

# **BUSINESS INTELLIGENCE**

**Técnicas, herramientas y  
aplicaciones**

**María Pérez Marqués**



Business Intelligence. Técnicas, herramientas y aplicaciones  
María Pérez Marqués

ISBN: 978-84-943055-2-8  
EAN: 9788494305528  
IBIC: UNH, UNC

Copyright © 2015 RC Libros  
© RC Libros es un sello y marca comercial registrados

### **Business Intelligence. Técnicas, herramientas y aplicaciones**

Reservados todos los derechos. Ninguna parte de este libro incluida la cubierta puede ser reproducida, su contenido está protegido por la Ley vigente que establece penas de prisión y/o multas a quienes intencionadamente reprodujeren o plagiaren, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución en cualquier tipo de soporte existente o de próxima invención, sin autorización previa y por escrito de los titulares de los derechos de la propiedad intelectual. La infracción de los derechos citados puede constituir delito contra la propiedad intelectual. (Art. 270 y siguientes del Código Penal). Dirijase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra a través de la web [www.conlicencia.com](http://www.conlicencia.com); o por teléfono a: 91 702 19 70 / 93 272 04 47.

RC Libros, el Autor, y cualquier persona o empresa participante en la redacción, edición o producción de este libro, en ningún caso serán responsables de los resultados del uso de su contenido, ni de cualquier violación de patentes o derechos de terceras partes. El objetivo de la obra es proporcionar al lector conocimientos precisos y acreditados sobre el tema tratado pero su venta no supone ninguna forma de asistencia legal, administrativa ni de ningún otro tipo, si se precisase ayuda adicional o experta deberán buscarse los servicios de profesionales competentes. Productos y marcas citados en su contenido estén o no registrados, pertenecen a sus respectivos propietarios.

RC Libros  
Calle Mar Mediterráneo, 2. Nave 6  
28830 SAN FERNANDO DE HENARES, Madrid  
Teléfono: +34 91 677 57 22  
Fax: +34 91 677 57 22  
Correo electrónico: [info@rclibros.es](mailto:info@rclibros.es)  
Internet: [www.rclibros.es](http://www.rclibros.es)

Diseño de colección, cubierta y pre-impresión: Grupo RC  
Impresión y encuadernación: Service Point  
Depósito Legal: M-32774-2014  
Impreso en España

18 17 16 15 14 (1 2 3 4 5 6 7 8 9 10 11 12)

# INTRODUCCIÓN

Podemos definir la Inteligencia de negocio o BI (*Business Intelligence*) como el conjunto de estrategias enfocadas a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o empresa.

El término **inteligencia de negocio** se refiere al uso de datos en una empresa para facilitar la toma de decisiones. Abarca tanto la comprensión del funcionamiento actual de la empresa, como la anticipación de acontecimientos futuros, con el objetivo de ofrecer conocimientos para respaldar las decisiones empresariales. Estamos así ante el proceso de **extracción del conocimiento** o KDD.

Las herramientas de inteligencia de negocios se basan en la utilización de un sistema de información de inteligencia que se forma con distintos datos extraídos de los datos de producción, con información relacionada con la empresa o sus ámbitos y con datos económicos. Mediante las herramientas y técnicas ELT (extraer, cargar y transformar), o actualmente ETL (extraer, transformar y cargar) se extraen los datos de distintas fuentes, se depuran y preparan (homogeneización de los datos) para luego cargarlos en un almacén de datos. La vida o el periodo de éxito de un software de inteligencia de negocios dependerá únicamente del éxito de su uso en beneficio de la empresa. Si esta empresa es capaz de incrementar su nivel financiero, administrativo y sus decisiones mejoran la actuación de la empresa, el software de inteligencia de negocios seguirá presente mucho tiempo, en caso contrario será sustituido por otro que aporte mejores y más precisos resultados.

Asimismo, las herramientas de inteligencia analítica posibilitan el modelado de las representaciones basadas en consultas para crear un cuadro de mando integral que sirve de base para la presentación de informes.

El conjunto de herramientas y metodologías para la inteligencia de negocios tiene en común las siguientes características:

- *Accesibilidad a la información.* Los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos con independencia de su procedencia.
- *Apoyo en la toma de decisiones.* Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- *Orientación al usuario final.* Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

Una clasificación muy simple de las soluciones de Business Intelligence podría ser la siguiente:

- Informes
  - Informes predefinidos
  - Informes a medida
  - Consultas (*Query*) / Cubos OLAP (*On-Line Analytic Processing*).
  - Alertas
- Análisis
  - Análisis estadístico
  - Pronósticos (*Forecasting*)
  - Modelado predictivo o Minería de datos (*Data Mining*)
  - Optimización

La estrategia de la inteligencia de negocios debe ser vista como un proceso creativo, buscar nuevas formas de hacer las cosas, de generar valor en el mundo de continuo cambio, y ser efectivo en el corto plazo, por lo cual se necesita:

- Inteligencia para crear y compartir el conocimiento.
- La habilidad para integrar y administrar este conocimiento.
- La imaginación para visualizar acciones alternativas a las usuales y analizar sus consecuencias.
- La pericia para manejar los recursos y atender las necesidades actuales sin dejar de construir el futuro deseable.

Con la globalización, la competencia se acentúa y hay que reaccionar con rapidez, sorpresa y anticipación. Las herramientas de inteligencia de negocios desempeñan un papel muy importante en este proceso de superación y competitividad en que está inmerso el mercado actual.

Este libro analiza las herramientas más habituales en inteligencia de negocios y sus posibilidades de trabajo. Se utilizará el software adecuado para apoyar las técnicas de Business Intelligence para una extracción óptima del conocimiento contenido en los datos.

A través de ejemplos totalmente resueltos a lo largo del libro se irán presentando las diferentes técnicas de inteligencia de negocios. Los archivos se encuentran en la página web: [www.rclibros.es](http://www.rclibros.es) en la sección *Zona de archivos*.

Se trata de exponer con sencillez y mediante una metodología interactiva los conceptos de extracción del conocimiento e inteligencia de negocios.

# INTELIGENCIA DE NEGOCIOS Y SISTEMAS DE INFORMACIÓN. INFORMES

## FINALIDAD DE LOS SISTEMAS DE INFORMACIÓN Y ORIGEN DEL BUSINESS INTELLIGENCE

La información *reduce nuestra incertidumbre* (sobre algún aspecto de la realidad) y, por tanto, nos permite tomar mejores decisiones.

Inicialmente la finalidad de los sistemas de información era recopilar información sobre una parcela del mundo para ayudar en la toma de decisiones y se basaba en recuentos, censos civiles y militares, libros contables, etc. Actualmente, con la informatización de las organizaciones y la aparición de aplicaciones software operacionales sobre el sistema de información, la finalidad principal de los sistemas de información es dar soporte a los procesos básicos de la organización (ventas, producción, personal, etc.).

Una vez satisfecha la necesidad de tener un soporte informático para los procesos básicos de la organización (**sistemas de información para la gestión**), las organizaciones exigen nuevas prestaciones de los sistemas de información (**sistemas de información para la toma de decisiones**). Es aquí donde aparece el Business Intelligence.

## HERRAMIENTAS PARA LA TOMA DE DECISIONES EN BUSINESS INTELLIGENCE. OLAP, INFORMES Y MINERÍA

Ante el problema de la toma de decisiones han aparecido diferentes herramientas de inteligencia de negocio o DSS que *coexisten*: EIS, OLAP, consultas & informes, minería de datos, etc.

Un EIS (*Executive Information System*) es un sistema de información y un conjunto de herramientas asociadas que tiene las siguientes características:

- Proporciona a los directivos acceso a la información de estado y sus actividades de gestión.
- Está especializado en analizar el estado diario de la organización (mediante indicadores clave) para informar rápidamente sobre *cambios* a los directivos.
- La información solicitada suele ser, en gran medida, numérica (*ventas semanales, nivel de stocks, balances parciales, etc.*) y representada de forma gráfica al estilo de las hojas de cálculo.

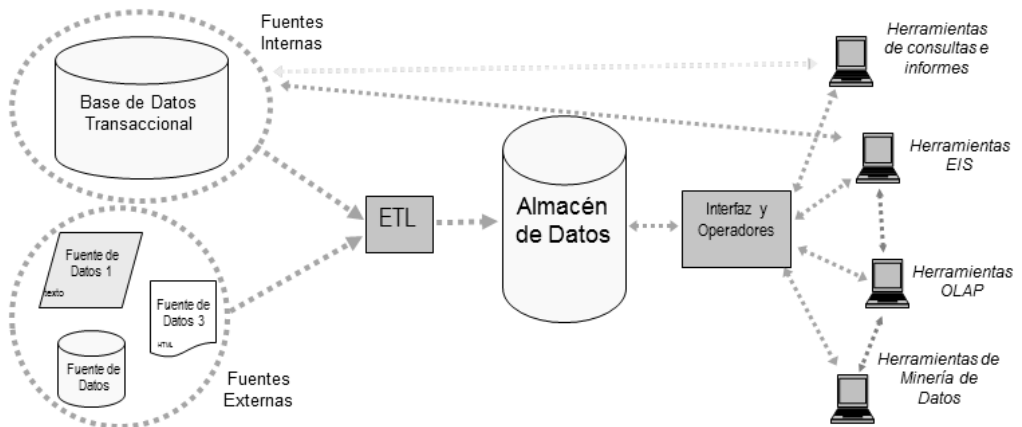
Las *herramientas OLAP (On-Line Analytical Processing)* son más genéricas:

- Funcionan sobre un sistema de información (transaccional o almacén de datos).
- Permiten realizar agregaciones y combinaciones de los datos de maneras más complejas y ambiciosas, con objetivos de análisis más estratégicos.
- Están basadas, generalmente, en sistemas o *interfaces multidimensionales*.
- Se utilizan operadores específicos (además de los clásicos): *drill, roll, pivot, slice & dice,...*
- El resultado se presenta de una manera matricial o híbrida.
- Proporcionan facilidades para “manejar” y “transformar” los datos.
- **Producen otros “datos”** (más agregados, combinados).
- Ayudan a analizar los datos porque producen *diferentes vistas* de los mismos.

Por otro lado, los *sistemas de informes* o *consultas avanzadas* están basados, generalmente, en sistemas relacionales u objeto-relacionales, utilizan los operadores clásicos como concatenación, proyección, selección, agrupamiento... (en SQL y extensiones) y el resultado se presenta de una manera tabular.

Instrumentos más avanzados para la toma de decisiones en inteligencia de negocios son las herramientas de Minería de Datos. Son muy variadas y permiten “extraer” patrones, modelos, descubrir relaciones, regularidades, tendencias, etc. También producen “reglas” o “patrones” (“conocimiento”).

La interrelación entre todas estas herramientas se presenta en el esquema siguiente:



Mediante las herramientas y técnicas ELT (extraer, cargar y transformar), o actualmente ETL (extraer, transformar y cargar) se extraen los datos de distintas fuentes externas e internas (bases de datos transaccionales), se depuran y preparan (homogeneización de los datos) para luego cargarlos en un almacén de datos. En el centro del esquema aparece el *almacén de datos*, que es el “**sistema de información central**” en todo este proceso. Un almacén de datos es una colección de datos orientada a un dominio, integrada, no volátil y variante en el tiempo para ayudar en la toma de decisiones. A partir del almacén de datos, mediante interfaces y operadores se utilizan las herramientas de informes, EIS, OLAP y Minería de Datos.

Los almacenes de datos y las técnicas OLAP son las maneras más efectivas y tecnológicamente más avanzadas para **integrar, transformar y combinar los datos para facilitar** al usuario o a otros sistemas **el análisis de la información**. La tecnología



OLAP generalmente se asocia a los almacenes de datos, aunque podemos tener almacenes de datos sin OLAP, y viceversa.

La minería de datos es solo una etapa del proceso de extracción de conocimiento a partir de datos. Consta de varias fases: Preparación de Datos (selección, limpieza y transformación), Análisis de Datos, Evaluación, Difusión y Uso de Modelos. Incorpora diferentes técnicas como árboles de decisión, regresión lineal, redes neuronales artificiales, técnicas predictivas, técnicas de segmentación, etc. Se aplica en campos diversos como el aprendizaje automático e inteligencia artificial, estadística, bases de datos, clasificación, categorización, estimación y regresión, agrupamiento, etc.

Los almacenes de datos no son *imprescindibles* para hacer extracción de conocimiento a partir de datos. Se puede hacer minería de datos sobre un simple fichero de datos. Las ventajas de organizar un almacén de datos para realizar minería de datos se amortizan sobradamente a medio y largo plazo cuando tenemos grandes volúmenes de datos, o estos aumentan con el tiempo, o provienen de fuentes heterogéneas o se van a combinar de maneras arbitrarias y no predefinidas.

## ALMACENES DE DATOS. DATA WAREHOUSE

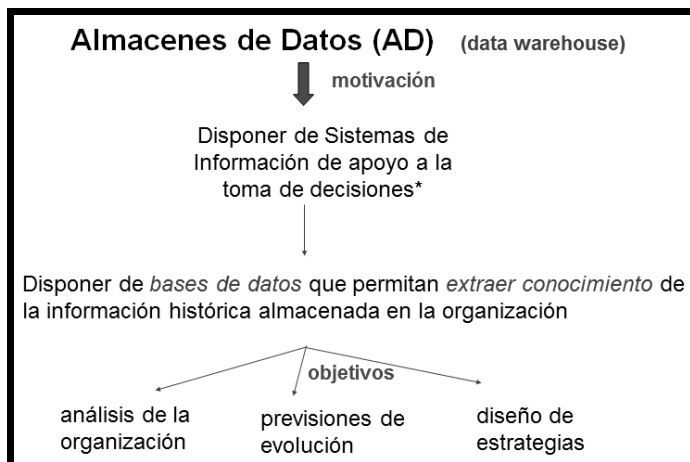
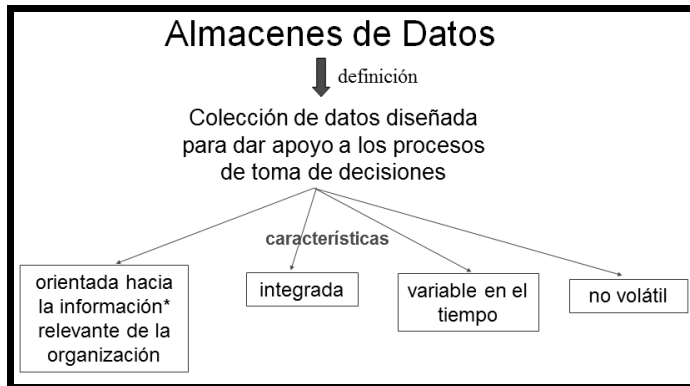
Generalmente, la información que se quiere investigar sobre un cierto dominio de la organización se encuentra en bases de datos y otras fuentes muy diversas, tanto internas como externas. Muchas de estas fuentes son las que se utilizan para el trabajo diario (*bases de datos operacionales*). Sobre estas mismas bases de datos de trabajo ya se puede extraer conocimiento (visión tradicional).

Una *base de datos transaccional* es una fuente de datos mediante la cual se mantiene el trabajo transaccional diario de los sistemas de información originales (conocido como OLTP, *On-Line Transactional Processing*). También se hacen análisis de los datos en tiempo real sobre la misma base de datos (conocido como OLAP, *On-Line Analytical Processing*).

Como problemas más comunes, la base de datos transaccional perturba el trabajo transaccional diario de los sistemas de información originales (“killer queries”). Se debe hacer por la noche o en fines de semana. Además, la base de datos está diseñada para el trabajo transaccional, no para el análisis de los datos. Generalmente no puede ser en tiempo real.

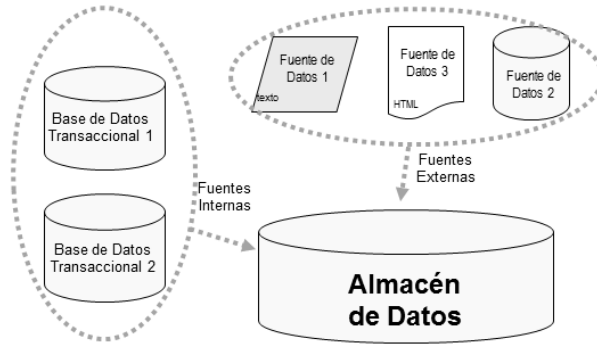
Para operar eficientemente con los datos, los costes de almacenamiento masivo y conectividad se han reducido drásticamente en los últimos años. Parece razonable recopilar los datos (información histórica) en un sistema separado y específico. Aparece así el *Data warehouse* (Almacén o Bodega de Datos).

Los esquemas siguientes muestran la definición, la motivación y los objetivos de los almacenes de datos.

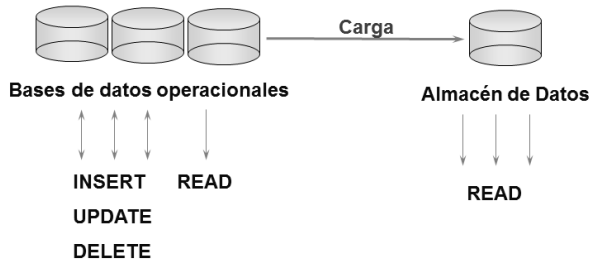


Un almacén de datos siempre está orientado hacia la información relevante de la organización. Se diseña para consultar eficientemente información relativa a las actividades (ventas, compras, producción...) básicas de la organización y no para soportar los procesos que se realizan en ella (gestión de pedidos, facturación, etc.).

Un almacén de datos integra datos recopilados de diferentes sistemas operacionales de la organización, incluyendo bases de datos transaccionales y/o fuentes externas.



Los datos en un almacén de datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente. Los datos son almacenados como fotos (*snapshots*) correspondientes a periodos de tiempo. Además, los datos almacenados no son actualizados, solo son incrementados. Las operaciones de inserción, actualización y borrado de los datos se realizan en la base de datos operacional antes de que sean cargados en el almacén de datos.



Los almacenes de datos presentan múltiples ventajas para las organizaciones entre las que destacan la rentabilidad de las inversiones realizadas para su creación, el aumento de la competitividad en el mercado y el aumento de la productividad de los técnicos de dirección. Pero también presentan problemas como la infravaloración del esfuerzo necesario para su diseño y creación, la infravaloración de los recursos necesarios para la captura, la carga y el almacenamiento de los datos, el incremento continuo de los requisitos de los usuarios y la privacidad de los datos. El esquema siguiente presenta las diferencias esenciales entre una base de datos operacional y un almacén de datos.

Sistema Operacional (OLTP)	Almacén de datos (DW)
<ul style="list-style-type: none"> <li>- almacena datos actuales</li> <li>- almacena datos de detalle</li> <li>-bases de datos medianas (100Mb-1Gb)</li> <li>- los datos son dinámicos (actualizables)</li> <li>- los procesos (transacciones) son repetitivos</li> <li>- el número de transacciones es elevado</li> <li>- tiempo de respuesta pequeño (segundos)</li> <li>- dedicado al procesamiento de transacciones</li> <li>- orientado a los procesos de la organización</li> <li>- soporta decisiones diarias</li> <li>- sirve a muchos usuarios (administrativos)</li> </ul>	<ul style="list-style-type: none"> <li>- almacena datos históricos</li> <li>- almacena datos de detalle y datos agregados a distintos niveles</li> <li>- bases de datos grandes (100Gb-1Tb)</li> <li>- los datos son estáticos</li> <li>- los procesos no son previsibles</li> <li>- el número de transacciones es bajo o medio</li> <li>- tiempo de respuesta variable (segundos-horas)</li> <li>- dedicado al análisis de datos</li> <li>- orientado a la información relevante</li> <li>- soporta decisiones estratégicas</li> <li>- sirve a técnicos de dirección</li> </ul>

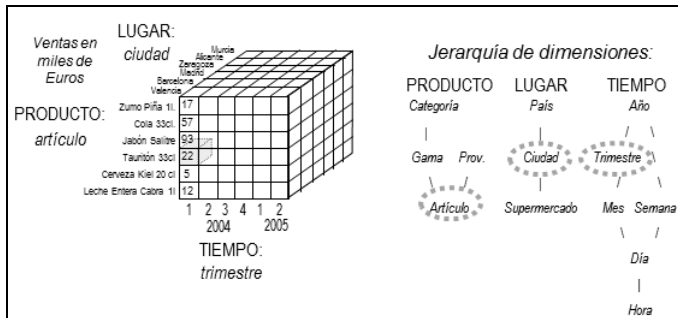
La *arquitectura* de un AD viene determinada por su situación central como fuente de información para las herramientas de análisis, tal y como se muestra en la primera ilustración de este capítulo (página 3).

Las *componentes* típicas de un almacén de datos pueden enumerarse como se indica a continuación:

- Sistema ETL (*Extraction, Transformation, Load*): realiza las funciones de *extracción* de las fuentes de datos (transaccionales o externas), *transformación* (limpieza, consolidación...) y la *carga* del AD, realizando:
  - extracción de los datos.
  - filtrado de los datos: limpieza, consolidación, etc.
  - carga inicial del almacén: ordenación, agregaciones, etc.
  - refresco del almacén: operación periódica que propaga los cambios de las fuentes externas al almacén de datos.
- Repositorio Propio de Datos: información relevante, metadatos.
- Interfaces y Gestores de Consulta: permiten acceder a los datos y sobre ellos se conectan herramientas más sofisticadas (OLAP, EIS, minería de datos).
- Sistemas de Integridad y Seguridad: se encargan de un mantenimiento global, copias de seguridad...

Las *herramientas de explotación* de los almacenes de datos han adoptado un **modelo multidimensional de datos**. En un esquema multidimensional se representa una actividad que es objeto de análisis (*hecho*) y las dimensiones que caracterizan la actividad (*dimensiones*). La información relevante sobre el hecho (*actividad*) se representa por un conjunto de indicadores (*medidas o atributos de hecho*). La información descriptiva de cada dimensión se representa por un conjunto de atributos (*atributos de dimensión*). Entre los atributos de una dimensión se definen *jerarquías*.

Se pueden obtener hechos a diferentes niveles de agregación. Es posible la obtención de medidas sobre los hechos parametrizadas por atributos de las dimensiones y restringidas por condiciones impuestas sobre las dimensiones. Un nivel de agregación para un conjunto de dimensiones se denomina *cubo*. El cubo que se muestra a continuación ilustra el hecho de las ventas en miles de euros de un artículo (PRODUCTO) en determinadas ciudades (LUGAR) en distintos momentos del tiempo (TIEMPO). De esta forma observamos la jerarquía de dimensiones PRODUCTO → LUGAR → TIEMPO.



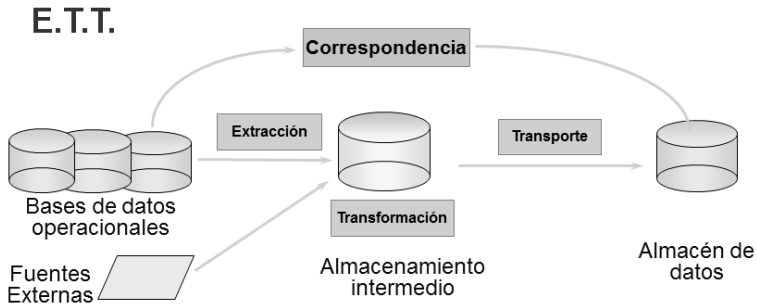
La información de un almacén de datos se recopila en varios esquemas, cada uno de los cuales se denomina *datamart*. Los datamarts se definen para satisfacer las necesidades de un departamento o sección de la organización y contienen menos información de detalle y más información agregada. El almacén de datos puede estar formado por varios datamarts y, opcionalmente, por tablas adicionales.

El sistema encargado de la carga y mantenimiento del almacén de datos es el Sistema E.T.T. (Extracción - Transformación -Transporte). La construcción del Sistema E.T.T. es responsabilidad del equipo de desarrollo del almacén de datos. El Sistema E.T.T. es construido específicamente para cada almacén de datos. Aproximadamente 50% del esfuerzo. En la construcción del E.T.T. se pueden utilizar herramientas del mercado o programas diseñados específicamente.

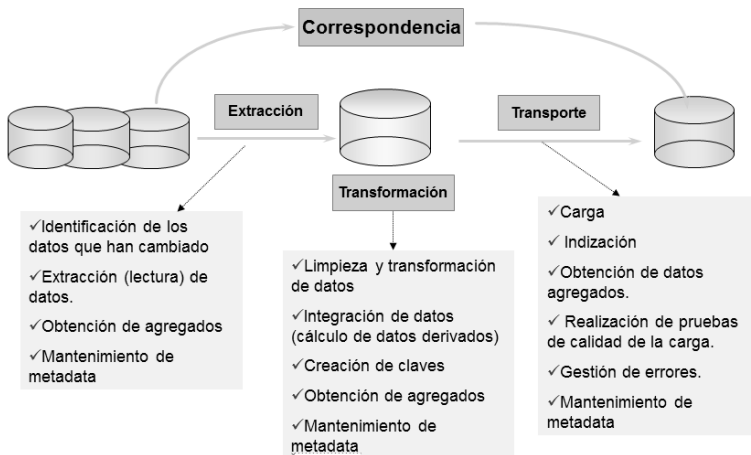
Las funciones del Sistema E.T.T. son la carga inicial (*initial load*) y el mantenimiento o *refresco* periódico: inmediato, diario, semanal, mensual... (*refreshment*).

El sistema E.T.T. es conocido también por E.T.L. (Extracción – Transformación – Load o carga).

En el proceso de carga y mantenimiento, a partir de fuentes externas de las que se realiza la extracción (que generalmente son bases de datos operacionales), se lleva a cabo el almacenamiento intermedio previo al transporte al almacén de datos. El almacenamiento intermedio permite realizar transformaciones sin paralizar las bases de datos operacionales y el almacén de datos, almacenar metadatos y facilitar la integración de fuentes externas. El esquema siguiente ilustra estas fases:



El esquema que se presenta a continuación especifica las tareas a realizar en cada una de las fases del proceso de carga y mantenimiento de un almacén de datos.



## HERRAMIENTAS OLAP, ROLAP Y MOLAP

Las herramientas de OLAP presentan al usuario una visión multidimensional de los datos (esquema multidimensional) para cada actividad que es objeto de análisis.

El usuario formula consultas a la herramienta OLAP seleccionando atributos de este esquema multidimensional sin conocer la estructura interna (esquema físico) del almacén de datos. Una consulta a un almacén de datos consiste generalmente en la obtención de medidas sobre los hechos parametrizadas por atributos de las dimensiones y restringidas por condiciones impuestas sobre las dimensiones. La herramienta OLAP genera la correspondiente consulta y la envía al gestor de consultas del sistema (p.ej. mediante una sentencia SELECT).

Por ejemplo, podemos plantearnos una consulta del tipo “Importe total de las ventas durante este año de los productos del departamento **bebidas**, por **trimestre** y por **categoría**”. En este caso, las restricciones son: productos del departamento bebidas y ventas durante este año, mientras que los parámetros de la consulta **son**: por categoría de producto y por trimestre.

Se pueden presentar en forma tabular (relacional) los datos seleccionados asumiendo dos categorías en el departamento de *bebidas*: refrescos y zumos.

Categoría	Trimestre	Ventas
Refrescos	T1	2000000
Refrescos	T2	1000000
Refrescos	T3	3000000
Refrescos	T4	2000000
Zumos	T1	1000000
Zumos	T2	1500000
Zumos	T3	8000000
Zumos	T4	2400000

Pero también se puede realizar una presentación matricial (multidimensional) de los datos seleccionados.

trimestre categoria	T1	T2	T3	T4
Refrescos	2000000	1000000	3000000	2000000
Zumos	1000000	1500000	8000000	2400000

Los parámetros de la consulta (“por trimestre” y “por categoría”) determinan los criterios de agrupación de los datos seleccionados (ventas de productos del departamento *Bebidas* durante este año). La agrupación se realiza sobre dos dimensiones (Producto, Tiempo).

Pero lo interesante no es poder realizar consultas que, en cierto modo, se pueden hacer con selecciones, proyecciones, concatenaciones y agrupamientos tradicionales. Lo realmente interesante de las herramientas OLAP son sus operadores de refinamiento o manipulación de consultas como DRILL, ROLL, SLICE & DICE y PIVOT.

El carácter agregado de las consultas en el Análisis de Datos, aconseja la definición de nuevos operadores que faciliten la agregación (consolidación) y la disgregación (división) de los datos. Para la agregación tenemos el operador ROLL, que permite eliminar un criterio de agrupación en el análisis, agregando los grupos actuales. Para la desagregación tenemos el operador DRILL, que permite introducir un nuevo criterio de agrupación en el análisis, disgregando los grupos actuales.

Por ejemplo, en el esquema siguiente, mediante una operación DRILL, cada grupo (categoría-trimestre) de la consulta original se disgrega en dos nuevos grupos (categoría-trimestre-ciudad) para las ciudades de León y Valencia.