

Presentación

Bases de datos II - Curso 2023-2024

Departamento de Sistemas Informáticos

E.T.S.I. de Sistemas Informáticos - UPM

8 de febrero de 2024



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

Es de esperar que los estudiantes tengan conocimientos previos en:

- Fundamentos de la programación
- Bases de datos relacionales

Pero bueno, se supone que ya sabéis todo esto, ¿no?

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
2. Bloque II: Bases de datos distribuidas
3. Bases de datos NoSQL: documentales
4. Bases de datos NoSQL: grafos
5. Bases de datos NoSQL: clave-valor
6. Bloque VI: Data streaming

Este año no hemos estado finos en organizar el temario

- Gracias E.T.S.I. Ingenieros Informáticos por proponer un temario tan claro

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. Cassandra - bases de datos basadas en columnas (key-value)
- ii. MongoDB - bases de datos basadas en documentos
- iii. Redis - bases de datos basadas en clave-valor
- iv. Neo4j - bases de datos basadas en grafos

4. Bloque IV: Data streaming

Horario

- Martes de 13:00 a 15:00
- Miércoles de 09:00 a 11:00

Evaluación progresiva

Prácticas (40%)

- Dos prácticas, 20% cada una
- En grupos
- Evaluación en el aula

Examen final (60%)

- Nota mínima de 4 sobre 10
- Todo el temario
- Viernes 30 de mayo de 2024 a las 18:00 en el aula 4303

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

- Un 5 en el examen **no es aprobar la asignatura** (evaluación progresiva)

Convocatoria extraordinaria

Prueba escrita el **viernes 8 de julio de 2024 a las 18:00** en el aula **4303**

- Incluye preguntas teórico-prácticas de todo el temario de la asignatura

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

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