

¿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>Autor del tutorial: Cristhian Kirs Herrera Basurto</p> <p>§ Lugar de residencia: 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>Cristhian.Herrera@gmail.com / cherrera@kruger.com.ec</p>	<p>NUEVO CATÁLOGO DE SERVICIOS DE AUTENTIA (PDF 6,2MB)</p> <p>www.adictosaltrabajo.com es el Web de difusión de conocimiento de www.autentia.com</p>  <p>autentia real business solutions</p> <p>Catálogo de cursos</p>
--	--

Descargar este documento en formato PDF [datawarehouse.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

Java j2ee

Curso Programador Java Plataforma J2EE. Gratuito Desempleados. Madrid cursos.maestroteca.com/java

Formación Oficial SUN

Formación y Certificación en Java y Solaris. Los líderes en Barcelona www.netmind.es

Programadores Java

Empleo para Informáticos Ofertas por Tecnología www.informaticos.com

Anuncios Google

Fecha de creación del tutorial: 2007-10-30

Apuntes Datawarehouse

D A T A W A R E H O U S E 3

[Sobre el presente documento 3](#)

[Definición 3](#)

[Objetivos del Datawarehouse 4](#)

[Características del Datawarehouse 4](#)

[Orientado a temas 5](#)

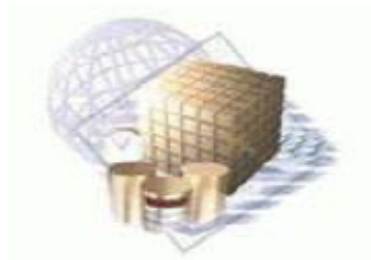
[Integrado 6](#)

[De tiempo variante \(Variable en el tiempo\) 8](#)

[No volátil 10](#)

[Por qué construir un Datawarehouse? 11](#)

D A T A W A R E H O U S E



Sobre el presente documento

El presente documento es una humilde recopilación que concentra los mejores documentos que he podido encontrar en internet y en español con referencia a la temática de datawarehouse, sumados mi propia experiencia y a la formación que en su momento recibí en las aulas universitarias, esperando que este esfuerzo de dar un sentido lógico y coherente a la información que he logrado reunir sea de utilidad para quienes se inicien en la temática.

Definición

De acuerdo con W. H. Inmon, quien es considerado como el padre del Data Warehouse: "Un Data Warehouse es un conjunto de datos integrados orientados a una material que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración."

De acuerdo con algunas organizaciones, el Data Warehouse es una arquitectura. Para otras, es un depósito semánticamente consistente en datos (separados y que no interfieren con los sistemas operativos y de producción existentes) que llenan por completo los diferentes requerimientos de acceso y reporte de datos. Para algunos otros, el Data Warehouse es un proceso continuo que mezcla los datos de varias fuentes heterogéneas, incluyendo datos históricos y adquiridos para soportar la constante necesidad de consultas estructuradas y/o ad hoc, reportes analíticos y soporte de decisiones.

Así como hay gran divergencia para establecer una definición precisa de un Data Warehouse, hay un claro consenso de que la tecnología del Data Warehouse es un ingrediente esencial en el conjunto de soluciones para el soporte de decisiones en una empresa.

Objetivos del Datawarehouse

- Hace que la información de la organización sea accesible: los contenidos del Data Warehouse son entendibles y navegables, y el acceso a ellos son caracterizado por el rápido desempeño. Estos requerimientos no tienen fronteras y tampoco límites fijos. Cuando hablamos de entendible significa, que los niveles de la información sean correctos y obvios. Y Navegables significa el reconocer el destino en la pantalla y llegar a donde queramos con solo un clic. Rápido desempeño significa, cero tiempo de espera. Todo lo demás es un compromiso y por consiguiente algo que queremos mejorar.
- Hacer que la información de la organización sea consistente: la información de una parte de la organización puede hacerse coincidir con la información de la otra parte de la organización. Si dos medidas de la organización tienen el mismo nombre, entonces deben significar la misma cosa. Y a la inversa, si dos medidas no significan la misma cosa, entonces son etiquetados diferentes. Información consistente significa, información de alta calidad. Significa que toda la información es contabilizada y completada. Todo lo demás es un compromiso y por consiguiente algo que queremos mejorar.
- Es información adaptable y elástica: el Data Warehouse esta diseñado para

cambios continuos. Cuando se le hacen nuevas preguntas al Data Warehouse, los datos existentes y las tecnologías no cambian ni se corrompen. Cuando se agregan datos nuevos al Data Warehouse, los datos existentes y las tecnologías tampoco cambian ni se corrompen. El diseño de Data Marts separados que hacen al Data Warehouse, deben ser distribuidos e incrementados. Todo lo demás es un compromiso y por consiguiente algo que queremos mejorar.

- Es un seguro baluarte que protege los valores de la información: el Data Warehouse no solamente controla el acceso efectivo a los datos, si no que da a los dueños de la información gran visibilidad en el uso y abusos de los datos, aún después de haber dejado el Data Warehouse. Todo lo demás es un compromiso y por consiguiente algo que queremos mejorar.
- Es la fundación de la toma de decisiones: el Data Warehouse tiene los datos correctos para soportar la toma de decisiones. Solo hay una salida verdadera del Data Warehouse: las decisiones que son hechas después de que el Data Warehouse haya presentado las evidencias. La original etiqueta que preside el Data Warehouse sigue siendo la mejor descripción de lo que queremos construir: un sistema de soporte a las decisiones.

Características del Datawarehouse

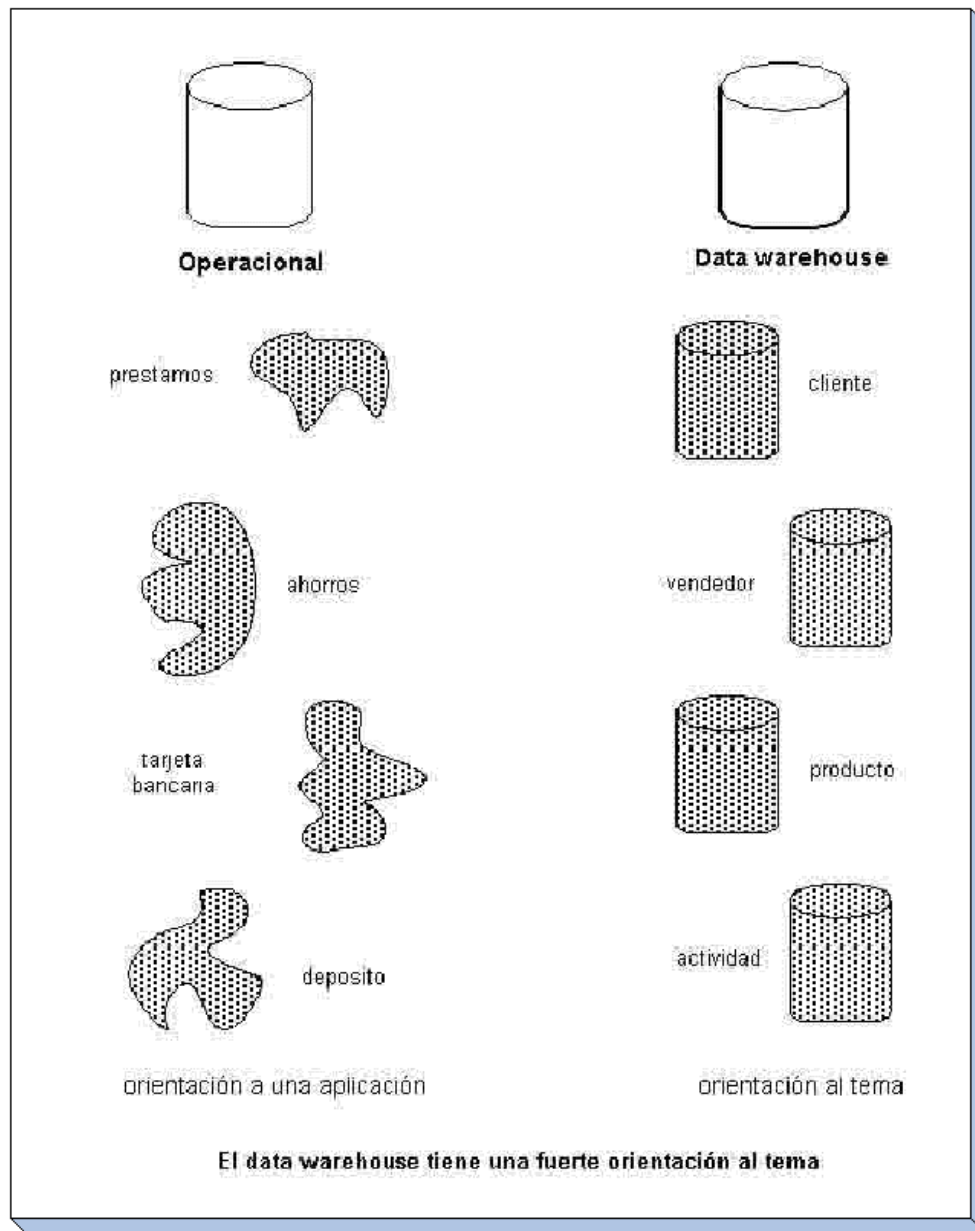
Entre sus principales características tenemos

- Orientado al tema
- Integrado
- De tiempo variante
- No volátil

Orientado a temas

Una primera característica del data warehouse es que la información se clasifica en base a los aspectos que son de interés para la empresa. Siendo así, los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones.

En la Figura se muestra el contraste entre los dos tipos de orientaciones.



El ambiente operacional se diseña alrededor de las aplicaciones y funciones tales como préstamos, ahorros, tarjeta bancaria y depósitos para una institución financiera. Por ejemplo, una aplicación de ingreso de órdenes puede acceder a los datos sobre clientes, productos y cuentas. La base de datos combina estos elementos en una estructura que acomoda las necesidades de la aplicación.

En el ambiente data warehousing se organiza alrededor de sujetos tales como cliente, vendedor, producto y actividad. Por ejemplo, para un fabricante, éstos pueden ser clientes, productos, proveedores y vendedores. Para una universidad pueden ser estudiantes, clases y profesores. Para un hospital pueden ser pacientes, personal médico, medicamentos, etc.

La alineación alrededor de las áreas de los temas afecta el diseño y la implementación de los datos encontrados en el data warehouse. Las principales áreas de los temas influyen en la parte más importante de la estructura clave.

Las aplicaciones están relacionadas con el diseño de la base de datos y del proceso. En data warehousing se enfoca el modelamiento de datos y el diseño de la base de datos.

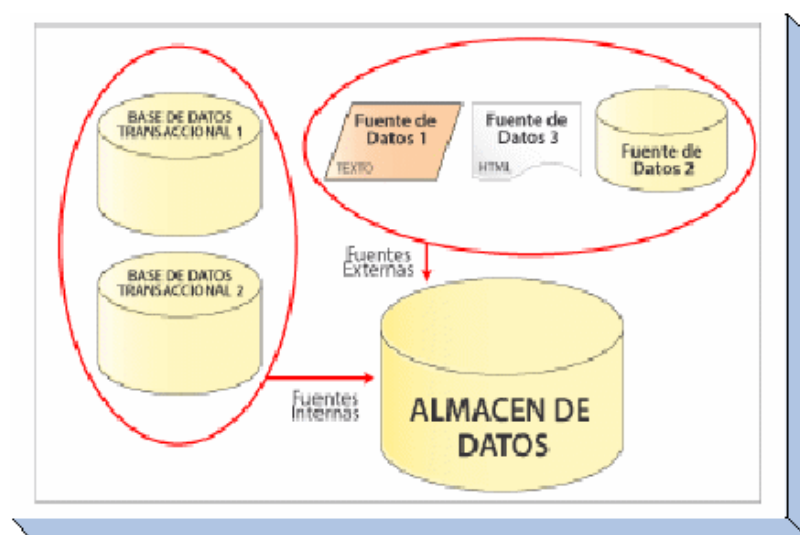
El diseño del proceso (en su forma clásica) no es separado de este ambiente.

Las diferencias entre la orientación de procesos y funciones de las aplicaciones y la orientación a temas, radican en el contenido de la data a escala detallada. En el data warehouse se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones, mientras que la información de las orientadas a las aplicaciones, contiene datos para satisfacer de inmediato los requerimientos funcionales y de proceso, que pueden ser usados o no por el analista de soporte de decisiones.

Otra diferencia importante está en la interrelación de la información. Los datos operacionales mantienen una relación continua entre dos o más tablas basadas en una regla comercial que está vigente. Las del data warehouse miden un espectro de tiempo y las relaciones encontradas en el data warehouse son muchas. Muchas de las reglas comerciales (y sus correspondientes relaciones de datos) se representan en el data warehouse, entre dos o más tablas.

Integrado

Integra datos recolectados de diferentes sistemas operacionales de la organización y o fuentes externas.



El aspecto más importante del ambiente data warehousing es que la información encontrada al interior está siempre integrada.

La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros.

El contraste de la integración encontrada en el data warehouse con la carencia de integración del ambiente de aplicaciones, se muestran en la figura, con diferencias bien marcadas.

A través de los años, los diseñadores de las diferentes aplicaciones han tomado sus propias decisiones sobre cómo se debería construir una aplicación. Los estilos y diseños personalizados se muestran de muchas maneras.

Se diferencian en la codificación, en las estructuras claves, en sus características físicas, en las convenciones de nombramiento y otros. La capacidad colectiva de muchos de los diseñadores de aplicaciones, para crear aplicaciones inconsistentes, es fabulosa. La Figura mencionada, muestra algunas de las diferencias más importantes en las formas en que se diseñan las aplicaciones.

Codificación. Los diseñadores de aplicaciones codifican el campo GENERO en varias formas. Un diseñador representa GENERO como una "M" y una "F", otros como un "1" y un "0", otros como una "X" y una "Y" e inclusive, como "masculino" y "femenino".

No importa mucho cómo el GENERO llega al data warehouse. Probablemente "M" y "F" sean tan buenas como cualquier otra representación. Lo importante es que sea de cualquier fuente de donde venga, el GENERO debe llegar al data warehouse en un estado integrado uniforme.

Por lo tanto, cuando el GENERO se carga en el data warehouse desde una aplicación, donde ha sido representado en formato "M" y "F", los datos deben convertirse al formato del data warehouse.

Medida de atributos. Los diseñadores de aplicaciones miden las unidades de medida de las tuberías en una variedad de formas. Un diseñador almacena los datos de tuberías en centímetros, otros en pulgadas, otros en millones de pies cúbicos por segundo y otros en yardas.

Al dar medidas a los atributos, la transformación traduce las diversas unidades de medida usadas en las diferentes bases de datos para transformarlas en una medida estándar común.

Cualquiera que sea la fuente, cuando la información de la tubería llegue al data warehouse necesitará ser medida de la misma manera.

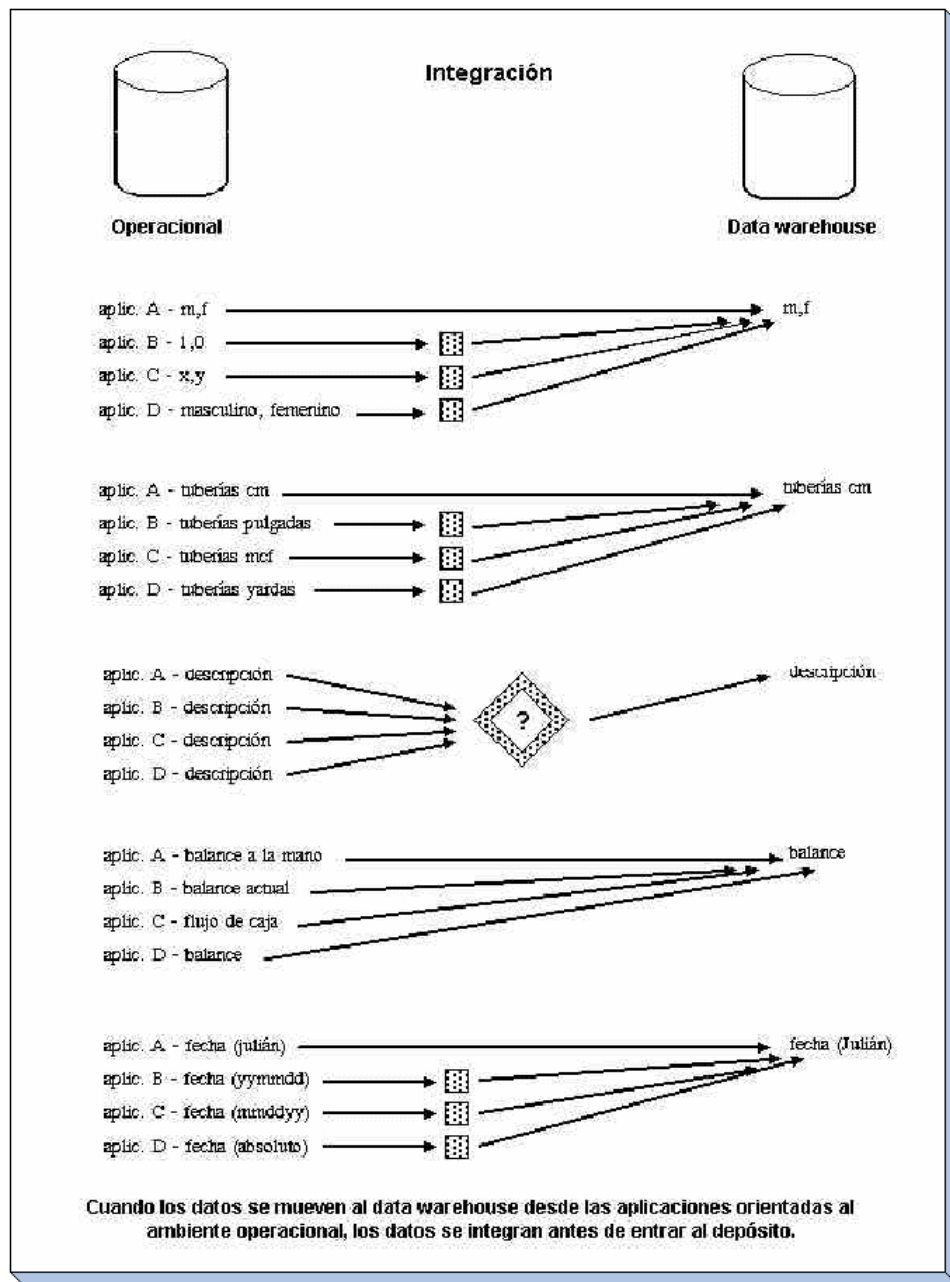
Convenciones de Nombramiento. El mismo elemento es frecuentemente referido por nombres diferentes en las diversas aplicaciones. El proceso de transformación asegura que se use preferentemente el nombre de usuario.

Fuentes Múltiples. El mismo elemento puede derivarse desde fuentes múltiples. En este caso, el proceso de transformación debe asegurar que la fuente apropiada sea usada, documentada y movida al depósito.

Tal como se muestra en la figura, los puntos de integración afectan casi todos los aspectos de diseño - las características físicas de los datos, la disyuntiva de tener más de una de fuente de datos, el problema de estándares de denominación inconsistentes, formatos de fecha inconsistentes y otros.

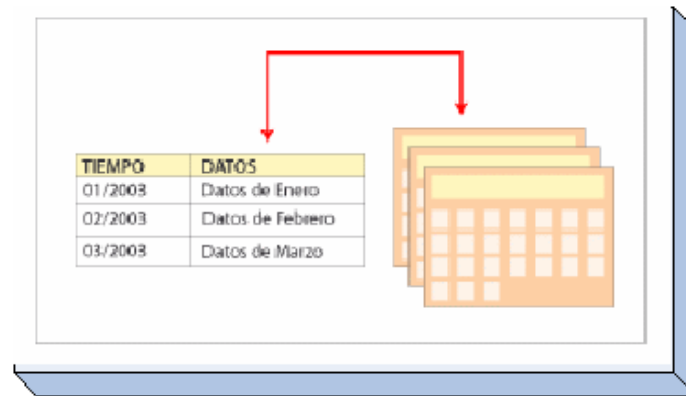
Cualquiera que sea la forma del diseño, el resultado es el mismo - la información necesita ser almacenada en el data warehouse en un modelo globalmente aceptable y singular, aun cuando los sistemas operacionales subyacentes almacenen los datos de manera diferente.

Cuando el analista de sistema de soporte de decisiones observe el data warehouse, su enfoque deberá estar en el uso de los datos que se encuentre en el depósito, antes que preguntarse sobre la confiabilidad o consistencia de los datos.



De tiempo variante (Variable en el tiempo)

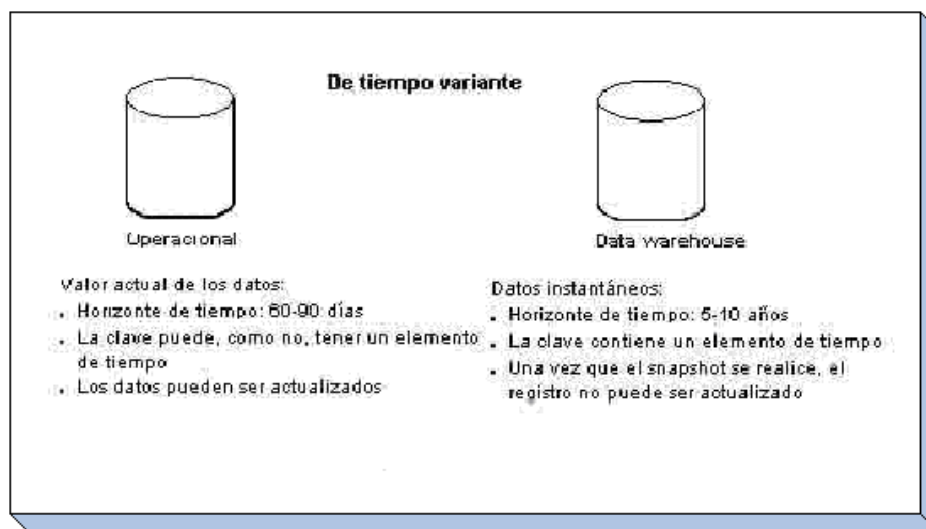
Los datos son relativos a un periodo de tiempo y estos deben ser integrados periódicamente, los mismos son almacenados como fotos que se corresponden a un periodo de tiempo.



Toda la información del data warehouse es requerida en algún momento. Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En éstos, la información se requiere al momento de acceder. En otras palabras, en el ambiente operacional, cuando usted accede a una unidad de información, usted espera que los valores requeridos se obtengan a partir del momento de acceso.

Como la información en el data warehouse es solicitada en cualquier momento (es decir, no "ahora mismo"), los datos encontrados en el depósito se llaman de "tiempo variante".

Los datos históricos son de poco uso en el procesamiento operacional. La información del depósito por el contraste, debe incluir los datos históricos para usarse en la identificación y evaluación de tendencias. (Ver Figura).



El tiempo variante se muestra de varias maneras:

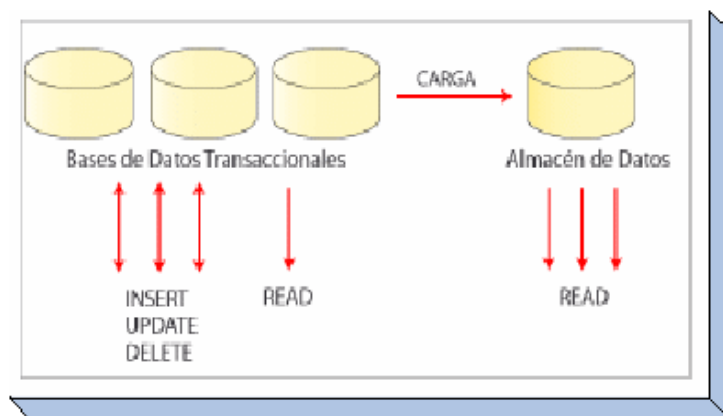
1. La más simple es que la información representa los datos sobre un horizonte largo de tiempo - desde cinco a diez años. El horizonte de tiempo representado para el ambiente operacional es mucho más corto - desde valores actuales hasta sesenta a noventa días. Las aplicaciones que tienen un buen rendimiento y están disponibles para el procesamiento de transacciones, deben llevar una cantidad mínima de datos si tienen cualquier grado de flexibilidad. Por ello, las aplicaciones

operacionales tienen un corto horizonte de tiempo, debido al diseño de aplicaciones rígidas.

2. La segunda manera en la que se muestra el tiempo variante en el data warehouse está en la estructura clave. Cada estructura clave en el data warehouse contiene, implícita o explícitamente, un elemento de tiempo como día, semana, mes, etc. El elemento de tiempo está casi siempre al pie de la clave concatenada, encontrada en el data warehouse. En ocasiones, el elemento de tiempo existirá implícitamente, como el caso en que un archivo completo se duplica al final del mes, o al cuarto.
3. La tercera manera en que aparece el tiempo variante es cuando la información del data warehouse, una vez registrada correctamente, no puede ser actualizada. La información del data warehouse es, para todos los propósitos prácticos, una serie larga de "snapshots" (vistas instantáneas). Por supuesto, si los snapshots de los datos se han tomado incorrectamente, entonces pueden ser cambiados. Asumiendo que los snapshots se han tomado adecuadamente, ellos no son alterados una vez hechos. En algunos casos puede ser no ético, e incluso ilegal, alterar los snapshots en el data warehouse. Los datos operacionales, siendo requeridos a partir del momento de acceso, pueden actualizarse de acuerdo a la necesidad.

No volátil

Los datos que son almacenados no sufren ninguna actualización solo son incrementados. El período cubierto para un DW va de 2 a 10 años.



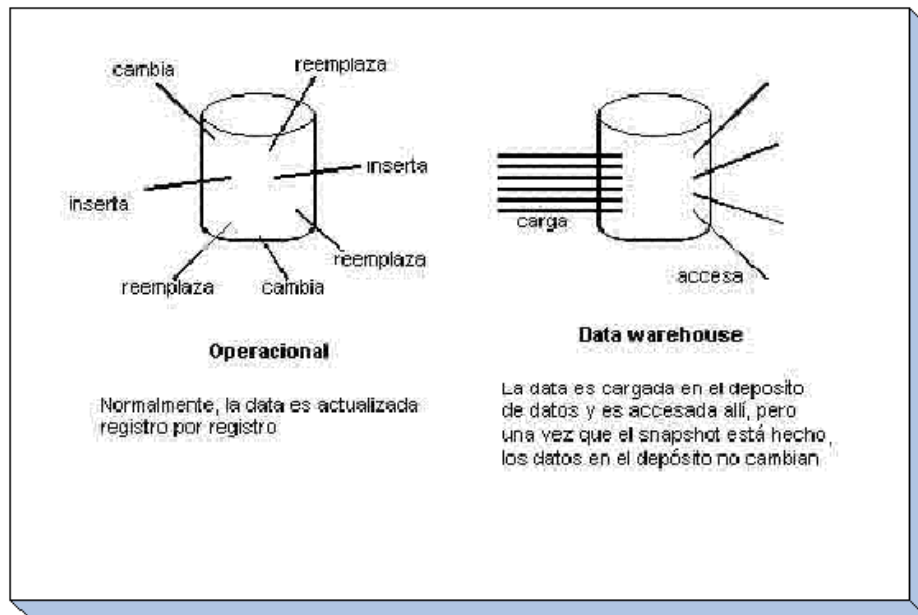
La información es útil sólo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable.

En la Figura se muestra que la actualización (insertar, borrar y modificar), se hace regularmente en el ambiente operacional sobre una base de registro por registro. Pero la manipulación básica de los datos que ocurre en el data warehouse es mucho más simple. Hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No hay actualización de datos (en el sentido general de actualización) en el depósito, como una parte normal de procesamiento.

Hay algunas consecuencias muy importantes de esta diferencia básica, entre el

procesamiento operacional y del data warehouse. En el nivel de diseño, la necesidad de ser precavido para actualizar las anomalías no es un factor en el data warehouse, ya que no se hace la actualización de datos. Esto significa que en el nivel físico de diseño, se pueden tomar libertades para optimizar el acceso a los datos, particularmente al usar la normalización y desnormalización física.

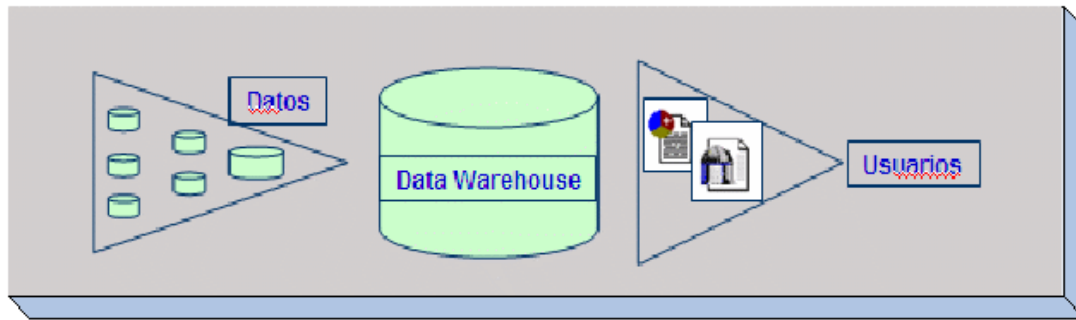
Otra consecuencia de la simplicidad de la operación del data warehouse está en la tecnología subyacente, utilizada para correr los datos en el depósito. Teniendo que soportar la actualización de registro por registro en modo on-line (como es frecuente en el caso del procesamiento operacional) requiere que la tecnología tenga un fundamento muy complejo debajo de una fachada de simplicidad.



Por qué construir un Datawarehouse?¹

Pueden darse algunas justificaciones para un emprendimiento de Datawarehouse:

- Sistemas no integrados
- Múltiples e incompatibles estructuras de datos
- Muchos puntos de entrada a los datos
- Manejo de información histórica
- Para facilitar las actividades de reporte y análisis de usuarios
- Proveer una vista única del negocio



[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

[Continuación del tutorial \(2ª parte\)](#)

Ing Crithian Herrera 64



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 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

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.

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

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

Nombre Corto	Descripción
Administración Web de MySQL	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
Administracion Web de MySQL	Os mostramos como instalar en vuestro PC phpMySQL, un potente gestor Web de MySQL utilizado en la mayoría de los Hostings.
Modelado de MySQL con herramientas gratuitas	Os mostramos otra alternativa de modelado gráfico de MySql.
Apache, MySQL y PHP	Os mostramos como configurar Apache, MySQL y PHP en vuestra máquina
Pool de conexiones a BBDD con struts	Os mostramos como configurar un pool de conexiones a base de datos en vuestras aplicaciones construidas con struts
Todo lo que querias saber sobre DatawareHouse (IV)	En este documento Christian nos enseña exhaustivamente qué es y como funciona un datawarehouse.
Modelado Gráfico de MySQL	Os mostramos como instalar y utilizar DeZing e Importer de Datanamic para crear el modelo lógico y físico de bustra base de datos MySQL
Todo lo que querias saber sobre DatawareHouse (III)	En este documento Christian nos enseña exhaustivamente qué es y como funciona un datawarehouse.
Todo lo que querias saber sobre DatawareHouse (II)	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 para su resolución.

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

