

20-21

MÁSTER UNIVERSITARIO EN
INGENIERÍA Y CIENCIA DE DATOS

GUÍA DE ESTUDIO PÚBLICA



GESTIÓN/ALMACENAMIENTO DE INFORMACIÓN NO ESTRUCTURADA

CÓDIGO 3111008-

UNED

20-21

GESTIÓN/ALMACENAMIENTO DE
INFORMACIÓN NO ESTRUCTURADA
CÓDIGO 3111008-

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	GESTIÓN/ALMACENAMIENTO DE INFORMACIÓN NO ESTRUCTURADA
Código	3111008-
Curso académico	2020/2021
Título en que se imparte	MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA DE DATOS
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	SEMESTRE 2
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

En los últimos años han surgido una gran cantidad de aplicaciones diferentes, que generan datos de forma masiva. Esta cantidad de datos ha generado problemas de escalabilidad y rendimiento en los sistemas de bases de datos tradicionales, basados en el estándar SQL. Para atender estos nuevos requerimientos de almacenamiento han surgido las bases de datos NoSQL, que pretenden acabar con este problema aportando una estructura de almacenamiento más versátil que SQL y recurriendo a estructuras de bases de datos no normalizadas para obtener una mayor escalabilidad, mejorando el rendimiento

Dentro de este Máster es importante adquirir una visión sólida de las herramientas más utilizadas en ese contexto, dado que las arquitecturas Big Data tienen como propósito mover y tratar datos masivos estructurados y no estructurados, de una manera rápida y con herramientas con alta escalabilidad.

En esta asignatura se proporciona una visión general de algunas de las diferentes soluciones existentes en el mercado, clasificándolas por su tipología de almacenamiento, y atendiendo a su uso en proyectos reales dentro del contexto Big Data. El objetivo está en comprender la estructura de estas herramientas y las técnicas desde el punto de vista de su uso y potencial utilidad, entrando en los detalles de su funcionamiento interno y en los criterios especializados para su selección.

Esta asignatura cuenta con 6 ECTS de carga docente y es obligatoria, pertenece a la materia "Infraestructuras y Sistemas para la gestión de datos", junto a las asignaturas "Infraestructuras computacionales para procesamiento de datos masivos" y "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.

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	PABLO RUIPEREZ GARCIA (Coordinador de asignatura)
Correo Electrónico	pablo@scc.uned.es
Teléfono	91398-7159
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL

Nombre y Apellidos	LUIS GRAU FERNANDEZ
Correo Electrónico	lgrau@scc.uned.es
Teléfono	91398-7153
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL

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

HORARIO DE ATENCIÓN AL ESTUDIANTE

La enseñanza a distancia utilizada para el seguimiento de esta asignatura, que garantiza la ayuda al estudiante, dispone de los siguientes recursos:

Entorno Virtual:

- A través del curso virtual el equipo docente de la asignatura pondrá a disposición de los estudiantes diverso material de apoyo al estudio, así como el enunciado de los ejercicios prácticos.
- Se dispone además de foros donde los estudiantes podrán plantear sus dudas para que sean respondidas por los tutores o por el propio equipo docente. Es el SOPORTE FUNDAMENTAL de la asignatura, y supone la principal herramienta de comunicación entre el equipo docente, los tutores y los estudiantes, así como de los estudiantes entre sí.

Contacto con el Equipo docente:

- Dr. Agustín C. Caminero (accaminero@scc.uned.es) Lunes de 11h a 13h y de 15h a 17h
- Dr. Andrés Duque (aduque@scc.uned.es) Lunes de 12h a 13h y de 15h a 18h
- Dr. Luis Grau (lgrau@scc.uned.es) Martes de 15h a 19h
- Dr. Pablo Ruipérez (pablo@scc.uned.es) Lunes de 15h a 19h

Dirección postal:

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 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

CE10 - Describir las diferentes alternativas de almacenamiento de información no estructurada, y los mecanismos disponibles para gestionar y manipular estos tipos de datos

RESULTADOS DE APRENDIZAJE

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

- Definir el concepto de base de datos NoSQL frente a bases de datos relacionales (SQL)
- Describir las principales características de los diferentes tipos de bases de datos NoSQL objeto de estudio.
- Instalar y manejar diferentes bases de datos NoSQL.
- Seleccionar qué base de datos NoSQL utilizar dependiendo del contexto del proyecto.

CONTENIDOS

Módulo 1: Fundamentos de bases de datos NoSQL.

En este módulo se verán los siguientes contenidos:

- Introducción a las bases de datos no estructuradas (NoSQL) y comparativa con bases de datos estructuradas (SQL).
- Tipologías de bases de datos no estructuradas (No Sql).
- Indexación en bases de datos no estructuradas.

Módulo 2: Bases de datos key-value

En este módulo se verán los siguientes contenidos:

- Ejemplo de base de datos key-value.
- Arquitectura.
- Características principales.
- Preparación del entorno para ejercicios prácticos.
- Lenguaje de consultas.

Módulo 3: Bases de datos documentales

En este módulo se verán los siguientes contenidos:

- Ejemplo de base de datos documental.
- Arquitectura.
- Características principales.
- Preparación del entorno para ejercicios prácticos.
- Lenguaje de consultas.

Módulo 4: Bases de datos de tipo columna (columnares)

En este módulo se verán los siguientes contenidos:

- Ejemplo de base de datos de tipo columna.
- Arquitectura.
- Características principales.
- Preparación del entorno para ejercicios prácticos.
- Lenguaje de consultas.

Módulo 5: Bases de datos orientadas a grafos.

En este módulo se verán los siguientes contenidos:

- Ejemplo de base de datos orientada a grafos.
- Arquitectura.
- Características principales.
- Preparación del entorno para ejercicios prácticos.
- Lenguaje de consultas.

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)
- Otras actividades evaluables (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. 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 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. Es necesario conseguir un mínimo de 4 puntos en la prueba presencial antes de ponderarla para poder aprobar la asignatura.

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

Se realizarán cuatro PEC:

PEC 1: Módulo 2: Bases de datos documentales

PEC 2: Módulo 3: Bases de datos key-value

PEC 3: Módulo 4: Bases de datos de tipo columna (columnares)

PEC 4: Módulo 5: Bases de datos orientadas a grafos

Cada PEC consistirá en un cuestionario con 10 preguntas tipo test.

En relación con la convocatoria extraordinaria, las calificaciones de estas pruebas de evaluación se mantendrán. Además, estas pruebas de evaluación tendrán un plazo de entrega extraordinario para dicha convocatoria.

Criterios de evaluación

Cada una de las PEC se evaluará sobre una calificación de 10 puntos. Para superarlas el alumno debe obtener una calificación mínima de 4 puntos en cada una de ellas.

El peso de cada PEC en la nota final es como sigue:

PEC 1: Módulo 2. 2 % de la nota.

PEC 2: Módulo 3. 3 % de la nota.

PEC 3: Módulo 4. 2 % de la nota.

PEC 4: Módulo 5. 3 % de la nota.

En relación con la convocatoria extraordinaria, las calificaciones de estas pruebas de evaluación se mantendrán.

Ponderación de la PEC en la nota final PEC1 y PEC3 valen 2% de la nota. PEC2 y PEC4 valen 3 % de la nota.

Fecha aproximada de entrega PEC1: semana 4 / PEC2: semana 7 / PEC3: semana 9 / PEC4: semana 12

Comentarios y observaciones

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? Si, no presencial

Descripción

El alumno tendrá que realizar cuatro trabajos prácticos:

Trabajo práctico 1: Módulo 2: Bases de datos documentales

Trabajo práctico 2: Módulo 3: Bases de datos key-value

Trabajo práctico 3: Módulo 4: Bases de datos de tipo columna (columnares)

Trabajo práctico 4: Módulo 5: Bases de datos orientadas a grafos

Criterios de evaluación

Trabajos prácticos: 50 % de la nota. Es obligatorio entregar los cuatro trabajos prácticos. Cada trabajo se evaluará sobre 10 puntos y es necesario obtener una calificación media mínima de 4 sobre 10 para poder superar la asignatura.

Trabajo práctico 1: Módulo 2. 12.5 % de la nota.

Trabajo práctico 2: Módulo 3. 12.5 % de la nota.

Trabajo práctico 3: Módulo 4. 12.5 % de la nota.

Trabajo práctico 4: Módulo 5. 12.5 % de la nota.

En relación con la convocatoria extraordinaria, las calificaciones de estas pruebas de evaluación se mantendrán. Además, estas pruebas de evaluación tendrán un plazo de entrega extraordinario para dicha convocatoria.

Ponderación en la nota final	Los trabajos prácticos suponen el 50 % de la nota final
Fecha aproximada de entrega	Trabajo1: semana 2 a la 4 / Trabajo2: semana 5 a la 7 / Trabajo3: semana 8 a la 9 / Trabajo4: semana 10 a la 12
Comentarios y observaciones	

¿CÓMO SE OBTIENE LA NOTA FINAL?

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

Para superar la asignatura es necesario obtener, al menos, el 40% de la puntuación en el examen presencial, en los trabajos y en las PEC (de manera individual por tipo cada actividad). En este caso se calculará la nota final sumando las diferentes pruebas de evaluación ponderadas con los porcentajes descritos anteriormente: prueba presencial 40%, trabajos prácticos 50% y PECs 10%.

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

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

Para la convocatoria extraordinaria, se guardan las calificaciones de PEC y trabajos prácticos. Además, los trabajos prácticos dispondrán de un plazo de entrega extraordinario.

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: Getting Started with NoSQL

Autores: Gaurav Vaish.

Editorial: Packt Publishing

Año: 2013

URL: <https://learning.oreilly.com/library/view/getting-started-with/9781849694988/>

Título: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence

Autores: Pramod J. Sadalage, Martin Fowler.

Editorial: Addison-Wesley Profesional

Año: 2012

URL: <https://learning.oreilly.com/library/view/nosql-distilled-a/9780133036138/>

Título: Seven NoSQL databases in a Week

Autores: Xun Wu, Sudarshan Kadambi, Devram Kandhare, Aaron Ploetz.

Editorial: Packt Publishing

Año: 2018

URL: <https://learning.oreilly.com/library/view/seven-nosql-databases/9781787288867/>

Título: A Deep Dive into NoSQL Databases: The Use Cases and Applications (2018).

Editorial: Academic Press.

Autores: Pethuru R., Deka, G. C.,

URL: <https://www.oreilly.com/library/view/a-deep-dive/9780128137871/>

Título: Learning Redis

Autores: Vinoo Das.

Editorial: Packt Publishing

Año: 2015

URL: <https://learning.oreilly.com/library/view/learning-redis/9781783980123/>

Título: Cassandra: The definitive guide, 2nd edition

Autores: Eben Hewitt, Jeff Carpenter.

Editorial: O'Reilly Media, Inc.

Año: 2016

URL: <https://learning.oreilly.com/library/view/cassandra-the-definitive/9781491933657/>

Título: MongoDB: The definitive guide, 3rd edition

Autores: Kristina Chodorow, Shannon Bradshaw

Editorial: O'Reilly Media, Inc.

Año: 2019

URL: <https://learning.oreilly.com/library/view/mongodb-the-definitive/9781491954454/>

Título: Learning Neo4j 3.x - Second edition

Autores: Rik Van Bruggen, Jérôme Baton.

Editorial: Packt Publishing

Año: 2017

URL: <https://learning.oreilly.com/library/view/learning-neo4j-3x/9781786466143/>

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 O'Reilly for Higher Education (New Safari), 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.