



Grado en Ingeniería Informática 30253 - Almacenes y minería de datos

Guía docente para el curso 2013 - 2014

Curso: 4, Semestre: 1, Créditos: 6.0

Información básica

Profesores

- **María Piedad Garrido Picazo** piedad@unizar.es
- **Sergio Ilarri Artigas** sillarri@unizar.es
- **Raquel Trillo Lado** raqueltl@unizar.es
- **Felipe Moliner Callao** fmoliner@unizar.es

Recomendaciones para cursar esta asignatura

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

Actividades y fechas clave de la asignatura

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 (<http://add.unizar.es>) de la Universidad. El alumnado que no disponga de acceso a la asignatura en dicha plataforma debe ponerse en contacto con el profesorado.

Inicio

Resultados de aprendizaje que definen la asignatura

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

- 1:**
Es capaz de describir los aspectos básicos de los almacenes de datos y de explicar los procesos necesarios para su diseño.
- 2:**
Es capaz de diseñar un almacén de datos de forma efectiva.

3: Conoce los aspectos básicos de los almacenes de datos y es capaz de buscar información adicional sobre otros temas relacionados avanzados.

4: Es capaz de comparar diversos diseños de almacenes de datos y de elegir un sistema comercial adecuado para su implantación.

5: Conocer las técnicas básicas de aprendizaje automático aplicado a la minería de datos.

6: Utilizar las técnicas básicas en problemas reales.

Introducción

Breve presentación de la asignatura

Hoy en día, el almacenamiento y la explotación de información digital constituyen un aspecto clave para el éxito de cualquier empresa y/o organización. Por ello, esta asignatura se ha diseñado con el objetivo de adquirir los conocimientos y habilidades necesarias para diseñar e implantar almacenes de datos, así como para aplicar técnicas básicas de minería de datos sobre los mismos.

Contexto y competencias

Sentido, contexto, relevancia y objetivos generales 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.

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.

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

- 1:**
Planificar, presupuestar, organizar, dirigir y controlar tareas, personas y recursos (CT2).
- 2:**
Combinar los conocimientos generalistas y los especializados de Ingeniería para generar propuestas innovadoras y competitivas en la actividad profesional (CT3).
- 3:**
Resolver problemas y tomar decisiones con iniciativa, creatividad y razonamiento crítico (CT4)
- 4:**
Aplicar las tecnologías de la información y las comunicaciones en la Ingeniería (CT11).
- 5:**
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 (CES1).
- 6:**
Participar activamente en la especificación, diseño, implementación y mantenimiento de los sistemas de información y comunicación (CESI3).

Importancia de los resultados de aprendizaje que se obtienen en la asignatura:

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 (por ejemplo, datos históricos, de clientes, etc.) 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*.

Evaluación

Actividades de evaluación

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

- 1:**
 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, serán 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 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 contrario la nota final será el máximo entre las notas que no superen el aprobado.

La calificación obtenida en la primera convocatoria en cada una de las partes (pruebas prácticas de laboratorio, realización o defensa de trabajos / proyectos prácticos en grupo, o prueba escrita) tendrá validez en todas las convocatorias del curso.

Actividades y recursos

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.

Actividades de aprendizaje programadas (Se incluye programa)

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

- 1:**
- En las clases impartidas en el aula se desarrollará el programa de la asignatura.
 - En las clases de problemas se resolverán problemas de aplicación de los conceptos y técnicas presentadas en el programa de la asignatura.
 - Las sesiones de prácticas se desarrollarán en un laboratorio informático. En dichas sesiones el alumno deberá realizar trabajos prácticos relacionados con la asignatura.

Planificación y calendario

Calendario de sesiones presenciales y presentación de trabajos

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.

Programa

Programa de la asignatura

El programa previsto comprende los siguientes bloques:

1. Introducción a los almacenes de datos

- Conceptos básicos
 - Análisis de requerimientos de usuario
 - Ciclo de vida
 - El problema de la reconciliación de fuentes de datos
 - Transacciones OLTP vs. OLAP
- Arquitectura de almacenes de datos
 - Diseño conceptual, lógico y físico
 - 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
 - Toma de decisiones (Banca-finanzas-seguros, marketing, políticas sanitarias/demográficas, etc.)
 - Procesos Industriales
 - Ingeniería Inversa (*Reverse Engineering*)

Bibliografía

Bibliografía de la asignatura

Bibliografía básica:

- "Multidimensional Databases and Data Warehousing", Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen, Morgan & Claypool Publishers, 2010.
- "The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling", Ralph Kimball, Margy Ross, John Wiley & Sons, 2011.
- "Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications", Elzbieta Malinowski, Esteban Zimányi, Springer, 2008.
- 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>).

Bibliografía complementaria:

- "The Data Warehouse Lifecycle Toolkit" (Second Edition), Ralph Kimball, John Wiley & Sons, 2008.
- "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data", Bing Liu, Springer, 2011.
- "Introduction to Data Mining and its Applications", S. Sumathi, S. N. Sivanandam, Studies in Computational Intelligence, volume 29, Springer, 2006.

Referencias bibliográficas de la bibliografía recomendada

Escuela Politécnica Superior

- Jensen, Christian. Multidimensional Databases and Data Warehousing / Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen Morgan & Claypool Publishers, 2010
- Kimball, Ralph. The Data Warehouse Lifecycle Toolkit / Ralph Kimball. - 2nd ed. John Wiley & Sons, 2008.
- Kimball, Ralph. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling / Ralph Kimball, Margy Ross, John Wiley & Sons, 2011.
- liu, Bing. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data / Bing Liu Springer, 2011.
- Malinowski, Elzbieta. "Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications", Elzbieta Malinowski, Esteban Zimányi Springer, 2008.
- Sumathi, S.. Introduction to Data Mining and its Applications / S. Sumathi, S. N. Sivanandam Springer, 2006.

Escuela Universitaria Politécnica

- Jensen, Ch. S. Multidimensional databases and data warehousing / Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen San Rafael, Calif.] : Morgan & Claypool Publishers, 2010
- Kimball, Ralph. The Data warehouse toolkit :the complete guide to dimensional modeling / Ralph Kimball, Margy Ross. New York : Wiley, cop. 2002
- Liu, B. Web data mining :exploring hyperlinks, contents, and usage data / Bing Liu. Heidelberg ; New York : Springer, cop. 2011
- 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
- Sumathi, S. Introduction to Data Mining and its Applications [Recurso electrónico] / S. Sumathi, S. N. Sivanandam Berlin, Heidelberg : Springer-Verlag Berlin Heidelberg, 2006
- The data warehouse lifecycle toolkit [Recurso electrónico] / Ralph Kimball ... [et al.]. Indianapolis, Ind. : Wiley Pub., 2008