

# ¿Qué ofrece Autentia Real Business Solutions S.L?

Somos su empresa de **Soporte a Desarrollo Informático**.  
Ese apoyo que siempre quiso tener...

## 1. Desarrollo de componentes y proyectos a medida



## 2. Auditoría de código y recomendaciones de mejora

## 3. Arranque de proyectos basados en nuevas tecnologías

1. Definición de frameworks corporativos.
2. Transferencia de conocimiento de nuevas arquitecturas.
3. Soporte al arranque de proyectos.
4. Auditoría preventiva periódica de calidad.
5. Revisión previa a la certificación de proyectos.
6. Extensión de capacidad de equipos de calidad.
7. Identificación de problemas en producción.



## 4. Cursos de formación (impartidos por desarrolladores en activo)

Spring MVC, JSF-PrimeFaces /RichFaces,  
HTML5, CSS3, JavaScript-jQuery

Gestor portales (Liferay)  
Gestor de contenidos (Alfresco)  
Aplicaciones híbridas

Tareas programadas (Quartz)  
Gestor documental (Alfresco)  
Inversión de control (Spring)

Control de autenticación y  
acceso (Spring Security)  
UDDI  
Web Services  
Rest Services  
Social SSO  
SSO (Cas)

JPA-Hibernate, MyBatis  
Motor de búsqueda empresarial (Solr)  
ETL (Talend)

Dirección de Proyectos Informáticos.  
Metodologías ágiles  
Patrones de diseño  
TDD

BPM (jBPM o Bonita)  
Generación de informes (JasperReport)  
ESB (Open ESB)



[Home](#) | [Quienes Somos](#) | [Empleo](#) | [Tutoriales](#) | [Contacte](#)



**CoNcept**



**Lanzado**

## **TNTConcept versión 0.6** ( 12/07/2007)

Desde [Autentia](#) ponemos a vuestra disposición el software que hemos construido (100% gratuito y sin restricciones funcionales) para nuestra gestión interna, llamado TNTConcept (auTeNTia).

Construida con las últimas tecnologías de desarrollo Java/J2EE (Spring, JSF, Acegi, Hibernate, Maven, Subversion, etc.) y disponible en licencia GPL, seguro que a muchos profesionales independientes y PYMES os ayudará a organizar mejor vuestra operativa.

**Las cosas grandes empiezan siendo algo pequeño .....** Saber más en: <http://tntconcept.sourceforge.net/>

 <p><b>Autor del tutorial:</b> Cristhian Kirs Herrera Basurto</p> <p>§ <b>Lugar de residencia:</b> Quito - Ecuador</p> <p><i>Cuento con experiencia en el área de desarrollo de software y en la docencia académica. Dentro de la construcción de software he manejado las etapas de: análisis, diseño, personalización e implementación de aplicaciones bajo ambientes Cliente / Servidor e Internet.</i></p> <p><a href="mailto:Cristhian.Herrera@gmail.com">Cristhian.Herrera@gmail.com</a> / <a href="mailto:cherrera@kruger.com.ec">cherrera@kruger.com.ec</a></p>	<p><b>NUEVO CATÁLOGO DE SERVICIOS DE AUTENTIA (PDF 6,2MB)</b></p> <p><a href="http://www.adictosaltrabajo.com">www.adictosaltrabajo.com</a> es el Web de difusión de conocimiento de <a href="http://www.autentia.com">www.autentia.com</a></p>  <p><b>autentia</b> real business solutions</p> <p><a href="#">Catálogo de cursos</a></p>
--	--

Descargar este documento en formato PDF [datawarehouse4.pdf](#)

[Firma en nuestro libro de Visitas](#) <-----> [Asociarme al grupo AdictosAlTrabajo en eConozco](#)

### Java

¿Cuanto te pagan? Mira nuestras Ofertas, Compara e Insíbete.  
[InfoJobs.net](http://InfoJobs.net)

### Java j2ee

Curso Programador Java Plataforma J2EE. Gratuito Desempleados. Madrid [cursos.maestroteca.com/java](http://cursos.maestroteca.com/java)

### Real-Time Embedded Java

Java productivity, C++ performance PERC:Europe's best selling solution [www.aonix.com](http://www.aonix.com)

### Formación Oficial SUN

Formación y Certificación en Java y Solaris. Los líderes en Barcelona [www.netmind.es](http://www.netmind.es)

Anuncios Google

**Fecha de creación del tutorial: 2007-10-30**

Apuntes Datawarehouse

[Viene de la 3ª parte](#)

[Ejemplo de un Datawarehouse 36](#)

[DataMart 38](#)

[Diferencias entre Data Mart y Datawarehouse 38](#)

[Razones para crear un Data Mart 38](#)

[Aspectos del Data Mart 39](#)

[Funcionalidad del Data Mart 39](#)

[Tamaño del Data Mart 39](#)

[Rendimiento de carga del Data Mart 39](#)

[Acceso de usuarios a datos en múltiples data marts 40](#)

[Acceso al Data Mart a través del Internet/Intranet 40](#)

[Administración del Data Mart 40](#)

[Instalación del Data Mart 40](#)

[Arquitectura de una aplicación de Data Mart 41](#)

[Construcción del datawarehouse 41](#)[Definición de una arquitectura. 41](#)

Arquitectura DW: 42

Resultados de la Arquitectura 42

[Generación del Datawarehouse 42](#)[Estrategia Bottom-up. 42](#)[Estrategia Top-down. 43](#)[Construcción en Incrementos, Datamarts. 44](#)

Construcción Incremental: 44

Resultados de Implementación: 44

[Consideraciones de Implementación mediante DataMarts. 44](#)[Metodología De Desarrollo de DW 45](#)[Diseño de la Arquitectura 46](#)

Arquitectura del Depósito 46

[Estructura Multidimensional 49](#)[Modelamiento Multidimensional 50](#)[Modelos de Datos 52](#)[Características del MER 52](#)[Características del Modelo Multidimensional 53](#)[Conceptos asociados al Datawarehouse 53](#)[Tabla de Hechos \(Tablas Fac\) 53](#)[Tablas Dimensionales \(Tablas Lock-up\) 53](#)[Esquema Estrella 54](#)[Esquema snowflake \(Copos de Nieve\) 55](#)[Pasos básicos del Modelamiento Dimensional 55](#)[Profundizaciones de Diseño 56](#)[Dimensión Tiempo 56](#)[Dimensiones que varían lentamente en el Tiempo 56](#)[Niveles 56](#)[Jerarquías 57](#)[Explotación del datawarehouse 57](#)[Potencial del datawarehouse 57](#)[Extracción y manipulación de datos del datawarehouse 57](#)[Procesamiento Informático 58](#)[Procesamiento Analítico 59](#)[Data Mining 61](#)[Glosario de Términos asociados con Datawarehouse 62](#)[Bibliografía 64](#)

## **Ejemplo de un Datawarehouse**

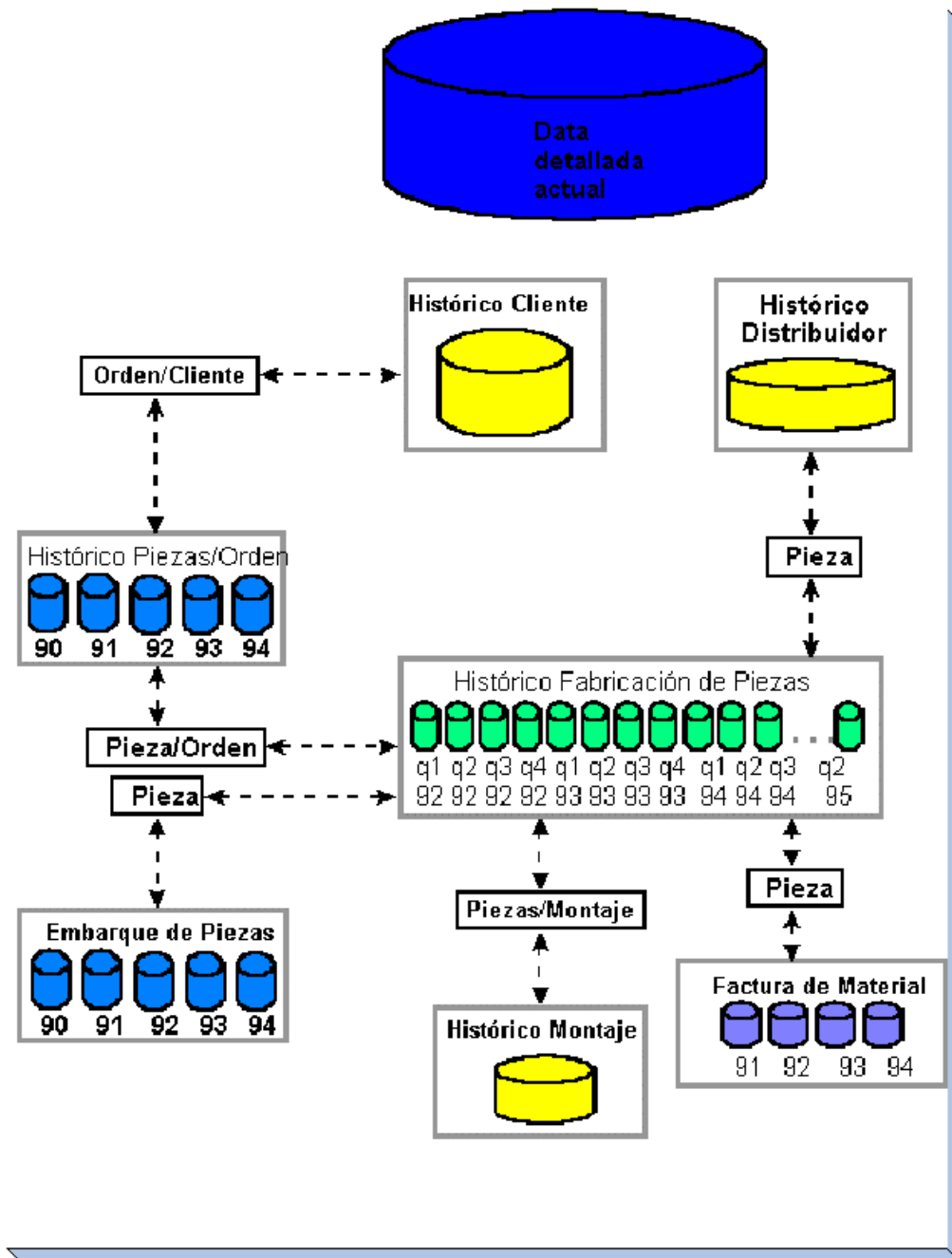
En la Figura se muestra un ejemplo hipotético de un data warehouse estructurado para un centro de producción industrial.

Se muestra sólo el detalle actual, no así los niveles de esquematización ni los archivos de detalle más antiguos.

Además, se observa que hay tablas del mismo tipo divididas a través del tiempo. Por ejemplo, para el histórico de la fabricación de las piezas, hay muchas tablas separadas físicamente, representando cada una un trimestre diferente. La estructura de los datos es consistente con la tabla de la elaboración de las piezas, aunque físicamente hay muchas tablas que lógicamente incluyen el histórico.

Para los diferentes tipos de tablas hay diferentes unidades de tiempo que físicamente dividen las unidades de información. El histórico de fabricación está dividido por trimestres, el histórico de la orden de piezas está dividido por años y el histórico de cliente es un archivo único, no dividido por el tiempo.

Así también, las diferentes tablas son vinculadas por medio de un identificador común, piezas u órdenes de piezas (la representación de la interrelación en el ambiente de depósito toma una forma muy diferente al de otros ambientes, tal como el ambiente operacional).



## **DataMart**

Data Mart, es un subconjunto de los datos de un datawarehouse, normalmente en la forma de información resumida que soporta los requerimientos de un departamento o función de negocio particular

Un data mart puede ser considerado como un centro de distribución, creado para servir más eficientemente a un segmento de usuarios. En algunos casos el data mart, como centro de distribución, es creado desde un data warehouse, aunque es mas común su creación directa sin enlace a un data warehouse.

Existen varios enfoques para construir data marts. Un enfoque es construir un dw corporativo que puede ser usado directamente por los usuarios y proveer los datos para otros data marts. Otro enfoque es construir varios data marts con una vista para la virtual integración en un warehouse, y el enfoque final es construir la infraestructura para un dw corporativo mientras al mismo tiempo se construye uno o más data marts para satisfacer las necesidades inmediatas del negocio.

## **Diferencias entre Data Mart y Datawarehouse**

- El Data Mart se centra solamente en los requerimientos de usuarios asociados con un departamento o función de negocio
- Los Data Marts normalmente no contienen datos operacionales detallados a diferencia de datawarehouse.
- Debido a que los data marts contienen menos información comparados con los datawarehouse, los data marts son más fácilmente entendibles y navegables.

## **Razones para crear un Data Mart**

El siguiente listado de razones proporciona algunos argumentos para la creación de Data Marts en un negocio

- Dar a los usuarios acceso a los datos que ellos necesitan para analizarlos mas a menudo
- Proveer los datos en una forma que concuerda la vista colectiva de los datos por un grupo de usuarios en un departamento o función de negocio
- Mejorar el tiempo de respuesta al usuario final debido a la reducción en el volumen de información a ser accedido.
- Proveer datos apropiadamente estructurados para satisfacer los requerimientos de las herramientas de acceso de usuario final.
- Un número de herramientas de acceso de usuario, particularmente especialistas en minería de datos o herramientas de análisis multidimensional pueden requerir sus propias estructuras internas de bases de datos.
- En la práctica, estas herramientas a menudo crean su propio data mart diseñado para soportar su funcionalidad específica
- Los Data Marts normalmente usan menos datos de tal manera que la tarea de

carga de datos es más fácil, y de esta manera la implementación de un data mart es más simple comparado con un datawarehouse corporativo

- El costo de implementación de un data mart normalmente es menor que el requerido para establecer un datawarehouse
- Los potenciales usuarios de un data mart son más claramente definidos y puede ser más fácilmente involucrados para obtener soporte para un proyecto de data mart

## Aspectos del Data Mart

Algunos de aspectos de consideración dentro de la implementación de un Data Mart:

- Funcionalidad del Data Mart
- Tamaño del Data Mart
- Rendimiento de Carga del Data Mart
- Acceso a usuarios en múltiples Data Mart
- Acceso al Data Mart desde Intranet / Internet
- Administración del Data Mart
- Instalación del Data Mart

## Funcionalidad del Data Mart

- Las capacidades del data mart se ha incrementado con el crecimiento de su popularidad. Más que ser simplemente base de datos pequeñas, fáciles de acceder, algunos data marts deben ser ahora escalables a cientos de gigabytes (GB).
- Proveen análisis sofisticado usando herramientas de minería de datos.
- Además cientos de usuarios deben ser capaces de acceder remotamente al data mart.
- La complejidad y tamaño de algunos data marts son concordantes con las características de data warehouses corporativos de pequeña escala.

## Tamaño del Data Mart

- Los usuarios esperan tiempos de respuesta más rápidos desde data marts que desde data warehouses, sin embargo, el deterioro del rendimiento de un data mart crece con el tamaño.
- Varios vendedores de data marts están investigando maneras para reducir el tamaño de data marts para ganar mejoras en el rendimiento; por ejemplo soportar dimensiones y jerarquías dinámicas para reducir el tamaño de la base de datos y tiempo de consolidación.
- Dimensiones dinámicas permiten agregaciones a ser calculadas sobre demanda

antes que precalculados y almacenados en los cubos de bases de datos multidimensionales (MDDDB).

## Rendimiento de carga del Data Mart

- Un data mart tiene que balancear dos componentes críticos: tiempo de respuesta al usuario final y rendimiento de carga de datos.
- Un data mart designado para respuesta rápida a usuarios tendrá un gran número de tablas resúmenes y valores agregados. Desgraciadamente, la creación de tales tablas y valores incrementan grandemente el tiempo del procedimiento de carga. MDDDBs, soportan actualización incremental de la base de datos de tal manera que la estructura completa del MDDDB no tiene que ser cambiada por cada actualización, lo cual es tradicionalmente requerido; únicamente las celdas afectadas por el cambio son actualizadas.

## Acceso de usuarios a datos en múltiples data marts

- Un enfoque es replicar los datos entre diferentes data marts o, alternativamente, construir *data marts virtuales*.
- Data marts virtuales son vistas de varios data marts físicos o el data warehouse corporativo ajustado para satisfacer los requerimientos de grupos específicos de usuarios. Los vendedores están desarrollando productos para gestionar data marts virtuales.

## Acceso al Data Mart a través del Internet/Intranet

- Tecnología Internet/Intranet ofrece a los usuarios acceso de bajo costo a data marts y data warehouses usando Web browsers tales como Netscape Navigator e Internet Explorer.
- Productos Data Mart para Internet/Intranet normalmente se sitúan entre un Web server y el producto de análisis de datos.

## Administración del Data Mart

- A medida que el número de data marts en una organización se incrementa, se tiene la necesidad de gestionar y coordinar centralmente las actividades del data mart. Una vez que los datos son copiados al data mart, los datos pueden llegar a ser inconsistentes a medida que los usuarios alteran sus propios data marts para permitirles analizar los datos en diferentes maneras.
- Las organizaciones no pueden ejecutar fácilmente la administración de múltiples data marts, dando surgimiento a aspectos tales como versionamiento de data marts, consistencia e integridad de datos y meta datos, seguridad de empresas amplias, y afinamiento de rendimiento.

- Existen herramientas administrativas de data marts en el mercado.

## Instalación del Data Mart

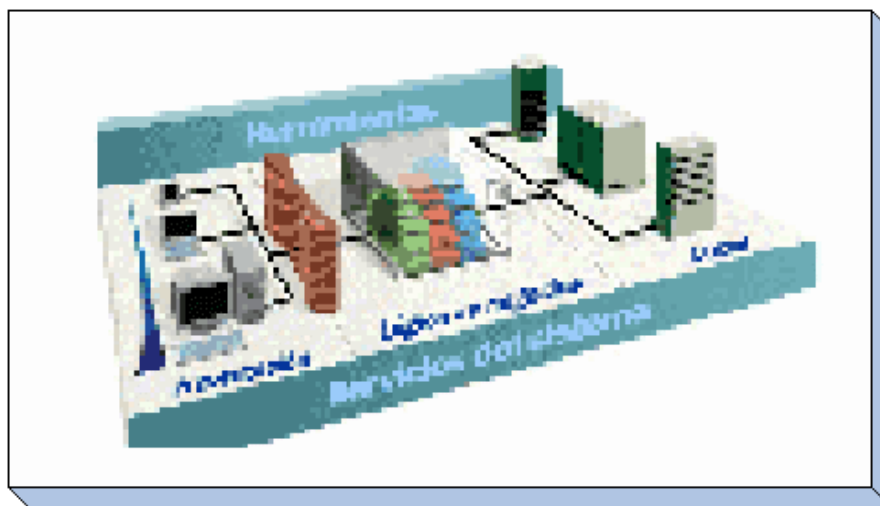
- Data marts están llegando a ser incrementalmente complejos de construir.
- Actualmente los vendedores están ofreciendo productos referidos como "data marts en una caja" que proveen una fuente de herramientas data mart de bajo costo.

## Arquitectura de una aplicación de Data Mart

Desde el punto de vista de arquitectura tecnológica, la arquitectura de una aplicación de Data Mart es la misma que si se desarrollara todo un Datawarehouse.

Es conveniente que una aplicación de data warehouse o data mart sea desarrollada considerando que va a ser explotada en una arquitectura de tres capas (aplicaciones distribuidas); las cuales son:

- Capa de presentación (los componentes de interfase de usuario y de despliegue)
- Capa lógica analítica (las consultas, procesos, y funciones de formateo)
- Capa de datos (data warehouse)



## Construcción del datawarehouse

Al iniciar un proyecto Datawarehousing no debemos olvidar establecer un marco de referencia de construcción del DW.



Podemos distinguir en dicha construcción dos etapas principales: la definición de una arquitectura DW y la construcción de los incrementos DW.

## **Definición de una arquitectura.**

### **Arquitectura DW:**

Arquitectura DW establece el marco de trabajo, estándares y procedimientos para el DW a un nivel empresarial. Los objetivos de las actividades de la arquitectura son simples, integrar al DW las necesidades de información empresarial.

### **Resultados de la Arquitectura**

Los principales resultados del desarrollo de la arquitectura DW incluyen:

- El modelo de datos fuente.
- El modelo de datos conceptual DW.
- Arquitectura tecnológica DW.
- Estándares y procedimientos DW.
- El plan de implementación incremental para el DW.

Los modelos de datos proveen una estructura para identificar, nombrar, describir y asociar los componentes de una base de datos. En general se necesitan modelos de datos para los datos fuente como para los datos seleccionados para existir en el DW.

Los estándares DW son una parte importante de la arquitectura DW. Sin estándares, oportunidades para reusar no son posibles y hay riesgos de que partes del desarrollo no ganen trabajando juntos.

El plan de implementación DW es la parte de la arquitectura de DW que identifica los incrementos del DW y describe la secuencia de desarrollo de estos incrementos.

## **Generación del Datawarehouse**

### **Estrategia Bottom-up.**

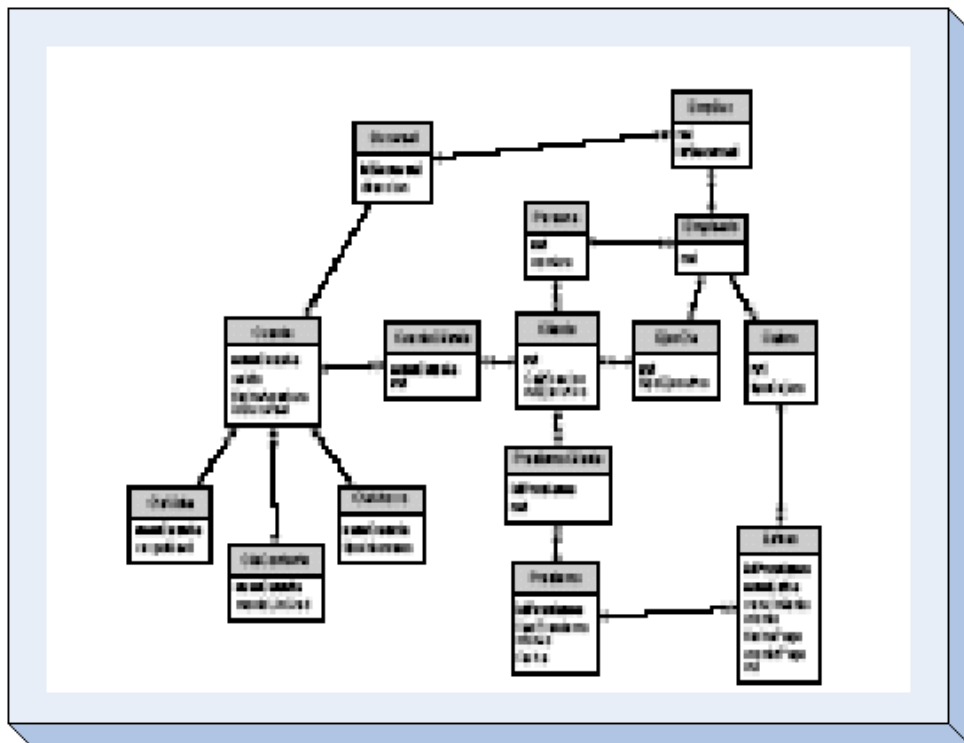
- Establece la construcción de un Datawarehouse exclusivamente desde la información contenida en los sistemas transaccionales.
- Genera todas las estructuras dimensionales que puedan ser alimentadas desde las fuentes de datos de los sistemas OLTP.

**Ventajas:**

- Se asegura la existencia de toda la información de los sistemas OLTP, que requiera el Datawarehouse.
- Posibilita la generación de mecanismos de carga automatizados que simplifiquen las operaciones de mantenimiento y administración de la información.

**Desventajas:**

- Generación de estructuras dimensionales innecesarias para la correcta toma de decisiones, con lo cuál se desperdician recursos valiosos y alargan los tiempos de implementación.
- Para una correcta toma de decisiones, no solo se requiere la información presente en los sistemas transaccionales, también es necesaria información externa a la empresa, información que queda fuera del modelo al ser utilizado este enfoque.

**Estrategia Top-down.**

- Establece como paso inicial la definición de todos los requerimientos de información para los ejecutivos de la organización.
- Se identifican las fuentes de información que serán utilizadas para satisfacer los requerimiento definidos, dichas fuentes pueden ser sistemas transaccionales, información no automatizada y fuentes externas a la organización.

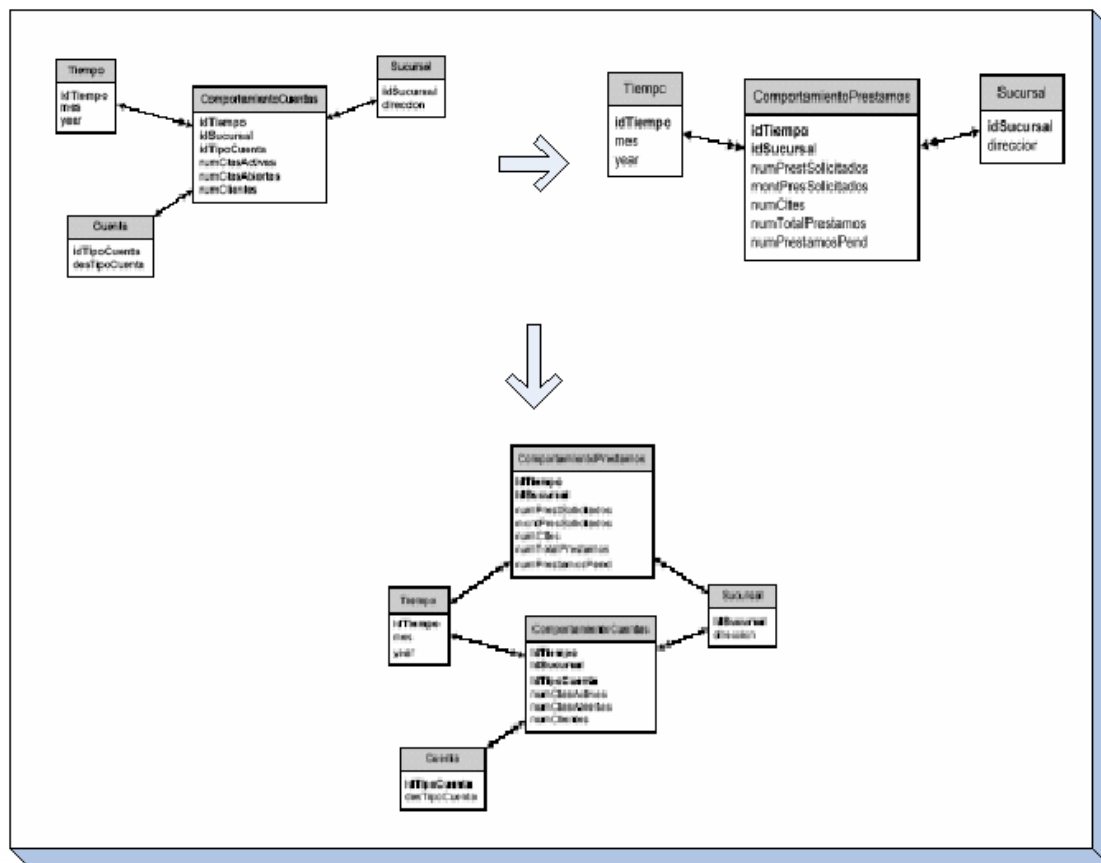
**Ventajas:**

- El datawarehouse resultante esta realmente enfocado a las necesidades de los

clientes y que apoya de forma más eficiente la toma de decisiones.

#### Desventajas:

- Aumenta la complejidad en la obtención de información necesaria para la carga de datos, especialmente cuando las fuentes no se encuentran automatizadas o están fuera de la organización.



## Construcción en Incrementos, Datamarts.

### Construcción Incremental:

La implementación en incrementos warehouse desarrolla y genera un subconjunto del DW total. La implementación incremental es una aproximación pragmática para construir un warehouse a un nivel empresarial en forma evolutiva.

### Resultados de Implementación:

Los resultados más significativos de la implementación de un incremento DW, incluyen:

- Las bases de datos warehouse.
- Programas y procedimientos para extraer, transformar y cargar datos.
- Instalar herramientas de acceso a los datos.
- Poblar el DW con los datos necesarios.
- Poblar el catálogo de metadatos con los datos necesarios.

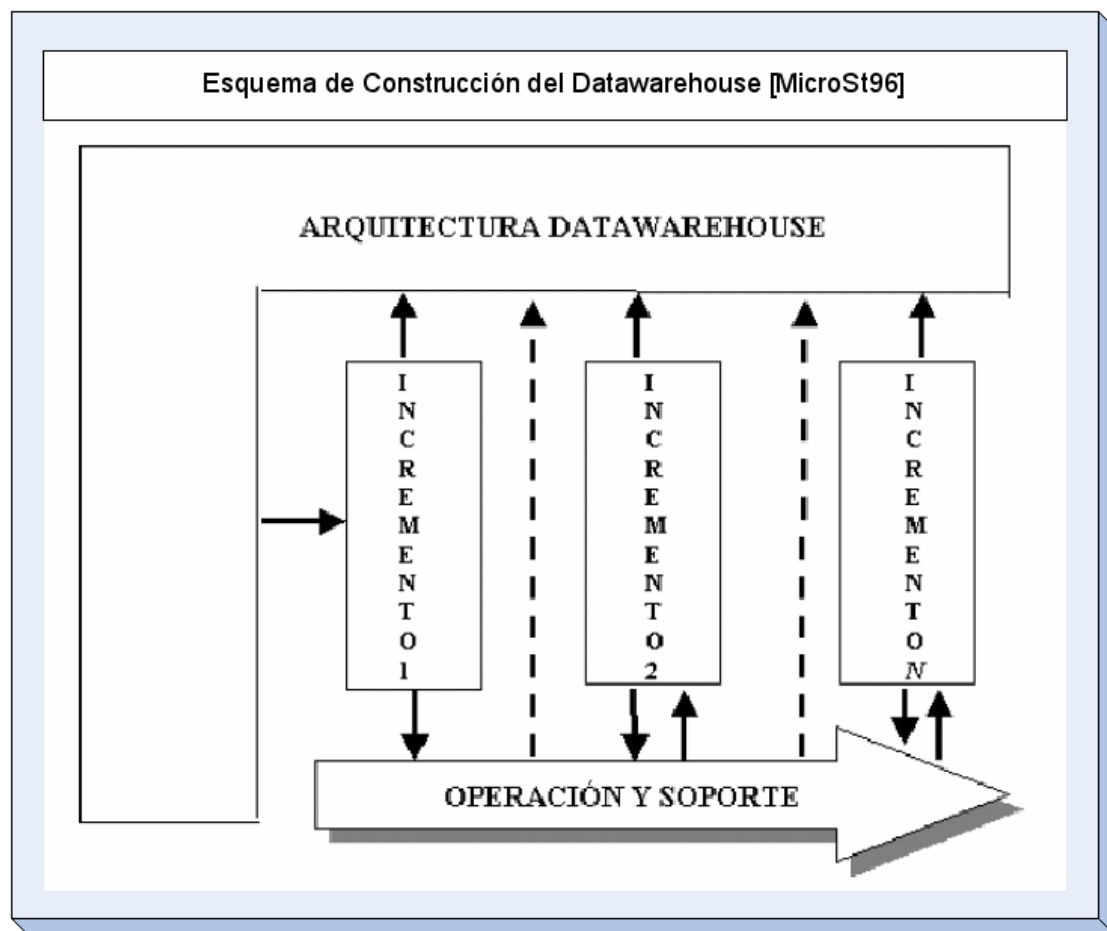
- Técnicas de uso y soporte el almacén

## Consideraciones de Implementación mediante DataMarts.

Debe tomar en cuenta lo siguiente;

- La arquitectura Datawarehouse se debe desarrollar al principio del proyecto.
- El primer incremento se desarrolla basándose en la arquitectura.
- La construcción del primer incremento puede causar cambios en la arquitectura.
- La operación del DW puede implicar la realización de cambios en la arquitectura.
- Cada incremento adicional puede extender el datawarehouse.
- Cada nuevo incremento puede causar ajustes en la arquitectura.
- La operación continua puede causar ajustes en la arquitectura.

La siguiente figura, intenta esquematizar el proceso de construcción del Datawarehouse.



## Metodología De Desarrollo de DW

De acuerdo a la forma de implementación analizada, se deben considerar y asociar metodologías congruentes con el desarrollo de incrementos dirigidos a grupos específicos en las organizaciones.

Al acercarse a la implementación de un DW con un conjunto de proyectos pequeños, altamente enfocados para implementar partes DW, resulta natural pensar en una metodología incremental para abordar su desarrollo. Pero no debemos olvidar la integración de cada incremento a la arquitectura Datawarehouse, así entonces el desarrollo evolutivo resulta ser una aproximación práctica de construcción de un DW.

De esta manera, estos proyectos pueden aprovechar los beneficios de la implementación incremental, que incluyen la contención de riesgos, oportunidades para aprender a desarrollar datamarts, entrega frecuente y la minimización del impacto sobre la comunidad de usuarios, y a la vez pueden ser organizados en forma secuencial, paralela o en una combinación de estructuras en serie y paralelo.

En tanto, el desarrollo evolutivo implica que cada vez que un incremento sea entregado, se debe operar y desarrollar simultáneamente el DW. De esta forma se logra integrar cada incremento a la estructura final DW.

El costo del desarrollo evolutivo queda dado por el incremento en la frecuencia de los cambios y la necesidad intensificada de realizar soporte del DW.

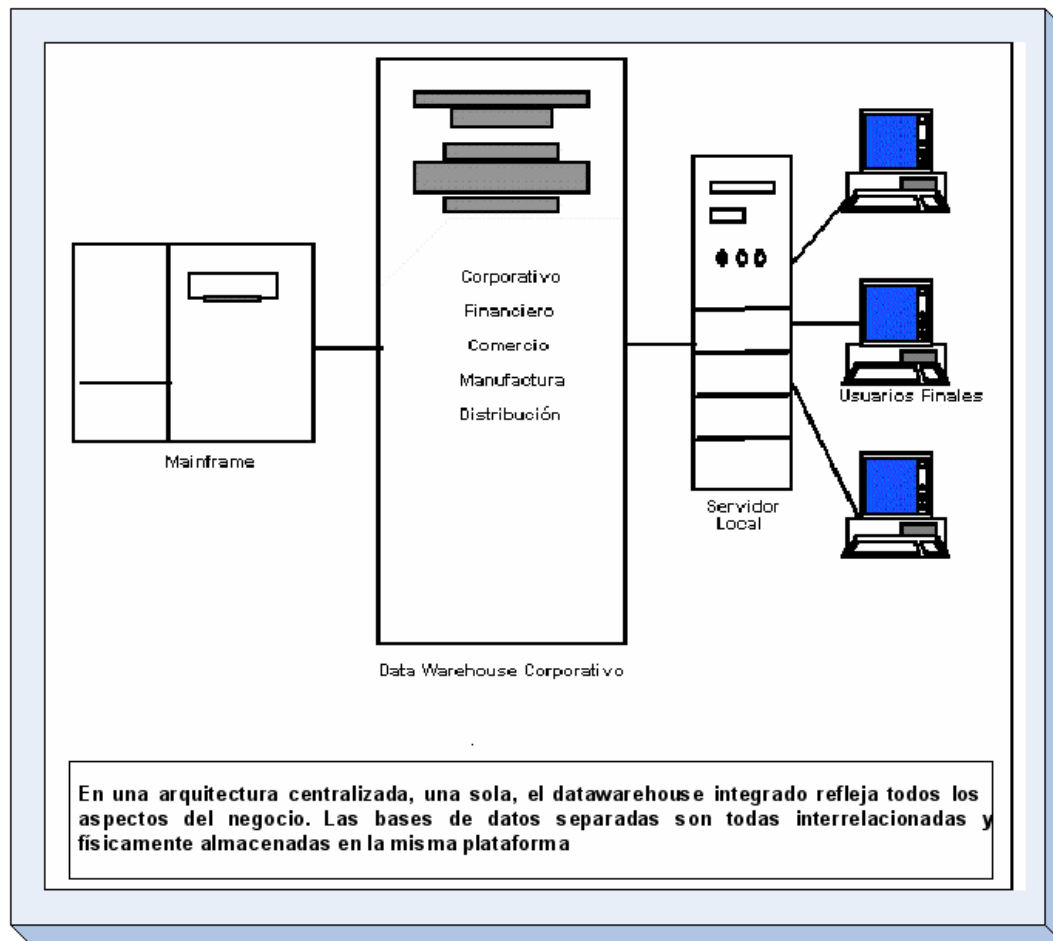
## **Diseño de la Arquitectura**

### **Arquitectura del Depósito**

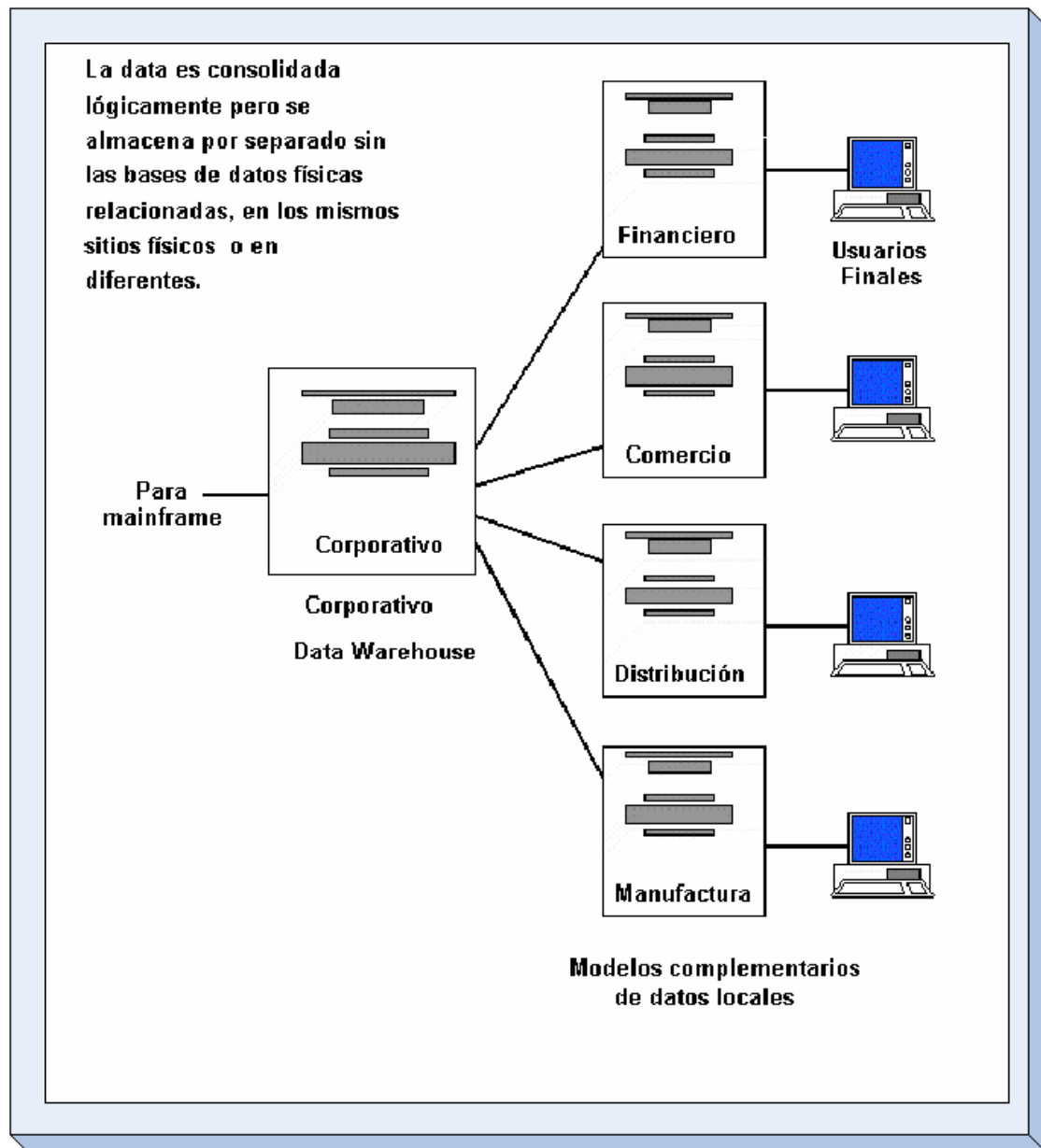
El desarrollo del data warehouse comienza con la estructura lógica y física de la base de datos del depósito más los servicios requeridos para operar y mantenerlo. Esta elección conduce a la selección de otros dos ítems fundamentales: el servidor de hardware y el DBMS.

La plataforma física puede centralizarse en una sola ubicación o distribuirse regional, nacional o internacionalmente. A continuación se dan las siguientes alternativas de arquitectura:

- Un plan para almacenar los datos de su compañía, que podría obtenerse desde fuentes múltiples internas y externas, es consolidar la base de datos en un data warehouse integrado. El enfoque consolidado proporciona eficiencia tanto en la potencia de procesamiento como en los costos de soporte. (Ver Figura).



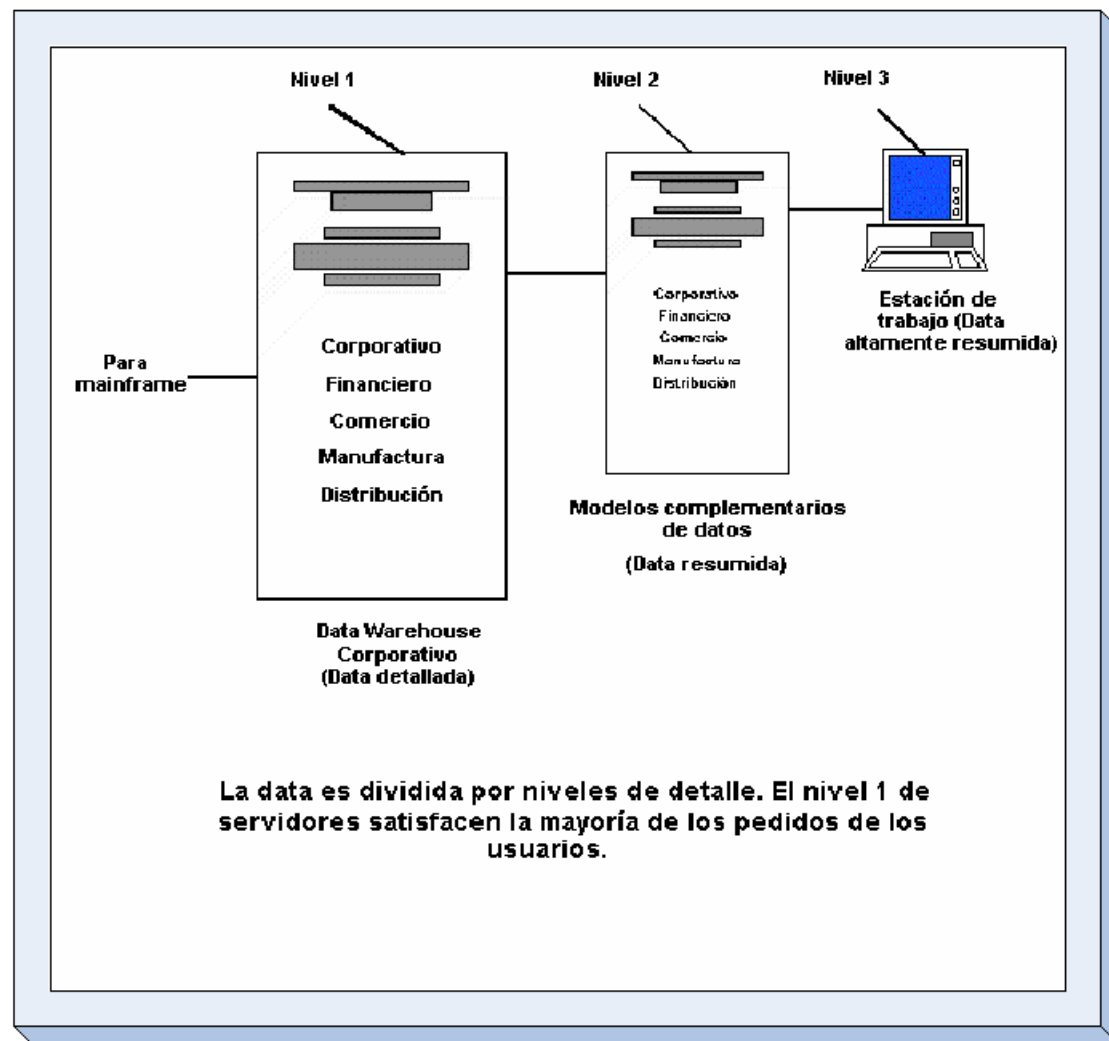
- La arquitectura global distribuye información por función, con datos financieros sobre un servidor en un sitio, los datos de comercialización en otro y los datos de fabricación en un tercer lugar. (Ver Figura )



- Una arquitectura por niveles almacena datos altamente resumidos sobre una estación de trabajo del usuario, con resúmenes más detallados en un segundo servidor y la información más detallada en un tercero.

La estación de trabajo del primer nivel maneja la mayoría de los pedidos para los datos, con pocos pedidos que pasan sucesivamente a los niveles 2 y 3 para la resolución.

Las computadoras en el primer nivel pueden optimizarse para usuarios de carga pesada y volumen bajo de datos, mientras que los servidores de los otros niveles son más adecuados para procesar los volúmenes pesados de datos, pero cargas más livianas de usuario. (Ver figura).



## Estructura Multidimensional

En contraste con las bases de datos relacionales, las multidimensionales están compuestas por dimensiones que son atributos estructurales de un cubo, organizadas con jerarquías de categorías que describen los datos en tablas. Por ejemplo se tienen las ventas de cada producto por región. Una compañía tiene tres productos (arandelas, tornillos, tuercas) que se venden en tres territorios (Este, Oeste, Central).

Producto	Region #	Ventas
Arandelas	Este	50000
Arandelas	Oeste	60000
Arandelas	Central	100000
Tornillos	Este	40000
Tornillos	Oeste	70000
Tornillos	Central	80000
Tuercas	Este	90000
Tuercas	Oeste	120000
Tuercas	Central	30000

	Este	Oeste	Central
Arandelas	50000	60000	100000
Tornillos	40000	70000	80000
Tuercas	90000	120000	30000

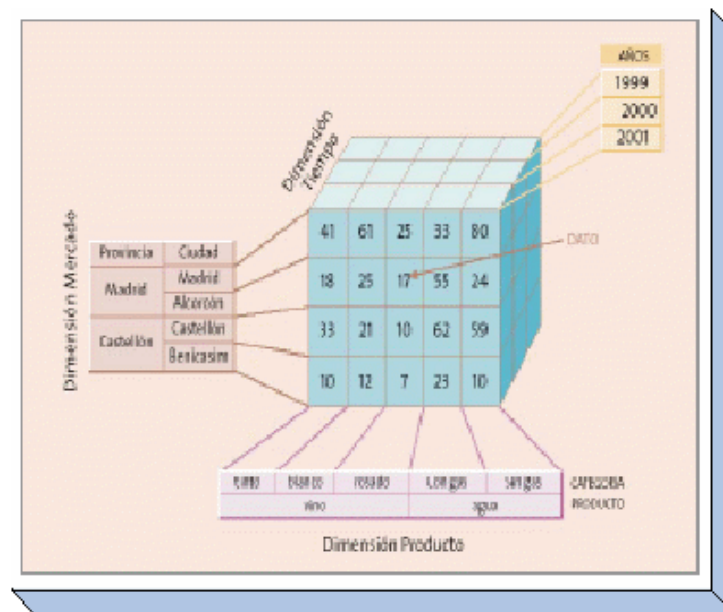
Un camino para representar esta tabla en una forma más óptima es a través de una



matriz de dos dimensiones. De esta forma se pueden realizar preguntas como ¿Cuáles fueron las ventas de arandelas en el Este?, ¿Cuáles fueron las ventas de Tornillos en el Oeste?.

Representa los datos como matrices N-Dimensionales denominadas Hipercubos.

- Las dimensiones definen dominios como geografía, producto, tiempo, cliente...
- Los miembros de una dimensión se agrupan de forma jerárquica ( dimensión geográfica: ciudad, provincia, autonomía, país ... ).
- Cada celda contiene datos agregados que relacionan los elementos de las dimensiones.



## Modelamiento Multidimensional<sup>5</sup>

La tecnología Datawarehousing debido a su orientación analítica, impone un procesamiento y pensamiento distinto, la cual se sustenta por un modelamiento de Bases de Datos propio, conocido como Modelamiento Multidimensional, el cual busca ofrecer al usuario su visión respecto de la operación del negocio.

Modelamiento Dimensional es una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre las distintas componentes del modelo.

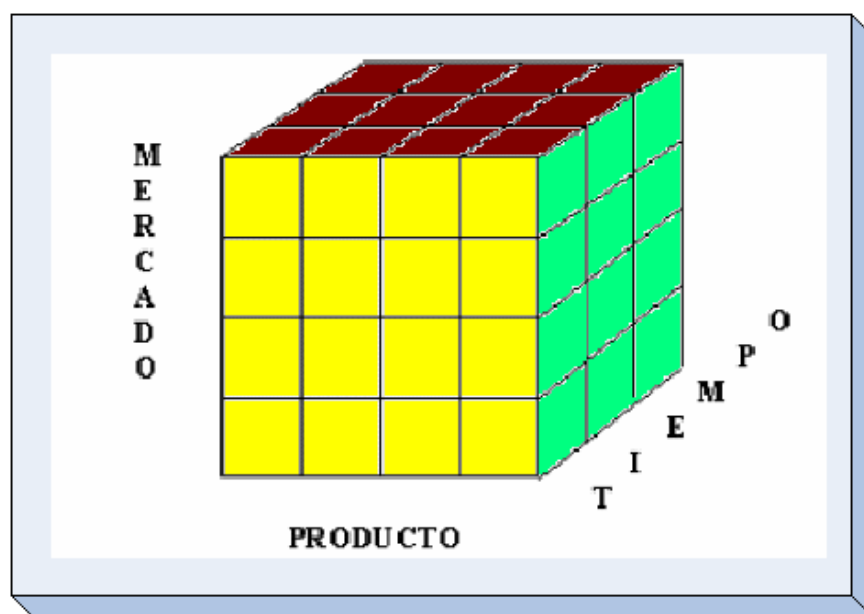
Consideremos un punto en el espacio. El espacio se define a través de sus ejes coordenados (por ejemplo X, Y, Z). Un punto cualquiera de este espacio quedará determinado por la intersección de tres valores particulares de sus ejes. Si se le asignan valores particulares a estos ejes. Digamos que el eje X representa Productos, el eje Y representa el Mercado y, el eje Z corresponde al Tiempo. Se podría tener por ejemplo, la siguiente combinación: producto = madera, mercado = Concepción, tiempo = diciembre-1998. La intersección de estos valores nos definirá un solo punto en nuestro espacio. Si el punto que buscamos, lo definimos como la cantidad de madera vendida, entonces se tendrá un valor específico y único para tal combinación.

En el modelo multidimensional cada eje corresponde a una dimensión particular. Entonces la dimensionalidad de nuestra base estará dada por la cantidad de ejes (o dimensiones) que le asociemos.

Cuando una base puede ser visualizada como un cubo de tres o más dimensiones, es más fácil para el usuario organizar la información e imaginarse en ella cortando y rebanando el cubo a través de cada una de sus dimensiones, para buscar la información deseada.

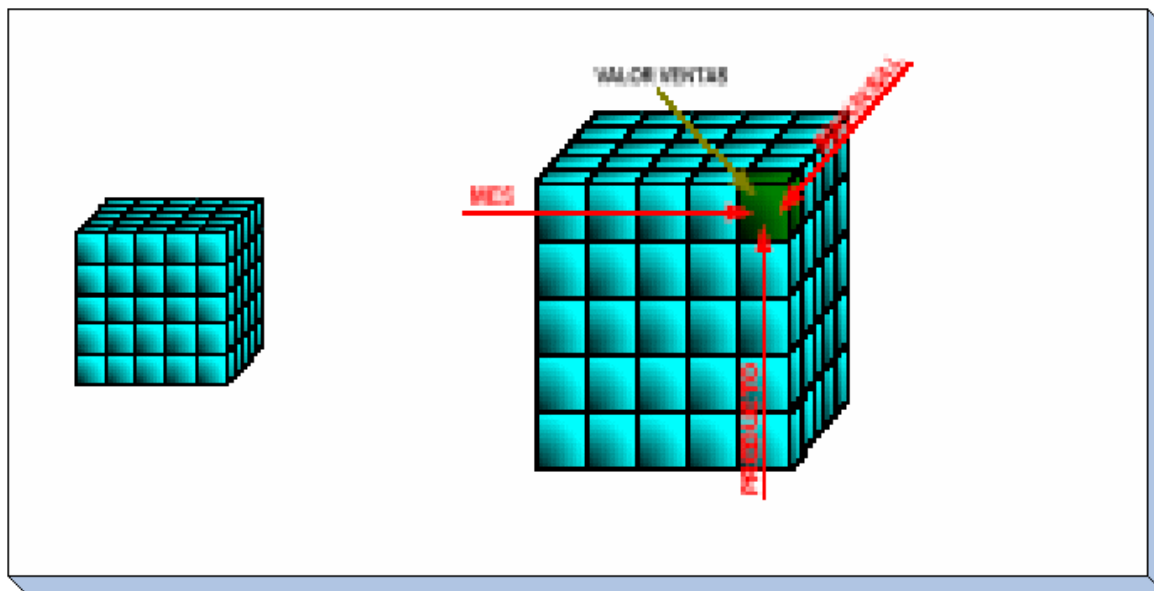
Para entender más el concepto, retomemos el ejemplo anterior. La descripción de una organización típica es: “Nosotros vendemos productos en varios mercados, y medimos nuestro desempeño en el tiempo”: Un diseñador dimensional lo verá como: “Nosotros vendemos productos en varios mercados, y medimos nuestro desempeño en el tiempo. Donde cada palabra subrayada corresponde a una dimensión.

Esto puede visualizarse como un cubo (Figura ), donde cada punto dentro del cubo es una intersección de coordenadas definidas por los lados de éste (dimensiones). Ejemplos de medidas son: unidades producidas, unidades vendidas, costo de unidades producidas, ganancias(\$) de unidades vendidas, etc.



- El punto de intersección de las tres dimensiones representa un valor particular denominado hecho.
- Cada una de las dimensiones son identificadores para ese hecho particular.
- Mientras mayor sea el número de dimensiones vinculadas a un hecho, mayor es el nivel de detalle de dicho hecho.

La siguiente figura representa la acción de medir el valor de las ventas mensuales de un producto en una sucursal.



## Modelos de Datos

Un factor clave presente durante todo el diseño de un DW, fue expresado por Codd en 1983: “Ustedes pueden pensar que el significado de los datos es simple.. pero no es así”.

Para construir un DW se debe primero tener claro que existe una diferencia entre la estructura de la información y la semántica de la información, y que esta última es mucho más difícil de abarcar y que también es precisamente con ella con la que se trabaja en la construcción de un DW.

Aquí se encuentra la principal diferencia entre los sistemas operacionales y el DW: Cada uno de ellos es sostenido por un modelo de datos diferente. Los sistemas operacionales se sustentan en el Modelo Entidad Relación (MER) y DDW trabaja con el Modelo Multidimensional.

## Características del MER

Tiene las siguientes características:

- Maneja la redundancia fuera de los datos. Por lo tanto realizar un cambio en la base significa tocarla en un solo lugar.
- Divide los datos en entidades, las que son representadas como tablas en una base de datos.
- Los MER crecen fácilmente, haciéndose más y más complejos.
- Se puede apreciar la existencia de muchos caminos para ir de una tabla a otra. Sería natural pensar que al tener diversos caminos para llegar desde una tabla a otra, cualquiera de ellos entregaría el mismo resultado, pero lamentablemente esto no siempre sucede así.

- El diagrama se visualiza simétrico, donde todas las tablas se parecen, sin distinguir a priori la importancia de unas respecto a otras. No es fácil de entender tanto para usuarios como para los diseñadores.

## Características del Modelo Multidimensional

En general, la estructura básica de un DW para el Modelo Multidimensional está definida por dos elementos: esquemas y tablas.

- **Tablas DW:** como cualquier base de datos relacional, un DW se compone de tablas. Hay dos tipos básicos de tablas en el Modelo Multidimensional:
- **Tablas Fact :** contienen los valores de las medidas de negocios, por ejemplo: ventas promedio en dólares, número de unidades vendidas, etc.
- **Tablas Lock\_up:** contienen el detalle de los valores que se encuentran asociados a la tabla Fact.
- **Esquemas DW:** la colección de tablas en el DW se conoce como Esquema. Los esquemas caen dentro de dos categorías básicas: esquemas estrellas y esquemas snowflake.

## Conceptos asociados al Datawarehouse

### Tabla de Hechos (Tablas Fac)

Es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. Estas medidas se hacen sobre el grano, o unidad básica de la tabla.

El grano o la granularidad de la tabla queda determinada por el nivel de detalle que se almacenará en la tabla. Por ejemplo, para el caso de producto, mercado y tiempo antes visto, el grano puede ser la cantidad de madera vendida “mensualmente”. El grano revierte las unidades atómicas en el esquema dimensional.

Cada medida es tomada de la intersección de las dimensiones que la definen. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. La razón de estas características es que así se facilita que los miles de registros que involucran una consulta sean comprimidos en unas pocas líneas en un set de respuesta.

La clave de la tabla fact recibe el nombre de clave compuesta o concatenada debido a que se forma de la composición (o concatenación) de las llaves primarias de las tablas dimensionales a las que está unida.

Así entonces, se distinguen dos tipos de columnas en una tabla fact: columnas fact y columnas key.

Donde la columna fact es la que almacena alguna medida de negocio y una columna

key forma parte de la clave compuesta de la tabla.

## **Tablas Dimensionales (Tablas Lock-up)**

Estas tablas son las que se conectan a la tabla fact, son las que alimentan a la tabla fact. Una tabla lock\_up almacena un conjunto de valores que están relacionados a una dimensión particular.

Tablas lock\_up no contienen hechos, en su lugar los valores en las tablas lock\_up son los elementos que determinan la estructura de las dimensiones. Así entonces, en ellas existe el detalle de los valores de la dimensión respectiva.

Una tabla lock\_up está compuesta de una primary key que identifica unívocamente una fila en la tabla junto con un conjunto de atributos, y dependiendo del diseño del modelo multidimensional puede existir una foreign key que determina su relación con otra tabla lock\_up.

Para decidir si un campo de datos es un atributo o un hecho se analiza la variación de la medida a través del tiempo. Si varía continuamente implicaría tomarlo como un hecho, caso contrario será un atributo.

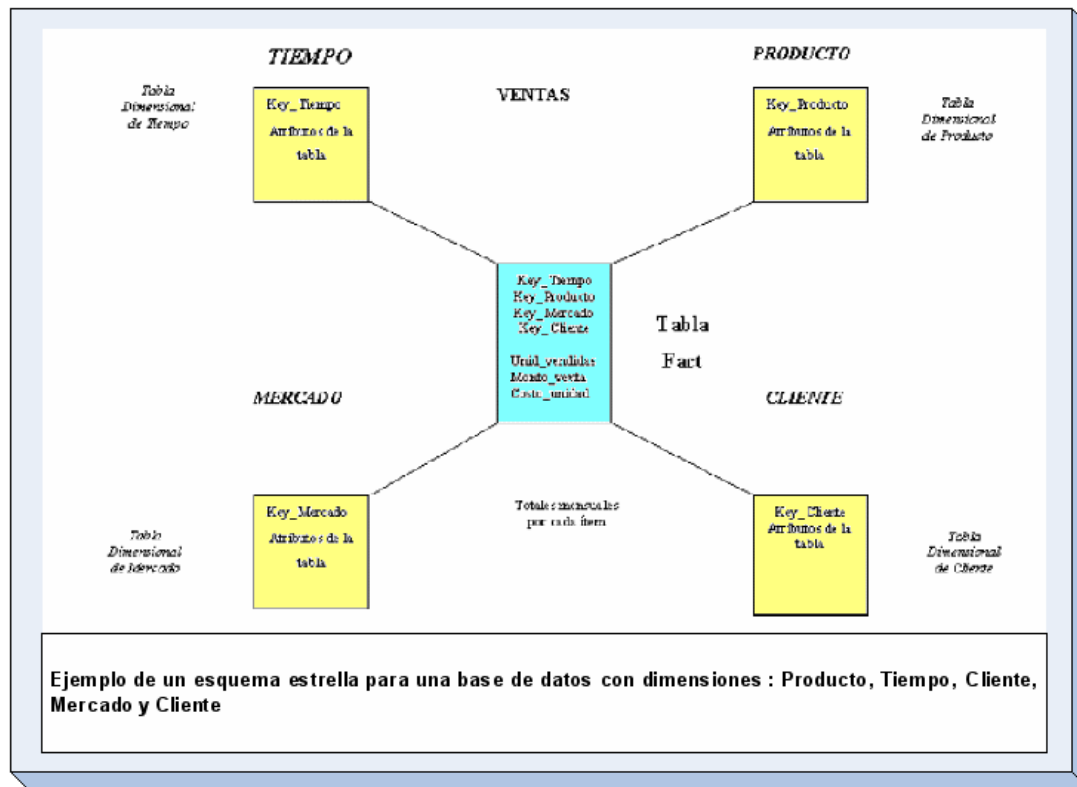
Los atributos dimensionales son un rol determinante en un DDW. Ellos son la fuente de todas las necesidades que debieran cubrirse. Esto significa que la base de datos será tan buena como lo sean los atributos dimensionales, mientras más descriptivos, manejables y de buena calidad, mejor será el DDW.

## **Esquema Estrella**

En general, el modelo multidimensional también se conoce con el nombre de esquema estrella, pues su estructura base es similar: una tabla central y un conjunto de tablas que la atienden radialmente. (ver Figura ).

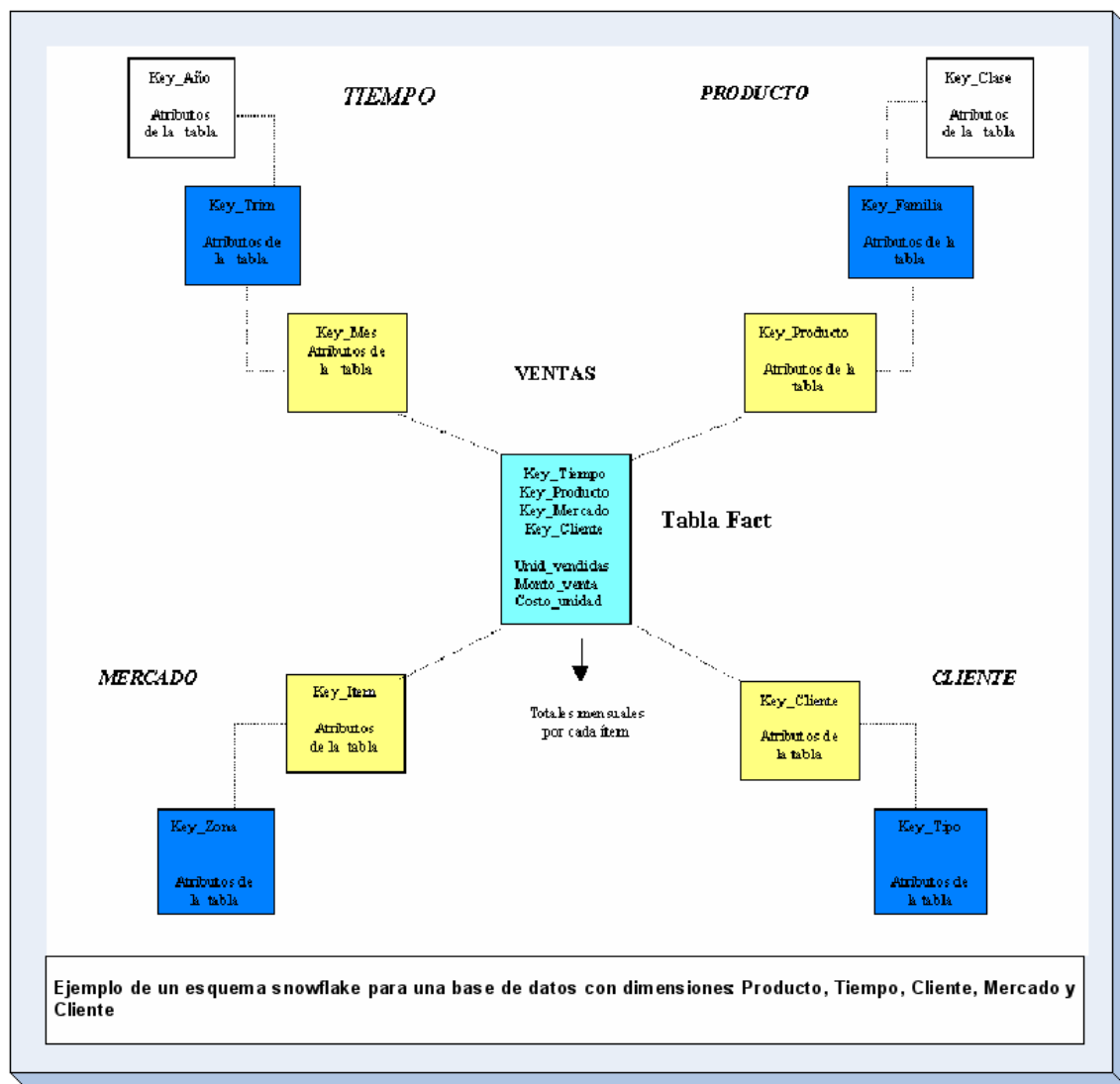
El esquema estrella deriva su nombre del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. El centro de la estrella consiste de una o más tablas fact, y las puntas de la estrella son las tablas lock\_up.

Este modelo entonces, resulta ser asimétrico, pues hay una tabla dominante en el centro con varias conexiones a las otras tablas. Las tablas Lock-up tienen sólo la conexión a la tabla fact y ninguna más.



## Esquema snowflake (Copos de Nieve)

La diferencia del esquema snowflake comparado con el esquema estrella, está en la estructura de las tablas lock\_up: las tablas lock\_up en el esquema snowflake están normalizadas. Cada tabla lock\_up contiene sólo el nivel que es clave primaria en la tabla y la foreign key de su parentesco del nivel más cercano del diagrama.



## Pasos básicos del Modelamiento Dimensional

1. Decidir cuáles serán los procesos de negocios a modelar, basándose en el conocimiento de éstos y de los datos disponibles. Ejemplo: Gastos realizados por cada mercado para cada ítem a nivel mensual. Productos vendidos por cada mercado según el precio en cada mes.
2. Decidir el Grano de la tabla Fact de cada proceso de negocio. Ejemplo: Producto x mercado x tiempo. En este punto se debe tener especial cuidado con la magnitud de la base de datos, con la información que se tiene y con las preguntas que se quiere responder. El grano decidirá las dimensiones del DDW. Cada dimensión debe tener el grano más pequeño que se pueda puesto que las preguntas que se realicen necesitan cortar la base en caminos precisos (aunque las preguntas no lo pidan explícitamente).
3. Decidir las dimensiones a través del grano. Las dimensiones presentes en la mayoría de los DDW son: tiempo, mercado, producto, cliente. Un grano bien elegido determina la dimensionalidad primaria de la tabla fact. Es posible usualmente agregar dimensiones adicionales al grano básico de la tabla fact, donde estas dimensiones adicionales toman un solo valor para cada combinación de las dimensiones primarias. Si se reconoce que una dimensión adicional deseada viola el grano por causar registros adicionales a los generados, entonces el grano debe ser revisado para acomodar esta dimensión adicional.
4. Elegir las mediciones del negocio para la tabla fact. Se deben establecer los ítems que quedarán determinados por la clave compuesta de la tabla Fac.

## Profundizaciones de Diseño

Se consideran los siguientes aspectos:

### Dimensión Tiempo

Virtualmente se garantiza que cada DDW tendrá una tabla dimensional de tiempo, debido a la perspectiva de almacenamiento histórica de la información. Usualmente es la primera dimensión en definirse, con el objeto de establecer un orden, ya que la inserción de datos en la base de datos multidimensional se hace por intervalos de tiempo, lo cual asegura un orden implícito.

### Dimensiones que varían lentamente en el Tiempo

Son aquellas dimensiones que se mantienen “casi” constantes en el tiempo y que pueden preservar la estructura dimensional independiente del tiempo, con sólo agregados menores relativos para capturar la naturaleza cambiante del tiempo.

Cuando se encuentra una de estas dimensiones se está haciendo una de las siguientes fundamentales tres elecciones. Cada elección resulta en un diferente grado de seguimiento sobre el tiempo:

**Tipo 1:** Sobreescribir el viejo valor en el registro dimensional y por lo tanto perder la capacidad de seguir la vieja historia.

**Tipo 2:** Crear un registro dimensional adicional (con una nueva llave) que permita registrar el cambio presentado por el valor del atributo. De esta forma permanecerían en la base tanto el antiguo como el nuevo valor del registro con lo cual es posible segmentar la historia de la ocurrencia.

**Tipo 3:** Crear un campo “actual” nuevo en el registro dimensional original el cual almacene el valor del nuevo atributo, manteniendo el atributo original también. Cada vez que haya un nuevo cambio en el atributo, se modifica el campo “actual” solamente. No se mantiene un registro histórico de los cambios intermedios.

### Niveles

Un nivel representa un nivel particular de agregación dentro de una dimensión; cada nivel sobre el nivel base representa la sumarización total de los datos desde el nivel inferior. Para un mejor entendimiento, veamos el siguiente ejemplo: consideremos una dimensión Tiempo con tres niveles: Mes, Semestre, Año. El nivel Mes representa el nivel base, el nivel Semestre representa la sumarización de los totales por Mes y el nivel Año representa la sumarización de los totales para los Semestres.



Agregar niveles de sumariazación otorga flexibilidad adicional a usuarios finales de aplicaciones EIS/ DSS para analizar los datos.

## **Jerarquías**

A nivel de dimensiones es posible definir jerarquías, las cuales son grupos de atributos que siguen un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, con cada nivel representando el total agregado de los datos del nivel inferior. Las jerarquías definen cómo los datos son sumariazados desde los niveles más bajos hacia los más altos. Una dimensión típica soporta una o más jerarquías naturales. Una jerarquía puede pero no exige contener todos los valores existentes en la dimensión.

Se debe evitar caer en la tentación de convertir en tablas dimensionales separadas cada una de las relaciones muchos-a-uno presentes en las jerarquías. Esta descomposición es irrelevante en el planeamiento del espacio ocupado en disco y sólo dificulta el entendimiento de la estructura para el usuario final, además de destruir el desempeño del browsing.

## ***Explotación del datawarehouse***

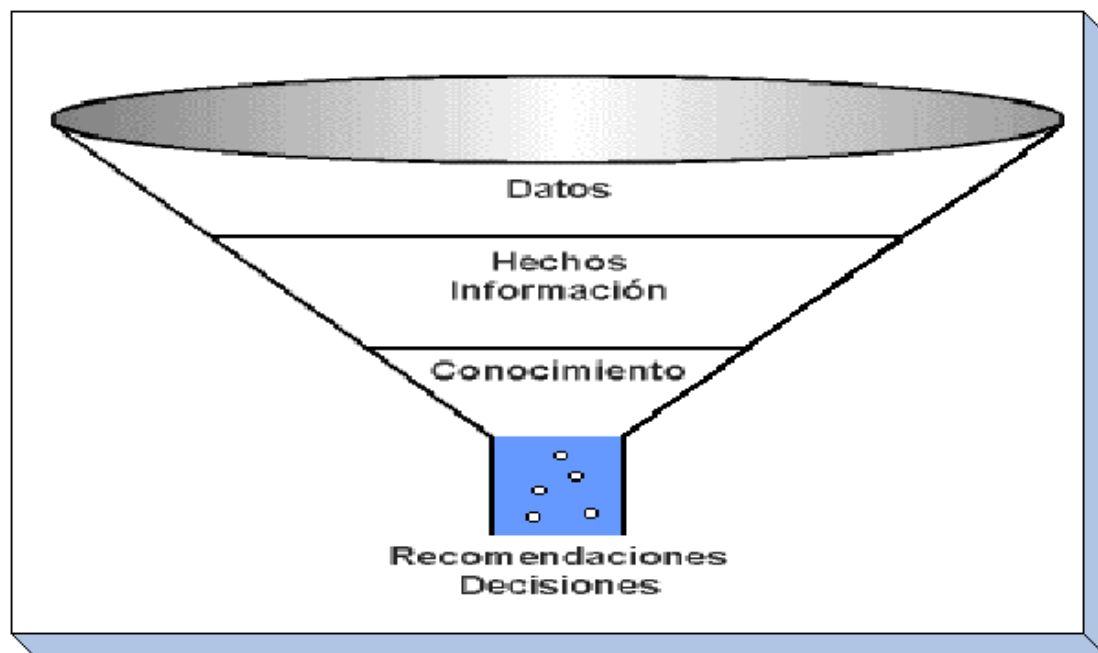
### **Potencial del datawarehouse**

El Data Warehouse no produce resultados en forma mágica. Los administradores de empresas y los analistas deben acceder y recuperar los datos del Data Warehouse y convertirlos en información y en hechos. Estos hechos conforman los cimientos de una base de conocimientos que sirve para determinar la salud de la empresa y la dirección futura del negocio. Como en las granjas, los usuarios sólo cosecharán la información que se pueda derivar de los datos que sembraron en el Data Warehouse, y sólo mediante el uso de las herramientas de cosecha adecuadas. Algunas de las herramientas de cosecha necesarias son: las de acceso y recuperación, las de reportes de base de datos, las de análisis y las de data mining.

Uno de los retos al cosechar un Data Warehouse consiste en no convertir montículos de información en montañas de datos. Es fácil caer en la trampa de "entre más, mejor". No es esencial conocer todos los hechos, sólo los cruciales. Como ejemplo, una campaña de ropa para niños necesita cosechar exacta y rentablemente sólo aquellas familias que tienen niños.

## **Extracción y manipulación de datos del dawarehouse**

Herramientas de soporte de decisiones es el término genérico para referirse a las aplicaciones y herramientas del Data Warehouse que se emplean para recuperar, manipular y analizar los datos, y para presentar después los resultados. Estas herramientas se usan en dos modalidades: verificación y descubrimiento. En la modalidad de verificación, el usuario empresarial crea una hipótesis -una cuestión empresarial- e intenta confirmarla accediendo a los datos en el Data Warehouse. Las herramientas que implementan la modalidad de verificación son de consulta, de sistemas de reporte y de análisis multidimensional. En la modalidad de descubrimiento, las herramientas intentan descubrir características en los datos como patrones de compra o la asociación entre la adquisición de artículos diferentes. En la modalidad de descubrimiento, o eureka, el usuario empresarial no conoce ni sospecha los patrones y asociaciones descubiertos. La herramienta de Data Mining es un ejemplo de la modalidad de descubrimiento.



Desde la perspectiva de disponibilidad de herramientas, las dos modalidades de verificación y descubrimiento se clasifican en tres enfoques: Procesamiento Informático, Procesamiento Analítico y Data Mining.

### Procesamiento Informático

La recuperación de la inversión en un Data Warehouse se basa en la capacidad de los usuarios empresariales para extraer los datos correctos del Data Warehouse, convertirlos en información y luego utilizar esa información para tomar mejores decisiones. Los usuarios empresariales pretenden extraer los datos correctos con una mínima inversión en tiempo y sin complicaciones.

A los ejecutivos y gerentes de alto nivel les interesa ver los resultados de una actividad de procesamiento informático en forma de reportes, cuadros y gráficas. Los gerentes desean acceder a consultas estándar de rutina. Un ambiente de 'consulta administrativa' cubre la mayoría de sus necesidades. El procesamiento informático apoya a la modalidad de verificación del soporte de decisiones. Comprende técnicas como análisis estadísticos básicos y de datos, consultas y reportes. Los datos que se acceden y procesan pudieran ser históricos o bastante recientes, y pudieran estar un poco o muy resumidos. Los resultados se presentan en forma de reportes o gráficas.

El procesamiento informático asiste a los usuarios empresariales en la búsqueda de respuestas a cuestiones empresariales, como las siguientes:

- ¿Cuáles fueron los ingresos por ventas en el fin de semana del Día de Acción de Gracias (nuestro mejor fin de semana de ventas) para todas las tiendas del medio oeste, con corte por departamento?
- ¿Cuáles fueron los diez artículos más rentables durante la venta posterior a la Navidad? ¿Cuáles fueron los diez menos rentables?
- ¿Cómo se comparan las ventas del Día de Acción de Gracias con las del mismo fin de semana, en los últimos cinco años, por departamento y tienda?

### Procesamiento Analítico

También el procesamiento analítico apoya a la modalidad de verificación del soporte de decisiones. Su meta consiste en hacer que los datos estén disponibles para el usuario

de la empresa en su perspectiva de las dimensiones empresariales. Se pueden responder e interpretar preguntas complejas como "¿Cuántos automóviles vendimos en Estados Unidos en el primer trimestre de 1995 que tuvieran un sistema de audio CD, con un precio de 25,000 dólares o menos?". Este procesamiento maneja capacidades de análisis de subconjuntos (*slice and dice*), profundización (*drill-down*) y condensación y adición (*roll-up*). Los datos que se emplean en el procesamiento analítico son, por lo general, históricos tanto a nivel de resumen como al de detalle.

Los gerentes y analistas empresariales requieren la funcionalidad del procesamiento analítico cuando deben responder preguntas complejas como las siguientes:

- ¿Cuántos esquís de nieve, fabricados por SpeedSkiDown, Inc., se vendieron a hombres en el mes de noviembre, en nuestras tiendas de las regiones del medio oeste, del noroeste y de la Montaña?
- ¿Cómo se compara lo programado con lo real del mismo mes en los dos últimos años?
- ¿Cuántas minivans azules teníamos en inventario (al fin del trimestre) con un reproductor de discos compactos y un tercer asiento, cuando la lista de precios era menor de \$19,995? Se requieren totales por condado para cada trimestre de los últimos cinco años, comparar lo real contra lo planeado, y comparar el inventario de cada trimestre con el del anterior y el del siguiente?

Los gerentes ejecutivos saben que "el futuro pertenece a quienes pueden verlo y llegar ahí primero". Por tal razón, los ejecutivos y gerentes empresariales no sólo comprenden "lo que pasa en el negocio", sino también "que va a suceder". El procesamiento analítico se utiliza tanto para análisis históricos complejos, con una extensa manipulación, como para la planeación a futuro y pronóstico -el pasado como prólogo del futuro.

Los datos empresariales son, de hecho, multidimensionales. Se encuentran relacionados y regularmente son jerárquicos; por ejemplo, los datos de ventas, los datos del inventario y los pronósticos de presupuestos están interrelacionados y dependen entre sí. En la práctica, para predecir las ventas de un nuevo producto específico, se requiere analizar los patrones de compras anteriores, la adopción de nuevos productos, las preferencias regionales y otros factores empresariales similares. La proyección de ventas para nuestra cuestión de los "esquís de SpeedSkiDown, Inc.", requiere comprender los patrones de ventas de los últimos años.

#### **Análisis Multidimensional**

Tanto para la eficiencia operativa como para la planeación a futuro, se deben analizar muchos datos empresariales interrelacionados. Esta necesidad empresarial se aborda mediante el procesamiento analítico. En éste, el enfoque está en el análisis de los datos, específicamente en el análisis multidimensional.

En el análisis multidimensional, los datos se representan mediante dimensiones como producto, territorio y cliente. Por lo regular las dimensiones se relacionan en jerarquías, por ejemplo, ciudad, estado, región, país y continente; o estado, territorio y región. El tiempo es también una dimensión estándar con su propia jerarquía como: día, semana, mes, trimestre y año; o día y año calendario.

#### **Procesamiento Analítico en Línea (OLAP)**

En un Data Warehouse se depositan datos para consulta, análisis y divulgación, a diferencia del procesamiento de transacciones en línea (OLTP por las siglas de *On-Line Transaction Processing*), en donde los datos se reúnen y almacenan para operación y control. OLAP es una tecnología de procesamiento analítica que crea nueva información empresarial a partir de los datos existentes, por medio de un rico conjunto de transformaciones empresariales y cálculos numéricos.

El procesamiento analítico en línea es una tecnología de análisis de datos que hace lo siguiente:

• Presenta una visión multidimensional lógica de los datos en el Data Warehouse. La visión es independiente de cómo se almacenan los datos.

• Comprende siempre la consulta interactiva y el análisis de los datos. Por lo regular la interacción es de varias pasadas, lo cual incluye la

profundización en niveles cada vez más detallados o el ascenso a niveles superiores de resumen y adición.

â€¢ Ofrece opciones de modelado analítico, incluyendo un motor de cálculo para obtener proporciones, desviaciones, etc., que comprende mediciones de datos numéricos a través de muchas dimensiones.

â€¢ Crea resúmenes y adiciones (también conocidas como consolidaciones), jerarquías, y cuestiona todos los niveles de adición y resumen en cada intersección de las dimensiones.

â€¢ Maneja modelos funcionales de pronóstico, análisis de tendencias y análisis estadísticos.

â€¢ Recupera y exhibe datos tabulares en dos o tres dimensiones, cuadros y gráficas, con un pivoteo fácil de los ejes. El pivoteo es fundamental ya que los usuarios empresariales necesitan analizar los datos desde perspectivas diferentes; y el análisis desde una perspectiva conduce a otra cuestión empresarial que se va a examinar desde otra perspectiva.

â€¢ Responde con rapidez a las consultas, de modo que el proceso de análisis no se interrumpe y la información no se desactualiza.

â€¢ Tiene un motor de depósito de datos multidimensional, que almacena los datos en arreglos. Estos arreglos son una representación lógica de las dimensiones empresariales.

La tecnología OLAP se aplica en muchas áreas funcionales de una empresa, tales como: Producción, ventas y análisis de rentabilidad de la comercialización; mezcla de manufacturas y análisis de logística; consolidaciones financieras, presupuestos y pronósticos, planeación de impuestos y contabilidad de costos.

## Data Mining

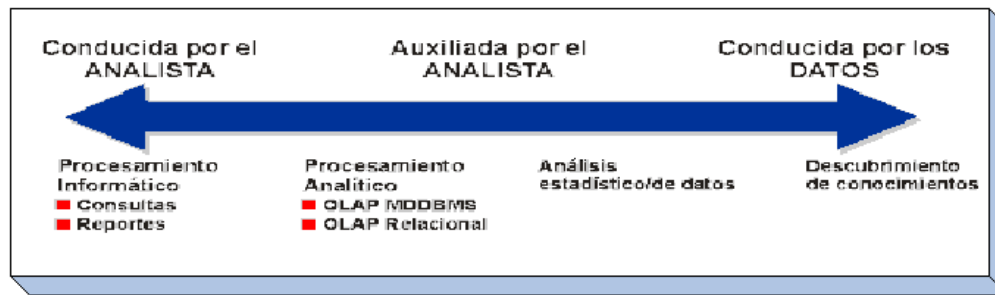
El Data Mining apoya la modalidad de descubrimiento del soporte de decisiones. Las herramientas de Data Mining recorren los datos detallados de transacciones para desenterrar patrones y asociaciones ocultos. Por lo regular los resultados generan extensos reportes o se les analiza con herramientas de visualización de datos descubiertos.

El procesamiento informático es excelente y rentable para el despliegue masivo de consultas, análisis y reportes de datos de dos o tres dimensiones. Las herramientas de procesamiento analítico permiten diversas visualizaciones de los datos, como ventas por marca, tienda, temporada y periodos de tiempo, las cuales se pueden definir, consultar y analizar. Las herramientas de Data Mining son esenciales para comprender el comportamiento de los clientes.

### Usuarios del Data Mining

Los usuarios clave en perspectiva del Data Mining son los analistas empresariales, los peritos en estadística y los profesionales en tecnología de la información que auxilian a los usuarios empresariales. Quienes obtienen beneficios de los resultados del Data Mining son los gerentes empresariales y los ejecutivos, que desean entender los factores de éxito del negocio con base en datos *completos* del cliente y, utilizar luego, este conocimiento para afinar las estrategias de producción, precios y comercialización; mejorar el nivel de éxito de las estrategias; e impulsar el balance.

Hasta la fecha, las empresas han dependido del procesamiento informático y analítico para medir y comprender la estabilidad de un negocio. El procesamiento informático -consultas y reportes- es más sencillo de usar, pero requiere de una estrecha dirección del analista (ver figura). Los analistas preguntan cuestiones específicas y verifican las cuestiones o hipótesis con los datos. Para este fin, los datos deben estar bien organizados. El procesamiento analítico (OLAP) requiere de menos dirección del analista, aunque los datos deben estar organizados en una forma especial (base de datos multidimensional), o accederse bien de manera especial (visión multidimensional). En ocasiones se utiliza una combinación de técnicas de consulta y OLAP para comprender el comportamiento del cliente o para construir perfiles de segmentos de mercado; pero el proceso de aplicar estas técnicas es conducido esencialmente por el analista empresarial. En estos casos, este proceso también se conoce como *Data Mining* y se define como la modalidad de descubrimiento del soporte de decisiones, la cual es conducida por los datos y no por el analista empresarial.



## ***Glosario de Términos asociados con Datawarehouse***

### **Agregación :**

Actividad de combinar datos desde múltiples tablas para formar una unidad de información más compleja, necesitada frecuentemente para responder consultas del DataWarehouse en forma más rápida y fácil.

### **Data Mart (DM).**

Es un conjunto de datos que son estructurados de una forma que facilite su posterior análisis. Un data mart contiene información referente a un área en particular, con datos relevantes que provienen de las diferentes aplicaciones operacionales. Los data mart pueden ser de diversas bases de datos relacionales o de diversas bases de datos OLAP dependiendo del tipo de análisis que se quiera desarrollar.

### **Data mining de pronóstico.**

Es un data mining que produce un modelo para ser usado con nuevos datos para pronosticar un valor o predecir un probable resultado basado en patrones en la data histórica.

### **Data mining descriptivo.**

Es un tipo de data mining (minería de datos) que produce un modelo para describir patrones de los datos históricos y requiere interacción humana para determinar la significación y el significado de estos patrones.

### **Datos no depurados.**

Datos que son ambiguos o que carecen de validez por ser inexistentes, ausentes, incorrectos o duplicados.

### **Data Warehouse (DW)**

Es un almacén o repositorio para los datos. Muchos expertos definen el data warehouse como un almacén de datos centralizados que introduce datos en un almacén de datos específico llamado data mart. Otros aceptan una amplia definición de data warehouse, como un conjunto integrado de data marts.

### **Descendiente.**

Algún miembro de algún nivel inferior en relación a otro miembro específico.

#### **Desktop online analytical processing (DOLAP).**

Es un tipo de almacenamiento OLAP que mantiene los datos en una máquina cliente y suministra análisis multidimensional de forma local.

#### **Dimensión.**

Es una vista de datos categóricamente consistente. Todos los miembros de una dimensión pertenecen a un mismo grupo. Entidad independiente dentro del modelo multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como mecanismo de selección de datos.

#### **Drill Down :**

Exponer progresivamente más detalle (dentro de un reporte o consulta), mediante selecciones de ítems sucesivamente.

**Drill-Up :** Es el efecto contrario a drill -down. Significa ver menos nivel de detalle, sobre la jerarquía significa generalizar o sumarizar, es decir, subir en el árbol jerárquico.

#### **DSS :**

Sistema de Soporte de Decisiones. Sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la información histórica.

#### **ETT (Extracción, Transformación y Transporte de datos) :**

Pasos por los que atraviesan los datos para ir desde el sistema OLTP ( o la fuente de datos utilizada) a la bodega dimensional.

- **Extracción**, se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original.
- **Transformación** (también conocida como limpieza) es la etapa por la que puede atravesar una base de datos para estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos.
- **Transporte**, que consiste básicamente en llevar los datos leídos y estandarizados a la bodega dimensional (puede ser remota o localmente). Generalmente, para un Data Mart no es necesario atravesar por todos estos pasos, pues al ser información localizada, sus datos suelen estar naturalmente estandarizados (hay una sola fuente).

#### **Granularidad:**

- Indica el grado de detalle asociado a un hecho particular.
- El primer gran factor decisivo de la granularidad es el tiempo, ya que mientras menor sea el intervalo de tiempo, mayor será el grado de detalle obtenido.
- La granularidad depende directamente del número de dimensiones que se asocian a la tabla de hechos.

- Se deben considerar otros factores como la carga del procesador, espacio de almacenamiento y satisfacer a cabalidad los requerimientos del cliente.

**Hechos (Medidas):**

Son medidas numéricas, las cuales describen los resultados (rendimiento) del negocio.

Un hecho debe ser un valor cuantificable ó medible. Por ejemplo el número de unidades vendidas, ingreso por ventas, etc.

**Jerarquía :**

Es un conjunto de atributos descriptivos que permite que a medida que se tenga una relación de muchos a uno se ascienda en la jerarquía. Por ejemplo : los Centros de Responsabilidad están asociados a un Tipo de Unidad, el cual pueden corresponder a una gerencia, subgerencia, superintendencia, etc.; por otro parte, cada CR está asociado a otro CR a nivel administrativo y, también existe una clasificación a nivel funcional.

**Olap (On-line Analytical Processing) :**

Conjunto de principios que proveen una ambiente de trabajo dimensional para soporte decisional.

**Oltp (On-line Transaction Processing) :**

Sistema transaccional diario (o en detalle) que mantiene los datos operacionales del negocio.

**Rollup:**

Comando propio del lenguaje Oracle Express, que simboliza las sumas agregadas de una variable a través de los niveles jerárquicos de las dimensiones que la sustentan.

**Snapshot:**

Imagen instantánea de los datos en un tiempo dado.

**Sumarización :**

Actividad de incremento de la granularidad de la información en una base de datos. La

sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.

**Tabla Dimensional:**

Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad a esta estrategia.

***Bibliografía***

1. BFSsystem Definición de Datawarehouse
2. Datawarehouse, Monografías.com, Damián Gutiérrez
3. Apuntes de Datawarehouse, Presentaciones Ing, Victor Aguilar, Escuela Politécnica Nacional Ecuador
4. Modelamiento Dimensional, Carmen Gloria Wolf
5. IBM DB2 OLAP Server, Maria Ibarra, Documentación Técnica.
6. Diseño y Optimización de Bases de Datos y Datawarehouse, Universidad Politécnica de Madrid, Dr. Eusebio Santos
7. Datawarehousing Fácil, Programación en Castellano, Claudio Casares.
8. DataMart Paso a Paso, <http://www.ruedatecnologica.com>
9. TODO BI, Características de Pentaho OPEN SOURCE
10. Manual para la Construcción de un Datawarehouse, <http://www.inei.gob.pe/biblioineipub/bancopub/Inf/Lib5084/3-1.HTM>
11. Business Intelligence, Transformando la Información en Decisiones, Octavio Licea

[1](#) Referencia [11] de la Bibliografía.

[2](#) Imagen perteneciente al sitio de Rueda Tecnológica. Referencia [8] de la Bibliografía

[3](#) Referencia 7 de Bibliografía, Datawarehousing Fácil.

[4](#) Información e imágenes tomadas del sitio de TODO BI.

[5](#) Sección basada en su mayor parte de la referencia [4] de la bibliografía: Modelamiento Dimensional, Carmen Wolf

---

**Ing Cristhian Herrera 64**



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 2.5](http://creativecommons.org/licenses/by-nc-nd/2.5/)



[License.](#)  
[Puedes opinar sobre este tutorial aquí](#)

## Recuerda

que el personal de [Autentia](#) te regala la mayoría del conocimiento aquí compartido ([Ver todos los tutoriales](#))

¿Nos vas a tener en cuenta cuando necesites consultoría o formación en tu empresa?

**¿Vas a ser tan generoso con nosotros como lo tratamos de ser con vosotros?**

[info@autentia.com](mailto:info@autentia.com)

Somos pocos, somos buenos, estamos motivados y nos gusta lo que hacemos .....  
**Autentia = Soporte a Desarrollo & Formación**

[Autentia S.L.](#) Somos expertos en:



**J2EE, Struts, JSF, C++, OOP, UML, UP, Patrones de diseño ..**  
y muchas otras cosas

## Nuevo servicio de notificaciones

Si deseas que te enviemos un correo electrónico cuando introduzcamos nuevos tutoriales, inserta tu dirección de correo en el siguiente formulario.

<b>Subscribirse a Novedades</b>	
<b>e-mail</b>	<input type="text"/>
	<input type="button" value="Enviar"/>

## Otros Tutoriales Recomendados ( [También ver todos](#) )

Nombre Corto	Descripción
<a href="#">Administración Web de MySQL</a>	En este tutorial se mostrará como administrar MySQL de forma rápida y muy sencilla a través de páginas webs implementadas con tecnología PHP, para ello se utilizará la herramienta PHPmyAdmin
<a href="#">Administracion Web de MySQL</a>	Os mostramos como instalar en vuestro PC phpMySQL, un potente gestor Web de MySQL utilizado en la mayoría de los Hostings.
<a href="#">Modelado de MySQL con herramientas gratuitas</a>	Os mostramos otra alternativa de modelado gráfico de MySql.
<a href="#">Apache, MySQL y PHP</a>	Os mostramos como configurar Apache, MySQL y PHP en vuestra máquina
<a href="#">Pool de conexiones a BBDD con struts</a>	Os mostramos como configurar un pool de conexiones a base de datos en vuestras aplicaciones construidas con struts
<a href="#">Modelado Gráfico de MySQL</a>	Os mostramos como instalar y utilizar DeZing e Importer de Datanamic para crear el modelo lógico y físico de bustra base de datos MySQL
<a href="#">Todo lo que querias saber sobre DatawareHouse (III)</a>	En este documento Christian nos enseña exhaustivamente qué es y como funciona un datawarehouse.
<a href="#">Todo lo que querias saber sobre DatawareHouse (II)</a>	En este documento Christian nos enseña exhaustivamente qué es y como funciona un datawarehouse.
<a href="#">Todo lo que querias saber sobre DatawareHouse (I)</a>	En este documento Christian nos enseña exhaustivamente qué es y como funciona un datawarehouse.

Nota: Los tutoriales mostrados en este Web tienen como objetivo la difusión del conocimiento.

Los contenidos y comentarios de los tutoriales son responsabilidad de sus respectivos autores.

En algún caso se puede hacer referencia a marcas o nombres cuya propiedad y derechos es de sus respectivos dueños. Si algún afectado desea que incorporemos alguna reseña específica, no tiene más que solicitarlo.

Si alguien encuentra algún problema con la información publicada en este Web, rogamos que informe al administrador [rcanales@adictosaltrabajo.com](mailto:rcanales@adictosaltrabajo.com) para su resolución.

[Patrocinados por enredados.com .... Hosting en Castellano con soporte Java/J2EE](#)

