

Guía docente

270221 - BDA - Bases de Datos Avanzadas

Última modificación: 11/07/2024

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 747 - ESSI - Departamento de Ingeniería de Servicios y Sistemas de Información.
Titulación: GRADO EN CIENCIA E INGENIERÍA DE DATOS (Plan 2017). (Asignatura obligatoria).
Curso: 2024 **Créditos ECTS:** 6.0 **Idiomas:** Catalán, Inglés

PROFESORADO

Profesorado responsable: ALBERTO ABELLO GAMAZO
Otros: Primer quadrimestre:
ALBERTO ABELLO GAMAZO - 11, 12, 13
BESIM BILALLI - 11, 12, 13

CAPACIDADES PREVIAS

Ser capaz de leer y entender materiales en inglés.
Ser capaz de enumerar las etapas que conforman el proceso de la ingeniería de software.
Ser capaz de entender esquemas conceptuales en UML.
Ser capaz de crear, consultar y manipular bases de datos con SQL.
Ser capaz de programar utilizando programación funcional tipo Spark.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CE7. Demostrar conocimiento y capacidad de aplicación de las herramientas necesarias para el almacenaje, el procesamiento y el acceso a los datos.

Genéricas:

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.
CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

Transversales:

CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Básicas:

CB2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
CB3. 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.

METODOLOGÍAS DOCENTES

La asignatura se compone de teoría, y sesiones de laboratorio.

Teoría: Se utilizarán técnicas de clase invertida que requieren que el estudiante trabaje materiales multimedia antes de la clase. Las clases de teoría consisten en explicaciones complementarias del profesor y resolución de problemas.

Laboratorio: Se utilizarán herramientas representativas para la aplicación de conceptos teóricos (por ejemplo, PostgreSQL, Talend, HDFS, MongoDB). Se harán también dos proyectos, en el que los estudiantes trabajarán en equipos: uno de análisis descriptivo de datos en un almacén de datos y el otro de análisis predictivo en un entorno Big Data. Consecuentemente, habrá dos entregables fuera del horario de clase, pero los estudiantes también serán evaluados individualmente en el aula sobre los conocimientos adquiridos durante cada uno de los proyectos.

El curso tiene una componente de aprendizaje autónomo, dado que los estudiantes tendrán que trabajar con diferentes herramientas de gestión y procesamiento de datos. Aparte del material de apoyo, los estudiantes deben ser capaces de resolver dudas o problemas de uso de estas herramientas.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1. Ser capaz de discutir y justificar con detalle los principios arquitectónicos y cuellos de botella de los gestores relacionales frente a sistemas de almacenamiento y procesamiento alternativos.
2. Ser capaz de obtener el esquema lógico de un almacén de datos a partir de un esquema conceptual expresado en UML, detectar y corregir sus defectos.
3. Ser capaz de explicar y utilizar los principales mecanismos de procesamiento paralelo de consultas en entornos distribuidos, y detectar cuellos de botella.
4. Ser capaz de justificar y utilizar NOSQL storage systems.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTE

Tipo	Horas	Porcentaje
Horas grupo grande	30,0	20.00
Horas grupo pequeño	30,0	20.00
Horas aprendizaje autónomo	90,0	60.00

Dedicación total: 150 h

CONTENIDOS

Introducción

Descripción:

Almacenamiento de datos en entornos analíticos y Big Data

Almacenamiento de datos

Descripción:

Almacenamiento de datos. Flujos de datos ETL. Integración de datos. Herramientas OLAP.

Bases de datos distribuidas

Descripción:

Taxonomía de bases de datos distribuidas. Arquitecturas. Diseño de bases de datos distribuidas (fragmentación y replicación). Paralelismo. Métricas de escalabilidad. Sistemas de ficheros distribuidos.

Procesamiento distribuido de datos

Descripción:

Importancia del acceso secuencial paralelo. Barreras de sincronización (Bulk Synchronous Parallel modelo). Arquitecturas de Big Data y sistemas NOSQL.

ACTIVIDADES

Introducción

Descripción:

Introducción de la asignatura, motivación y visión general de las herramientas de gestión de datos existentes actualmente, sus ventajas e inconvenientes

Objetivos específicos:

1

Competencias relacionadas:

CB2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

CB3. 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.

CE7. Demostrar conocimiento y capacidad de aplicación de las herramientas necesarias para el almacenaje, el procesamiento y el acceso a los datos.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesamiento de la señal.

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Dedicación: 2h

Grupo grande/Teoría: 2h

Estudio de los almacenes de datos

Objetivos específicos:

2

Competencias relacionadas:

CB2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

CB3. 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.

CE7. Demostrar conocimiento y capacidad de aplicación de las herramientas necesarias para el almacenaje, el procesamiento y el acceso a los datos.

CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Dedicación: 62h

Aprendizaje autónomo: 38h

Grupo grande/Teoría: 10h

Grupo pequeño/Laboratorio: 14h

Estudio de bases de datos distribuidas

Descripción:

Aprendizaje de los principios de bases de datos distribuidas y su aplicación en sistemas NOSQL

Objetivos específicos:

1, 3

Competencias relacionadas:

CB2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

CB3. 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.

CE7. Demostrar conocimiento y capacidad de aplicación de las herramientas necesarias para el almacenaje, el procesamiento y el acceso a los datos.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Dedicación: 14h

Aprendizaje autónomo: 4h

Grupo grande/Teoría: 6h

Grupo pequeño/Laboratorio: 4h

Estudio del procesamiento distribuido de datos y sistemas NOSQL

Descripción:

Aprendizaje de técnicas de procesamiento distribuido de datos y sistemas NOSQL

Objetivos específicos:

1, 3, 4

Competencias relacionadas:

CB2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

CB3. 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.

CE7. Demostrar conocimiento y capacidad de aplicación de las herramientas necesarias para el almacenaje, el procesamiento y el acceso a los datos.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Dedicación: 60h

Aprendizaje autónomo: 38h

Grupo grande/Teoría: 10h

Grupo pequeño/Laboratorio: 12h

Examen final

Descripción:

Examen global de la asignatura

Objetivos específicos:

1, 2, 3, 4

Competencias relacionadas:

CB2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

CB3. 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.

CE7. Demostrar conocimiento y capacidad de aplicación de las herramientas necesarias para el almacenaje, el procesamiento y el acceso a los datos.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

Dedicación: 12h

Aprendizaje autónomo: 10h

Actividades dirigidas: 2h

SISTEMA DE CALIFICACIÓN

Nota final = $\max(20\%EP + 40\%EF ; 60\% EF) + 40\% P$

EP = nota del examen parcial

EF = nota del examen final

P = nota de proyecto, como promedio ponderado de los proyectos del curso

Para los estudiantes que puedan concurrir a la reevaluación, la nota del examen de reevaluación sustituirá EF.



BIBLIOGRAFÍA

Básica:

- Garcia-Molina, Hector; Ullman, Jeffrey D; Widom, Jennifer. Database systems : the complete book [en línea]. Second edition, Pearson new international edition. Essex: Pearson Education, 2013 [Consulta: 14/03/2025]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=5174436>. ISBN 9781292024479.
- Database Technologies and Information Management. Slides on Advanced Databases course.
- Golfarelli, M.; Rizzi, S. Data warehouse design: modern principles and methodologies. New York [etc.]: McGraw Hill, 2009. ISBN 9780071610391.
- Vaisman, A.; Zimányi, E. Data warehouse systems: design and implentation. Second edition. Berlin: Springer, 2022. ISBN 9783662651667.
- Özsu, M.T.; Valduriez, P. Principles of distributed database systems. 4th ed. New York: Springer, 2020. ISBN 9783030262525.
- Sadalage, P.J.; Fowler, M. NoSQL distilled: a brief guide to the emerging world of polygot persistence. Addison-Wesley, 2013. ISBN 9780321826626.
- Badia, Antonio. SQL for data science : data cleaning, wrangling and analytics with relational databases. Springer, 2020. ISBN 9783030575915.
- Abelló, Alberto; Jovanovic, Petar. Data Warehousing and OLAP.
- Abelló, Alberó; Nadal, Sergi. Big Data Management.

Complementaria:

- Exercises Big Data Management.
- Exercises Data Warehousing.

RECURSOS

Enlace web:

- <https://bdma.ulb.ac.be/bdma>- <https://cs.ulb.ac.be/conferences/ebiss.html>