive/9781491901687/>

Título: Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools

Autor: Deepak Vohra

Editorial: Apress

Año: 2016

ISBN-13 en papel: 978-1-4842-2198-3

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/practical-hadoop-ecosystem/9781484221990/>

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: Hadoop with Python

Autores: Zachary Radtka, Donald Miner

Editorial: O'Reilly

Año: 2015

ISBN: 978-1-491-94227-7

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/hadoop-with-python/9781492048435/>

Título: Fast Data Processing with Spark 2 -Third Edition

Autor: Krishna Sankar Editorial: Packt Publishing

Año:2016 ISBN-13 en papel:978-1-78588-927-1

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/fast-data-processing/9781785889271/>

Título: Sams Teach Yourself Apache Spark™ in 24 Hours

Autor: Jeffrey Aven.

Editorial: Sams Año:2016.

ISBN-13 en papel:978-0-672-33851-9.

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/sams-teach-yourself/9780134445786/>

Título: Mastering Apache Spark 2.x -Second Edition

Autor: Romeo Kienzler

Editorial: Packt Publishing

Año:2017

ISBN-13 en papel:978-1-78646-274-9

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/mastering-apache-spark/9781786462749/>

Título: Apache Spark 2.x Cookbook

Autor: Rishi Yadav

Editorial: Packt Publishing

Año: 2017

ISBN-13 en papel:978-1-78712-726-5

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/apache-spark-2x/9781787127265/>

Título: Spark for Python Developers

Autor: Amit Nandi

Editorial: Packt Publishing

Año: 2015

ISBN-13 en Web: 978-1-78439-737-1

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/spark-for-python/9781784399696/>

Título: Machine Learning with Spark -Second Edition

Autor:Rajdeep Dua; Manpreet Singh Ghotra; Nick Pentreath

Editorial: Packt Publishing

Año: 2017

ISBN-13 en papel:978-1-78588-993-6

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/machine-learning-with/9781785889936/>

Título: Spark GraphX in Action

Autor:Michael S. Malak and Robin East

Editorial: Manning Publications

Año:2016

ISBN-13: 978-1-61729-252-1

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/spark-graphx-in/9781617292521/>

Título: Spark in Action

Autor: Petar Zeevi, Marko Bonai

Editorial: Manning Publications

Año: 2016

ISBN-13: 978-1-61729-260-6

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/spark-in-action/9781617292606/>

Título: Streaming Systems

Autor: Reuven Lax, Slava Chernyak, Tyler Akidau

Editorial: O'Reilly Media, Inc.

Año: 2018

ISBN-13: 978-1-49198-387-4

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/streaming-systems/9781491983867/>

Título: Kafka: The Definitive Guide

Autor: Gwen Shapira, Neha Narkhede, Todd Palino

Editorial: O'Reilly Media, Inc.

Año: 2017

ISBN-13: 978-1-49193-616-0

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/kafka-the-definitive/9781491936153/>

Título: Stream Processing with Apache Spark

Autor: Francois Garillot, Gerard Maas

Editorial: O'Reilly Media, Inc.

Año: 2019

ISBN-13: 978-1-49194-424-0

URL (solamente funciona tras autenticarse en UNED.es):

<https://learning.oreilly.com/library/view/stream-processing-with/9781491944233/>

RECURSOS DE APOYO Y WEBGRAFÍA

Los/as estudiantes dispondrán de los siguientes recursos de apoyo al estudio:

- **Guía de la asignatura.** Incluye el plan de trabajo y orientaciones para su desarrollo. Esta guía será accesible desde el curso virtual.
- **Curso virtual.** A través de esta plataforma los/as estudiantes tienen la posibilidad de consultar información de la asignatura, realizar consultas al Equipo Docente a través de los foros correspondientes, consultar e intercambiar información con el resto de los compañeros/as.
- **Biblioteca.** El estudiante tendrá acceso tanto a las bibliotecas de los Centros Asociados como a la biblioteca de la Sede Central, en ellas podrá encontrar un entorno adecuado para el estudio, así como de distinta bibliografía que podrá serle de utilidad durante el proceso de aprendizaje. Además, desde la biblioteca digital de la UNED, el estudiante tendrá acceso a Safari Books Online, una biblioteca digital con más de 30.000 libros técnicos en constante actualización.

ADENDA AL SISTEMA DE EVALUACIÓN CON MOTIVO DE LA PANDEMIA COVID 19

<https://app.uned.es/evacaldos/asignatura/adendasig/31110056>

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.