



Guía docente 270952 - DW - Almacenes de Datos

Última modificación: 23/11/2023

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 747 - ESSI - Departamento de Ingeniería de Servicios y Sistemas de Información.

Titulación: MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (Plan 2021). (Asignatura obligatoria).

Curso: 2023 **Créditos ECTS:** 6.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: ALBERTO ABELLO GAMAZO

Otros: Primer quadrimestre:
ALBERTO ABELLO GAMAZO - 11, 12
PETAR JOVANOVIC - 11, 12

CAPACIDADES PREVIAS

Conocimiento básico de bases de datos relacionales y SQL.

Específicamente, se asumira conocimientos de:

- Diagramas de classes UML
- Algebra relacional
- Consultas SQL
- Vistas relacionales
- Operaciones de árboles-B (inserciones y splits)
- Conceptes bàsics d'optimització física de consultes

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

- CE2. Aplicar los fundamentos de la gestión y procesamiento de datos en un problema de ciencia de datos
- CE3. Aplicar métodos de integración de datos para dar solución a problemas de ciencia de datos en entornos heterogéneos
- CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos
- CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

Genéricas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

Transversales:

CT1. ESPÍRITU EMPRENDEDOR E INNOVADOR: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio. Conocer y entender los mecanismos en que se basa la investigación científica, así como los mecanismos e instrumentos de transferencia de resultados entre los diferentes agentes socioeconómicos implicados en los procesos de I+D+i.

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.



Básicas:

CB10. 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.

CB6. 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.

CB7. 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

CB8. 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.

CB9. 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.

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 los materiales multimedia antes de la clase. Las clases de teoría consisten en explicaciones complementarias del profesor y resolución de problemas.

Laboratorio: Se utilizarán algunas herramientas representativas para la aplicación de conceptos teóricos (por ejemplo, Indyco Builder, PotgreSQL, Oracle, Pentaho Data Integration, Tableau). El curso incluye prácticas continuas a través de un proyecto del curso, dividido en tres bloques lógicos: modelado de almacén de datos, integración y migración de datos (ETL), y visualización descriptiva, en los que los estudiantes trabajarán en equipos. Habrá tres entregables del proyecto fuera del horario de clase, pero los estudiantes también serán evaluados individualmente en el aula sobre los conocimientos adquiridos durante cada bloque del proyecto.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1. Ser capaz de modelar almacenes de datos multidimensionales y analizar sus datos de forma visual
2. Ser capaz de aplicar técnicas específicas de diseño físico para sistemas decisionales
3. Ser capaz de diseñar e implementar procesos de migración de datos (ETL)

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	96,0	64.00
Horas grupo pequeño	27,0	18.00
Horas grupo grande	27,0	18.00

Dedicación total: 150 h

CONTENIDOS

Introducción

Descripción:

Comparación entre sistemas operacionales y decisionales; Metadatos.

Arquitectura de almacenamiento de datos.

Descripción:

Factoria de Información Corporativa. DW 2.0



Modelado multidimensional, herramientas OLAP

Descripción:

Estructura; restricciones de integridad, operaciones, conceptos avanzados.

Diseño físico de bases de datos para consultas analíticas

Descripción:

Star-join e índices-join; Bitmaps; Vistas materializadas; Datos espacio-temporales

Extracción, transformación y carga

Descripción:

Calidad de los datos; Integración de datos y esquemas; Gestión de ETL

Visualización y análisis descriptivo de la información

Descripción:

Indicadores de Rendimiento Empresarial; Cuadros de Mando

ACTIVIDADES

Clases de teoría

Descripción:

En estas actividades, el profesor introducirá los conceptos teóricos básicos de esta asignatura. Además de las exposiciones, se usarán técnicas de aprendizaje cooperativo. Esto requerirá la participación activa de los estudiantes y, consecuentemente, serán evaluados.

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CE2. Aplicar los fundamentos de la gestión y procesamiento de datos en un problema de ciencia de datos

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE3. Aplicar métodos de integración de datos para dar solución a problemas de ciencia de datos en entornos heterogéneos

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CT1. ESPÍRITU EMPRENDEDOR E INNOVADOR: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio. Conocer y entender los mecanismos en que se basa la investigación científica, así como los mecanismos e instrumentos de transferencia de resultados entre los diferentes agentes socioeconómicos implicados en los procesos de I+D+i.

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CB9. 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.

CB10. 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 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

CB8. 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.

CB6. 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.

Dedicación: 50h

Grupo grande/Teoría: 25h

Aprendizaje autónomo: 25h



Clases de laboratorio

Descripción:

Al estudiante se le pedirá que practique los diferentes conceptos introducidos en las clases teóricas. Esto incluye resolver problemas en el ordenador o en papel.

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CE2. Aplicar los fundamentos de la gestión y procesamiento de datos en un problema de ciencia de datos

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE3. Aplicar métodos de integración de datos para dar solución a problemas de ciencia de datos en entornos heterogéneos

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CT1. ESPÍRITU EMPRENDEDOR E INNOVADOR: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio. Conocer y entender los mecanismos en que se basa la investigación científica, así como los mecanismos e instrumentos de transferencia de resultados entre los diferentes agentes socioeconómicos implicados en los procesos de I+D+i.

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CB9. 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.

CB10. 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 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

CB8. 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.

CB6. 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.

Dedicación: 81h

Grupo pequeño/Laboratorio: 27h

Aprendizaje autónomo: 54h



Examen

Descripción:

Examen escrito de los conceptos teóricos introducidos durante el curso.

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CG1. Identificar y aplicar los métodos y procesos de gestión de datos más adecuados para gestionar el ciclo de vida de los datos, incluyendo datos estructurados y no estructurados

CE2. Aplicar los fundamentos de la gestión y procesamiento de datos en un problema de ciencia de datos

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE3. Aplicar métodos de integración de datos para dar solución a problemas de ciencia de datos en entornos heterogéneos

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CT1. ESPÍRITU EMPRENDEDOR E INNOVADOR: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio. Conocer y entender los mecanismos en que se basa la investigación científica, así como los mecanismos e instrumentos de transferencia de resultados entre los diferentes agentes socioeconómicos implicados en los procesos de I+D+i.

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, 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.

CB9. 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.

CB10. 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 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

CB8. 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.

CB6. 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.

Dedicación: 19h

Grupo grande/Teoría: 2h

Aprendizaje autónomo: 17h

SISTEMA DE CALIFICACIÓN

Final grade = $\min(10 ; \max(20\%EP+40\%EF ; 60\% EF) + 40\% P + 10\% C)$

EP = partial (mid term) exam mark

EF = final exam mark

P = Weighted average of the marks of the project deliverables

C = participation in the class



BIBLIOGRAFÍA

Básica:

- Golfarelli, M.; Rizzi, S. Data warehouse design: modern principles and methodologies. New York [etc.]: McGraw Hill, 2009. ISBN 9780071610391.
- Jensen, C.S.; Pedersen, T.B.; Thomsen, C.W. Multidimensional databases and data warehousing. San Rafael, Calif: Morgan & Claypool, 2010. ISBN 9781608455379.
- Inmon, William H; Imhoff, Claudia; Sousa, Ryan. Corporate information factory. 2nd ed. New York: John Wiley, 2001. ISBN 0471399612.
- Kimball, R. [et al.]. The data warehouse lifecycle toolkit. 2nd ed. Indianapolis: Wiley publishing, 2008. ISBN 9780470149775.
- Garcia-Molina, H.; Ullman, J.D.; Widom, J. Database systems: the complete book. 2nd ed. Harlow, Essex: Pearson Education Limited, 2013. ISBN 9781292037301.
- Vaisman, A.; Zimanyi, E. Data warehouse systems: design and implementation. Berlin: Springer, 2014. ISBN 9783642546549.
- Taniar, D.; Rahayu, W. Data warehousing and analytics: fueling the data engine. Cham: Springer, 2021. ISBN 9783030819781.

Complementaria:

- Teorey, Toby ... [et al.]. Database modeling and design: logical design. 5th ed. Burlington, MA: Morgan Kaufmann Publishers/Elsevier, 2011. ISBN 9780123820204.
- Lightstone, S.; Teorey, T.J.; Nadeau, T. Physical database design: the database professional's guide to exploiting indexes, views, storage, and more. Amsterdam [etc.]: Morgan Kaufmann Publishers, 2007. ISBN 9780123693891.
- Ramakrishnan, R.; Gehrke, J. Database management systems. 3rd ed. Boston [etc.]: McGraw-Hill, 2003. ISBN 0071151109.
- Lewis, J. Cost-based oracle fundamentals. Berkeley, CA: Apress, 2006. ISBN 9781590596364.

RECURSOS

Enlace web:

- <http://cs.ulb.ac.be/conferences/ebiss.html>- <http://tdwi.org>- <https://deds.ulb.ac.be>