

30253 - Almacenes y minería de datos

Información del Plan Docente

Año académico	2018/19
Asignatura	30253 - Almacenes y minería de datos
Centro académico	110 - Escuela de Ingeniería y Arquitectura 326 - Escuela Universitaria Politécnica de Teruel
Titulación	439 - Graduado en Ingeniería Informática 443 - Graduado en Ingeniería Informática
Créditos	6.0
Curso	4
Periodo de impartición	Semestral
Clase de asignatura	
Módulo	---

1. Información Básica

1.1. Objetivos de la asignatura

La asignatura y sus resultados previstos responden a los siguientes planteamientos y objetivos:

En la asignatura el alumnado se familiarizará con el diseño de almacenes de datos y su posterior explotación utilizando técnicas de minería de datos. Se abordarán los conceptos teóricos, las distintas técnicas de diseño de almacenes de datos existentes, y las diferentes aproximaciones que pueden adoptarse para luego formular preguntas interesantes sobre el almacén de datos. Cuando sea posible, se recurrirá a ejemplos reales para ilustrar los diversos elementos tratados. También se tratarán herramientas informáticas existentes que dan soporte a la solución de este tipo de problemas.

1.2. Contexto y sentido de la asignatura en la titulación

Esta asignatura persigue que el alumnado comprenda el diseño y la explotación de almacenes de datos. De ahí su relación con diversas asignaturas de la titulación, como "Bases de Datos", "Bases de Datos 2" y "Sistemas de Información":

- En relación a la materia de Bases de Datos, es necesario señalar que en las asignaturas previas se trata el diseño de bases de datos tradicionales (adecuadas para el procesamiento de transacciones) pero no el diseño de almacenes de datos (adecuados para el procesamiento analítico), por lo que existe una complementariedad con las asignaturas previas.
- En relación a la asignatura de "Sistemas de Información", es necesario incidir en que las empresas que manejen un gran volumen de datos cuentan con almacenes de datos para realizar procesamiento analítico mediante técnicas de minería de datos.

Por último, señalar la relación de esta asignatura con la asignatura "Sistemas de Ayuda a la Toma de Decisiones". Esta asignatura se centra en el almacenamiento de datos y la extracción de información a partir de ellos, mientras que "Sistemas de Ayuda a la Toma de Decisiones" se centra en el diseño y explotación de sistemas que emplean la información extraída para la consecución de los objetivos empresariales.

1.3. Recomendaciones para cursar la asignatura

El alumno que curse esta asignatura ha de contar con conocimientos básicos de Sistemas de Información y Bases de Datos.

2.Competencias y resultados de aprendizaje

2.1.Competencias

Al superar la asignatura, el estudiante será más competente para...

- Planificar, presupuestar, organizar, dirigir y controlar tareas, personas y recursos (CT2).
- Combinar los conocimientos generalistas y los especializados de Ingeniería para generar propuestas innovadoras y competitivas en la actividad profesional (CT3).
- Resolver problemas y tomar decisiones con iniciativa, creatividad y razonamiento crítico (CT4)
- Aplicar las tecnologías de la información y las comunicaciones en la Ingeniería (CT11).
- Integrar soluciones de Tecnologías de la Información y las Comunicaciones y procesos empresariales para satisfacer las necesidades de información de las organizaciones, permitiéndoles alcanzar sus objetivos de forma efectiva y eficiente, dándoles así ventajas competitivas (CESI1).
- Participar activamente en la especificación, diseño, implementación y mantenimiento de los sistemas de información y comunicación (CESI3).

2.2.Resultados de aprendizaje

El estudiante, para superar esta asignatura, deberá demostrar los siguientes resultados...

- R1: Es capaz de describir los aspectos básicos de los almacenes de datos y de explicar los procesos necesarios para su diseño.
- R2: Es capaz de diseñar un almacén de datos de forma efectiva.
- R3: Conoce los aspectos básicos de los almacenes de datos y es capaz de buscar información adicional sobre otros temas relacionados avanzados.
- R4: Es capaz de comparar diversos diseños de almacenes de datos y de elegir un sistema comercial adecuado para su implantación.
- R5: Conocer las técnicas básicas de aprendizaje automático aplicado a la minería de datos.
- R6: Utilizar las técnicas básicas en problemas reales.

2.3.Importancia de los resultados de aprendizaje

El conjunto de los resultados de aprendizaje se puede resumir diciendo que el alumnado será capaz de diseñar e implementar almacenes de datos y aplicar técnicas básicas de minería de datos sobre los mismos. Los almacenes de datos (*data warehouses*) son un tipo especial de base de datos especialmente diseñadas para almacenar grandes cantidades de datos (datos históricos) sobre los que se pueden ejecutar consultas complejas de tipo analítico. En particular, los almacenes de datos facilitan la minería de datos, que permite extraer información que puede ser útil para la toma de decisiones en las diferentes organizaciones y entidades. Por ello, se considera que los almacenes y minería de datos son la base de la denominada *Business Intelligence*.

3.Evaluación

3.1.Tipo de pruebas y su valor sobre la nota final y criterios de evaluación para cada prueba

El estudiante deberá demostrar que ha alcanzado los resultados de aprendizaje previstos mediante las siguientes actividades de evaluación

1. Trabajo práctico en el laboratorio (40%): Se realizarán trabajos que podrán ser en grupo, y se realizará un seguimiento del progreso del aprendizaje del alumnado durante el cuatrimestre. Se valorará el funcionamiento según especificaciones, la calidad de su diseño y su presentación, la adecuada aplicación de los métodos de resolución, así como la capacidad para explicar y justificar el trabajo realizado. Los alumnos que hayan cumplido con los plazos de entrega fijados para los trabajos prácticos, y hayan demostrado en ellos un nivel de aprovechamiento y calidad de resultados adecuados, obteniendo en la valoración de su trabajo práctico al menos la nota mínima de 5, estarán exentos de la realización de un examen práctico en el laboratorio.
2. Prueba escrita (40%). En esta prueba se plantearán cuestiones y/o problemas relacionados con el programa

30253 - Almacenes y minería de datos

impartido en la asignatura. Su tipología y complejidad será similar a los presentados en las sesiones de aula y laboratorio. En general, se valorará la calidad y claridad de las respuestas, así como las estrategias de solución planteadas por el alumnado.

3. Realización de un informe sobre temas teórico/prácticos relacionados con los contenidos impartidos en la asignatura (20%). Deberá entregarse un documento escrito con el trabajo realizado y además deberá hacerse una presentación pública. Se valorará la calidad de los contenidos, así como la organización y estructura del documento escrito y la presentación, la expresión oral, las fuentes de referencia utilizadas y las respuestas a las preguntas planteadas al final de la presentación.

La calificación final se obtendrá mediante la media ponderada de los apartados anteriores. Hay que aprobar cada prueba por separado. En caso de no alcanzar ese mínimo en alguna de las partes, la calificación global de la asignatura será la mínima entre 4.0 y el resultado de ponderar con los porcentajes de cada parte.

Los alumnos que no entreguen las prácticas o trabajos de forma regular en las fechas de entrega parciales establecidas (o las hayan entregado de forma deficiente) podrán realizar una única entrega antes del día del examen de la convocatoria en la que quieran entregarlas. Sin embargo, en ese caso (además de defender la entrega), tendrán que superar también una prueba relacionada consistente en la resolución de un enunciado.

4. Metodología, actividades de aprendizaje, programa y recursos

4.1. Presentación metodológica general

El proceso de aprendizaje que se ha diseñado para esta asignatura se basa en lo siguiente:

1. La presentación de los contenidos de la asignatura por parte de los profesores, así como la realización de ejercicios en clase.
2. El estudio personal de la asignatura por parte de los alumnos y la participación en clase en la resolución de los ejercicios planteados.
3. El desarrollo de trabajos prácticos por parte de los alumnos, guiados por los profesores, que desarrollan los conocimientos teóricos.
4. La realización y defensa de trabajos prácticos en grupo, tutorizados por los profesores de la asignatura.
5. La atención personalizada al alumno a través de tutorías con el objetivo de revisar y discutir los materiales y temas presentados en las sesiones presenciales.

Se debe tener en cuenta que, aunque la asignatura tiene una orientación fundamentalmente práctica, es necesario adquirir los conocimientos teóricos previos. Por ello, el proceso de aprendizaje pone énfasis tanto en los conceptos teóricos y en el estudio individualizado como en la realización de los ejercicios prácticos planteados.

4.2. Actividades de aprendizaje

El programa que se ofrece al estudiante para ayudarle a lograr los resultados previstos comprende las siguientes actividades...

- En las clases impartidas en el aula se desarrollará el programa de la asignatura.
- En las clases de problemas se resolverán problemas y ejercicios de aplicación de los conceptos y técnicas presentadas en el programa de la asignatura, realizándose en algunos casos la lectura y discusión de textos de interés.
- Las sesiones de prácticas se desarrollarán en un laboratorio informático. En dichas sesiones el alumno deberá realizar trabajos prácticos relacionados con la asignatura.

4.3. Programa

1. Introducción a los almacenes de datos

- Conceptos básicos:
 - o Análisis de requerimientos de usuario.

30253 - Almacenes y minería de datos

- o Ciclo de vida.
- o El problema de la reconciliación de fuentes de datos.
- o Transacciones OLTP vs. OLAP.
- Arquitectura de almacenes de datos:
 - o Diseño conceptual, lógico y físico.
 - o Procedimientos ETL.
- Sistemas comerciales.

2. Introducción a la minería de datos:

- Descubrimiento de conocimiento y datos.
- Minería Web.
- Herramientas de análisis de datos.
- Áreas de aplicación, por ejemplo:
 - o Toma de decisiones (Banca-finanzas-seguros, marketing, políticas sanitarias/demográficas, etc.).
 - o Procesos Industriales.
 - o Ingeniería Inversa (*Reverse Engineering*).

4.4. Planificación de las actividades de aprendizaje y calendario de fechas clave

El calendario de clases, prácticas y exámenes, así como las fechas de entrega de trabajos de evaluación, se anunciará con suficiente antelación, de acuerdo con las sesiones y fechas establecidas por el centro.

El calendario de clases, prácticas y exámenes, así como las fechas de entrega de trabajos de evaluación, se anunciará con suficiente antelación en la plataforma de apoyo a la enseñanza presencial Moodle 2 del Anillo Digital Docente de la Universidad (<http://add.unizar.es>).

4.5. Bibliografía y recursos recomendados

[BB: Bibliografía básica / BC: Bibliografía complementaria]

Zaragoza:

- [BB] Adamson, Christopher. Star schema : the complete reference / Christopher Adamson . New York : McGraw-Hill, 2010
- [BB] Jensen, Christian S. Multidimensional databases and data warehousing / Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen . [San Rafael (California)] : Morgan & Claypool Publishers, cop. 2010
- [BB] Kimball, Ralph. The data warehouse toolkit : the definitive guide to dimensional modeling / Ralph Kimball, Margy Ross . 3rd ed. Indianapolis : John Wiley & sons, cop. 2013
- [BC] Malinowski, Elzbieta. Advanced data warehouse design : from conventional to spatial and temporal applications / Elzbieta Malinowski, Esteban Zimányi . [1st ed.], 2nd corr. print. Berlin : Springer, cop. 2009
- [BC] Liu, Bing. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data / Bing Liu Springer, 2011.
- [BC] Sumathi, S.. Introduction to Data Mining and its Applications / S. Sumathi, S. N. Sivanandam Springer, 2006.

Listado de URL

- Transparencias, bibliografía comentada, enunciados de problemas, casos de estudio y guiones de prácticas que los profesores de la asignatura pondrán a disposición del alumnado mediante la plataforma Moodle 2 del Anillo Digital Docente. [<http://add.unizar.es>]

30253 - Almacenes y minería de datos

Teruel:

- [BB] Jensen, Christian S.. Multidimensional databases and data warehousing / Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen . [San Rafael (California)] : Morgan & Claypool Publishers, cop. 2010
- [BB] Malibowski, E. Advanced Data Warehouse Design [Recurso electrónico] :From Conventional to Spatial and Temporal Applications / Elzbieta Malinowski, Esteban Zimányi. Berlin, Heidelberg : Springer-Verlag Berlin Heidelberg, 2008
- [BC] Liu, B. Web data mining :exploring hyperlinks, contents, and usage data / Bing Liu. Heidelberg ; New York : Springer, cop. 2011
- [BC] Sumathi, S. Introduction to Data Mining and its Applications [Recurso electrónico] / S. Sumathi, S. N. Sivanandam Berlin, Heidelberg : Springer-Verlag Berlin Heidelberg, 2006
- [BC] The data warehouse lifecycle toolkit [Recurso electrónico] / Ralph Kimball ... [et al.]. Indianapolis, Ind. : Wiley Pub., 2008