

BASES DE DATOS II

Curso 2022-2023

License CC BY-NC-SA 4.0

Esta obra está bajo una [licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/). Icono diseñado por Flaticon

Descripción de la asignatura

En esta asignatura el alumno aprenderá los conceptos fundamentales, tanto teóricos como prácticos, necesarios para conocer el funcionamiento de las bases de datos NoSQL. Se estudiarán diferentes tipos de tipología de bases de datos diferentes a las que siguen un modelo relacional como son las bases de datos basadas en ficheros o las basadas en grafos.

Durante la asignatura se estudiará las técnicas estandarizadas para preparar los modelos de datos para que se adapten a la forma de manejo de datos de cada tipología de Bases de Datos NoSQL. Del mismo modo se mostrará al alumno la metodología que permite la transición entre las diferentes tipologías.

La asignatura ha sido diseñada con un elevado contenido práctico, de tal manera que los alumnos puedan aplicar los contenidos teóricos aprendidos en clase en entornos reales de trabajo con herramientas ampliamente utilizadas en el sector productivo.

Profesorado

- Edgar Talavera Muñoz(**Coordinador**). e.talavera@upm.es.
Despacho 1222.
- Alberto Díaz Álvarez. alberto.diaz@upm.es. *Despacho 4122.*
- Ángel Panizo Lledot. angel.panizo@upm.es *Despacho 1214.*

Conocimientos previos recomendados

- Fundamentos de la programación
- Bases de Datos I
- Programación para Ciencia de Datos

Otros conocimientos recomendados

- Bases de datos relacionales
- Programación

Competencias

CB03 | Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB04 | Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

CE05 | Capacidad de diseñar e implementar los procesos de selección, limpieza, transformación, integración y verificación de la calidad de los datos de cara a su posterior tratamiento.

CE07 | Capacidad de diseñar e implementar sistemas de información (incluyendo modelos de datos y estrategias de gestión de datos) dimensionados para gestionar el volumen, velocidad y variedad de los datos, de forma adecuada para su almacenamiento, procesamiento y acceso para tratamientos posteriores.

Resultados de aprendizaje

RA129 - Ser capaz de implementar y gestionar una base de datos en un gestor no relacional.

RA89 - Usar lenguajes de programación y de descripción de datos, comunes en Ciencia de Datos.

RA76 - RA-APID-5 Configuración, administración, uso y optimización de sistemas gestores de bases de datos relacionales.

RA77 - RA-APID-6 Diseño, creación, consulta y manipulación de repositorios de datos, e integración con aplicaciones del sistema.

RA134 - RA114 - RA-APID-18 Ser capaz de utilizar las tecnologías de información para preparar los conjuntos de datos.

Temario guía docente

1. Bloque I: Bases de datos avanzadas
 - i. Bases de datos basadas en ficheros
 - ii. Bases de datos federadas
2. Bloque II: Gobernanza de datos
 - i. Gobernanza de datos
 - ii. Data provenance
 - iii. Curación y limpieza de datos
 - iv. Transformación y serialización de datos
3. Bloque III: Bases de datos NoSQL
 - i. Fundamentos de bases de datos NoSQL
 - ii. Comparativa de bases de datos relacionales vs NoSQL
 - iii. Tipos y características de bases de datos NoSQL
 - iv. Bases de datos documentales
 - v. Bases de datos basadas en grafos
 - vi. Bases de datos basadas en clave-valor
 - vii. Otras bases de datos (dispersas, columnares, en memoria, ...)
4. Bloque IV: Data streaming
 - i. Procesamiento de datos continuos en memoria
 - ii. Motores de data streaming

Temario

1. Bloque I: Bases de datos avanzadas
 - i. Introducción a las bases de datos NoSQL (tipos de bases de datos No-SQL)
 - ii. Comparativa entre bases de datos relacionales vs bases de datos NoSQL
 - iii. Tipos de características de bases de datos NoSQL (Bases de datos basadas en ficheros)
2. Bloque II: Gobernanza de datos
 - i. Gobernanza de datos y data provenance
 - ii. Curación y limpieza de datos
 - iii. Transformación y serialización de datos
3. Bloque III: Bases de datos NoSQL
 - i. MongoDB - bases de datos basadas en documentos
 - ii. Cassandra - bases de datos basadas en columnas (key-value)
 - iii. Neo4j - bases de datos basadas en grafos
4. Bloque IV: Data streaming
 - i. Programación contra bases de datos
 - ii. Motores de data streaming

Horario

Día Lectivo	Tema o clase	Evento
1 Febrero (Miércoles)	Presentación	
2 Febrero (Jueves)	Bloque I: Introducción	
8 Febrero (Miércoles)	Bloque I: Introducción	
9 Febrero (Jueves)	Bloque II: Gobernanza	
15 Febrero (Miércoles)	Bloque II: Gobernanza -> Labo	
16 Febrero (Jueves)	Bloque II: Gobernanza	
22 Febrero (Miércoles)	Bloque II: Gobernanza -> Labo	
23 Febrero (Jueves)	Bloque III: MongoDB	
1 Marzo (Miércoles)	Bloque III: MongoDB	
2 Marzo (Jueves)	Bloque III: MongoDB	
8 Marzo (Miércoles)	Bloque III: MongoDB	
9 Marzo (Jueves)	Bloque III: MongoDB	
15 Marzo (Miércoles)	Bloque III: MongoDB	
16 Marzo (Jueves)	Bloque III: Cassandra	
22 Marzo (Miércoles)	Bloque III: Cassandra	
23 Marzo (Jueves)	Bloque III: Cassandra	
29 Marzo (Miércoles)	Bloque III: Cassandra	
30 Marzo (Jueves)	Corrección de la práctica	Entrega práctica 1
5 Abril (Miércoles)		
6 Abril (Jueves)		
12 Abril (Miércoles)	Bloque III: Cassandra	
13 Abril (Jueves)	Bloque III: Cassandra	
20 Abril (Miércoles)	Bloque III: Neo4j	
21 Abril (Jueves)	Bloque III: Neo4j	
27 Abril (Miércoles)	Bloque III: Neo4j	
28 Abril (Jueves)	Bloque III: Neo4j	
3 Mayo (Miércoles)		
4 Mayo (Jueves)	Bloque III: Neo4j	Examen Ingles
10 Mayo (Miércoles)	Bloque III: Neo4j	
11 Mayo (Jueves)	Bloque IV: Programación	
17 Mayo (Miércoles)	Bloque IV: Programación	
18 Mayo (Jueves)	Corrección de la práctica	Entrega Práctica 2

Evaluación progresiva

Prácticas (40%)

- En grupos
- Evaluación en el aula

Examen final (60%)

- Nota mínima de 4
- Todo el temario
- Viernes 9 de junio de 2023 a las 10:00 en el aula 1302

Será necesario alcanzar una nota total ≥ 5 para aprobar la asignatura.

Evaluación prueba global

Prueba escrita el **viernes 9 de junio de 2023 a las 10:00** en el aula **1302** incluyendo preguntas teórico-prácticas de todo el temario de la asignatura.

Es necesario obtener una nota mínima de 4 sobre 10.

Para aprobar sin haber seguido la evaluación progresiva será necesario obtener una nota total ≥ 5 (la prueba global cuenta un 60% del total de la nota, por lo que sacar un 5 en esta prueba no equivale a aprobar la asignatura).

Convocatoria extraordinaria

Prueba escrita el **jueves 6 de julio de 2023 a las 10:00** incluyendo preguntas teórico-prácticas de todo el temario de la asignatura.

Es obligatorio alcanzar una nota mínima de 5 puntos sobre 10.

Recursos didácticos

1. **Moodle de la asignatura.**
2. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program (English Edition) 2o Edición. John Ladley.
3. Data Provenance A Complete Guide - 2020 Edition (English Edition). Gerardus Blokdyk
4. Pandas 1.x Cookbook: Practical recipes for scientific computing, time series analysis, and exploratory data analysis using Python, 2nd Edition (English Edition). Matt Harrison; Theodore Petrou
5. SQL & NoSQL Databases: A Comprehensive Introduction To Relational (SQL) And Non-relational (NoSQL) Databases (English Edition). Danniella Hardin
6. SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management (English Edition). Andreas Meier; Michael Kaufmann
7. Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications. Fabian Hueske, Vasiliki Kalavri