

# **Análisis y empleo de herramientas OLAP en el sistema de información corporativa de la CFE**

Liliana Argotte, Norma Jácome, Manuel Mejía

Instituto de Investigaciones Eléctricas, Gerencia de Sistemas Informáticos,  
Ap.Postal 1-475, 62001, Cuernavaca, Morelos, MÉXICO,  
Tel / Fax. (01 777) 3 18 26 58  
{largotte, nejacome, mlavalle}@iie.org.mx

**Resumen.** Se presentan, aplican y comparan las herramientas comerciales Oracle Express y Oracle Discoverer. Ambas fueron aplicadas en el contexto del Almacén de Datos que la CFE desarrolla denominado Sistema de Información Corporativa (SICORP), el cual involucra el manejo de grandes volúmenes de información histórica y actual. Dada la importancia y el auge que actualmente tiene el tema de Almacenes de Datos, las experiencias que se describen en el artículo son de gran utilidad para futuros desarrollos.

## **1 Introducción**

### **1.1 Los sistemas de información corporativa**

El estar bien informado a veces resulta una meta difícil de alcanzar, sobre todo cuando las empresas se desarrollan y crecen al ritmo de los tiempos actuales. Es partir de esta necesidad que surgen los Sistemas de Información Corporativa (SIC). Un SIC es el enfoque de la informática actual que busca apoyar al ejecutivo moderno en la toma de decisiones estratégicas de la empresa, proporcionándole información de calidad, resumida y oportuna a partir de los grandes volúmenes de datos actuales e históricos con que ya cuenta la organización; un SIC se integra con varias tecnologías y aplicaciones, como las redes locales y amplias, los manejadores de bases de datos, las herramientas de visualización de datos y otras, cuyo empleo dependerá de cada negocio, por lo que un SIC tendrá características muy particulares para cada organización. Lo que hace importante el concepto es que hoy día su implementación es perfectamente factible debido a los cambios tecnológicos en materia de hardware y software ocurridos en los últimos diez años. Entre otros, estos avances incluyen [1]:

- El surgimiento del Procesamiento Analítico en Línea (On Line Analytical Processing - OLAP) y su aplicación más allá de las áreas tradicionales de marketing y finanzas.
- El crecimiento de la computación Cliente/Servidor, que ha creado servidores de hardware y software más poderosos y sofisticados que nunca. Los servidores de

- hoy compiten con los mainframes de ayer y ofrecen arquitecturas de memoria tecnológicamente superiores, con procesadores de alta velocidad y capacidades de almacenamiento masivas.
- La aparición de modernos Sistemas de Gestión de Base de Datos (Data Base Management Systems – DBMS’s) que proporcionan mayor soporte para las estructuras de datos complejas.
  - El uso generalizado del Internet e Intranet, acompañado de nuevas y más potentes aplicaciones.

## 1.2 El sistema de información corporativa de la CFE

La idea de crear un SIC para la Comisión Federal de Electricidad (CFE) surge en la Subdirección Técnica de la CFE. La Subdirección Técnica, a través de la Gerencia de Informática y Telecomunicaciones (GIT), es responsable de apoyar los objetivos estratégicos de la CFE, mediante la implementación de soluciones corporativas en materia de informática, que provean de información oportuna, integral y confiable a los ejecutivos de la institución, a fin de mejorar los procesos de planeación, operación y gestión de la misma. En virtud de lo anterior, la GIT elaboró un plan maestro de sistemas informáticos que integró acciones a mediano plazo y en el que se identifican proyectos que requieren de un gran esfuerzo y de la participación de todas las áreas de la CFE.

Tal es el caso del Sistema de Información Corporativa (SICORP), con el que se está creando un esquema de acceso a los datos de las diversas Subdirecciones de la CFE, concentrando información histórica y actual que apoye al proceso de toma de decisiones y que contribuya al incremento de la productividad en las áreas usuarias. Así, el SICORP será una herramienta indispensable para la planeación y gestión de la empresa en el nuevo milenio.

## 1.3 Almacenes de datos y herramientas de análisis de la información

Dentro de los distintos elementos o tecnologías que pueden conformar a un SIC destaca de manera importante el del Almacén de Datos (Data Warehouse), porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales ya existentes años atrás [2]. Un Almacén de Datos (AD) es una colección de datos en la cual se encuentra concentrada e integrada la información institucional y se usa como soporte para el proceso de toma de decisiones gerenciales. Esta integración de información no se logra mediante una simple copia de los datos operacionales, sino que se obtiene a través de un modelado especial de los datos y de procesos específicos de extracción, limpieza, transformación e inserción de datos [3].

Estrechamente ligadas al concepto de SIC y AD, se encuentran las modernas herramientas para interrogar, analizar, presentar, visualizar y descubrir conocimiento en los grandes volúmenes de información contenidas en estas tecnologías. En efecto, en tanto que el enfoque de SIC nos proporciona un marco general para la

conceptualización de sistemas de soporte en la toma de decisiones estratégicas y AD es el medio para la concentración de los datos de interés corporativo, las Herramientas para el Análisis de la Información (HAI) son el software que permite a los ejecutivos interrogar, buscar y visualizar, de manera sencilla, aquella información que apoye la toma de decisiones gerenciales.

Las HAI son una pieza fundamental en todo SIC, pues es la parte que el usuario final ve y con la que interactúa: la evaluación y elección de la (o las) HAI puede representar el éxito o el fracaso de un proyecto de SIC, pues el grado de aceptación del sistema depende en gran medida de que el usuario encuentre útiles y entendibles las HAI que se le proporcionen para trabajar. Por ser el componente con la cual ejecutivo accesa la información se le conoce como el front-end del sistema (en forma análoga, el back-end se refiere a los procesos que el usuario final no ve y que son los responsables de extraer, limpiar, transformar e insertar los datos en el AD).

Dentro de las HAI se encuentran distintas categorías de acceso: los reportes, aplicaciones, las de análisis y las herramientas para consultas-reportes a la medida.

En cuanto a los reportes estándar, son vistas oficiales (institucionales) de información, en formatos prediseñados y que son emitidos de manera periódica para distribución a un grupo generalmente extenso de usuarios. Esta es la forma más tradicional de proporcionar información en la empresa, sin embargo el contenido estos reportes es muy distinto a los reportes generados por los sistemas operacionales, pues están dirigidos a la alta gerencia, más que a los mandos medios encargados de la operación cotidiana de la organización.

Las aplicaciones están constituidas por software con interfaz del tipo apropiado para su uso con el ratón de la computadora (aplicaciones de acceso push-button) generan un conjunto relativamente pequeño de reportes clave orientados a una unidad usuaria muy específica de ejecutivos de alto nivel, que realizan consultas en la pantalla de su computador.

Las herramientas para consultas-reportes a la medida permiten a los usuarios crear sus propios reportes usando manejadores de consultas basadas en SQL como QMF, QBE, MS-Query o como las que contienen productos como Excel o Access.

La categoría de acceso denominada de análisis incluye aplicaciones sofisticadas para búsqueda de información y son las de más reciente integración a los SIC. minería de datos es una de estas aplicaciones, con la que se busca la información importante que se encuentra oculta e implícita en los datos, sin realizar interrogaciones específicas: se deja que el algoritmo minero encuentre qué relaciones, tendencias o patrones de comportamiento existen o existirán, y que permitirían mejorar los procedimientos que actualmente sigue la organización. Algunas de las técnicas computacionales usadas para realizar la minería son las redes neuronales artificiales, la inducción de árboles de decisión, las reglas de decisión, y los métodos basados casos similares (nearest neighborhood). También en esta categoría se encuentran facilidades de visualización, los paquetes de análisis estadístico avanzado, las facilidades de navegación por los datos (data-surfing), la inteligencia artificial, las simulaciones y las tecnologías para el procesamiento analítico en línea (OLAP). Esta última es típica en los AD, pues cuenta con la habilidad de examinar la información en varios niveles de resúmenes a lo largo de múltiples dimensiones del negocio

definidas en el AD) mediante una interfaz flexible e intuitiva; permite al usuario explorar distintos enfoques de búsqueda, dejándolo ir de lo muy general a lo muy particular (drilling down) o visceversa, lográndose realizar consultas no predefinidas y complejas, con gran facilidad, pues el usuario actúa como si rebanara y o uniera un cubo de información, capacidad que se conoce como pivoteo del dato o slice and dice.

La tecnología OLAP es el tema central del presente artículo. En las siguientes secciones se presentarán y compararán dos herramientas comerciales que fueron aplicadas en el SICORP, resaltándose las experiencias obtenidas, las cuales son de gran utilidad para desarrollos futuros.

## 2 La herramienta Oracle Express

### 2.1 Descripción

Oracle Express [4] es a la vez una base de datos multidimensional y un ambiente de desarrollo para construir aplicaciones OLAP que faciliten el análisis y la toma de decisiones. Almacena datos como objetos multidimensionales y puede acceder a una o más bases de datos Express para presentar información o bien a una base de datos relacional (RDBMS) cuando es necesario (Figura 1). Se manejan 2 tipos de conexiones a bases de datos relacionales: Direct Oracle Call Interface (OCI) y Open Database Connectivity Interface (ODBC).

El lenguaje utilizado en esta tecnología se denomina Express Language, el cual es un lenguaje de procedimientos almacenados (Stored Procedure Language, SPL), usado por las aplicaciones de desarrollo para almacenar, procesar, recuperar y navegar en los datos de las bases de datos Express.

La tecnología Express es implementada en 2 productos:

- Oracle Express Server 6: desarrollado para plataformas Microsoft Windows NT y Unix; servicio multithreaded; soporta múltiples instancias, usuarios y sesiones simultáneas; provee acceso a clientes remotos y locales.
- Personal Express 6: desarrollado para plataformas Windows 95 o Windows NT; aplicación multithreaded; atiende sólo un usuario; provee acceso sólo a clientes locales; soporta sesiones múltiples simultáneas.

### 2.2 Aplicación en el SICORP

Debido a que la CFE cuenta con licencias corporativas de Oracle, se optó por trabajar con Oracle Express Server 6.2.0 para el desarrollo de la aplicación OLAP, además de que Express Server y las aplicaciones Express, constituyen una inversión segura, que tendrá el soporte, mantenimiento y disponibilidad por muchos años de Oracle, con la seguridad de poder migrar la aplicación hacia los servicios OLAP que Oracle vaya abriendo en el mercado.

En SICORP se utilizó Oracle Express como Multidimensional OLAP (MOLAP), sin embargo es un producto híbrido que permite ambos modos: MOLAP y Relational OLAP (ROLAP) [5].

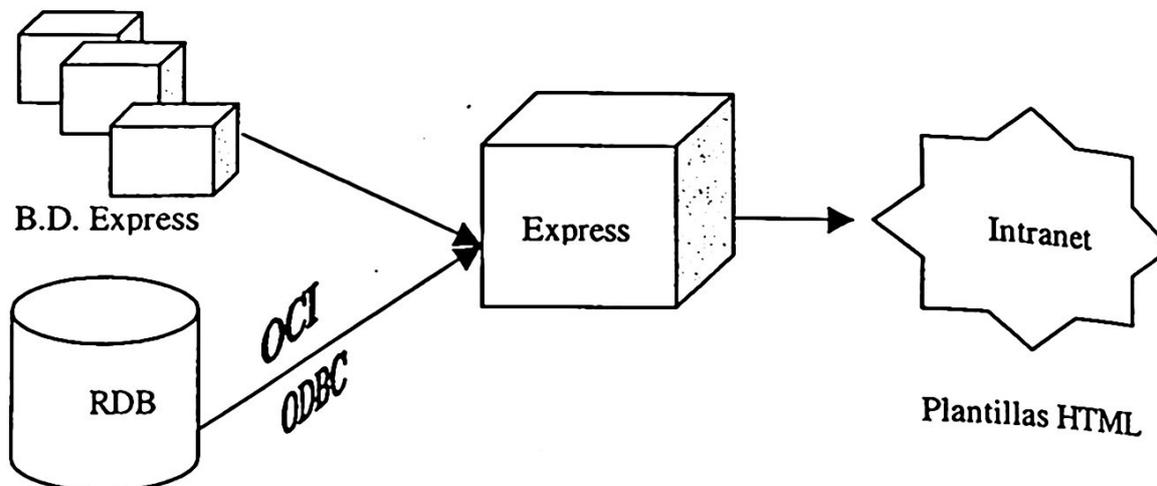


Fig. 1. Flujo de información en Oracle Express

El módulo que se utilizó para crear y cargar las dimensiones y hechos (variables en el lenguaje de Oracle) fue Express Administrator. Una vez creados los cubos de datos en Express, se hicieron plantillas html con código embebido del lenguaje Express para crear vistas en forma de tablas y gráficas que pudieran ser desplegadas en Web. Las facilidades que el usuario tiene en el manejo del cubo son las que Express proporciona, como el drilling down y slice and dice.

Empleando Express, durante el año 2001, se crearon alrededor de un centenar de cubos de áreas de la CFE como: Generación, Distribución, Comercial, Recursos Humanos, Capacitación y Finanzas.

Actualmente, la tecnología de bases de datos de Oracle y Express Server convergen para formar los Servicios OLAP Oracle 9i y llegar a ser la plataforma líder de aplicaciones analíticas en Internet [6].

Express Server 6.3 provee poderosas aplicaciones analíticas con capacidades para predecir, planear y dirigir. Por más de 20 años, Express Server y sus predecesores, Classic Express y Express MDB han sido líderes en la industria de servidores OLAP, por su desempeño y capacidades analíticas. Express Server ha mantenido su liderazgo, porque ha evolucionado con las necesidades analíticas a lo largo del tiempo y con las plataformas de desarrollo (primero en mainframe, después en Cliente/servidor y actualmente en Internet) [9].

### 2.3 Ventajas y desventajas

#### Ventajas:

- El módulo Express Administrator provee una interfaz sencilla para crear y mantener bases de datos de Express.

- Es fácil de usar y entender (se requiere de una a dos semanas de capacitación).
- Permite crear dimensiones a partir de archivos planos o de tablas de una base de datos.
- Permite crear jerarquías en las dimensiones.
- Puede tener múltiples variables dimensionadas.
- Permite aplicar fórmulas tanto en las agregaciones, ó sumarios, como para cargar datos en una variable.
- Realiza automáticamente sumarios (Rollup's).
- Permite aplicar criterios en el formato (color) de las cifras, para llamar la atención del ejecutivo para casos de especial interés.
- Se pueden realizar consultas a la base de datos Express mediante los siguientes módulos:
  - Express Analyser,
  - Express Web Publisher,
  - Express Objects,
  - Escribiendo un programa SPL.

#### Desventajas:

- No fue posible cargar datos mediante un programa OCI en una variable con más de 6 dimensiones, ya que no se podían ver los valores aún después de ejecutar los sumarios (Rollup's), y aún cuando Oracle ofrece que la herramienta está diseñada para soportar más de 30 dimensiones.
- El tiempo de respuesta de Express al presentar los datos de los cubos, tanto en Express Administrator como al momento de crear las vistas con el lenguaje Express es lento, del orden de 3 minutos en una máquina conectada a la red local.
- A la fecha no se ha encontrado la forma para evitar la carga de applets en Web, al momento de desplegar un cubo en la máquina del cliente.
- Permite aplicar sólo una fórmula en los agregados, siendo que existen casos en la CFE que se requiere aplicar distintas fórmulas dependiendo del nivel de agregación..
- Actualmente no se ha logrado configurar un esquema de seguridad (privilegios de acceso a los datos), es decir, la herramienta no prevé esta necesidad.
- Tiene un costo elevado respecto a competidores como SQL Server.

## 3 La herramienta Oracle Discoverer

### 3.1 Descripción

Discoverer es una herramienta de consulta orientada al negocio. Permite al usuario, dentro de los diferentes niveles de la organización, obtener acceso inmediato a la información del negocio para la toma de decisiones. Soporta análisis multidimensional de consulta rápida y poderosas búsquedas de datos. Tiene una capa de metadatos que simplifica y mejora la visualización de los datos de los usuarios.

Está formado por tres componentes: un módulo para administradores, un módulo para usuarios finales y una capa de metadatos. El módulo de administradores sirve para crear y mantener los metadatos que tienen como finalidad facilitar al usuario final el análisis y la construcción de vistas e informes. En el módulo para usuarios finales se puede llevar a cabo el análisis y desglose multidimensional para tener una información más detallada sin conocer la estructura de la base de datos fuente. Se puede tener un mapa visual de la base de datos y llevar a cabo consultas sofisticadas sin programación. Los datos se desglosan a través de diferentes niveles de resumen para descubrir información detallada del negocio. Los resultados de las consultas son desplegados en una variedad de vistas (tablas y gráficos) fáciles de comprender y sólo con la información necesaria para cada nivel directivo. Y por último la capa de metadatos sirve para ocultar la complejidad de la base de datos de tal manera que el usuario interactúa con sus datos pero en términos de su negocio, en su propio idioma. Dentro del módulo de administradores se crea esta capa.

Aunque algunos autores [5] consideran a Discoverer como una herramienta del tipo Desktop OLAP (DOLAP), nuestra experiencia en la explotación del SICORP nos hace incluirla dentro de las herramientas de tipo ROLAP de acuerdo a la definición que hace de ella Oracle [7,8].

### 3.2 Aplicación en el SICORP

Discoverer versión 3.1 Cliente/Servidor [7,8] fue la herramienta utilizada como ROLAP para la explotación del SICORP de la CFE. Se partió de una base de datos relacional diseñada en Oracle 8i que contiene diferentes estrellas (modelado multidimensional) que modelan cada una de las áreas que conforman a la CFE. Las áreas que se trabajaron con Discoverer fueron: Generación Hidroeléctrica, Gerencia Comercial, Recursos Humanos y Administración de la Distribución. Para cada una de ellas se implementaron una o más estrellas.

Cada área de la CFE fue creada en Discoverer como un área de negocio diferente a la que sólo los usuarios involucrados pueden acceder. Utilizando el módulo de administradores de Discoverer se creó para cada una de ellas toda la capa de metadatos, esto es, se definieron todos los elementos necesarios para ocultar la complejidad de las estructuras de la base de datos al usuario, y éste pudiera hacer todos los análisis y consultas de información en su propio lenguaje y sin hacer programación, teniendo diferentes niveles de detalle de la misma. También con este módulo se crearon todos los sumarios necesarios.

Para cada área se definieron, con el módulo de usuario final, una serie de vistas predefinidas que le permiten al usuario revisar y analizar toda la información extraída y resumida de la base de datos. Estas pueden ser modificadas fácilmente por él mismo. Discoverer cuenta con una interfaz muy intuitiva que le permite al usuario crear otras vistas, además de las predefinidas, a la medida de sus necesidades. Esto se logra con un buen diseño de la capa de metadatos a través del módulo de administradores.

### 3.3 Ventajas y Desventajas

A continuación se describen las ventajas y desventajas que se tienen al utilizar Discoverer como una herramienta ROLAP.

#### Ventajas:

- Provee una perspectiva de negocios mediante la agrupación de información en áreas de negocios lógicas.
- Las complejas estructuras de la base de datos son vistas como una lista de objetos familiares para el usuario.
- Presenta la información a través de distintos formatos (tablas y gráficos), fáciles de comprender y sólo con la información necesaria para cada nivel directivo.
- El usuario puede acceder a vistas predefinidas, de la información de las diferentes áreas del negocio, y modificarlas.
- Permite la creación y parametrización de nuevas vistas de la información.
- El usuario puede manipular los resultados de las consultas de diferentes maneras: cambiando los ejes (dimensiones), moviéndose entre jerarquías, manejando excepciones, condiciones y cálculos.
- En una misma vista se pueden tener distintas variables (data point's) dimensionadas.
- El usuario puede hacer formateo de datos, exportación de datos, y creación de gráficos.
- Se pueden tener datos sumariados.
- Proporciona al usuario final acceso de sólo lectura para proteger la importante información de la base de datos fuente.

#### Desventajas:

- Su forma de presentar los datos depende completamente de la estructura que éstos tengan en la base de datos relacional de la cual parten.
- No construye de manera natural jerarquías, si éstas no existen modeladas en la base de datos como copos de nieve (tablas jerarquizadas).
- No se pueden hacer análisis complejos de los datos. Soporta preguntas tácticas más no estratégicas.
- Si no se usan datos sumariados (previamente), el tiempo de recuperación de los datos dentro de las vistas del usuario es lento, del orden de segundos a un minuto.
- El control de acceso (seguridad) lo proporciona a través de las áreas de negocio, no a nivel de celdas.
- En las consultas que contienen un promedio de 10 dimensiones el tiempo de recuperación de los datos es lento, del orden de segundos a un minuto.
- Se tiene que adquirir una licencia por cada usuario que lo utilice, pues es una herramienta Cliente/Servidor.

## 4 Cuadro Comparativo

Tabla 1. Cuadro Comparativo

Característica	Oracle Express	Oracle Discoverer
Uso de la herramienta en Sicorp	MOLAP	ROLAP
Necesidad de Bases de Datos intermedias para dar velocidad a las consultas	Se crean cubos MOLAP	Se crean Bases de Datos de Agregados.
Accede información de otras Bases de Datos	B.D. Express / RDBMS	RDBMS
Plataforma requerida	Win NT / Unix	Win 95 / NT
Responde a cuestionamientos	Estratégicos	Tácticos
Vistas	Tablas y gráficos en Web.	Tablas y gráficos.
Creación de vistas	Predefinidas, no las puede alterar el usuario.	Predefinidas, el usuario puede realizar cambios y puede crear nuevas vistas.
Navegación en las vistas	Drilling down, Slice and dice	Drilling down, Slice and dice
Presentación de los datos	Depende de la estructura de la base de datos Multidimensional.	Depende de la estructura de la base de datos relacional.
Licencias	Únicamente para el administrador de la B.D.	Se requiere licencia por cada usuario.
Máximo número real de dimensiones	Hasta 6	Se probaron hasta 6
Esquema de seguridad	No lo prevé	Sólo por vista (no por celdas)
Tiempo para presentar las vistas al usuario	Hasta 3 minutos	Hasta 1 minuto

## 5 Conclusiones

El vertiginoso crecimiento de las empresas modernas ha exigido a la informática crear nuevas herramientas para el apoyo a la toma de decisiones, surgiendo así, entre otras, una gran variedad de herramientas OLAP, desarrolladas por distintas empresas de software.

Una de las empresas líderes en el ramo es Oracle. En este artículo se analizaron dos herramientas OLAP de Oracle, detectándose que si bien son herramientas poderosas, aún no cumplen en su totalidad con las necesidades actuales de una organización de gran tamaño, como lo es la CFE.

La apreciación del usuario sobre Oracle Express es que realmente modela una base de datos del tipo multidimensional con todas las ventajas de análisis en línea que esto tiene, a diferencia de otras herramientas (como Brío, de Brío Intelligence) que aunque están orientadas al mismo tipo de uso, en realidad solo crean tablas

pivote sin la funcionalidad completa de un verdadero OLAP. Lo anterior significa que el usuario aprecia tener Oracle Express pues puede realizar muy diversos análisis, a nivel macro y micro, de manera sencilla, con intranet en línea, lo cual asegura la oportunidad de la obtención de respuestas a los cuestionamientos estratégicos.

## 6 Agradecimientos

Los autores desean agradecer la importante labor de apoyo, organización y promoción del proyecto realizada por los Ings. Jorge Hernández Aguilar, Gerente de la GIT, Ricardo Hernández Rodríguez, Subgerente de Informática de la GIT y Oscar Miranda. También se reconoce la participación en las distintas fases del SICORP del personal del IIE: Dr. Gustavo Arroyo, Gerente de la GSI, Dr. Guillermo Rodríguez, Jorge González, Ricardo Molina, Israel Galván, Rosa Mondragón, Alfredo Martínez, Víctor García, Hilda Escobedo, Felipe López, Javier Martínez, Octavio Reynaga, Humberto Hernández, Norma García, Gustavo Bravo, Raúl García, Luz Ocampo, Yazmín Hernández, Miguel Pérez, Carlos Herrera, Benjamín Zayas y Reynaldo Alanís.

## Referencias

- [1] Gill, H. S., Data warehousing, Prentice Hall Hispanoamericana, S.A., 1996.
- [2] Weldon, J. L., "Data warehouse building blocks", Byte, enero de 1997, pp. 82-102.
- [3] Kimball, R., The data warehouse toolkit, John Wiley & Sons, Inc., 1996.
- [4] Personal Express, Installation and configuration Guide, Release 6.2, 1998.
- [5] (<http://www.olapreport.com/Segments.htm>).
- [6] (<http://www.rockportsoft.com/oraclesod.html>)
- [7] Oracle Discoverer for Administrators, Release 3.1, 1998.
- [8] Oracle Discoverer for End Users, Release 3.1, 1998.
- [9] [http://www.oracle.com/ip/analyze/warehouse/collateral/olap\\_sod.html](http://www.oracle.com/ip/analyze/warehouse/collateral/olap_sod.html).

## Sitios Web de interés:

- Olfalito, "Construcción de un data warehouse",  
192.41.50.238/rincon/ficheros/dataware.zip
- The data warehousing information center, [www.dwinfocenter.org/](http://www.dwinfocenter.org/)
- Welch, J.D., An introduction to data warehousing,  
[www.datawing.com/intro2dw/intro.htm](http://www.datawing.com/intro2dw/intro.htm)