

Data warehouse

Àngels Rius Gavídia
Montse Serra Vizern
Alberto Abelló Gamazo
José Samos Jiménez
Josep Vidal Portolés
Josep Curto Díaz

PID_00203539

Material docente de la UOC


Àngels Rius Gavidia

Ingeniera en Informática por la UPC y doctora en Sociedad de la Información y Conocimiento por la UOC. Actualmente es profesora de los Estudios de Informática, Multimedia y Telecomunicación de la UOC y directora académica del máster de Tecnologías .NET. Su actividad docente se centra en el área de las bases de datos y proyectos fin de carrera. Sus intereses de investigación se enmarcan dentro del ámbito de la representación formal de procesos para entornos de aprendizaje, las especificaciones de servicios para la implementación de procesos de e-learning y la automatización de dichas especificaciones.


Montse Serra Vizern

Ingeniera en Informática y máster en Combinatoria y Comunicación Digital por la Universidad Autónoma de Barcelona. Profesora del programa especial Citius de Iniciación Profesional a la Empresa de la Universidad Autónoma de Barcelona. Su ámbito de investigación se enmarca dentro del campo de la enseñanza en CTS (ciencia, tecnología y sociedad). Actualmente, es profesora propia de los Estudios de Informática y Multimedia de la Universitat Oberta de Catalunya.


Alberto Abelló Gamazo

Doctor e ingeniero en Informática por la Universidad Politécnica de Cataluña. Profesor asociado al Departamento de Lenguajes y Sistemas Informáticos de esta universidad.


José Samos Jiménez

Doctor en Informática por la Universidad Politécnica de Cataluña. Profesor titular del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Granada, asignado a la Escuela Técnica Superior de Ingeniería Informática.


Josep Vidal Portolés

Ingeniero técnico en Informática de Gestión y diplomado profesor de EGB. Profesor asociado a la Universidad de Girona, en la especialidad de Sistemas de información. Jefe de proyectos y desarrollo del gestor de datos de una entidad financiera (Caixa de Girona).


Josep Curto Díaz

MBA por la IE Business School, máster en *Business intelligence*, máster en Dirección y gestión en sistemas y tecnologías de la información por la Universitat Oberta de Catalunya y licenciado en Matemáticas por la Universidad Autónoma de Barcelona. Consultor en la Universitat Oberta de Catalunya especializado en inteligencia de negocio (*business intelligence*). Analista sénior en IDC Research Spain.

La revisión de este material docente ha sido coordinada por los profesores: Jordi Conesa Caralt, Àngels Rius Gavidia (2013)

Primera edición: octubre 2013

© Àngels Rius Gavidia, Montse Serra Vizern, Alberto Abelló Gamazo, José Samos Jiménez, Josep Vidal Portolés, Josep Curto Díaz

Todos los derechos reservados

© de esta edición, FUOC, 2013

Av. Tibidabo, 39-43, 08035 Barcelona

Diseño: Manel Andreu

Realización editorial: Oberta UOC Publishing, SL

Depósito legal: B-18.120-2013



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Introducción

Como consecuencia de la evolución de las nuevas tecnologías e Internet, en esta asignatura se presenta un tipo de bases de datos que va más allá de las bases de datos convencionales.

En primera instancia, se define el concepto de almacén de datos a partir de las propiedades que lo caracterizan y que a la vez lo diferencian respecto de las bases de datos convencionales. La característica principal que distingue el almacén de datos de la base de datos tradicional es el objetivo por el cual ha sido concebido: ofrecer apoyo a la toma de decisiones. Para que la información contenida en el almacén pueda ser utilizada como ayuda en la toma de decisiones, es necesario que dicho almacén sea diseñado siguiendo un modelo de datos específico: el modelo de datos multidimensional.

Un sistema de información que tenga como objetivo ofrecer apoyo a la toma de decisiones, sea cual sea la naturaleza de la organización, no debe basarse únicamente en el almacén de datos. Es precisa la integración de un conjunto de componentes que giren alrededor de un eje central: el almacén de datos. Todos estos elementos que hemos citado anteriormente configuran lo que se denomina factoría de la información corporativa (FIC).

Construir una FIC no resulta nada fácil: es costoso tanto en recursos como en su propia concepción, y requiere un cambio cultural dentro de la corporación. Sin embargo, si se consigue llevar a cabo hasta el final y se ha diseñado correctamente, la gran inversión inicial se llegará a traducir en beneficios, que colocarán a la institución en una posición competitiva dentro del mercado actual. Una vez construida la FIC, el almacén de datos puede ofrecer apoyo a nuevas estrategias de negocio (*business intelligence*, ERP, CRM), y también a aplicaciones (OLAP, EIS, minería de datos, *webhousing*) destinadas a la explotación de los datos con un nivel de detalle muy preciso.

Esperamos que el contenido de esta asignatura os permita adentraros en nuevos horizontes dentro del ámbito de las bases de datos que os puedan ser útiles en vuestro futuro profesional.

Objetivos

Los objetivos que se pretende que los estudiantes logren con esta asignatura son los siguientes:

- 1.** Llegar a un nivel de conocimientos alto para diseñar un almacén de datos, elegir la mejor arquitectura que lo soporte y saberlo administrar.
- 2.** Conocer los diferentes tipos de aplicaciones y herramientas para una óptima explotación del almacén de datos (EIS, OLAP).
- 3.** Saber elegir las mejores técnicas para extraer información del almacén de datos, lo que servirá para llegar a conclusiones importantes que permitirán emprender un nuevo plan de acción que lleve a la empresa a ser más competitiva.
- 4.** Saber analizar el entorno de negocio de la empresa: qué módulos funcionales necesitarían, como debería ser la arquitectura del sistema, qué bases de datos se deberían compartir, tipos de interfaz, etc.
- 5.** Saber cómo deben implementarse procesos de negocio de alto nivel en las empresas para que proporcionen un rendimiento competitivo.
- 6.** Ser capaz de proponer soluciones integradas y lograr hacer entender a los clientes que la implantación de un ERP es un proyecto complejo debido a su profundo impacto en los procesos de la empresa (cambio cultural de la empresa, cambio en los procesos de negocio, cambio en la organización y cambio en la disciplina de trabajo).
- 7.** Adquirir los conocimientos necesarios para seleccionar de manera adecuada proveedores y productos teniendo en cuenta la problemática de la empresa.

Contenidos

Módulo didáctico 1

Introducción al almacenamiento de datos

Àngels Rius Gavídia, Montse Serra Vizern y Josep Curto Díaz

1. ¿Qué es un almacén de datos?
2. Evolución histórica
3. Características de un almacén de datos
4. Objetivos de un almacén de datos
5. Comparativa: almacén de datos y bases de datos operacionales

Módulo didáctico 2

La factoría de información corporativa

Alberto Abelló Gamazo, José Samos Jiménez y Josep Curto Díaz

1. Usuarios y fuentes de información de los almacenes de datos
2. Los almacenes de datos departamentales
3. El almacén de datos corporativo
4. El almacén de datos operacional
5. El componente de integración y transformación
6. Los metadatos
7. La factoría de información corporativa

Módulo didáctico 3

Construcción de la factoría de información corporativa

José Samos Jiménez y Josep Curto Díaz

1. Punto de partida: apoyo a la toma de decisiones en las organizaciones
2. Variantes de la FIC
3. Estrategias para la construcción de la FIC
4. Desarrollo del componente de integración y transformación
5. Desarrollo del almacén de datos corporativo
6. Incorporación del almacén de datos operacional

Módulo didáctico 4

Diseño multidimensional

Alberto Abelló Gamazo

1. Las necesidades de los analistas y las herramientas OLAP
2. Componentes del modelo multidimensional
3. Diseño conceptual
4. Diseño lógico
5. Consultas con SQL'99
6. Diseño físico

Módulo didáctico 5

Uso de la factoría de información corporativa

José Curto Díaz y Josep Vidal Portolés

1. Beneficios de la factoría de información corporativa
2. Evaluación de la factoría de información corporativa
3. Usos de la factoría de información corporativa
4. Estrategias de negocio

Bibliografía

Devlin, B. (1997). *Data Warehouse from Architecture Implementation*. EE. UU.: Addison Wesley Longman Inc.

Inmon, W. H. (1999). *Building the Operational Data Store*. EE. UU.: John Wiley & Sons Inc.

Inmon, W. H. (2005). *Building the Data Warehouse* (4.^a ed.). EE. UU.: John Wiley & Sons Inc.

Inmon, W. H.; Hackathorn, R. D. (1994). *Using the Data Warehouse*. Nueva York: Wiley.

Inmon, W. H.; Imhoff, C.; Sousa, R. (1998). *Corporate Information Factory*. EE. UU.: John Wiley & Sons Inc.

Inmon, W. H.; Strauss, D.; Neushloss, G. (2008). *DW 2.0: The Architecture for the next generation of Data Warehousing*. EE. UU.: Morgan Kaufman Series.

Kimball, R. (2009). *Data Warehouse Toolkit Classics: The Data Warehouse Toolkit; The Data Warehouse Lifecycle Toolkit; The Data Warehouse ETL* (2.^a ed.). Nueva York: John Wiley & Sons Inc.

