

**Guía docente de la asignatura**

Asignatura	DISEÑO FÍSICO DE GRANDES ALMACENES DE DATOS ORIENTADOS A LA REPRESENTACIÓN DE CONOCIMIENTO		
Materia	SISTEMAS INTELIGENTES Y BASADOS EN CONOCIMIENTO		
Módulo	TECNOLOGÍAS INFORMÁTICAS		
Titulación	MÁSTER EN INGENIERÍA INFORMÁTICA		
Plan	510	Código	53180
Periodo de impartición	2º CUATRIMESTRE	Tipo/Carácter	OPTATIVA
Nivel/Ciclo	MÁSTER	Curso	1º
Créditos ECTS	3 ECTS		
Lengua en que se imparte	CASTELLANO		
Profesor/es responsable/s	JOAQUÍN ADIEGO RODRÍGUEZ		
Datos de contacto (E-mail, teléfono...)	TELÉFONO: 983 423000 ext. 5646 E-MAIL: jadiego@infor.uva.es		
Horario de tutorías	Véase www.uva.es → Centros → Campus de Valladolid → Escuela Técnica Superior de Ingeniería Informática → Tutorías		
Departamento	INFORMÁTICA		



1. Situación / Sentido de la Asignatura

1.1 Contextualización

La asignatura Diseño Físico de Grandes Almacenes de Datos Orientados a la Representación de Conocimiento se encuadra dentro de la materia Sistemas Inteligentes y Basados en Conocimiento y ofrece al alumno los conocimientos fundamentales para afrontar el diseño físico de un almacén de datos (data warehouse) en un escenario tecnológico gobernando por la generación de colecciones de datos de tamaños cada vez más grandes (Big Data).

Esta asignatura se centra en dotar al alumno con los conocimientos y habilidades necesarias para afrontar el diseño físico de un almacén de datos en diferentes situaciones del mundo real, acordes con el escenario tecnológico actual. Los contenidos se distribuyen en cuatro temas que presentarán los conceptos fundamentales del área, las técnicas tradicionales de diseño físico (estructuras específicas de almacenamiento, indexación, particionamiento de datos, etc.) y su aplicación para soluciones base relacional, así como en otras de naturaleza no relacional (NoSQL) especializadas en la gestión de datos a gran escala.

1.2 Relación con otras materias

Métodos Avanzados de Razonamiento y Representación del Conocimiento.

Métodos Avanzados de Aprendizaje y Minería de Datos.

Web Semántica y Extracción de Información.

1.3 Prerrequisitos

Estructuras de Datos.

Bases de Datos.

Administración de Bases de Datos.



2. Competencias

2.1 Generales

Código	Descripción
CG1	Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la ingeniería informática.
CG4	Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática.

2.2 Específicas

Código	Descripción
CET5	Capacidad para analizar las necesidades de información que se plantean en un entorno y llevar a cabo en todas sus etapas el proceso de construcción de un sistema de información.
CET9	Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

3. Objetivos

Código	Descripción
CET5.1	Entender la problemática de representar, consultar y consumir grandes colecciones de datos, evaluando su influencia en diferentes áreas de la sociedad actual.
CET5.2	Conocer los mecanismos tradicionales para la optimización física de bases de datos relacionales.
CET5.3	Apreciar la necesidad de utilizar nuevos modelos para el diseño de almacenes de datos (NoSQL).
CET5.4	Conocer y entender cómo la compresión y la distribución de los datos pueden ser utilizadas para la optimización física de almacenes de datos.
CET9.1	Entender cómo el diseño físico de los almacenes de datos afecta directamente a la eficiencia de los algoritmos utilizados para la extracción de información y, por consiguiente, al rendimiento de los sistemas construidos sobre estos grandes volúmenes de datos.

**4. Tabla de dedicación del estudiante a la asignatura**

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teórico-prácticas (T/M)	16	Estudio y trabajo autónomo individual	30
Clases prácticas de aula (A)		Estudio y trabajo autónomo grupal	15
Laboratorios (L)	10		
Prácticas externas, clínicas o de campo			
Seminarios (S)	4		
Tutorías grupales (TG)			
Evaluación (fuera del periodo oficial de exámenes)			
Total presencial	30	Total no presencial	45



5. Bloques temáticos

Bloque 1: Diseño físico de grandes almacenes de datos

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

El almacén de datos (*data warehouse*) es un valor principal dentro de una organización ya que recoge los datos sobre los cuáles dicha organización soporta su toma de decisiones. En general, estos almacenes de datos combinan varias fuentes de información que son accedidas mediante diferentes tipos de consultas y sirven de base para la generación de diferentes tipos de informes. Tradicionalmente, la implementación de los almacenes de datos se ha llevado a cabo utilizando tecnología relacional apoyándose en la madurez de los sistemas gestores de bases de datos disponibles así como en la flexibilidad que proporciona SQL como lenguaje de consulta. En este caso, los conceptos tradicionales de diseño físico aplican de forma específica a la implementación del almacén. Sin embargo, el escenario tecnológico actual muestra un crecimiento sin precedentes en la generación de datos de interés para la organización. Este gran volumen de información compromete seriamente la escalabilidad del almacén, siendo necesaria la aplicación de otros conceptos más avanzados y otras tecnologías de naturaleza no relacional (NoSQL).

b. Objetivos de aprendizaje

CET5.1	Entender la problemática de representar, consultar y consumir grandes colecciones de datos, evaluando su influencia en diferentes áreas de la sociedad actual.
CET5.2	Conocer los mecanismos tradicionales para la optimización física de bases de datos relacionales.
CET5.3	Apreciar la necesidad de utilizar nuevos modelos para el diseño de almacenes de datos (NoSQL).
CET5.4	Conocer y entender cómo la compresión y la distribución de los datos pueden ser utilizadas para la optimización física de almacenes de datos.
CET9.1	Entender cómo el diseño físico de los almacenes de datos afecta directamente a la eficiencia de los algoritmos utilizados para la extracción de información y, por consiguiente, al rendimiento de los sistemas construidos sobre estos grandes volúmenes de datos.

c. Contenidos

TEMA 1: Almacenes de datos

- 1.1 Data Warehouses
- 1.2 Big Data
- 1.3 Data Warehouses para Big Data

TEMA 2: Fundamentos de diseño físico

- 2.1 Jerarquía de memoria
- 2.2 Diseño físico centralizado
- 2.3 Diseño físico distribuido



TEMA 3: Diseño físico de almacenes relacionales

- 3.1 Arquitecturas de datos OLAP
- 3.2 Ubicación de datos distribuidos

TEMA 4: Diseño físico de almacenes no relacionales

- 4.1 Arquitectura de datos para Big Data
- 4.2 Concepto de NoSQL
- 4.3 Almacenes NoSQL
- 4.4 Procesamiento de Big Data

d. Métodos docentes

- Véase **Anexo: Métodos Docentes**

e. Plan de trabajo

La asignatura está formada por un único bloque dividido en 4 temas. Se estudiarán diferentes técnicas de diseño físico de Bases de Datos tanto SQL como no SQL. En la parte práctica, el alumno deberá presentar un trabajo que podrá ser de carácter práctico y/o teórico/bibliográfico.

f. Evaluación

Véase el apartado 7.

g. Bibliografía básica

- S. Lightstone, T. Teorey y T. Nadeu, *Physical Database Design*. Morgan Kaufmann Publishers, 2007. ISBN: 978-0123693891
- N. Marz y J. Warren, *Big Data: Principles and best practices of scalable realtime data systems*. Manning, 2013. ISBN: 978-1617290343
- E. Redmon y J.R. Wilson, *Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement*. Pragmatic Programmers, 2012. ISBN: 978-1934356920

h. Bibliografía complementaria

- S. Tiwari, *Professional NoSQL*. Wrox, 2011. ISBN: 978-0470942246
- I.H. Witten, A. Moffat y T.C.Bell, *Managing Gigabytes: Compressing and Indexing Documents and Images*. Morgan Kaufmann, 1999. ISBN: 978-1558605701

i. Recursos necesarios

<http://www.infor.uva.es/~jadiesgo>

**6. Temporalización (por bloques temáticos)**

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Bloque 1: Diseño físico de grandes almacenes de datos	3 ECTS	Semanas 1 a 8

7. Sistema de calificaciones – Tabla resumen

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Entrega práctica	25%	Aproximadamente semana 8
Examen final escrito	75%	Periodo de exámenes

8. Anexo: Métodos docentes

Actividad	Metodología
Clase de teoría	<ul style="list-style-type: none">• Cada clase de teoría está diseñada como una actividad completa y autocontenida compuesta de diversas actividades dirigidas a facilitar la adquisición de conocimientos, habilidades y competencias.• La asignatura combina la exposición de temas y realización de ejercicios por parte del profesor, con la realización de ejercicios individuales o en grupo por parte de los alumnos. Podrá haber sesiones específicas donde los alumnos expondrán sus soluciones propuestas. Estas sesiones serán anunciadas previamente por el profesor.• La teoría básica necesaria será expuesta en clase por el profesor de la asignatura, con ayuda de la pizarra y/o algún método de proyección, utilizando ejemplos variados tanto para introducir conceptos como para asimilar los ya introducidos.• Se suministrará al alumno una colección de documentos o enlaces a los mismos que contienen, ocasionalmente en forma ampliada, la documentación básica relacionada con el problema a resolver en la clase. Se desarrollarán ejemplos ilustrativos de la metodología de solución de pequeños problemas relacionados con el problema principal a resolver. El alumno debe utilizar la documentación extra para realizar las tareas encargadas.• Será importante que el alumno intente resolver los ejercicios propuestos en la documentación entregada al comienzo del curso, y así se le hará saber.• Así+mismo, los estudiantes conocerán con antelación los ejercicios que serán resueltos en cada clase práctica y el profesor solicitará su colaboración para responder diferentes cuestiones sobre los problemas.
Clase práctica	<ul style="list-style-type: none">• Durante la semana previa a la sesión o sesiones de prácticas de laboratorio el alumno estudiará de manera personal o en grupo la documentación relativa a las tareas correspondientes a las sesiones de laboratorio. Las horas presenciales de laboratorio incluirán para su desarrollo clase magistral participativa y la realización de un proyecto guiado por el profesor, que encargará y guiará el trabajo que se realizará de manera o individual o en grupos (2/3 alumnos), siguiendo un enfoque colaborativo.
Seminarios	<ul style="list-style-type: none">• Durante el curso se podrán celebrar varios seminarios, con el objeto de afianzar y completar algunos aspectos muy relacionados con la misma y para facilitar el desarrollo de algunas competencias genéricas. Estos seminarios tendrán un carácter teórico y práctico.• Los alumnos podrán ser distribuidos en un grupo de trabajo (el número de integrantes puede variar según circunstancias), cada uno de los cuales junto con el profesor llevará a cabo los seminarios previstos.• En estos seminarios el profesor orientará la actividad de los alumnos en



	<p>relación con la asignatura, exponiendo estos sus problemas con el aprendizaje de la materia. El profesor, previamente a cada seminario, propondrá a cada grupo de trabajo la resolución de varias cuestiones o problemas que deberán ser entregadas en el mismo y sobre los que los alumnos tendrán que debatir.</p> <ul style="list-style-type: none"> • Cada alumno entregará en cada seminario una hoja al empezar con su propuesta de solución, y otra al terminar con la nueva solución que propone y los comentarios que recojan de forma esquemática su aprendizaje en el seminario. El objetivo de esta actividad es que el alumno reconozca su propio aprendizaje y detecte posibles errores en el mismo, así como que el profesor esté informado de la marcha del curso, lo cual puede facilitar una reorientación de actividades o la recomendación de actuaciones particulares para mejorar el aprendizaje individual. En la calificación final se tendrá en cuenta la participación en los seminarios, y las soluciones propuestas.
Tutoría	<ul style="list-style-type: none"> • Las tutorías individualizadas podrán ser atendidas en las seis horas oficiales que se podrán consultar en la web de la Universidad de Valladolid a principio de curso o a cualquier otra hora, previa cita con el profesor. Como alternativa, se propondrá el uso de alguna plataforma de e-learning para la resolución de dudas y creación de debates relacionados con los temas que se están estudiando.
Actividades no presenciales	<ul style="list-style-type: none"> • Los alumnos deben realizar una serie de actividades fuera del aula, aprendiendo a gestionar su tiempo y organizar su trabajo. Incluyen tanto encargos específicos como actividades generales: <ul style="list-style-type: none"> ○ Preparación de sesiones. Los alumnos reciben el encargo de leer bibliografía y preparar dudas previamente a una sesión. Para ello se les suministrarán referencias, enlaces a documentos y/o material extra. ○ Repaso de conceptos y ejercicios de consolidación. El alumno debe dedicar al menos dos horas por cada sesión para repasar y afianzar los conceptos presentados. Puede utilizar ejercicios y problemas extras suministrados por el profesor para comprobar su progreso. ○ Laboratorio personal. El profesor pondrá a disposición de los alumnos el material necesario para que en su casa (si disponen de ordenador) o en el laboratorio de la facultad puedan realizar programas similares a los que se realizan en las sesiones presenciales. El entorno, metodología y herramientas serán los mismos que se utilizan en clase. De esta forma, el alumno podrá comprobar si la experiencia adquirida en las clases se traduce en un aumento correspondiente de su destreza en la materia.

9. Anexo: Cronograma de actividades previstas

Nota: La información mostrada en el siguiente cronograma es aproximada y provisional, por lo que puede sufrir cambios. Consulte periódicamente la página web de la asignatura para acceder a información actualizada.

Semana	Contenido	Teoría	Laboratorio
1	<u>Tema 1:</u> Almacenes de datos	2	
2		2	2
3	<u>Tema 2:</u> Fundamentos de diseño físico	2	2
4		2	2
5	<u>Tema 3:</u> Diseño físico de almacenes relacionales	2	2 (Seminario)
6		2	2
7	<u>Tema 4:</u> Diseño físico de almacenes no relacionales	2	2 (Seminario)
8		2	2