

22-23

MÁSTER UNIVERSITARIO EN  
INGENIERÍA Y CIENCIA DE DATOS

# GUÍA DE ESTUDIO PÚBLICA



## INFRAESTRUCTURAS COMPUTACIONALES PARA PROCESAMIENTO DE DATOS MASIVOS

CÓDIGO 31110056

Aviso: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el  
Código Seguro de Verificación (CSV) en la dirección <https://sede.uned.es/valida/>

A03DB3D3A74C846AD7012BF28AE7A0A2

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                              |  |



|                           |                                                                      |
|---------------------------|----------------------------------------------------------------------|
| Nombre de la asignatura   | INFRAESTRUCTURAS COMPUTACIONALES PARA PROCESAMIENTO DE DATOS MASIVOS |
| Código                    | 31110056                                                             |
| Curso académico           | 2022/2023                                                            |
| Título en que se imparte  | MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA DE DATOS                |
| Tipo                      |                                                                      |
| Nº ETCS                   | 0                                                                    |
| Horas                     | 0.0                                                                  |
| Periodo                   | SEMESTRE                                                             |
| Idiomas en que se imparte |                                                                      |

## 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.

Dentro de este Máster 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.

### CONTEXTUALIZACIÓN

La asignatura de "Infraestructuras Computacionales para Procesamiento de Datos másivos" 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. Pertenecce a la materia "Infraestructuras y Sistemas para la gestión de datos" a la que también pertenecen las asignaturas "Gestión/almacenamiento de información no estructurada" y "Seguridad de la gestión de datos". Además de éstas, también guarda relación con las siguientes asignaturas también disponibles en el mismo Máster:

- Programación en entorno de datos
- Visualización de datos

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/validar>



A03DB3D3A74C846AD7012BF28AE7A0A2

## 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 practicas 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

Nombre y Apellidos  
Correo Electrónico  
Teléfono  
Facultad  
Departamento

ANGEL MIGUEL GARCIA VICO  
amgarcia@scc.uned.es  
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA  
SISTEMAS DE COMUNICACIÓN Y CONTROL

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/>



## 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.

| <b>Profesor</b>             | <b>Horario de atención</b>                 | <b>Correo electrónico</b>     | <b>Teléfono de contacto</b> |
|-----------------------------|--------------------------------------------|-------------------------------|-----------------------------|
| Agustín C. Caminero Herráez | Lunes lectivos de 11h a 13h y de 15h a 17h | <b>accaminero@scc.uned.es</b> | 91 398 9468                 |
| Andrés Duque Fernández      | Lunes lectivos de 12h a 13h y de 15h a 18h | <b>aduque@scc.uned.es</b>     | 91 398 7162                 |
| Rafael Pastor Vargas        | Lunes lectivos de 16h a 20 h               | <b>rpastor@scc.uned.es</b>    | 91 398 8383                 |
| Salvador Ros Muñoz          | Lunes y martes lectivos de 16h a 18h.      | <b>sros@scc.uned.es</b>       | 91 398 7205                 |

La dirección postal del equipo docente es:  
ETSI Informática. UNED  
C/Juan del Rosal 16.  
28040. Madrid

## 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

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/>



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 multidisciplinarios 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 multidisciplinarios.

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.

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/validar/A03DB3D3A74C846AD7012BF28AE7A0A2>



## CONTENIDOS

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

Contenidos:

- Tema 1: Introducción a Big Data y Hadoop.
- Tema 2: Programación MapReduce.
- Tema 3: Programación MapReduce con lenguajes de alto nivel: Hive y Pig.
- Tema 4: Herramientas de serialización/deserialización e inyección/extracción de datos.

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

Contenidos:

- Tema 1: Introducción e instalación de Apache Spark
- Tema 2: Programación de aplicaciones en Spark
- Tema 3: Librerías/Componentes de Spark
- Tema 4: Configuración, monitorización y optimización de Spark

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

Contenidos:

- Tema 1: Introducción a las arquitecturas de procesamiento de streams: Lambda y Kappa
- Tema 2: Componentes tecnológicos de adquisición y transmisión/distribución de eventos: Kafka.
- Tema 3: Procesamiento de Streams: Spark Streaming.

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

Contenidos:

- Tema 1. Soluciones en la nube para la gestión del almacenamiento de datos.
- Tema 2. Soluciones en la nube para el procesamiento de datos.
- Tema 3. Soluciones en la nube para la implementación de arquitecturas de Streaming.

## 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

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/A03DB3D3A74C846AD7012BF28AE7A0A2>



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         |                  |

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/A03DB3D3A74C846AD7012BF28AE7A0A2>



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

|                                                                  |    |
|------------------------------------------------------------------|----|
| % del examen sobre la nota final                                 | 40 |
| Nota del examen para aprobar sin PEC                             | 4  |
| Nota máxima que aporta el examen a la calificación final sin PEC | 4  |
| Nota mínima en el examen para sumar la PEC                       | 4  |
| Comentarios y observaciones                                      |    |

#### CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

|                         |    |
|-------------------------|----|
| Requiere Presencialidad | No |
| Descripción             |    |

El estudiante debe realizar tres prácticas de laboratorio (trabajos prácticos) que están asociados a tres de los cuatro módulos que componen la asignatura. Las tres prácticas están ponderadas con el 60% de la nota final:

Práctica de laboratorio 1: El estudiante deberá realizar un desarrollo consistente en el procesamiento de un dataset utilizando tanto MapReduce como una herramienta de programación de alto nivel.

Práctica de laboratorio 2: El estudiante deberá realizar un desarrollo analítico usando Spark y usar alguna de las librerías explicadas en el módulo: Graphx o MLLib (aprendizaje máquina).

Práctica de laboratorio 3: El estudiante deberá emplear un proveedor de la nube de los analizados en el módulo para desplegar una solución de procesamiento de datos masivos en la nube.

#### Criterios de evaluación

El equipo docente publicará una guía para su realización, especificando los criterios de evaluación. Es obligatorio entregar las tres prácticas de laboratorio. Cada práctica se evaluará sobre 10 puntos y es necesario obtener una calificación media mínima de 4 sobre 10 para poder superar la asignatura.

|                                                                       |                                                                                                                                                                      |
|-----------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ponderación de la prueba presencial y/o los trabajos en la nota final | Práctica de laboratorio 1 (Módulo 1): 20 % de la nota. Práctica de laboratorio 2 (Módulo 2): 20 % de la nota. Práctica de laboratorio 3 (Módulo 4): 20 % de la nota. |
| Fecha aproximada de entrega                                           | Practica 1: 20 de noviembre; Practica 2: 10 de enero; Practica 3: 25 de enero                                                                                        |
| Comentarios y observaciones                                           |                                                                                                                                                                      |

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/A03DB3D3A74C846AD7012BF28AE7A0A2>



Se podrán entregar además en la convocatoria extraordinaria, con la fecha que indique el equipo docente.

**En caso de haber aprobado las prácticas pero no haber aprobado el examen, las notas de la prácticas se guardarán para la convocatoria extraordinaria de septiembre en el curso presente.**

#### PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC? No

Descripción

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? No

Descripción

Criterios de evaluación

Ponderación en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

#### ¿CÓMO SE OBTIENE LA NOTA FINAL?

La nota final se calcula con la siguiente fórmula:

**Nota Final = 40% NE + 60% PS**

**donde NE es Nota del examen (de 0 a 10) y PS es nota de las prácticas de laboratorio (calificadas de 0 a 10 cada uno).**

**Se deben tener en cuenta las siguientes observaciones:**

Si en el examen presencial o en cada práctica de laboratorio (**de manera individual**) no se obtiene al menos el 40% de la puntuación individual total de cada una de ellas, entonces el/la estudiante estará suspenso.

En otro caso (se tiene más del 40% de la nota total para el examen y para cada una de las prácticas obligatorias), se calculará la nota final sumando las diferentes pruebas de evaluación ponderadas con los porcentajes descritos más arriba.

Aprobarán la asignatura los estudiantes que consigan al menos 5 puntos en la nota final calculada con las ponderaciones definidas más arriba.

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/A03DB3D3A74C846AD7012BF28AE7A0A2>



## 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/>

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/>



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 SankarEditorial: Packt Publishing

Año:2016ISBN-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

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/validar/A03DB3D3A74C846AD7012BF28AE7A0A2>



Autor: Jeffrey Aven.

Editorial: SamsAño: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/>

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/A03DB3D3A74C846AD7012BF28AE7A0A2>



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

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/>



A03DB3D3A74C846AD7012BF28AE7A0A2

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.

---

## 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.

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/A03DB3D3A74C846AD7012BF28AE7A0A2>

