

# ¿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)



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 [datawarehouse3.pdf](#)

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

**¿Eres Informático?**

No Pierdas la Oportunidad de un Empleo Mejor. ¡No Pierdes Nada!  
[InfoJobs.net](http://InfoJobs.net)

**Java j2ee**

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

**Java Reporting ReportMill**

Great Java Report Tool - Free  
Eval! PDF, HTML, Excel, XML,  
Swing & more  
[www.reportmill.com](http://www.reportmill.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 2ª parte](#)

[Planificación Tecnológica 22](#)

[Estrategia de la Base de Datos 22](#)

[Estrategia de la Aplicación 22](#)

[Estrategia de la Explotación 23](#)

[Arquitectura del Datawarehouse 23](#)

[Fuentes de Datos 25](#)

[El motor del datawarehouse 25](#)

[Gestor de Carga 25](#)

[Confiabilidad de los datos 26](#)

[Meta Datos 28](#)

[Agregaciones 29](#)

[Gestor del Datawarehouse 29](#)

[Gestor de Respaldos 29](#)

- [Repositorio del Datawarehouse 29](#)
  - [Rendimiento de carga 29](#)
  - [Procesamiento de carga 29](#)
  - [Gestión de calidad de los datos 30](#)
  - [Rendimiento de consultas 30](#)
  - [Escalabilidad de terabytes 30](#)
  - [Escalabilidad de usuarios en masa 30](#)
  - [Trabajo en red 30](#)
  - [Facilidad de administración 31](#)
  - [Análisis dimensional integrado 31](#)
  - [Funcionalidad de consulta avanzada 31](#)
- [Cubos Dimensionales 31](#)
- [Herramientas de Acceso ó Herramientas de Usuario Final 32](#)
  - [Herramientas de Reportes y Consultas 32](#)
  - [Herramientas de desarrollo de aplicaciones 33](#)
  - [Herramientas de Procesamiento Analítico en Línea \(OLAP\) 33](#)
    - [Operaciones Analíticas Básicas de Sistemas OLAP 34](#)
    - [Vistas de Datos 35](#)
  - [Herramientas de Mineo de Datos 36](#)
- [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](#)

## **Planificación Tecnológica**

### **Estrategia de la Base de Datos**

Se trata de la creación de la base de datos. Entre otras cosas incluye

- Contenido: Qué datos e información se requieren para solucionar las preguntas y necesidades de los usuarios
- Fuentes: Cuáles son los fuentes de la información y donde se encuentran las fuentes.
- Extracción: Cómo se extraen los datos y con que periodicidad se cargan en el datawarehouse.
- Preparación: Qué se requiere para depurar y validar los datos fuentes
- Diseño: Cuál es el diseño apropiado para la base de datos
- Afinamiento: Qué aspectos de afinamiento y rendimiento se van a considerar
- Plataforma: Como será la plataforma en la que residirá el datawarehouse, como se compone la red, cuales son los componentes de hardware y software.
- Administración: Qué se requiere para administrar el datawarehouse en términos de seguridad, procesos de actualización, gestión de metadatos, aseguramiento de la calidad, etc.

### **Estrategia de la Aplicación**

La estrategia de aplicación trata con la tecnología en dos puntos: la capa de lógica analítica y la capa de presentación. Identificando acceso a los datos y análisis de requerimientos define el conjunto de requerimientos básicos del usuario. Algunas preguntas de los usuarios pueden ser respondidas simplemente recuperando los datos desde el warehouse, pero muchas mas preguntas requieren algún tipo de rutinas analíticas a ser ejecutadas sobre los datos. Estas rutinas analíticas pueden ser clasificadas desde algo tal simple como cálculo del porcentaje de cambio del volumen de ventas hasta la creación de un modelo matemático complejo.

Se identifican las funciones de análisis de datos que se necesitan para satisfacer las necesidades de los usuarios.

- Acceso; Identificar que usuarios van a tener acceso a la información y también que nivel de información podrá ver cada uno de ellos.
- Análisis: Qué funciones de análisis de información serán necesarias para satisfacer los requerimientos.
- Modelamiento; Requerimientos para análisis estadísticos de datos, minería de datos, u otro soporte de modelamiento matemático
- Aplicaciones; Necesidades para aplicaciones específicas del negocio
- Procesos: Cómo ayuda el datawarehouse a los procesos de negocio, Qué mejoras en los procesos de negocio se logran con el datawarehouse.
- Soporte: Cómo los usuarios recibirán soporte y capacitación en el datawarehouse.

## **Estrategia de la Explotación**

En la estrategia de explotación se consideran los siguientes aspectos

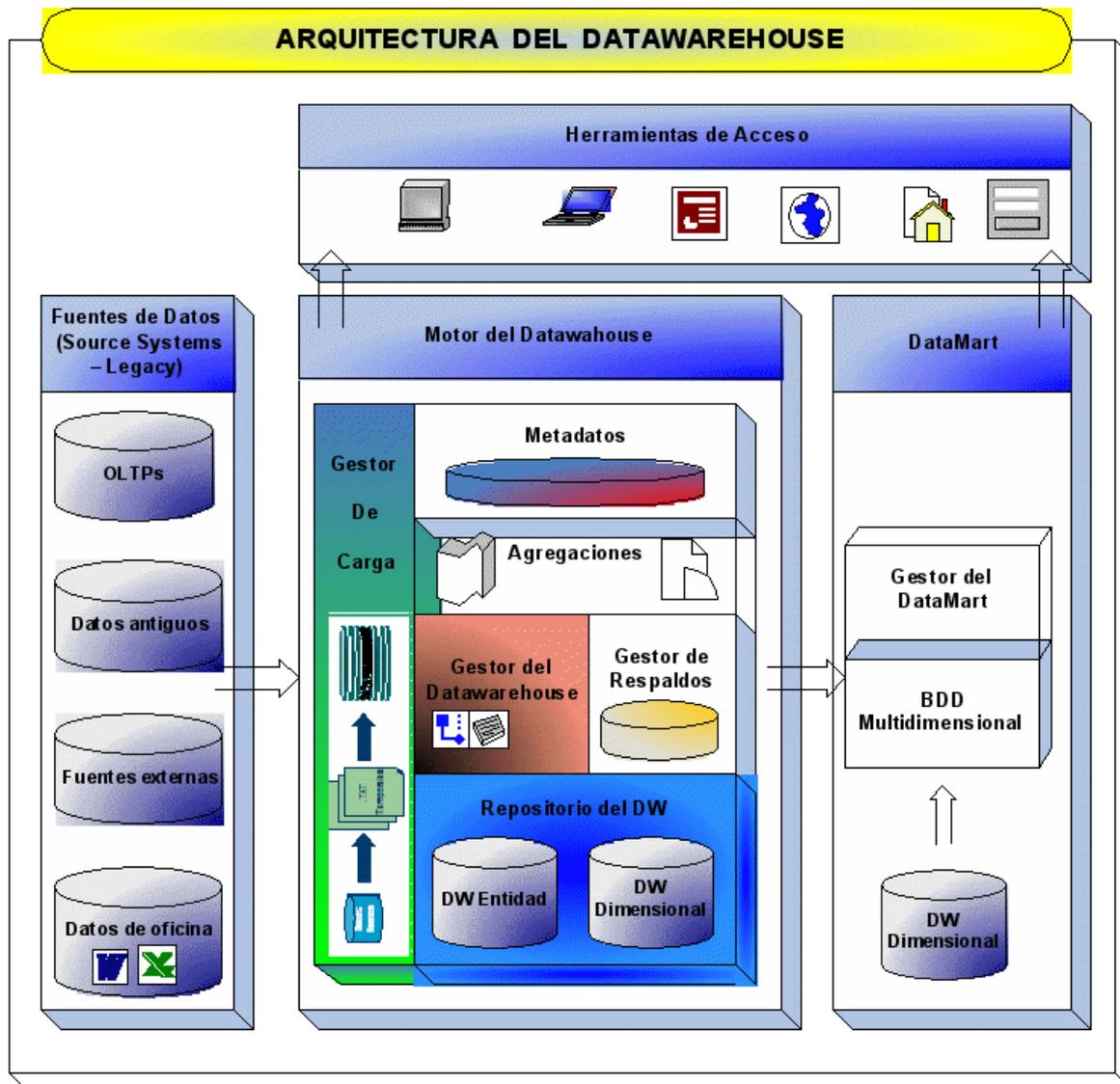
- Interfaz: Cuales usuarios usarán aplicaciones cliente servidor y cuales accederán a través de clientes web (browser)
- Colaboración; Como se promoverá la colaboración entre los usuarios.
- Agentes: Cómo se automatizarán los procesos de análisis y reportes.
- Motor de búsqueda; Cómo los recursos del datawarehouse serán registrados en motores de búsqueda.
- Seguridad: Cómo será garantizada la seguridad de la información y de la base de datos.

## **Arquitectura del Datawarehouse**

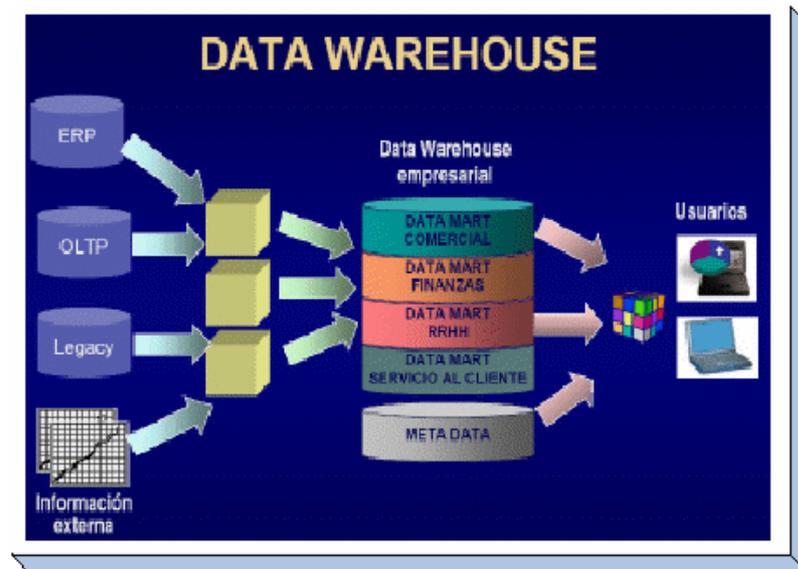
El siguiente gráfico muestra la arquitectura clásica de un Datawarehouse, compuesto por:

- Fuentes de Datos
- Motor del Datawarehouse

- Gestor de Carga
- Metadatos
- Agregaciones
- Gestor del Datawarehouse
- Gestor de Respaldos
- DW Repositorio
- DataMart
  - BDD Dimensional
  - Gestor del DataMart
- Herramientas de Acceso



En forma resumida la arquitectura puede verse expresada en la siguiente figura<sup>2</sup>:



## Fuentes de Datos

Cualquier origen de información que pueda ser considerado para el datawarehouse, aquí se incluyen los siguientes elementos:

- Los sistemas OLTP's que son los sistemas de Legacy que actualmente operan en la empresa.
- Datos antiguos provenientes de migraciones.
- Fuentes externas como otros sistemas de la compañía, sistemas de otras empresas, sistemas de gobierno, internet, etc.
- Datos de oficina, archivos en formato de Word, Excel, archivos planos, PDF™s, mails, etc.

## El motor del datawarehouse

Está integrado por los siguientes componentes

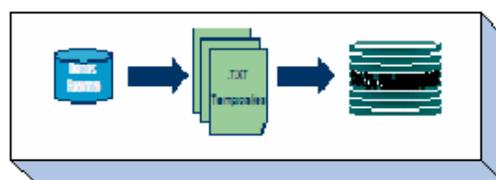
### Gestor de Carga

Quizá sea uno de los elementos más importantes para el datawarehouse, generalmente incluye las operaciones de

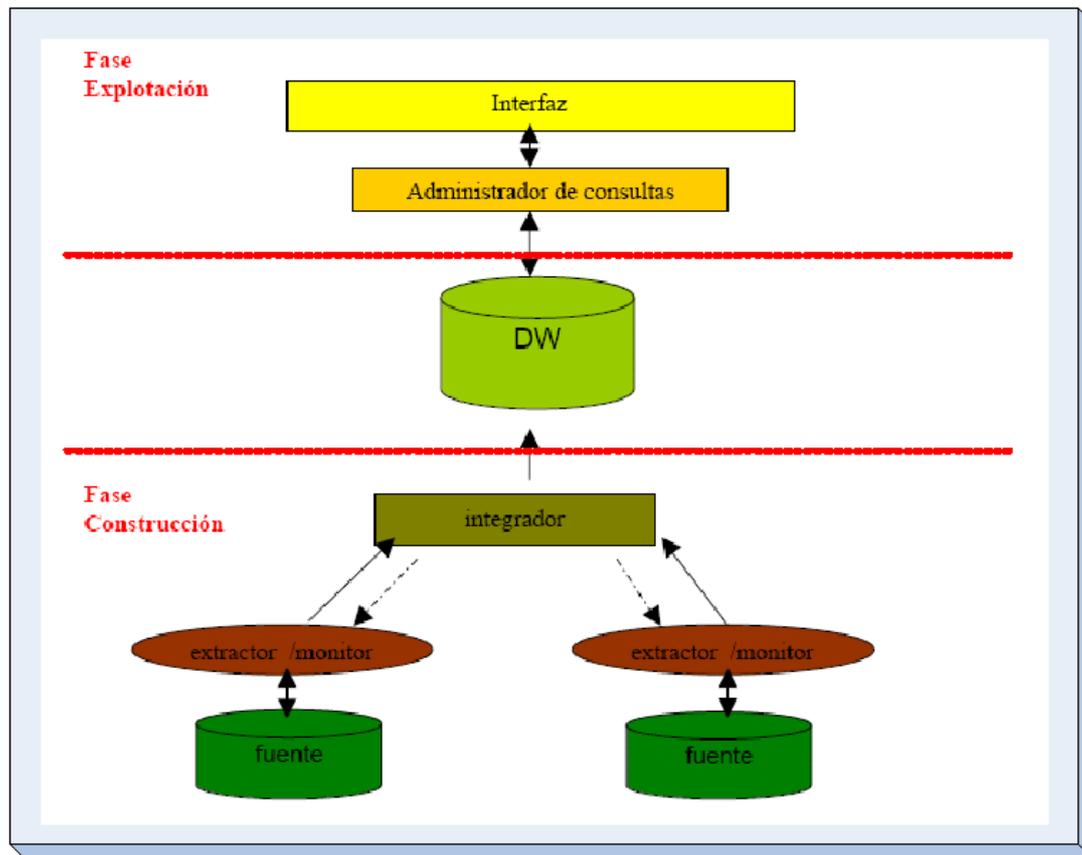
- **Extracción:** Es el proceso que accesa a los datos OLTP existentes, en cualquier forma que exista, desde cualquier DBMS en que exista. Típicamente, extracción y el siguiente paso, propagación, son administrados por el mismo producto. No todas las herramientas de extracción y propagación soportan todas las plataformas, de tal manera que una faceta importante de la selección de herramientas es si la herramienta soporta los sistemas operativos y las bases de datos que se esté usando para el datawarehouse.
- **Propagación:** Es el proceso de mover datos desde los sistemas fuente hacia el

sistema objetivo que contendrá el data warehouse. El proceso de propagación toma lugar en tiempo real, o en un calendario predeterminado (batch), o sobre demanda, y puede efectuar un refresco total del warehouse o justo un cambio neto. Cuando se selecciona una herramienta de propagación, se aspira que ésta ofrezca la gestión de cambios netos como también refresco total y permitirá tanto actualizaciones en tiempo real y calendarizadas (batch).

- **Depuración (Limpieza):** El nivel lógico cubre problemas de valores de datos que son inconsistentes dentro de la información importada (ejemplo, clientes con estado casado, pero con una edad de 3 años). El nivel técnico evalúa problemas de información tales como campos no inicializados o valores inválidos en los datos importados (ejemplo, valor de la fecha Febrero 31).
- **Transformación:** Convierte datos desde su formato OLTP al apropiado formato del datawarehouse ejecutando funciones tales como desnormalización de datos, traduciendo códigos hacia texto significativo, convirtiendo una variedad de formatos de fechas hacia un formato estándar, convirtiendo texto tal como nombres de ciudades hacia texto estándar y renombrando campos desde nombres técnicos no significativos hacia nombres significativos que un usuario final entenderá.
- **Carga:** Los datos fuentes normalmente son extraídos y almacenados en archivos temporales tipo texto, los mismos que deben ser cargados a la base de datos del datawarehouse. La figura resume el proceso de carga, los archivos temporales finalmente son colocados en la base de datawarehouse de destino.



El módulo de Gestor de Carga también es conocido como Integrador, y es muy importante tanto en la Fase de Construcción como en la Fase de Explotación de un DataWarehouse.



### Confiabilidad de los datos<sup>14</sup>

La data "sucia" es peligrosa. Las herramientas de limpieza especializadas y las formas de programar de los clientes proporcionan redes de seguridad.

No importa cómo esté diseñado un programa o cuán hábilmente se use. Si se alimenta mala información, se obtendrá resultados incorrectos o falsos. Desafortunadamente, los datos que se usan satisfactoriamente en las aplicaciones de línea comercial operacionales pueden ser basura en lo que concierne a la aplicación data warehousing.

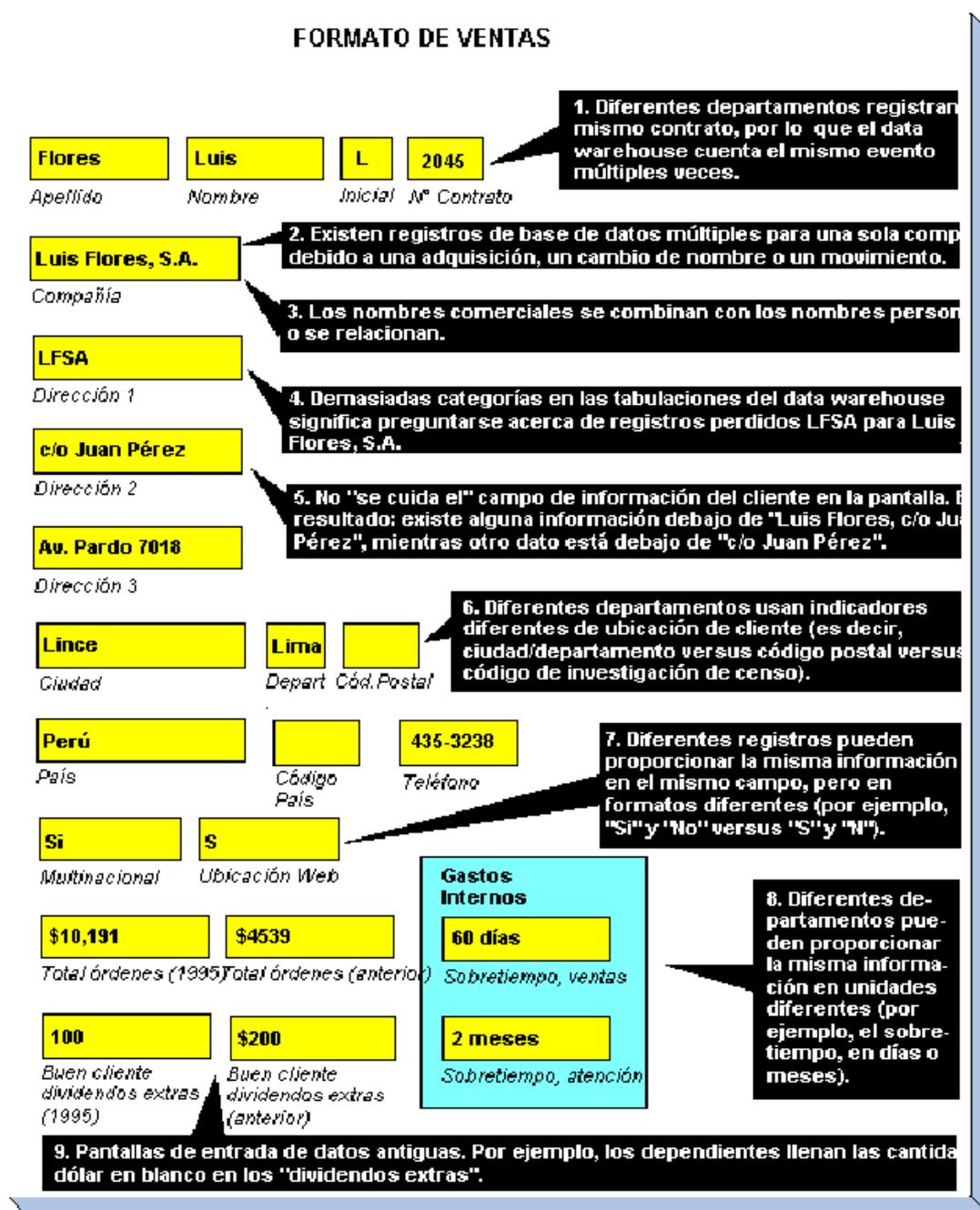
Los datos "sucios" pueden presentarse al ingresar información en una entrada de datos (por ejemplo, "Sistemas S. A." en lugar de "Sistemas S. A.") o de otras causas. Cualquiera que sea, la data sucia daña la credibilidad de la implementación del depósito completo. A continuación, en la Figura se muestra un ejemplo de formato de ventas en el que se pueden presentar errores.

Afortunadamente, las herramientas de limpieza de datos pueden ser de gran ayuda. En algunos casos, puede crearse un programa de limpieza efectivo. En el caso de bases de datos grandes, imprecisas e inconsistentes, el uso de las herramientas comerciales puede ser casi obligatorio.

Decidir qué herramienta usar es importante y no solamente para la integridad de los datos. Si se equivoca, se podría malgastar semanas en recursos de programación o cientos de miles de dólares en costos de herramientas.

La limpieza de una data "sucia" es un proceso multifacético y complejo. Los pasos a seguir son los siguientes:

1. Analizar sus datos corporativos para descubrir inexactitudes, anomalías y otros problemas.
2. Transformar los datos para asegurar que sean precisos y coherentes.
3. Asegurar la integridad referencial, que es la capacidad del data warehouse, para identificar correctamente al instante cada objeto del negocio, tales como un producto, un cliente o un empleado.
4. Validar los datos que usa la aplicación del data warehouse



## Meta Datos

Esta área del warehouse almacena todas las definiciones de los meta datos (datos acerca de los datos) usados por todos los procesos en el warehouse. Los meta datos son usados para una variedad de propósitos incluyendo:

- Los procesos de extracción, transformación y carga (meta datos es usado para mapear las fuentes de datos a una vista común de la información dentro del warehouse).
- Los procesos de gestión del warehouse (cada tabla es descrita incluyendo su estructura, índices, vistas; meta datos es usado también para automatizar la producción de tablas resumen).
- Como parte de los procesos de gestión de consulta (meta datos es usado para dirigir una consulta a la fuente de datos más apropiada)

## **Agregaciones**

Este componente del warehouse almacena todos los datos agregados, predefinidos y generados por el gestor del warehouse.

El propósito de información resumida es para mejorar el rendimiento de las consultas. Aunque hay costos operacionales incrementados asociados con la agregación inicial de los datos, esto debería ser compensado eliminando el requerimiento para ejecutar continuamente operaciones de agregación (tales como clasificación o agrupación) en las respuestas a las consultas de los usuarios. El dato agregado es actualizado continuamente en la medida que nuevos datos son cargados en al warehouse.

## **Gestor del Datawarehouse**

En algunos casos el gestor del warehouse también genera perfiles de consultas para determinar qué índices y agregaciones son apropiadas. Un perfil de consulta puede ser generado para cada usuario, grupo de usuario, o el data warehouse y está basada en la información que describe las características de las consultas tales como la frecuencia, tablas objetivo, y tamaño de los results set.

## **Gestor de Respaldo**

Es el componente que se encarga de respaldar constantemente la información del repositorio del datawarehouse.

## **Repositorio del Datawarehouse**

Es el repositorio en si o la base de datos física donde se almacena la información del datawarehouse.

Un DBMS para trabajar con un sistema de Datawarehouse debe cumplir con los siguientes requerimientos

### ***Rendimiento de carga***

- Datawarehouse requiere de carga incremental de nuevos datos en una base periódica dentro de ventanas de tiempo pequeñas
- El rendimiento de procesos de carga debería ser medido en cientos de millones de filas o gigabytes de datos por hora y no debería haber un límite máximo que restrinja al negocio

### ***Procesamiento de carga***

- Muchos pasos deben ser dados para cargar un dato nuevo o actualizado hacia el datawarehouse incluyendo conversión de datos, filtrado, reformateado, chequeos de integridad, almacenamiento físico, indexación y actualización de los meta datos
- Aunque cada paso en la práctica puede ser atómico, el proceso de carga debería parecer que se ejecuta como una unidad de trabajo única.

### ***Gestión de calidad de los datos***

- El datawarehouse debe asegurar consistencia local, consistencia global e integridad referencial a pesar de las fuentes "sucias" y tamaños masivos de bases de datos
- la preparación y carga son pasos necesarios, ellos no son suficientes. La habilidad para responder a las consultas de los usuarios finales es la medida del éxito para una aplicación de datawarehouse.
- Mientras más preguntas son respondidas, los analistas tienden a solicitar preguntas más complejas y creativas

### ***Rendimiento de consultas***

- Gestión basada en hechos y análisis ad hoc no deben ser retardadas o inhibidas por el rendimiento del RDBMS datawarehouse.
- Consultas complejas y grandes para operaciones claves del negocio deben ser completadas en un período de tiempo razonable.

### ***Escalabilidad de terabytes***

- El tamaño de data warehouse está creciendo a enormes tasas con tamaños en el rango de cientos de gigabytes hasta los terabytes y petabytes ( $10^{15}$ ).
- Los RDBMS no deben tener ninguna limitación arquitectural para el tamaño de la base de datos y deberían soportar gestión modular y paralela. En el evento de fallas, el RDBMS debería soportar disponibilidad continua, y proveer mecanismos para recuperación. El RDBMS debe soportar dispositivos de almacenamiento en masa tales como discos ópticos y dispositivos de gestión de almacenamiento jerárquico.
- Finalmente, rendimiento de consultas no debería ser dependiente del tamaño de la base de datos, sino más bien de la complejidad de la consulta.

### ***Escalabilidad de usuarios en masa***

- El pensamiento actual es que el acceso a un datawarehouse es limitado a un número de usuarios gerenciales relativamente bajo.
- Se predice que en el futuro, el datawarehouse RDBMS debería ser capaz de soportar cientos, o aún miles de usuarios concurrentes mientras se mantiene un aceptable rendimiento en las consultas

### ***Trabajo en red***

- Sistemas de datawarehouse deberían ser capaces de cooperación en un gran red de datawarehouse
- El datawarehouse debe incluir herramientas que coordinen el movimiento de subsets de datos entre datawarehouse
- Los usuarios deberían ser capaces de mirar en y trabajar con múltiples datawarehouses desde una estación de cliente única

### ***Facilidad de administración***

- La muy grande escala y naturaleza cíclica en el tiempo del datawarehouse demanda facilidad y flexibilidad administrativa
- El RDBMS debe proveer control para límites de recursos en la implementación

### ***Análisis dimensional integrado***

- La potencia de vistas multidimensionales es ampliamente aceptada y soporte dimensional debe ser inherente en el RDBMS warehouse para proveer el más alto rendimiento para herramientas OLAP relacional.
- El RDBMS debe soportar creación rápida y fácil de resúmenes comunes precalculados en grandes datawarehouses, y proveer herramientas de

mantenimiento para automatizar la creación de estos agregados precalculados

### ***Funcionalidad de consulta avanzada***

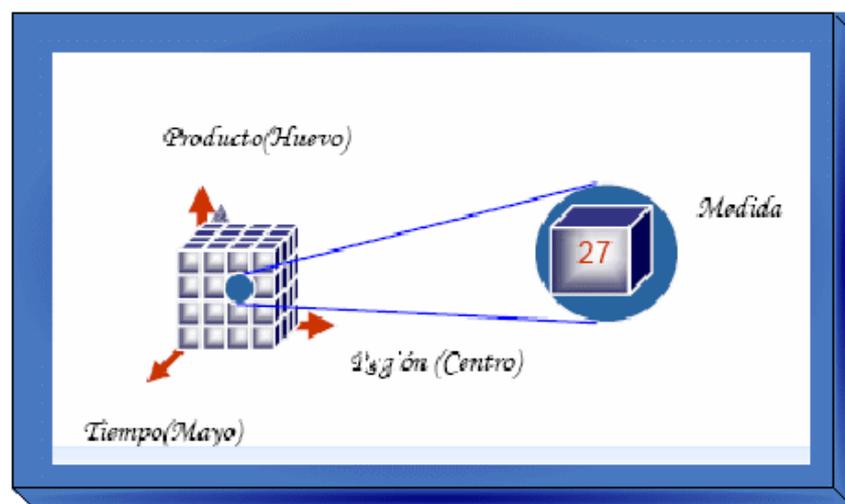
- Usuarios finales requieren cálculos analíticos avanzados, análisis secuencial y comparativo, y accesos consistentes a datos detallados y resumidos.
- Usando SQL en un ambiente cliente servidor y herramientas apunte y dispare puede algunas veces ser no práctico o aún imposible debido a la complejidad de las consultas de los usuarios.
- El RDBMS debe proveer un completo y avanzado set de operaciones analíticas.

### **Cubos Dimensionales**

Un modelo de datos multidimensional soporta el manejo de una basta cantidad de datos empresariales y temporales. De esta forma surge la instancia del modelo multidimensional, también conocido como cubo o hipercubo.

Para clarificarlo un poco se puede imaginar un cubo con tres dimensiones: producto, tiempo, región; donde cada dimensión tiene diferentes niveles o hechos, para finalmente intersectar estos valores y obtener una medida.

La medida es el índice de un producto como puede ser el huevo en el mes de mayo y en la zona centro del país.

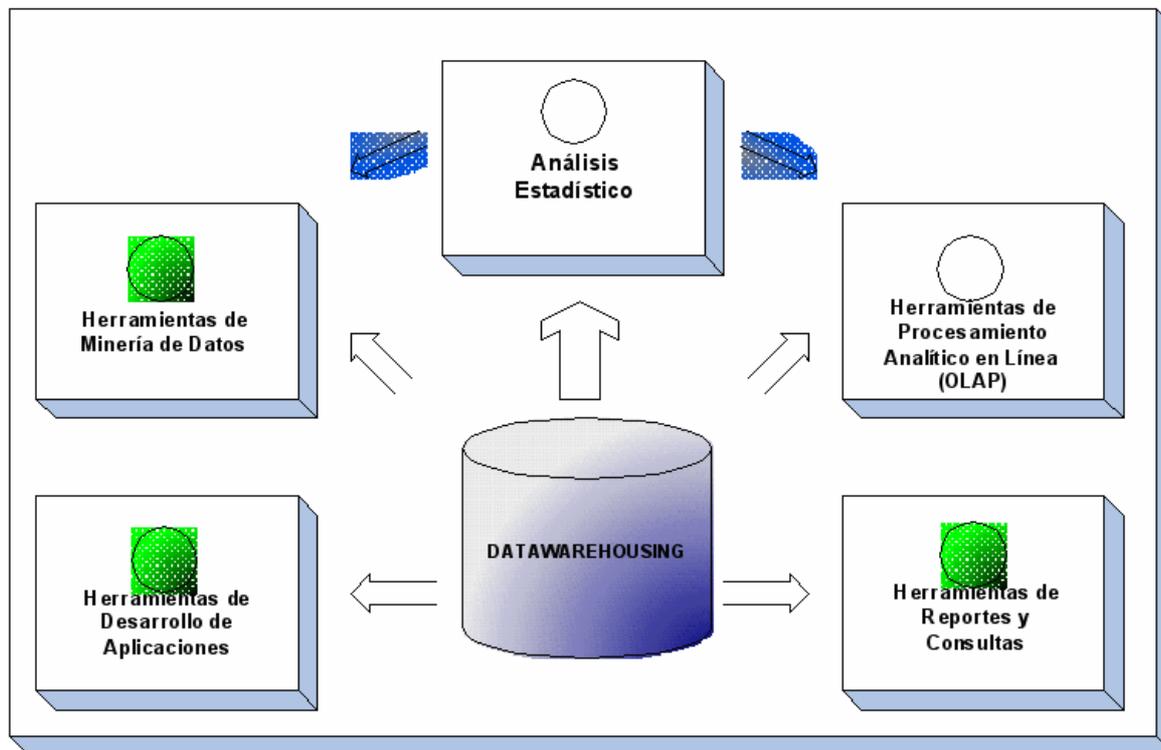


### **Herramientas de Acceso ó Herramientas de Usuario Final**

Estas herramientas pueden reunirse en 4 grupos

- Herramientas de Minería de Datos

- Herramientas de Procesamiento Analítico en Línea (OLAP)
- Herramientas de Desarrollo de Aplicaciones
- Herramientas de Reportes y Consultas



## Herramientas de Reportes y Consultas

Simplifican la generación de código SQL. Los usuarios formulan consultas a la base de datos sin tener que interactuar con el lenguaje de programación de bases de datos (SQL).

## Herramientas de desarrollo de aplicaciones

Los requerimientos de usuarios finales pueden ser tales que las capacidades pre-construidas de las herramientas de reportes y consultas son inadecuadas ya sea debido a que el análisis requerido no puede ser ejecutado o debido a que la interacción del usuario requiere un no razonable alto nivel de expertise por parte del usuario

En esta situación el acceso de usuarios puede requerir el desarrollo de aplicaciones usando ambientes gráficos de acceso de datos desarrollados principalmente para ambientes cliente/ servidor.

Algunas de estas plataformas de desarrollo de aplicaciones se integran con herramientas OLAP populares, y pueden acceder todos los sistemas de bases de datos principales, incluyendo Oracle, SQL Server, Sybase, e Informix. Ejemplos de estos ambientes de desarrollo de aplicaciones incluyen PowerBuilder de PowerSoft,

VisualBasic de Microsoft, y Bussines Objects.

## Herramientas de Procesamiento Analítico en Línea (OLAP)

Típicas aplicaciones de negocios para estas herramientas incluyen evaluación de la efectividad de una campaña de marketing, previsión de ventas de productos, y planificación de la capacidad. Estas herramientas asumen que los datos están organizados en un modelo multidimensional, el cual es soportado por una base de datos multidimensional especial (MDDBMS) o por una base de datos relacional diseñada para permitir consultas multidimensionales.

Este tipo de herramientas permiten a los usuarios ingresar a un data warehouse desde cualquier dimensión simple para empezar el análisis, luego navegar a otra dimensión para un mayor análisis de la información.

- **OLAP es online analytical processing.** Se trata de una forma de almacenar la información en una Base de Datos que permita realizar de forma más efectiva las queries. Es una definición abreviada, claro esta, la realidad es más compleja.
- **MOLAP, Multidimensional OLAP.** Tanto los datos fuente como los datos agregados o precalculados residen en el mismo formato multidimensional. Optimiza las queries, pero requiere más espacio de disco y diferente software. El primer punto esta dejando ser un problema: el espacio de disco cada vez es más barato.
- **ROLAP, Relational OLAP.** Tanto los datos precalculados y agregados como los datos fuente residen en la misma base de datos relacional. Si el DataWarehouse es muy grande o se necesita rapidez por parte de los usuarios puede ser un problema.
- **HOLAP, Hybrid OLAP:** Es una combinación de los dos anteriores. Los datos agregados y precalculados se almacenan en estructuras multidimensionales y los de menor nivel de detalle en el relacional. Requiere un buen trabajo de análisis para identificar cada tipo de dato.

Los 6 elementos básicos de un sistema OLAP son: Dimensiones, Valores, Jerarquías, Niveles, Atributos e Indicadores.

Veamos un ejemplo<sup>4</sup>:

Los datos se almacenan en una estructura de cubo (es como si estuviera totalmente indexado) y la velocidad de acceso se hace mucho mas eficiente.

En este caso, la primera consulta nos muestra los puntos que han acumulado los titulares de la tarjeta clásica, en todos los tipos de cabinas y en todos los meses. Mientras, que en la segunda filtramos el cubo para los que volaron en turista. Este tipo de consultas devuelve los datos de forma instantánea.



### *Operaciones Analíticas Básicas de Sistemas OLAP*

Los sistemas OLAP soportan las siguientes operaciones

- **Consolidación:** este comprende el conjunto de datos. Esto puede involucrar acumulaciones simples o agrupaciones complejas que incluyen datos interrelacionados.
- **Drill-Down:** OLAP puede moverse en la dirección contraria y presentar automáticamente datos detallados que abarcan datos consolidados.
- **Slicing and Dicing:** se refiere a la capacidad de visualizar a la BD desde diferentes puntos de vista. También se la conoce como operación de **Pivotaje**.

VENTAS			
	Productos	Store 1	Store 2
Q1	Electronics	\$5,20	\$5,60
	Toys	\$1,90	\$1,40
	Clothing	\$2,30	\$2,60
Q2	Cosmetics	\$1,10	\$1,10
	Electronics	\$8,90	\$7,20
	Toys	\$0,75	\$0,40
	Clothing	\$4,60	\$4,60
	Cosmetics	\$1,50	\$0,50

→ SLICE y DICE

VENTAS		
	Productos	Store 1
Q1	Electronics	\$5,20
	Toys	\$1,90
Q2	Electronics	\$8,90
	Toys	\$0,75

Presupuestos			
	Measures		
	Inicial		
	Tiempo		
	2004	2005	2006
-CAPITULO 1 - GASTOS DE PERSONAL	151,601,484.00	302,153,628.00	407,219,592.00
+ARTICULO 10. ALTOS CARGOS	21,402,816.00	48,059,948.00	54,823,044.00
+ARTICULO 11. PERSONAL EVENTUAL	45,437,520.00	90,504,864.00	118,141,764.00
+ARTICULO 12. FUNCIONARIOS	66,407,748.00	126,277,008.00	183,482,724.00
+ARTICULO 13. LABORALES	10,061,472.00	20,856,384.00	27,028,620.00
+ARTICULO 14. OTRO PERSONAL	142,968.00	321,240.00	415,992.00
+ARTICULO 15. INCENTIVOS AL RENDIMIENTO	7,202,736.00	15,708,504.00	20,629,416.00
+ARTICULO 16. CUOTAS, PRESTACIONES Y GASTOS SOC. A CARGO DEL EMPLEADOR	946,224.00	2,425,680.00	2,698,032.00

**Columnas**

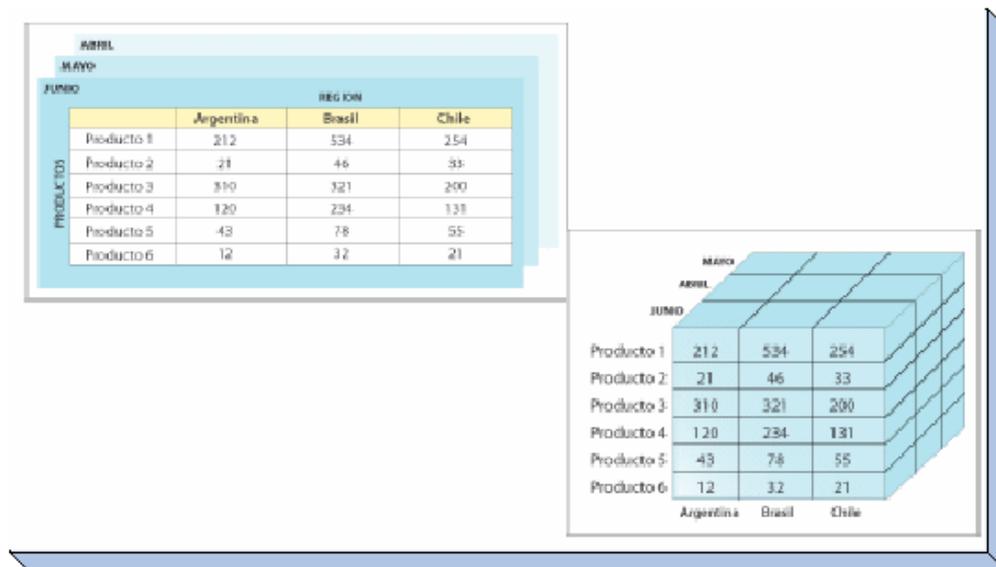
- Measures
- Tiempo
- Rows
- Presupuestos
- Filter
- Geografia ((All)=All Geografia)
- Organos Gestores ((All)=All Organos)

**Tiempo**

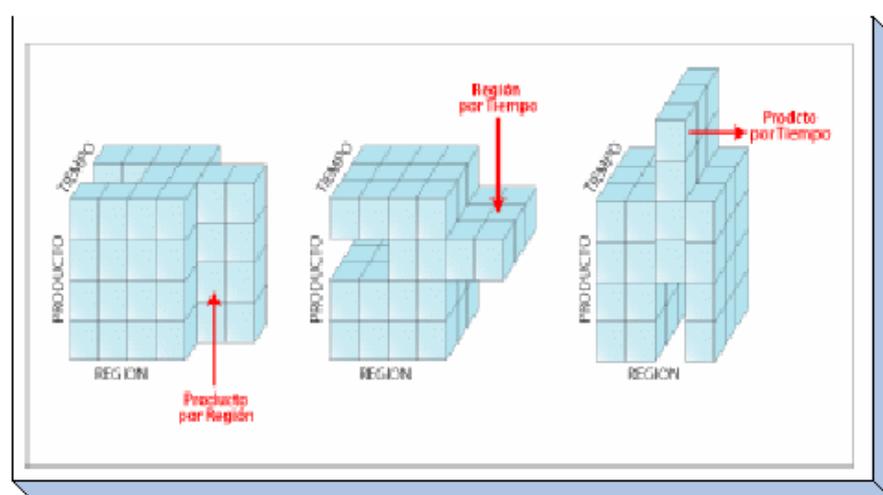
- All Geografia
- 2004
- 2005
- 2006

## Vistas de Datos

La vista de datos como cubos es una extensión de la manera normal en que los usuarios de negocios interactúan con los datos. Por Ejemplo: la mayoría de los usuarios desearía ver como se desarrollan las ventas a lo largo del tiempo. Para ello se necesitaría ver varias planillas de cálculo.



Debido a su representación pueden ser tomadas rebanadas de datos de las mismas, para responder diversas preguntas.



## Herramientas de Mineo de Datos

Mineo de datos es el proceso de descubrir nuevas correlaciones significativas, patrones y tendencias por medio del mineo de grandes cantidades de datos almacenados en un datawarehouse ó en un data mart, usando técnicas estadísticas, reconocimiento de patrones y algoritmos de aprendizaje para identificar relaciones entre los elementos de datos.

- [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 \(4ª parte\)](#)

---

Ing Cristhian Herrera 64



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 2.5](https://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.

Subscribirse a Novedades	
<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="#">Todo lo que querias saber sobre DatawareHouse (IV)</a>	En este documento Christian nos enseña exhaustivamente qué es y como funciona un datawarehouse.
<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 (I)</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.

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](http://www.enredados.com) .... [Hosting en Castellano con soporte Java/J2EE](#)



www.AdictosAlTrabajo.com Opimizado 800X600