

# Diseño y construcción de un almacén de datos

Àngels Rius Gavídia  
Montse Serra Vizern  
Alberto Abelló Gamazo  
José Samos Jiménez  
Josep Curto Díaz  
Juan Vidal Gil  
Carles Llorach Rius

PID\_00236068

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 profesora de los Estudios de Informática, Multimedia y Telecomunicación de la UOC. Sus intereses de investigación se centran en la representación formal de procesos y servicios en entornos de aprendizaje y su automatización.


**Montse Serra Vizern**

Ingeniera en Informática y doctora ingeniera en Informática por la UAB. Profesora de los Estudios de Informática, Multimedia y Telecomunicación de la UOC y profesora asociada del departamento de Arquitecturas y Sistemas Operativos de la UAB. Su investigación se centra en cuestiones tales como: ética profesional y responsabilidad social de las TIC; género y equidad dentro de las TIC; metodologías y herramientas de aprendizaje a distancia.


**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. Coordina en la UPC el programa de doctorado Erasmus Mundus IT4BI-DC. Sus intereses de investigación se centran en el área de bases de datos, Business Intelligence, gestión de *Big Data*, flujos de datos y gestión de metadatos.


**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.


**Jospe Curto Díaz**

Licenciado en Matemáticas por la Universidad Autónoma de Barcelona, máster en Business Intelligence y Dirección y Gestión de las Tecnologías de la Información por la Universitat Oberta de Catalunya, y MBA por el Instituto de Empresa Business School. Trabaja en los ámbitos de *Business Intelligence*, *Business Analytics* y *Big Data*. Desde 2014 en Delfos Research, empresa de la que es fundador, compagina esta actividad con colaboraciones docentes en IE Business School, UOC, EOI, U-TAD, IEB y Kschool.


**Juan Vidal Gil**

Licenciado en Físicas por la Universidad Complutense de Madrid. Experiencia en soluciones tecnológicas de *Business Intelligence* y *Data Warehouse*, como jefe de proyectos en importantes compañías y como formador especializado en empresas del sector. Profesor colaborador de la UOC.


**Carles Llorach Rius**

Máster en Gestión de Empresas - MBA por la Universidad Rovira i Virgili e ingeniero en Informática por la Universidad Politécnica de Cataluña. Profesor colaborador en la Universitat Oberta de Catalunya.

La revisión de este material docente ha sido coordinada por la profesora: Àngels Rius Gavidia (2016)

Primera edición: septiembre 2016

© Àngels Rius Gavidia, Montse Serra Vizern, Alberto Abelló Gamazo, José Samos Jiménez, Josep Curto Díaz, Juan Vidal Gil, Carles Llorach

Todos los derechos reservados

© de esta edición, FUOC, 2016

Av. Tibidabo, 39-43, 08035 Barcelona

Diseño: Manel Andreu

Realización editorial: Oberta UOC Publishing, SL

Depósito legal: B-14.785-2016



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

En esta asignatura vamos a estudiar los almacenes de datos, partiendo de su definición, necesidad y función dentro de las organizaciones hasta los aspectos relativos a su diseño y construcción.

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.

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 configuran lo que se denomina factoría de la información corporativa (FIC). Así mismo, dentro de la FIC existen diferentes tipos de almacenes de datos según la función y el ámbito de información y usuarios que abarque.

Construir una FIC no resulta nada fácil: es costoso tanto en recursos como en su propia concepción. Existen diferentes enfoques en su planteamiento y diferentes tipos de proyectos para su realización, cada uno de ellos con una serie de ventajas e inconvenientes que hay que considerar. Por otro lado, cada elemento de la FIC tiene sus propias particularidades de cara a su diseño y construcción.

El diseño y construcción de la FIC debe enmarcarse dentro de las actividades de gobernanza de datos de la organización y debe coordinarse con otras actividades de gestión de información, como la gestión de datos maestros, la gestión de la calidad de los datos y la gestión de los metadatos.

Como se ha dicho, el almacén sirve de apoyo a la toma de decisiones y es, por tanto, un almacén orientado a la consulta. Esta orientación condiciona mucho su diseño y modelización, lo que da lugar al llamado modelo multidimensional. Un correcto diseño conceptual, lógico y físico de un modelo multidimensional supone realizar una serie de pasos que facilitarán posteriormente la consulta desde herramientas de explotación de datos.

Esperamos que el contenido de esta asignatura sea útil para aprender a construir, con rigor y a través de la práctica, un almacén de datos adecuado para una organización dada.

## Objetivos

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

- 1.** Conocer en profundidad las características de un almacén de datos y sus diferencias respecto de una base de datos operacional.
- 2.** Conocer en profundidad los diferentes componentes de la FIC y las diferentes arquitecturas y tipos de proyectos que permiten implementarla conforme a las necesidades de la organización y a los recursos disponibles.
- 3.** Ser capaz de desarrollar el componente de integración y transformación de datos de la FIC, así como los distintos almacenes que la conforman.
- 4.** Comprender cómo se enmarca la FIC dentro de la estrategia de gobierno de datos de la organización y cómo interacciona con el resto de actividades de gestión de datos.
- 5.** Conocer y saber aplicar las técnicas para realizar un correcto diseño conceptual, lógico y físico de un modelo multidimensional.
- 6.** Conocer los fundamentos de los lenguajes de consulta para la explotación del almacén de datos.

## Contenidos

### Módulo didáctico 1

#### **Introducción al *Data Warehouse***

Alberto Abelló Gamazo, Josep Curto Díaz, Àngels Rius Gavídia, Montse Serra Vizern, José Samos Jiménez y Juan Vidal Gil

1. Qué es un *Data Warehouse*. Características
2. Objetivos de un *Data Warehouse*
3. Comparativa entre *Data Warehouse* y bases de datos operacionales
4. La factoría de información corporativa
5. El almacén de datos dentro de un sistema de *Data Warehouse*
6. Tendencias actuales

### Módulo didáctico 2

#### **Construcción de la FIC**

Alberto Abelló Gamazo, Josep Curto Díaz, José Samos Jiménez y Juan Vidal Gil

1. Transformación de datos desde el entorno operacional al decisional
2. Estrategias en la construcción de la FIC
3. Desarrollo del componente de integración y transformación
4. Construcción de almacén de datos: departamental, corporativo y operacional

### Módulo didáctico 3

#### **Gestión de datos en un *Data Warehouse***

Juan Vidal Gil

1. Integración de datos
2. Gestión de datos maestros
3. Calidad del dato
4. Gestión de metadatos

### Módulo didáctico 4

#### **Diseño e implementación multidimensional de un *Data Warehouse***

Alberto Abelló Gamazo y Carles Llorach Rius

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
7. Ejemplo de diseño multidimensional
8. Consultas con MDX
9. Nuevas tendencias

## Bibliografía

**Berson, A.; Dubov, L.** (2010). *Master Data Management and Data Governance*. McGraw-Hill/Osborne Media.

**Davenport, T.; Harris, J.** (2008). *Competing on Analytics*. Boston: Harvard Business School Press.

**Devlin, B.** (1997). *Data Warehouse from Architecture to Implementation*. Reading, Mass.: Addison Wesley Longman, Inc.

**English, L. P.** (1999). *Improving Data Warehouse and Business Information Quality*. Nueva York: John Wiley & Sons, Inc.

**Fan, W.; Geerts, F.** (2012). *Foundations of Data Quality Management*. Morgan & Claypool Publishers.

**Franco, J. M.; EDS-Institut Prométhéus** (1997). *El Data Warehouse - El Data Mining*. Barcelona: Gestión 2000.

**Gill, H. S.; Rao, P. C.** (1996). *Data Warehousing. La integración para la mejor toma de decisiones*. México: Prentice Hall.

**Inmon, W. H.** (1996). *Building the Data Warehouse* (2.<sup>a</sup> ed.). Nueva York: John Wiley & Sons, Inc.

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

**Inmon, W. H.** (2005). *Building the Data Warehouse* (4.<sup>a</sup> ed.). Nueva York: 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*. Nueva York: John Wiley & Sons, Inc.

**Inmon, W. H.; Strauss, D.; Neushloss, G.** (2010). *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. Burlington, Mass.: Morgan Kaufman Series in Data Management Systems.

**Inmon, W. H.; Welch, J. D.; Glassey, K. L.** (1997). *Managing the Data Warehouse*. Nueva York: John Wiley & Sons, Inc.

**Jarque, M.; Lenzerini, M.; Vassiliou, Y.; Vassiliadis, P.** (2000). *Fundamentals of Data Warehouses*. Berlín: Springer Verlag.

**Kelly, S.** (1997). *Data Warehousing in Action*. Nueva York: John Wiley & Sons, Inc.

**Kimball, R.** (2002). *The Data warehouse toolkit: the complete guide to dimensional modeling*. Nueva York: John Wiley & Sons, Inc.

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

**Kimball, R.** (2010). *The Kimball Group Reader; Relentlessly Practical Tools for data warehousing and Business Intelligence*. Indianapolis: John Wiley & Sons, Inc.

**Mendack, S.** (2008). *OLAP without cubes: Data Analysis in non-cube Systems*. Hoboken: VDM Verlag.

**Mattison, R.** (1996). *Data Warehousing: Strategies, Technologies and Techniques*. Computing McGraw-Hill.

**Silverston, L.; Inmon, W. H.; Graziano, K.** (1997). *The Data Model Resource Book*. Nueva York: John Wiley & Sons, Inc.

**Webb, C. y otros** (2006). *MDX Solutions: With Microsoft SQL Server Analysis Services 2005 and Hyperion Essbase*. Hoboken: John Wiley & Sons.

**Wrembel, R.** (2006). *Datawarehouses and OLAP: Concepts, Architectures and Solutions*. Hershey: IGI Globals.

