



**INSTITUTO UNIVERSITARIO
DE LA EMPRESA**

**DATA WAREHOUSE
ANTECEDENTES, SITUACIÓN ACTUAL Y TENDENCIAS**

GIL SOTO, ESPERANZA

SERIE ESTUDIOS 2001/44

SANTA CRUZ DE TENERIFE, SEPTIEMBRE 2001



**UNIVERSIDAD DE
LA LAGUNA**

RESUMEN

Los actuales sistemas de información presentan múltiples problemas debido a la falta de integración de los datos. Los usuarios de estos sistemas de información se enfrentan a problemas relacionados con el elevado tiempo que tienen que dedicar a la obtención de información, en detrimento del que le deberían dedicar al análisis de la misma. En este contexto surge la necesidad de modificar los sistemas actuales de recogida y tratamiento de la información. Se trata en definitiva de proporcionar a los directivos de las empresas unas herramientas adecuadas para la explotación y análisis de los datos que les permitan obtener el conocimiento necesario en el proceso de toma de decisiones estratégicas. Un data warehouse proporciona una base de datos global y única en la empresa, aportando una gran flexibilidad en la gestión de los datos, de manera que sea el usuario quien seleccione la información que necesita y la forma de explotarla. En este trabajo pretendemos hacer una revisión del concepto data warehouse y su evolución.

DATA WAREHOUSE
ANTECEDENTES, SITUACIÓN ACTUAL Y TENDENCIAS

GIL SOTO, ESPERANZA

SERIE ESTUDIOS 2001/44

SANTA CRUZ DE TENERIFE, SEPTIEMBRE 2001

Esperanza Gil Soto. Departamento de Economía y Dirección de Empresas. Facultad de Ciencias Económicas y Empresariales. Campus de Guajara. Universidad de La Laguna. Camino de la Hornera, s/n. 38071 La Laguna Tenerife Islas Canarias (España).

Teléfono: 922 31 70 65

Fax: 922 31 70 77

Correo electrónico: egilsoto@ull.es

DATA WAREHOUSE

ANTECEDENTES, SITUACIÓN ACTUAL Y TENDENCIAS

1. INTRODUCCIÓN
2. ANTECEDENTES Y SITUACIÓN ACTUAL
3. IMPLICACIONES DE UN DATA WAREHOUSE
4. DATA WAREHOUSE E INTERNET
5. CONCLUSIONES

1. INTRODUCCIÓN

Las bases de datos existentes actualmente en la mayoría de las empresas mantienen la información necesaria para la actividad diaria de la organización, ya que dichas bases de datos alimentan a los sistemas de información corporativos. Podemos afirmar, por lo tanto, que las bases de datos representan una herramienta imprescindible, aunque no suficiente, para el correcto funcionamiento de los sistemas de información de cualquier organización.

Es importante por lo tanto, desde un punto de vista de la organización empresarial, que la estrategia en el área de sistemas y tecnologías de la información vaya encaminada hacia una correcta política de mantenimiento, actualización y gestión de las bases de datos, por un lado, y por el otro, que se lleve a cabo una adecuada política de I+D+I, así como una progresiva adaptación a las nuevas tendencias tecnológicas.

En este sentido, y puesto que los cambios que se producen actualmente en las tecnologías y sistemas de información son demasiado rápidos, en este trabajo hacemos una revisión de una herramienta llamada data warehouse. Esta tecnología de la información representa el último avance dentro de las bases de datos, y se configura como el entorno idóneo para la consulta y el análisis de la información procedente tanto de los sistemas transaccionales internos, como de las fuentes de información externas de interés para la empresa.

La finalidad última del data warehouse consiste en convertir los datos contenidos en las bases de datos corporativas de las organizaciones, en información y ésta a su vez, en conocimiento útil en el proceso de toma de decisiones estratégicas. El data warehouse es una herramienta que va a permitir a los directivos de las organizaciones formular preguntas, realizar consultas y analizar los datos en el momento, forma y cantidad que precisen, sin necesidad de tener que acudir al personal informático de la empresa.

Como ejemplo de la capacidad de información de un data warehouse, citamos el comentario que hizo el Director Ejecutivo de Operaciones, de una cadena de supermercados en EEUU, dirigiéndose a la asamblea general de accionistas, (<http://www.map.es/csi/silice/DW21.html>):

“Hace diez años les pude decir cuantos Doritos vendimos al Oeste del Mississipi. Hoy no sólo les puedo decir eso mismo, sino cuantos vendimos en California, en el Condado de Orange, en la ciudad de Irvine, en el supermercado local Von`s, en una promoción especial, al final del pasillo 4, los jueves”.

Además de la facilidad de uso y la gran capacidad de información, el data warehouse se caracteriza por estar basado en una tecnología diseñada expresamente para el almacenamiento, consulta y análisis de datos, y por tener una estructura de la información orientada plenamente al negocio y al usuario final. Desde un punto de vista estratégico, una de las ventajas que aporta esta herramienta a la empresa, es la de permitir a los diferentes usuarios añadir valor a los datos contenidos en el data warehouse, con la finalidad de adquirir conocimiento y lograr una ventaja competitiva frente a sus competidores.

2. ANTECEDENTES Y SITUACIÓN ACTUAL

2.1. Historia

Devlin (1997) distingue cuatro etapas en la historia del data warehouse (DW):

- Prehistoria, a principios de los ochenta.

Durante estos años las tecnologías de la información se esforzaban en automatizar los procesos de tipo repetitivo o administrativo, haciendo uso de los sistemas de información operacio-

nales o transaccionales en línea (OLTP). La función de estos sistemas consiste en recoger y almacenar los datos generados a partir de las aplicaciones informáticas que se utilizan en los distintos departamentos de la organización. Por lo tanto, reúnen datos de la actividad diaria, tales como facturación de clientes y proveedores, nóminas, contabilidad, etc., con el objetivo de registrar información y permitir la impresión de facturas, documentos y algunos informes predefinidos que satisfagan las necesidades operacionales de la compañía.

Técnicamente los sistemas de información aludidos se caracterizan por permitir una captura rápida y segura de grandes volúmenes de información detallada, en los que la actualización de los datos, así como el tiempo de respuesta, constituyen conceptos importantes. El diseño e implantación de estos sistemas ha supuesto, para la mayoría de las empresas, un gran esfuerzo a lo largo de mucho tiempo, así como elevadas inversiones (Espinoza, 2000).

Desde la visión de la dirección estratégica, los sistemas de información de tipo OLTP están diseñados para resolver los problemas de carácter operativo. Esto significa que el tipo de decisiones que soportan están relacionadas con la ejecución y apoyo de las tareas básicas del negocio. No son útiles, por tanto para realizar análisis más avanzados de tipo estratégico (Borrajó, 1999).

- Edad media, de mediados a finales de los ochenta.

Con el tiempo, los requerimientos de información en las organizaciones empezaron a cambiar, ya que los responsables de los distintos departamentos realizaban cada vez más consultas y precisaban de más informes. Los departamentos de informática no eran capaces de satisfacer la creciente demanda de información, solicitada siempre con carácter urgente. Surge entonces la necesidad de crear unas nuevas bases de datos centrales, con información resumida y detallada.

Los problemas técnicos que surgieron inicialmente fueron de diversa índole. En primer lugar, para cargar la nueva base de datos había que centralizar la información que se encontraba dispersa, duplicada, así como en distintas versiones y formatos, ya que cada usuario mantenía sus propias aplicaciones. En segundo lugar, cómo cargar esa información en la nueva base de datos para llegar a obtener una base de datos central y fiable. Y en tercer lugar, cómo acceder a ella de forma ágil y eficiente.

A mediados de 1985 aparece el primer DW, su finalidad era la de integrar y organizar los datos que estaban dispersos en múltiples bases de datos, dentro de la compañía. Este primer almacén de datos central, lo construyó Inmon (1985), para un banco en Colorado.

- La revolución de los datos a principios de los noventa.

Las organizaciones empiezan a considerar la importancia de los datos, los cuales representan la materia prima de la información, y ésta a su vez es considerada como un activo intangible en la empresa. Es precisamente esta circunstancia la que convierte a la información, junto con otros activos intangibles, en la única generadora real de ventajas competitivas en la empresa, (Cornellá, 1994).

- La actual era de la información, desde finales de los noventa.

En los últimos años, el concepto de DW ha ido perfeccionándose (capacidad de almacenamiento, nuevas herramientas de consulta de datos, conexión a internet o a una intranet, etc.), y adaptándose a las necesidades crecientes de información en las empresas, de forma que los actuales macro almacenes de datos pueden proporcionar soluciones a todo tipo de usuarios.

2.2. Definición

La definición más generalizada y aceptada de un DW, es la que propuso Inmon (1996):

“Un DW, es una colección de datos orientados al tema, integrados, no volátiles e históricos, cuyo objetivo es el de servir de apoyo en el proceso de toma de decisiones gerenciales”.

Un DW representa el motor de los sistemas de información de la década de los noventa, ya que permite integrar datos, organizarlos y almacenarlos para su posterior tratamiento analítico, y siempre bajo una amplia perspectiva temporal. Se trata pues, de una tecnología que transforma los datos operacionales de una organización en una herramienta competitiva, ya que los usuarios finales pueden examinar los datos de un modo más estratégico, así como lograr alcanzar el conocimiento oculto en los enormes volúmenes de datos almacenados en el DW. Es tal la capacidad de almacenamiento en un DW, que hablar de gigabytes¹, o de terabytes, resulta de lo más habitual en los DW actuales. No obstante, existen ya algunas grandes compañías comerciales que empiezan a trabajar con volúmenes de datos del orden de los petabytes, (Diché, 2001).

En la tabla número 1 aparece un resumen con las principales diferencias entre los datos que una organización tiene almacenados en un DW y los datos almacenados en bases de datos tradicionales.

Tabla nº 1: Diferencias entre un DW y una base de datos tradicional

BASE DE DATO TRADICIONAL	DATA WAREHOUSE
Predomina la actualización	Predomina la consulta
La actividad más importante es de tipo operativo (día a día)	La actividad más importante es el análisis y la decisión estratégica
Predomina el proceso puntual	Predomina el proceso masivo
Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Importancia del dato actual	Importancia del dato histórico
Importancia del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva
Estructura relacional	Visión multidimensional
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos
Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa de todo el negocio

Fuente: www.map.es/csi/silice/DW21.html

Las bases de datos relacionales no representan una adecuada solución como herramientas para el soporte directo de sistemas de información para la toma de decisiones. El análisis de la información en un sistema DW parte de la visión de la información como dimensiones del negocio (zonas geográficas, productos, tiempo). A su vez, cada dimensión se organiza en diferentes jerarquías. Un ejemplo de cómo se organiza la información en un DW aparece en la tabla número dos.

Tabla n° 2: Esquema de cómo se organizaría la información en el DW central de una cadena de hoteles

ZONA GEOGRÁFICA	Norte	Puerto de la Cruz Bajamar	Litoral Centro
	Sur	Cristianos Américas	
PRODUCTO	Hoteles	5 estrellas 4 estrellas 3 estrellas	
	Apartamentos	5 llaves 4 llaves 3 llaves	
AÑO	Año 2001	Trimestre 1 Trimestre 2 Trimestre 3 Trimestre 4	Enero 2001 Febrero 2001 Marzo 2001 Etc.
	Año 2000		
	Etc.		

Fuente: Elaboración propia

Esta forma de organizar la información en un DW permite al usuario, realizar consultas del tipo: ¿cuántas habitaciones individuales se ocuparon en un determinado hotel, durante cualquier mes de un año en particular?, o al cliente Sr. Pérez, ¿qué tipo de habitación le gusta?, ¿cómo prefiere el desayuno?, ¿qué servicios utiliza en el hotel?, etc. Y todo ello en un tiempo de respuesta corto.

2.3. Características de un data warehouse

Inmon (1996) definió las siguientes características:

Temáticas: los datos se organizan por temas, en contraposición con los tradicionales sistemas que se organizaban generalmente por procesos funcionales, nóminas, recursos humanos, etc. Es decir, para una universidad, los temas o sujetos pueden ser estudiantes, clases y profesores. Para una empresa, los clientes, proveedores, empleados, etc.

Esta forma de agrupar los datos, permite a los responsables de las distintas áreas, disponer de los datos que consideren más oportunos para definir sus estrategias.

Integrados: se trata de una base de datos única, de uso global para todos los departamentos de la empresa. Esta visión única y transversal se encuentra en contraposición con los sistemas transaccionales, que satisfacen necesidades de información en sentido vertical.

No volátiles: es una característica derivada de la anterior, de forma que una vez cargado un dato en la DW, debe mantenerse invariable. En un sistema operacional, los datos se actualizan regularmente, mediante procesos que consisten en insertar, borrar o modificar registros. En el DW, sólo se realizan dos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No existe el proceso de actualización de datos, sólo de carga.

Históricas: los datos no se actualizan, sino que se almacena el historial de los datos, es decir, el conjunto de valores que el dato ha tenido a lo largo de su historia, asociando a cada dato una referencia de tiempo, con el fin de poder identificar los distintos valores que dicho dato ha ido tomando a lo largo de su ciclo de vida. En los sistemas transaccionales, los datos representan la situación actual a la que se refieren, por ejemplo la dirección actual de un cliente, su estado civil, número de hijos, profesión, etc. En un DW, los datos guardan información de un período que oscila entre los 5 y 10 años, (en contraposición con el horizonte temporal de las bases de datos operacionales, que oscila entre 60 y 90 días). Durante ese período, los clientes pueden cambiar de dirección, casarse, tener hijos, cambiar de trabajo, etc. Si se desea, la empresa puede analizar la evolución, cambios y tendencias de sus clientes.

2.4. Etapas en la elaboración del data warehouse

Martín y Ureña (2000), definen tres etapas en el proceso de elaboración de un DW.

Primera: *carga de datos* desde fuentes internas de la organización o bien de fuentes externas, estos últimos datos permiten validar o completar los datos existentes en el interior de la empresa. La carga de datos puede ser manual, a través de programas diseñados para este fin o por medio de herramientas específicas. En esta etapa se llevan a cabo complejos procesos de transformación, limpieza, depuración, filtrado e integración de datos.

Es importante tener en cuenta que los datos del DW no son los mismos que los de las bases de datos operacionales. En el proceso de carga de datos al DW, éstos sufren una profunda transformación, y el resultado es un DW que contiene un resumen de la información que no se encuentra en el ambiente operacional, con un grado de redundancia en los datos entre los dos depósitos de datos (el operacional y el DW), inferior al 1%, (Pascot, 1998).

Segunda: *almacenamiento de datos y creación de los metadatos*². Éstos últimos forman parte del DW, ya que contienen la semántica de los datos, es decir, su significado actual y sus distintas definiciones a lo largo del ciclo de vida del DW. También son necesarios los metadatos porque explican las interrelaciones entre los datos, necesarias para conocer las reglas del negocio.

Tercera: *explotación de los datos* por parte del usuario, con el fin de obtener información útil y necesaria para la toma de decisiones. La explotación de los datos puede realizarse con distintas herramientas, que van desde no muy complejas hojas de cálculo, herramientas de consulta, de visualización, etc., hasta la utilización de procesos analíticos en línea, utilizando herramientas de tipo OLAP (On Line Analytical Processing). Las herramientas OLAP más potentes incluyen sistemas llamados de minería de datos o Data Mining, los cuales representan procesos que se basan en la aplicación de técnicas (semi-) automáticas para la obtención de información no representada explícitamente en los datos y de mayor utilidad estratégica que éstos, (Martín y Ureña, 2000).

Para Lee y Siau (2001), el éxito de un sistema basado en un DW depende de las herramientas de análisis que se utilicen. Existen multitud de aplicaciones válidas, en función del tipo de información que se requiera, de las características del usuario e incluso del proceso decisional

en el que se vea inmerso. Esto significa que la empresa no debería elegir una única herramienta de análisis para todos los usuarios, sino que debería de analizar las distintas alternativas para cada caso y finalmente escoger la que mejor se adapte a sus necesidades en términos de eficacia.

Otro aspecto a considerar en esta etapa de explotación de los datos, es el tipo de consulta que se desea realizar al DW. Diché (2001), describe cuatro categorías de análisis de datos, que abarcan desde las simples consultas de datos, hasta técnicas de análisis muy avanzadas. Las cuatro categorías de análisis de datos representan una forma gradual de explotación del DW, y son las siguientes:

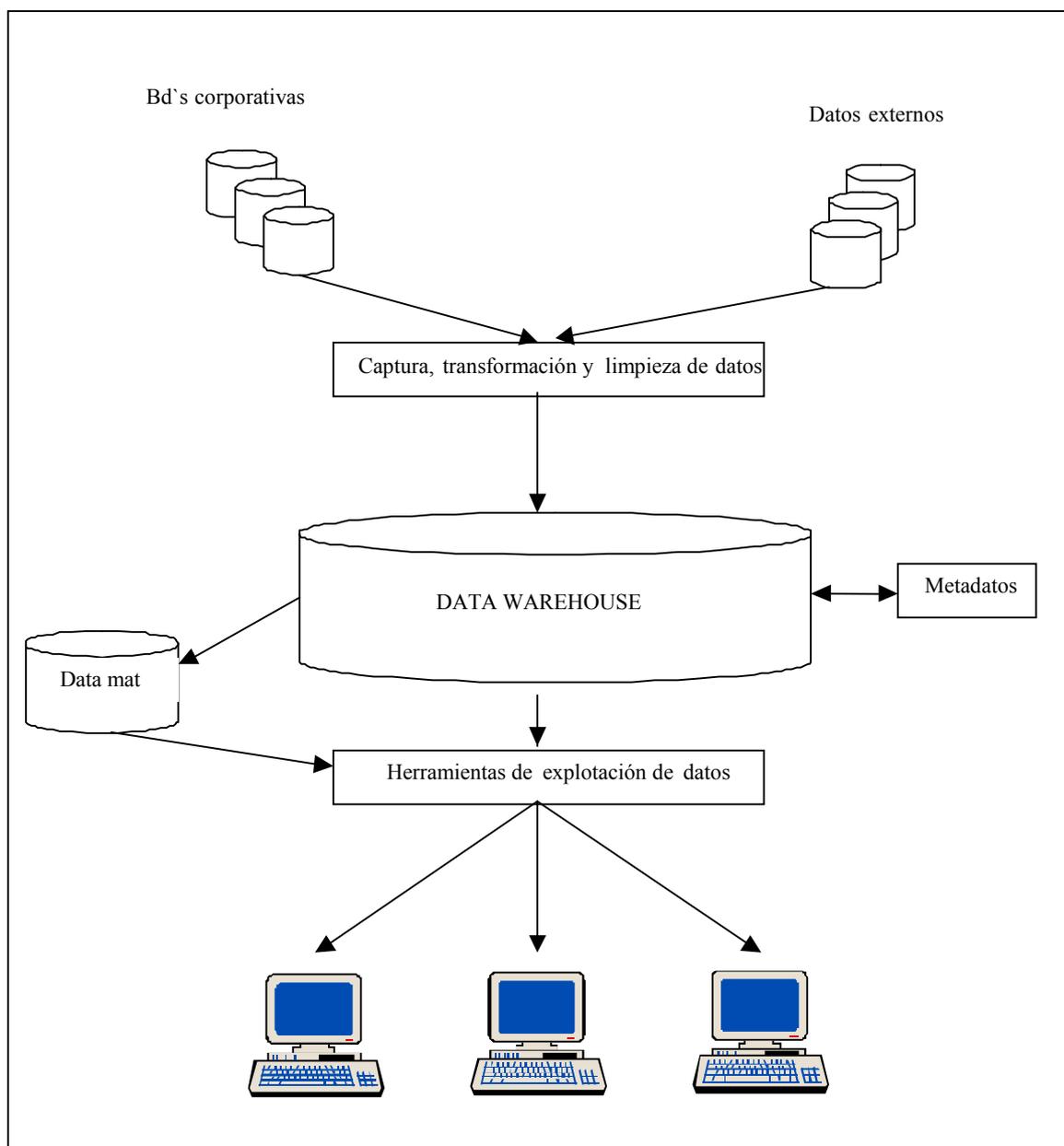
1. Consultas estándar: representan la forma más sencilla de análisis de datos, y consiste en hacer una pregunta a una base de datos. Pueden estar predefinidas y se representan por un icono en el escritorio del PC, o bien pueden ser consultas específicas de un usuario particular o para un problema concreto (consultas ad hoc). Este tipo de consultas permite a la empresa ser más ágil e incrementar su capacidad de respuesta.
2. Análisis multidimensional: permite a los usuarios del DW disponer de distintas dimensiones o perspectivas de los datos. Se trata de realizar análisis más profundos en datos específicos. Las herramientas usadas para este tipo de análisis son similares a las utilizadas para realizar consultas estándar, normalmente se trata de software de interfaz amigable.
3. Modelado y segmentación: se basan en técnicas de análisis más sofisticadas, a través de un software especializado. Un modelo es una colección de patrones para una característica dada, y se representa gráficamente o bien mediante un conjunto de reglas y notaciones, que explotan la información histórica y de gran nivel de detalle centralizada en el DW.
4. Descubrimiento del conocimiento: a través de algoritmos muy poderosos que buscan patrones en grandes bases de datos, y que se diferencia del modelado en que estos patrones no están especificados de antemano.

2.5. Arquitectura de un data warehouse

La arquitectura de un DW define la estructura y los elementos que contendrá el DW. Devlin (1997), señala los siguientes componentes dentro de la arquitectura de un DW.

- *Fuentes de datos:* pueden ser sistemas operacionales corporativos, con información relativa a la actividad rutinaria de la empresa, sistemas operacionales departamentales, fuentes externas, etc.
- *Extracción y transformación:* proceso que permite que la información pase desde las fuentes de datos, hasta la DW propiamente dicha.
- *Servidor o gestor de los datos:* su misión es la de mantener, distribuir, aportar seguridad y monitorización.
- *Herramientas de acceso y análisis:* son las llamadas business intelligence tool (BIT), cada usuario debe seleccionar qué herramientas se adaptan mejor a sus necesidades y a su DW.
- *Metadatos:* contienen información acerca de los datos contenidos en el DW. Entre sus funciones destacan: catalogar y describir la información, especificar el propósito de la misma, indicar las relaciones entre los datos, establecer quién es el propietario de la información, relacionar las estructuras técnicas de datos con la información de negocio, establecer las relaciones con los datos operacionales y las reglas de transformación y por último, limitar la validez de la información.
- *Data Marts:* son almacenes de datos especializados, diseñados para soportar necesidades de análisis específicas para un único departamento o área funcional de la empresa, por ejemplo marketing, finanzas, producción, etc. Estos almacenes soportan menos usuarios y menos cantidades de datos que un DW centralizado, y por lo tanto pueden ser optimizados para cargar y recuperar la información de forma más rápida y eficaz que un DW.

Tabla nº 3: Arquitectura de un data warehouse



Fuente: Martín y Ureña (2000)

2.6. Data Mining

La aplicación de tecnologías y sistemas de información en las empresas, de forma tan generalizada hoy en día, implica disponer de enormes cantidades de datos. En este sentido, podemos afirmar que el problema no reside pues en la falta de información, sino en las herramientas que se utilizan para la explotación eficiente de la misma (Lee y Siau, 2001). Estos autores, definen el data mining (DM), como un conjunto de técnicas, procesos y herramientas, que se utilizan para el análisis de los datos, lo cual conduce a las organizaciones a descubrir la información oculta en los mismos y añadirles valor.

DM representa un nuevo enfoque para el análisis y descubrimiento de la información o del conocimiento oculto en grandes bases de datos corporativas o en la DW. Se trata de una técnica que, apoyándose en el uso de tecnologías de información y de Internet, puede generar en las organizaciones ventajas competitivas.

Las herramientas que utiliza el DM se basan básicamente en técnicas estadísticas, y realizan entre otras, las siguientes funciones (García et al., 1998 y Borrajo, 1999):

- **Agrupamiento.** Esta herramienta permite la identificación de tipologías o grupos en los cuales los elementos guardan similitud entre sí y se diferencian de los de otros grupos. También recibe el nombre de segmentación. A través de este tipo de herramientas se puede segmentar el colectivo de clientes, el conjunto de valores e índices financieros, el conjunto de empleados,... Esto permite el tratamiento particularizado de cada uno de estos grupos.
- **Asociación.** Permiten establecer las posibles relaciones entre acciones o sucesos aparentemente independientes. Así, se puede reconocer cómo la ocurrencia de un determinado suceso puede inducir la aparición de otro u otros. Este tipo de herramientas es particularmente útil, por ejemplo, para comprender los hábitos de compra de los clientes y para la concepción de ofertas, de ventas cruzadas y del "merchandising".

- **Secuenciamiento.** Es un concepto similar al anterior, pero incluyendo también el factor tiempo. Es decir, permite reconocer el tiempo que transcurre o suele transcurrir entre el suceso inductor y los sucesos inducidos.
- **Reconocimiento de patrones.** Permiten la asociación de una señal o información de entrada con aquella o aquellas con las que guarda mayor similitud, y que están ya catalogadas en el sistema. En el ámbito del DM, estas herramientas pueden ayudar en la identificación de problemas e incidencias y de sus posibles soluciones, toda vez que se disponga de la base de información necesaria donde buscar.
- **Previsión.** La previsión permite establecer el comportamiento futuro más probable de una variable o una serie de variables a partir de la evolución pasada y presente de esas variables o de otras de las cuales dependan. Las técnicas asociadas a esta herramienta tienen ya un elevado grado de madurez.
- **Simulación.** Estas herramientas permiten comparar la situación actual de la empresa y su posible evolución futura con otras posibles situaciones, para a partir de ahí decidir si se deben o no modificar las políticas de actuación en determinados ámbitos de la misma.
- **Optimización.** La optimización resuelve el problema de la minimización o maximización de una función que depende de una serie de variables, encontrando los valores de éstas que satisfacen la condición de máximo, típicamente beneficios, o mínimo, típicamente costes. Normalmente suele haber unas restricciones, que hacen que no todas las posibles soluciones sean aceptables, de modo que el universo de búsqueda se reduce a aquellas soluciones que satisfagan las restricciones.
- **Clasificación.** Agrupa a todas las herramientas que permiten asignar a un elemento la pertenencia a un determinado grupo o clase. Esto se lleva a cabo a través de la dependencia de la pertenencia a cada clase en los valores de una serie de atributos o variables. Se establece un perfil característico de cada clase y su expresión, en términos de un algoritmo o reglas, en función de las distintas variables. Se establece también el grado de discriminación o influencia de estas últimas. Con ello, es posible clasificar un nuevo elemento una vez conocidos los valores de las variables presentes en él.

Vemos, por tanto, que el Data Mining se nos presenta como una potente herramienta en el tratamiento de la información para las organizaciones y las personas que forman parte de las mismas, por su aportación a la flexibilidad de la empresa, en tanto que la capacita para anticiparse en cierta medida a la evolución del entorno y las necesidades del cliente, y su capacidad de potenciar aquellas actividades de mayor valor como la creatividad, la innovación y la imaginación de las personas.

3. IMPLICACIONES DE UN DATA WAREHOUSE

3.1. Usuarios del data warehouse

En la tabla número cuatro aparece un esquema que muestra las características de los diferentes usuarios de los datos contenidos en bases de datos operacionales y los datos del DW.

Tabla 4: Usos de las diferentes bases de datos

USOS DE BASES DE DATOS OPERACIONALES	USOS DE DATA WAREHOUSE
Muchos usuarios concurrentes	Pocos usuarios concurrentes
Consultas predefinidas y actualizables	Consultas complejas, frecuentemente no anticipadas
Cantidades pequeñas de datos detallados	Cantidades grandes de datos detallados

Fuente: <http://www.timagazine.net/magazine/1299/dw.cfm>

Cada vez es más amplia la gama de usuarios finales de un DW. En general, se pueden considerar tres grandes categorías:

- Ejecutivos y gerentes
- "Power users" (analistas financieros y de negocios, ingenieros, etc.)
- Usuarios de soporte (de oficina, administrativos, etc.).

Cada una de estas categorías diferentes de usuario tienen su propio conjunto de requerimientos para los datos, acceso, flexibilidad y facilidad de uso.

Funcionalmente, un DW permite transformar los datos operacionales de una organización, en información útil para la toma de decisiones. De esta forma, los usuarios finales pueden examinar los datos desde una perspectiva estratégica, es decir, podrán realizar análisis de tendencias, producir informes con mayor rapidez, disponer de un acceso más fácil, más flexible y más intuitivo a la información que se necesite en cada momento.

3.2. Ventajas e inconvenientes

Las principales ventajas del DW son (Pascot, 1998):

- Menor coste en la toma de decisiones, en términos de tiempo y fiabilidad en la información obtenida.
- Mayor flexibilidad a la hora de realizar análisis.
- Mejor servicio al cliente en cuanto a que se incrementa la capacidad de atender sus necesidades.
- Se descentraliza la responsabilidad en la toma de decisiones.
- Crea una arquitectura de datos única para todas las aplicaciones analíticas
- Resuelve los problemas de integridad y calidad de datos

Entre los principales inconvenientes, Pascot (1998) señala:

- Complejidad y necesidad de anticipación
- Su puesta en funcionamiento requiere de un largo plazo
- Coste elevado

3.3. El data warehouse en las estrategias empresariales

Una de las finalidades del DW consiste en reunir y consolidar las diferentes bases de datos que existen en los distintos departamentos de la empresa. Estas bases de datos departamentales contienen datos repetidos y no homogéneos. Por otra parte, el DW también recoge datos o información que procede del exterior de la empresa, como por ejemplo información de sus clientes, proveedores, productos, canales, competencia, el mercado, la coyuntura económica, etc.

El objetivo es el de convertir los datos operacionales de la empresa y los que proceden de sus relaciones con el entorno en información relacionada y estructurada, homogénea y de mayor calidad, identificada de forma adecuada a través de los meta datos.

Las aplicaciones de DW no se compran, se construyen en un largo proceso, que requiere el establecimiento de unos objetivos claros, en cuanto a cuáles son las necesidades de información de los diferentes usuarios para la correcta toma de decisiones, (Chou et al. 2000). Una de las claves del éxito en la construcción de un DW, consiste en ir desarrollándolo de forma gradual, empezando por un departamento con pocos usuarios y en el que se pueda medir el resultado a corto plazo. Más adelante se irá expandiendo su implantación hacia otros departamentos y usuarios.

El DW se convierte en una herramienta fuente de ventajas competitivas en la empresa desde el momento en el que permite el análisis de la información, de forma sencilla y dinámica, para convertirla en conocimiento.

El DW proporciona a las organizaciones una base global y única con una gran flexibilidad de gestión, de manera que sea el usuario quien seleccione la información que necesita y su explotación. La información estará estructurada de acuerdo con las necesidades de cada negocio o departamento, Dyché (2001). Por ejemplo, el área de marketing necesitará datos sociodemográficos, económico-patrimoniales de los clientes actuales y potenciales, con lo que este departamento utilizará el almacén de datos para conocer el comportamiento de los clientes y determinar quiénes son los más adecuados para ofrecerles un producto determinado. Con las técnicas de explotación de la información (data mining) podrán anticipar o detectar qué clientes tienen intención de abandonar o no están satisfechos.

4. DATA WAREHOUSE E INTERNET

Llacer, et al. (1998), señalan que las tendencias actuales apuntan hacia una progresiva sustitución de las tecnologías basadas en las consultas de tipo cliente/servidor a las tecnologías Internet y Web. En este sentido, un DW conectado a Internet o a una intranet se convierte en un instrumento con un enorme potencial para la empresa, en cuanto a la cantidad de información, muchas veces gratuita, que puede adquirir del exterior.

Se trata de favorecer e incrementar el flujo de información, tanto desde la empresa hacia el exterior, como a la inversa. Con el desarrollo de tecnologías del tipo Web/OLAP, y con una conexión del servidor de la empresa a su DW, el usuario, interno o externo a la empresa, realizará las peticiones de información desde su navegador, el servidor transmitirá la petición al DW para que allí se procese, enviando los resultados al servidor Web, el cual los formateará como páginas html que se enviarán en forma de respuesta, al navegador de dicho usuario.

La combinación entre los DW e Internet / Intranet / Extranet está dando lugar a un nuevo concepto: los DW cibernéticos o DW activos (Dyché, 2001). Las empresas que no posean unos sistemas de información de gestión eficaces, se verán con dificultades para competir en los próximos años, ya que desde la perspectiva de la dirección estratégica, el recurso información es un activo intangible de gran valor y fuente de ventajas competitivas (Cornellá, 1994). Las necesidades de información y los cambios en el entorno cada vez mayores, requieren estar en el carro de las nuevas tecnologías y sistemas de información, en caso contrario, las empresas se encontrarán en una clara situación de desventaja competitiva.

Los conceptos relacionados con el DW, como datos orientados al tema o la integración de datos, pueden resultar de gran ayuda a la hora de estructurar y facilitar la búsqueda en Internet. De hecho, los buscadores más importantes en Internet permiten realizar búsquedas por jerarquías de temas y son capaces de clasificar la información según lo reciente de su fecha, su forma, contenido, etc.

El DW e Internet son probablemente, las dos tendencias recientes que las empresas deberán implantar y utilizar cada vez más para ser competitivas. Actualmente, casi todas las herramientas de análisis para la toma de decisiones, llevan algún componente que les permite visualizar una consulta a través de Internet, Dyché (2001).

Se estima que en los próximos años, las empresas dedicarán como media la tercera parte de su presupuesto de informática en la implantación y el uso de sistemas de decisión, (McKnight, 2000). Uno de los ámbitos donde el DW tiene mayor impacto es en la gestión de las relaciones de la empresa con sus clientes. Es probable que los sistemas informáticos sigan cada vez con mayor precisión los comportamientos de compra de cada cliente y utilicen la información para efectuar un seguimiento individual. Hoy, este tipo de aplicaciones no siempre resultan simples de implantar, porque a las bases de datos les falta información precisa, histórica y detallada de cada uno de sus clientes. En esta línea, el DW se convierte en el elemento central de un sistema de información orientado al cliente.

5. CONCLUSIONES

El DW representa una herramienta muy potente para el diseño y desarrollo de los actuales sistemas de información en las empresas. Dichos sistemas de información se han ido adaptando a los requerimientos y necesidades informacionales que los gestores empresariales precisaban en cada momento, es decir, mayor cantidad, calidad y velocidad en el proceso de obtención de información.

Este planteamiento es el resultado de una actuación empresarial cada vez más competitiva dentro de un entorno caracterizado por los rápidos cambios tecnológicos que se están produciendo en las últimas décadas y que influyen y modifican comportamientos tanto en el ámbito empresarial como en el cultural, vivimos en la era de la información. Las crecientes inversiones en I+D+I que se llevan a cabo en el seno de las grandes instituciones públicas y privadas, están propiciando tales cambios a ritmos cada vez más vertiginosos.

En este contexto disponer de los últimos avances en sistemas y tecnologías para la comunicación, no garantiza ni mucho menos el éxito empresarial, pero en lo que sí parecen estar de acuerdo los directivos de las empresas, es en que dichos sistemas y tecnologías representan un soporte adecuado para lograr los objetivos fijados en la planificación estratégica.

Dicho proceso de planificación estratégica está ligado al de la toma de decisiones, en este sentido y para que estas decisiones sean las más adecuadas, los directivos precisan de información útil, oportuna, veraz y relevante. El DW permite aprovechar el enorme volumen de

datos que las organizaciones registran diariamente en sus procesos transaccionales tanto internamente como con el exterior de la misma. El objetivo por tanto de esta tecnología es el de convertir los datos contenidos en las bases de datos corporativas de las organizaciones en información útil para el proceso de toma de decisiones. Por otra parte, la información será transformada posteriormente en conocimiento a través de un proceso de análisis y explotación de los datos contenidos en el DW, a partir de un proceso denominado Data mining.

BIBLIOGRAFÍA

BORRAJO DIZ, M.L. (1999), "Data warehouse".

[Http://www.timagazine.net/magazine/1299/dw.cfm?id=1299/dw&clave=sistemas%20OR%20gestion%20R%20conocimiento%20R%20data%20OR%20warehouse%20OR%20minin](http://www.timagazine.net/magazine/1299/dw.cfm?id=1299/dw&clave=sistemas%20OR%20gestion%20R%20conocimiento%20R%20data%20OR%20warehouse%20OR%20minin)

CORNELLÁ, A. (1994), Los recursos de información. Ventaja Competitiva de las Empresas, McGraw-Hill, Madrid.

DEVLIN, B. (1997), Data Warehouse: From Architecture to Implementation. Addison Wesley, Reading, MA.

DYCHÉ, J. (2001), E-Data: Transformando datos en información con Data Warehousing. Prentice Hall, 2001

ESPINOZA, M.T. (2000), Data warehouse: metodología.

<http://www.bi-magazine.com/bi-magazine/htm/desembre-00-artMetodologia.htm>

FORCHT, K.A.; COCHRAN, K. (1999), "Usind data mining and datawarehousing techniques", *Industrial Management & Data Systems*, vol. 99, nº 5, pp 189-196

GARCÍA CABRERA, C.; PASCUAL MIGUEL, F.J. Y CHAPARRO PELÁEZ, J. (1998), "Sistemas de información y comunicaciones para la gestión del conocimiento", *VIII Congreso Nacional de ACEDE en Las Palmas de Gran Canaria*.

INMON, W. H., (1996), Building the data warehouse. Jhon Wiley & Sons

LLACER RUBIO, E.; LUNA HUERTAS, P.; MARTÍNEZ LÓPEZ, F.J.; SALMERÓN SILVERA, J.L.; CASTAÑEDA BARRENA, R., (1998), "Data Warehousing, un paso más hacia la gestión del conocimiento en las empresas", *VIII Congreso Nacional de ACEDE, Las Palmas de Gran Canaria*

MA, CHATERINE; CHOU, D. C.; YEN, D.C.; (2000), "Data Warehousing, technology assessment and management". *Industrial Management & Data Systems*, vol. 100, nº 3, pp 125-134.

MARTÍN VALDIVIA, M.T.; UREÑA LÓPEZ, L.A., (2000), “Data Warehosue y la toma de decisiones en la empresa”. *Actas del I encuentro Iberoamericano de Finanzas y Sistemas de Información, vol. II, Jerez de la Frontera, nov.-dic. 2000*

PASCOT, D., (1998), “Les entrepôts de données clé de voûte de l’informatique décisionnelle”
<http://www.fsa.ulaval.ca/personnel/pascotd>

SANG JUN LEE; KENG SIAU, (2001), “A review of data mining techniques”. *Industrial Management & Data Systems*, vol. 101, nº 1, pp 41-46

WILLIAM MCKNIGHT, (2000), “The CRM-Ready Data Warehouse”, julio 2000.
<http://www.datawarehouse.com>

Algunas direcciones útiles de internet

<http://www.timagazine.net/buscador/mostrar.cfm>

http://www.sqlmax.com/manual_dw.htm

<http://www.datawarehouse.com>

<http://www.dmreview.com/master.cfm?NavID=106>

<http://www.dw-institute.com>¹

<http://www.map.es/csi/silice/DW21.html>

NOTAS

¹ Un gigabyte representa mil millones de bytes, un terabyte, un billón de bytes y un petabyte, mil billones de bytes.

² Los metadatos recogen la semántica de los datos (su definición actual y sus distintos significados en el ciclo de vida del DW), así como sus interrelaciones, lo cual permite conocer las reglas del negocio, Matín y Ureña (2000).