

Explotación y administración de sistemas de *data* *warehouse*

Alberto Abelló Gamazo
Josep Curto Díaz
Carles Llorach Rius
José Samos Jiménez

PID_00236083

Material docente de la UOC


Alberto Abelló Gamazo

Doctor e ingeniero en Informática por la Universidad Politécnica de Cataluña. Profesor asociado al Departamento de Ingeniería de Servicios y Sistemas de Información 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.


Josep 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 el 2014 en Delfos Research, empresa de la que es fundador, compaginando esta actividad con colaboraciones docentes en la IE Business School, UOC, EOI, U-TAD, IEB y Kschool.


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.


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.

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

Primera edición: septiembre 2016

© Alberto Abelló Gamazo, Josep Curto Díaz, Carles Llorach Rius, José Samos Jiménez

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-14.786-2016



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

Los sistemas de inteligencia de negocio (BI) son los encargados de transformar los datos en información completa, correcta y consistente, de lo que resulta una ventaja competitiva que nos permitirá formular una sólida estrategia corporativa.

Para que la información que obtenemos del sistema responda a los propósitos anteriores, deberemos tratar las fuentes de datos para que los datos provenientes de estas y que alimentan el sistema cumplan las características exigidas para su procesamiento. De ahí, la importancia de los datos y su adecuado tratamiento.

Teniendo en cuenta la complejidad de los proyectos de sistemas *data warehouse*, antes de invertir recursos y esfuerzos, se hace necesaria la evaluación de su viabilidad hasta determinar el retorno de la inversión que ha generado en la organización.

Su éxito se basa en la explotación de los datos que provee el sistema, de tal modo que se conviertan en una ventaja sobre sus competidores y permitan resolver problemas al unir a personas y tecnología.

La visualización de datos es una preocupación clave para las organizaciones y los profesionales de la inteligencia de negocio y analistas, ya que afecta de manera determinante al valor que se obtiene de la información.

La apropiada gestión de los almacenes de datos vendrá acompañada de un conjunto de actividades que se deben llevar a cabo para poner en funcionamiento un sistema de *data warehouse* y para, posteriormente, mantenerlo a lo largo de su vida útil, garantizando su accesibilidad, fiabilidad y adaptación a nuevas necesidades que se generen.

Esperamos que el contenido de esta asignatura sirva para aprender a determinar la viabilidad de un proyecto de creación de un sistema de *data warehouse* y que, una vez creado, os ayude a administrarlo y a explotar adecuadamente sus datos para conseguir máximo provecho de él.

Objetivos

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

1. Tomar conciencia de la importancia de los datos de un *data warehouse* por sí mismos, y de la necesidad de su adecuada gestión.
2. Conocer el marco legal existente y los principios que deben regir nuestra actuación en torno a los datos almacenados en los sistemas de *data warehouse*.
3. Comprender la importancia de la visualización de datos en los procesos de toma de decisiones de una organización, usando aquellas técnicas y herramientas idóneas para la explotación de datos requerida.
4. 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.
5. Conocer los diferentes tipos de aplicaciones y herramientas para una óptima explotación del almacén de datos (*reporting*, cuadros de mando, OLAP, etc.).
6. Conocer e interpretar las principales métricas económicas de evaluación de proyectos de almacenes de datos, así como los riesgos asociados a este tipo de proyectos.
7. Conocer ampliamente las actividades que se deben llevar a cabo para poner en funcionamiento y, posteriormente, mantener un almacén de datos a lo largo de su vida útil, garantizando su accesibilidad, fiabilidad y adaptación a las nuevas necesidades que se generen.
8. Adquirir los conocimientos necesarios para seleccionar de manera adecuada a proveedores y productos, teniendo en cuenta la problemática de la empresa.

Contenidos

Módulo didáctico 1

Los datos en sistemas de *data warehouse*

Alberto Abelló Gamazo, Josep Curto Díaz, Carles Llorach Rius y José Samos Jiménez

1. La importancia de los datos
2. Aspectos legales y éticos de los datos
3. Usuarios del sistema
4. Explotación y administración del sistema

Módulo didáctico 2

Evaluación de la viabilidad de los sistemas de *data warehouse*

Josep Curto Díaz y Carles Llorach Rius

1. Necesidad de evaluación antes de la construcción del sistema
2. Evaluación

Módulo didáctico 3

Explotación de datos

Carles Llorach Rius

1. Beneficios de una adecuada presentación de datos
2. Consideraciones para la presentación de datos (riesgos)
3. Formatos de presentación
4. Herramientas de apoyo a la presentación de datos

Módulo didáctico 4

Administración de los sistemas de *data warehouse*

Carles Llorach Rius

1. Desarrollo del sistema de *data warehouse*
2. Administración del sistema
3. Roles en la administración del sistema

Bibliografía

Davis, K.; Patterson, D. (2012). *Ethic of Big Data, Balancing Risk and Innovation*. O'Reilly Media.

Inmon, W. H.; Linstedt, D. (2014). *Data Architecture: a primer for the Data Scientist*. Burlington: Morgan Kaufman Series.

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

Kimball Group (2009). *The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems*.

Ladley, J. (2011). *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*. The Morgan Kaufmann Series on Business Intelligence.

Naciones Unidas (2011). *Cómo hacer comprensibles los datos*. Ginebra.