

Bibliografía Comentada

La bibliografía está dividida en varios grupos temáticos. La sección *Introducción a Data Warehousing* incluye todos los libros y artículos que tratan globalmente el tema, en especial tutoriales. La sección *Características Generales* incluye los artículos que describen las características teóricas de los sistemas, o aspectos generales de los mismos. La sección *Diseño* abarca la bibliografía relativa a la construcción de sistemas de Data Warehousing, en especial los diversos modelos de datos. La sección *Soluciones Tecnológicas* tiene los artículos que describen productos de Data Warehousing, o discusiones sobre su arquitectura. La sección *Extracción y Limpieza* es específica de los mecanismos y herramientas de extracción, limpieza, carga e integración de datos. La sección de *Front End* tiene las publicaciones relacionadas a las herramientas de consulta de datos, en especial herramientas OLAP. Por último, la sección *Otros* tiene publicaciones no relacionadas directamente con Data Warehousing, pero de interés para este proyecto.

En cada sección las publicaciones se hallan ordenadas alfabéticamente.

Introducción a Data Warehousing

[FORO-97] – **Foro de Data Warehousing**, Montevideo, Noviembre 1997. Organizado por Instituto de Computación, UNESCO, Tecnología Informática.

Tutorial de Data Warehousing y presentaciones de empresas sobre el tema Data Warehousing.

[PB-97] – Peter Brooks, “**March of the Data Marts**”, *DBMS*, Marzo 1997.

Cambios en la tecnología de los Data Marts en los últimos años (especialmente aumento en tamaño); diferencias entre Data Marts y DataWarehouses.

[PR-95] – Prism Solutions, “**Tech Topic: What is a Data Warehouse**”.

Artículo de W.H. Inmon para la empresa Prism Solutions, donde se detallan las principales características de un Data Warehouse.

[TM-96] – Terry Moriarty y Richard P. Greenwood, “**Data’s Quest – From source to query**”, *Database Programming & Design*, Octubre 1996.

Introducción sobre la tecnología, la estructura, el flujo de datos y el uso de un Data Warehousing. Hace especial énfasis en cómo son almacenados internamente los datos (older detail data, current detail data, lightly summarized, highly summarized y meta-data).

[WI-93] – William H. Inmon, “**Building the Data Warehouse**” 2ª Edición, *John Wiley & Sons*, 1993.

Libro inicial (de una serie de 3), donde W. Inmon explica la noción de Data Warehouse, cual es su utilidad y en que difiere de los sistemas tradicionales.

Características Generales

[AS-95] – Alan Simon, “**I want a Data Warehouse. So, What is it again?**”, *Database Programming & Design*, Marzo 1996.

Divide los Sistemas de Data Warehousing en una taxonomía según el objetivo, modelo, cliente, “middleware” y tipo de servidor.

[BC-96] – Brad Carlile, “**Seeking the Balance: Large SMP Warehouses**”, *Database Programming & Design*, Agosto 1996.

Elabora medidas de performance para un sistema de Data Warehouse basado en un ambiente SMP.

[CW-96] – Colin White, “**Data Warehouse: What’s in a Name?**”, *Database Programming & Design* - Marzo 1996.

Se resaltan las características fundamentales del Data Warehouse, por sobre otras características (como ser por ejemplo si se implementa relacional o multidimensionalmente).

[IDWA-96] – Doug Laney (Prism Solutions), “**All roads lead to the Data Warehouse**”, URL: <http://www.idwa.org/spr96/roads.htm>

Compara dos tipos de sistemas: los que corren el negocio y los que los reportean.

[KP-95] – Kamran Parsaye, “**The Sandwich Paradigm**”, *Database Programming & Design*, Abril 1995.

Propone un nuevo paradigma (“sandwich paradigm”), para sustituir el “data dump paradigm” que caracterizó los primeros proyectos de DW. En el “data dump”, hay dos pasos: guardar los datos, y luego investigarlos. En el “sandwich”, hay 3: preinvestigar los datos, guardarlos y luego investigarlos. Con este paso extra se logra una mayor comprensión de los datos. Para esto, no es bueno el modelo cascada, sino que es necesario tener feedback, por lo que propone usar modelo con prototipación.

[NA-95] – Nagraj Alur, “**Missing Links in Data Warehousing**”, *Database Programming & Design*, Setiembre 1995.

Lista los principales problemas de administración en los sistemas de Data Warehousing.

[RK-96c] – Ralph Kimball, “**Dangerous Preconceptions**”, *DBMS*, Agosto 1996.

Nombra 3 presunciones sobre los Data Warehouses y explica porque son erróneas. Las presunciones son: los ODS son un paso anterior al DW y no deben ser consultados por no estar listos para uso; los Data Marts son Data Warehouses rápidos y sucios y se pueden construir sin ninguna arquitectura general; el star join schema no es extensible.

[RK-97] – Ralph Kimball, “**Hackers, Crackers and Spooks**”, *DBMS*, Abril 1997.

Propone que haya un arquitecto de seguridad en el Data Warehouse, y que se elabore un plan de seguridad, puesto que no se suele dar importancia a la seguridad en estos sistemas.

[RK-97b] – Ralph Kimball, “**What does the Central Team Do?**”, *DBMS*, Junio 1997.

Define las responsabilidades del equipo central de Data Warehouse. El equipo central es el que se encarga de: definir y publicar las dimensiones corporativas (de esta forma se asegura la consistencia pasando de Data Mart a Data Mart); proveer aplicaciones de varias áreas (que exceden a un Data Mart); definir un esquema de seguridad (Ver [RK97]).

[SA-95] – Sid Adelman, “**People Oriented-Issues**”, *Database Programming & Design*, Junio 1995.

Sugiere prestar más atención a problemas organizacionales en la creación de un Data Warehouse.

[TM-95] – Terry Moriarty, “**Part of the Whole**”, *Database Programming & Design*, Octubre 1995.

Da las 3 principales fases de el manejo de un negocio (planificar, ejecutar y verificar lo planeado), y adapta a ellas el manejo de un Data Warehouse.

Diseño

[DM-98] – Dario Maio, Matteo Golfarelli, Stefano Rizzi, “**Conceptual Design of Data Warehouses from E/R Schemes**”, URL: <http://www.deis.unibo.it>.

Introduce una metodología para obtener el diseño conceptual de un DW a partir de un esquema entidad relación. Utiliza una nueva notación para representar esquemas multidimensionales y consultas sobre éstos.

[FG-96] – Frank McGuff, “**Data Modeling for Data Warehouses**”, URL: <http://members.aol.com/fmcguff/dwmodel/dwmodel.htm>. Octubre 1996.

Discute como transformar un esquema relacional en una arquitectura estrella para un Data Warehouse.

[IBM-95] – IBM Almaden Research Center, “**Modeling Multidimensional Databases**”, URL: <http://www.almaden.ibm.com/cs/people/ragrowal/pubs.html#olap>, 1995.

Indice de publicaciones de OLAP.

[JB-94] – Joyce Bischoff, “**Achieving Warehouse Success**”, *Database Programming & Design*, Julio 1994.

Principales puntos que hay que tener en cuenta en el diseño de una arquitectura para un Data Warehouse. Enfocado en los productos Information Warehouse y DB2 de IBM.

[KS-95] – Kenan Systems, “**An Introduction to Multidimensional Database Technology**”, URL: <http://www.kenan.com>

Una introducción muy interesante y fácil de comprender a los sistemas multidimensionales, sus utilidades y principales problemas.

[MM-96b] – Mark Madsen, “**Warehouse Design in the Aggregate**”, *Database Programming & Design*, Julio 1996.

¿Cómo elegir los agregados a crear? Si son muchos, se complica la actualización y hay una explosión de tamaño; si son pocos, se disminuye la performance de consulta.

[RK-95] – Ralph Kimball, “**Is ER Modeling Hazardous to DSS?**”, *DBMS*, Octubre 1995.

Compara los modelos ER (entidad-relación) y DM (modelación dimensional – o “star join schema”) para ser usados en un Data Warehouse. Argumenta a favor de DM por mayor velocidad y simplicidad de acceso.

[RK-96] – Ralph Kimball, “**The Datawarehouse Toolkit**”, *John Wiley & Sons*, 1996.

Este libro presenta ejemplos de diseños de Data Warehouses en varias clases de empresas, y de ellos va derivando principios de diseño generales de un Data Warehouse. Además, presenta el modelo Star-Schema en extensión (Kimball es uno de los creadores del mismo). Resulta especialmente interesante el capítulo 14 (“The Back Room”), donde se muestra un Data Warehouse desde la óptica del equipo que lo crea.

[RK-96e] – Ralph Kimball, “**It’s time for Data Compression**”, *DBMS*, Octubre 1996.

Beneficios de la compresión en un Data Warehouse, y donde se puede aplicar.

Soluciones Tecnológicas

[BG-97] – Butler Group, “**Business Case for Data Warehousing**”, URL: <http://www.butlergroup.co.uk>. 1997.

Describe las diferentes tecnologías utilizadas en la implementación de Data Warehouses. Incluye descripciones de productos comerciales (Brio, Holistic, NCR, Oracle, Pilot, SAS Institute, Soft Tooltrack, Software AG).

[CW-94] – Colin White, “**The Key to a Data Warehouse**”, *Database Programming & Design*, Febrero 1995.

Resalta las ventajas de disponer de un “information directory” para el DW. Estos tienen 3 grandes componentes: información técnica, información para el usuario gerencial, un navegador de la información.

[IBM-VW] – Visual Warehouse de IBM, **Papers de IBM**, URL: <http://www.software.ibm.com/data/warehouse/vw>

White paper sobre Data Warehouse y especificaciones y componentes del Visual Warehouse.

[JW-97] – Joseph Williams, “**Sagent Data Mart Solution**”, *DBMS*, Febrero 1997.

Describe el producto de Data Mart de Sagent (Sagent Data Mart Server, Sagent Admin, Sagent Design Studio, Sagent Information Studio).

[MM-96] – Mary Edie Meredith y Aslam Khader, “**Designing Large Warehouses**”, *Database Programming & Design*, Junio 1996.

Presenta estrategias para mejorar la performance a pesar del tamaño de los Data Warehouses. Las dos estrategias principales son: HTP (“horizontal table partition”), VTP (“vertical table partition”).

[NA-96] – Nagraj Alur, “**Data Warehousing & Parallel Technology**”, *Database Programming & Design*, Enero 1996.

Aplicaciones de la tecnología MPP de vendedores de hardware y software para los Data Warehouses gigantes.

[RF-96] – Richard Finkelstein, “**Sybase IQ: Expressly for the Warehouse**”, *Database Programming & Design*, Diciembre 1996.

Principales beneficios del Sybase IQ, un DBMS hecho especialmente para los sistemas de DW. Este sistema cuenta con una arquitectura que facilita las lecturas, penalizando los updates, permite threads en ambiente SMP, e implementa índices especiales para DW (“bitmaps”).

[SA-95b] – Shaku Atre y Peter Storer, “**Data Distribution and Warehousing**”, *DBMS*, Octubre 1995.

Compara Data Warehousing y bases distribuidas (sobre todo data marts). Enfocado en la tecnología de Sybase (reglas de distribución, replicación y acceso a los datos).

[TDWI-97] – **The Data Warehousing Institute Lessons from the experts:**

(<http://www.dw-institute.com/lessons/index.htm>)

John Bernardi (EMC Corp.), “**Data Warehousing's Hidden Cornerstone: Information Storage and Retrieval**”

Problemas de grabación de datos en los DW: “storage extensibility”, escalabilidad.

Harry Zimmer (AT&T Global Info Solutions) – “**Data Warehousing: Are You on a Path to Success or Failure?**”

Enumeración de varios de los problemas asociados a la creación de DW. Definición de distintos niveles de DW.

Tierry Costa (Hewlett Packard) – “**Three Issues in Data Warehousing**”

Problemas de los Data Warehouses: “manageability” (Hay herramientas para crearlos, pero no para manejarlos), facilidad de uso, escalabilidad en la empresa. Presenta el IW (Intelligent Warehouse de Hewlett Packard).

Chuck Ballard (IBM) – “**Strategies to Make Your Data Warehouse a Success**”

Habla de una implementación de un Data Warehouse con MPP (Massive parallel processing) con el IBM SP2 e IBM DB2 Paralelo.

Neal Hill (Cognos) – “**The Warehouse Is Not the ‘Answer’**”

Presenta la arquitectura de Cognos, basada en un enfoque “outside-in”, que se diferencia del tradicional en tener 1º el DW y luego las herramientas front-end.

Extracción y Limpieza

[Carleton] – Carleton MetaCenter, Carleton Passport, Software AG SourcePoint e Intellidex's Warehouse Control Center, **Papers de Carleton Corporation**, URL: <http://www.carleton.com>

Especificaciones y principales características de estos productos. Carleton MetaCenter es un manejador de meta-datos para un Data Warehouse, integrado por herramientas líderes de varias empresas. Passport es la herramienta de extracción; SourcePoint automatiza los procesos; Warehouse Control Center maneja la administración y el acceso, mientras que QDB se encarga de la interfase de usuario.

[JW-97b] – Joseph Williams, “**Tools for traveling Data**”, *DBMS*, Junio 1997.

Discusión sobre herramientas para cargar información al Data Warehouse: Calidad de datos, extracción y limpieza.

[KB-97] – Kathy Bohn, “**Converting Data for Warehouses**”, *DBMS*, Junio 1997.

El proceso más complicado de un sistema de Data Warehousing es el obtener datos limpios. Este artículo se centra en los distintos pasos necesarios para realizar este proceso.

[LG-97] – Larry Greenfield, “**Data Warehousing Information Center**”, URL: <http://pwp.starnetinc.com/larryg/>.

Página en Internet que recopila direcciones sobre investigaciones y trabajos de Data Warehousing, y material propio. Se puede destacar “**An (informal) taxonomy of data warehouse data errors**” (Taxonomía que divide los problemas en los datos en 4 categorías: Incompletos, Incorrectos, Incomprensibles, Inconsistentes. Si bien la división es informal, permite poder atacar a los problemas con una metodología basada en el tipo de problema); **Herramientas de Extracción y Limpieza** (se hace un relevamiento de herramientas existentes y sus características).

[RK-96b] – Ralph Kimball, “**Automating Data Extraction**”, *DBMS*, Julio 1996.

Basado en los 13 pasos del proceso de carga que menciona en el libro [RK-96], compara las 3 principales herramientas del mercado para extracción de datos (Carleton Passport, ETI Extract, Prism Warehouse Manager).

[RK-96d] – Ralph Kimball, “**Dealing with Dirty Data**”, *DBMS*, Setiembre 1996.

Presenta el problema de la limpieza de los datos del Data Warehouse. Da principales características de 3 productos líderes que realizan la limpieza de registros de clientes.

[RR-96] – Romain Le Roux, “**Construire son datawarehouse**”, *Informatiques Magazine*, Abril 1996.

Artículo donde se comparan varios productos de ayuda para la alimentación de un Data Warehouse. Se dan ocho criterios importantes para decidir por un producto.

Front End

[JW-96] – Jay Louise Weldon, “**Choosing tools for Multidimensional Data**”, *Database Programming & Design*, Febrero 1996.

Guía para la elección de productos. Discute sobre arquitectura de las OLAP, implementación usando bases relacionales, herramientas para manejar datos multidimensionales: consulta (Impromptu, BrioQuery, GQL), multidimensionales (PowerPlay, Commander, Pablo), o usando objetos (Illustra).

[NewsGroups] – USENET Newsgroups de **OLAP** (*comp.databases.olap*), **ORACLE** (*comp.databases.oracle.misc*)

[OLAP-97] – OLAP Council, “**OLAP Council White Paper**”, URL: <http://www.olapcouncil.org/research/whtpaply.htm>

Una introducción sobre OLAP: Qué es, cómo se usa, vistas multidimensionales, cálculos complejos y beneficios. Glosario de términos usados en OLAP.

[RK-97c] – Ralph Kimball, “**Features for Query Tools**”, *DBMS*, Febrero 1997.

Argumenta las características que tienen que tener las herramientas de consulta de un Data Warehouse.

Otros

[EF-96] – Ing. Eduardo Fernandez (Ce.Cal.), “**Estudios Estadísticos del Comportamiento del Alumnado en la Carrera 70**”

Se plantea un modelo estocástico que representa la evolución del alumnado de la carrera 70.

[PC-76] – P.P. Chen, “**The Entity-Relationship Model: Toward a Unified View of Data**”, *ACM Transactions on Database Systems*, Vol. 1. Num. 1, Enero 1976.

Definición del modelo Entidad-Relación.

Índice Bibliográfico

[AS-95] – Alan Simon, “I want a Data Warehouse. So, What is it again?”, *Database Programming & Design*, Marzo 1996.

[BC-96] – Brad Carlile, “Seeking the Balance: Large SMP Warehouses”, *Database Programming & Design*, Agosto 1996.

[BG-97] – Butler Group, “Business Case for Data Warehousing”, URL: <http://www.butlergroup.co.uk>. 1997.

[Carleton] – Carleton MetaCenter, Carleton Passport, Software AG SourcePoint e Intellidex's Warehouse Control Center, Papers de Carleton Corporation, URL: <http://www.carleton.com>

[CW-94] – Colin White, “The Key to a Data Warehouse”, *Database Programming & Design*, Febrero 1995.

[CW-96] – Colin White, “Data Warehouse: What's in a Name?”, *Database Programming & Design* - Marzo 1996.

[DM-98] – Dario Maio, Matteo Golfarelli, Stefano Rizzi, “Conceptual Design of Data Warehouses from E/R Schemes”, URL: <http://www.deis.unibo.it>.

[EF-96] – Ing. Eduardo Fernandez (Ce.Cal.), “Estudios Estadísticos del Comportamiento del Alumnado en la Carrera 70”

[FG-96] – Frank McGuff, “Data Modeling for Data Warehouses”, URL: <http://members.aol.com/fmcguff/dwmodel/dwmodel.htm>. Octubre 1996.

[FORO-97] – Foro de Data Warehousing, Montevideo, Noviembre 1997. Organizado por Instituto de Computación, UNESCO, Tecnología Informática.

[IBM-98] – IBM Almaden Research Center, “Modeling Multidimensional Databases”, URL: <http://www.almaden.ibm.com/cs/people/ragrowal/pubs.html#olap>.

[IBM-VW] – Visual Warehouse de IBM, Papers de IBM, URL: <http://www.software.ibm.com/data/warehouse/vw>

[IDWA-96] – Doug Laney (Prism Solutions), “All roads lead to the Data Warehouse”, URL: <http://www.idwa.org/spr96/roads.htm>

[JB-94] – Joyce Bischoff, “Achieving Warehouse Success”, *Database Programming & Design*, Julio 1994.

[JW-96] – Jay Louise Weldon, “Choosing tools for Multidimensional Data”, *Database Programming & Design*, Febrero 1996.

[JW-97] – Joseph Williams, “Sagent Data Mart Solution”, *DBMS*, Febrero 1997.

[JW-97b] – Joseph Williams, “Tools for traveling Data”, *DBMS*, Junio 1997.

[KB-97] – Kathy Bohn, “Converting Data for Warehouses”, *DBMS*, Junio 1997.

[KP-95] – Kamran Parsaye, “The Sandwich Paradigm”, *Database Programming & Design*, Abril 1995.

[KS-95] – Kenan Systems, “An Introduction to Multidimensional Database Technology”, URL: <http://www.kenan.com>

[LG-97] – Larry Greenfield, “Data Warehousing Information Center”, URL: <http://pwp.starnetinc.com/larry/>.

[MM-96] – Mary Edie Meredith y Aslam Khader, “Designing Large Warehouses”, *Database Programming & Design*, Junio 1996.

- [MM-96b] – Mark Madsen, “Warehouse Design in the Aggregate”, Database Programming & Design, Julio 1996.
- [NA-95] – Nagraj Alur, “Missing Links in Data Warehousing”, Database Programming & Design, Setiembre 1995.
- [NA-96] – Nagraj Alur, “Data Warehousing & Parallel Technology”, Database Programming & Design, Enero 1996.
- [NewsGroups] – USENET Newsgroups de OLAP (comp.databases.olap), ORACLE (comp.databases.oracle.misc)
- [OLAP-97] – OLAP Council, “OLAP Council White Paper”, URL: <http://www.olapcouncil.org/research/whtpaply.htm>
- [PB-97] – Peter Brooks, “March of the Data Marts”, DBMS, Marzo 1997.
- [PC-76] – P.P. Chen, “The Entity-Relationship Model: Toward a Unified View of Data”, ACM Transactions on Database Systems, Vol. 1. Num. 1, Enero 1976.
- [PR-95] – Prism Solutions, “Tech Topic: What is a Data Warehouse”.
- [RF-96] – Richard Finkelstein, “Sybase IQ: Expressly for the Warehouse”, Database Programming & Design, Diciembre 1996.
- [RK-95] – Ralph Kimball, “Is ER Modeling Hazardous to DSS?”, DBMS, Octubre 1995.
- [RK-96] – Ralph Kimball, “The Datawarehouse Toolkit”, John Wiley & Sons, 1996.
- [RK-96b] – Ralph Kimball, “Automating Data Extraction”, DBMS, Julio 1996.
- [RK-96c] – Ralph Kimball, “Dangerous Preconceptions”, DBMS, Agosto 1996.
- [RK-96d] – Ralph Kimball, “Dealing with Dirty Data”, DBMS, Setiembre 1996.
- [RK-96e] – Ralph Kimball, “It’s time for Data Compression”, DBMS, Octubre 1996.
- [RK-97] – Ralph Kimball, “Hackers, Crackers and Spooks”, DBMS, Abril 1997.
- [RK-97b] – Ralph Kimball, “What does the Central Team Do?”, DBMS, Junio 1997.
- [TM-96] – Terry Moriarty y Richard P. Greenwood, “Data’s Quest – From source to query”, Database Programming & Design, Octubre 1996.
- [RK-97c] – Ralph Kimball, “Features for Query Tools”, DBMS, Febrero 1997.
- [RR-96] – Romain Le Roux, “Construire son datawarehouse”, Informatiques Magazine, Abril 1996.
- [SA-95] – Sid Adelman, “People Oriented-Issues”, Database Programming & Design, Junio 1995.
- [SA-95b] – Shaku Atre y Peter Storer, “Data Distribution and Warehousing”, DBMS, Octubre 1995.
- [TDWI-97] – The Data Warehousing Institute Lessons from the experts. URL: <http://www.dw-institute.com/lessons/index.htm>.
- [TM-95] – Terry Moriarty, “Part of the Whole”, Database Programming & Design, Octubre 1995.
- [WI-93] – William H. Inmon, “Building the Data Warehouse” 2ª Edición, John Wiley & Sons, 1993.