

Universidad Nacional del Nordeste
Facultad de Ciencias Exactas y
Naturales y Agrimensura

Data

Warehouse

Rojas, Mariana Isabel
Prof. Director: Mgter. David Luis La Red Martinez
Licenciatura en Sistemas de Información
Corrientes-Argentina
2009

Introducción

- En la actualidad, el dinámico mundo de los negocios plantea la necesidad de disponer de un acceso rápido y sencillo a información para la toma de decisiones.
- Los niveles gerenciales necesitan a menudo tomar decisiones de alto nivel, cruciales para el funcionamiento de la empresa.

Introducción

- Frecuentemente se basan en su experiencia, utilizando un enfoque subjetivo del proceso decisorio.
- Este enfoque no es apto para las condiciones del mundo actual en el que los sistemas de gestión de calidad vigentes han demostrado la importancia de la toma de decisiones basada en cifras, datos y hechos.

Introducción

- El Data Warehouse permite que los gerentes tomen decisiones siguiendo un enfoque racional, basados en información confiable y oportuna.
- Consiste básicamente en la transformación de los datos operacionales en información útil para decidir.
- El uso del Data Warehouse permite también encontrar relaciones ocultas entre los datos y predecir el comportamiento futuro bajo condiciones dadas.

Introducción

- Los datos operacionales que sirven de entrada al Data Warehouse generalmente están dispersos en distintos sistemas de la organización, desarrollados en diferentes entornos de desarrollo, por diferentes personas y en diferentes momentos.

Introducción

- Es tarea fundamental del Data Warehouse recolectarlos, unificarlos y depurarlos según las necesidades del negocio, eliminando inconsistencias y conservando sólo la información útil para los objetivos empresariales.
- Esto se lleva a cabo mediante procesos que se ejecutan periódicamente y conducen a mantener la información actualizada.

¿Qué es un Data Warehouse?

- Un **almacén de datos** (del inglés *data warehouse*) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

¿Qué es un Data Warehouse?

- Data warehousing es el centro de la arquitectura para los sistemas de información desde la década de los '90.
- Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis.

¿Qué es un Data Warehouse?

- Facilita la integración de sistemas de aplicación no integrados.
- Organiza y almacena los datos que se necesitan para el procesamiento analítico, informático, sobre una amplia perspectiva de tiempo.

Sistemas de Información

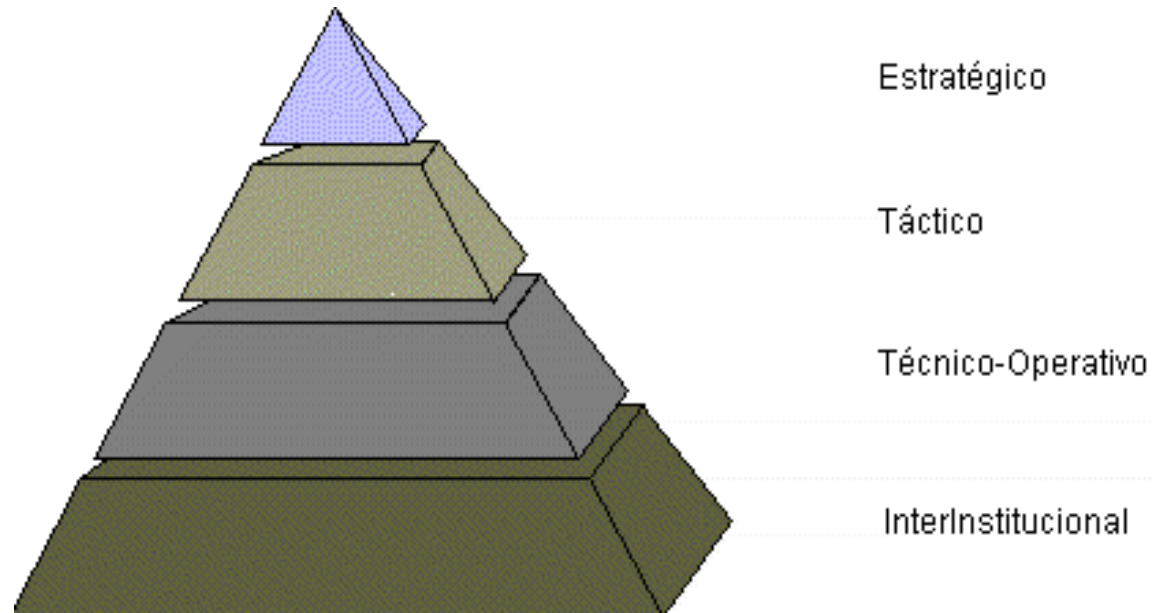
- Estos sistemas están relacionados con el análisis de los datos y la toma de decisiones, frecuentemente, decisiones importantes sobre cómo operará la empresa, ahora y en el futuro.
- Estos sistemas no sólo tienen un enfoque diferente al de los operacionales, sino que, por lo general, tienen un alcance diferente.

Sistemas de Información

- Mientras las necesidades de los datos operacionales se enfocan normalmente hacia una sola área, los datos para el soporte de decisiones, con frecuencia, toman un número de áreas diferentes y necesitan cantidades grandes de datos operacionales relacionadas.
- Son estos sistemas sobre los que se basa la tecnología Data Warehousing.

Sistemas de Información

- Los sistemas de información se han dividido de acuerdo al siguiente esquema:



Sistemas de Información

- **Sistemas Estratégicos**, orientados a soportar la toma de decisiones, facilitan la labor de la dirección, proporcionándole un soporte básico, en forma de mejor información, para la toma de decisiones.
- Se caracterizan porque son sistemas sin carga periódica de trabajo, es decir, su utilización no es predecible.

Sistemas de Información

- Destacan entre estos sistemas:
 - Sistemas de Información Gerencial (MIS).
 - Sistemas de Información Ejecutivos (EIS).
 - Sistemas de Simulación de Negocios (BIS y que en la práctica son Sistemas Expertos o de Inteligencia Artificial-AI).

Sistemas de Información

- **Sistemas Tácticos**, diseñados para soportar las actividades de coordinación de actividades y manejo de documentación, definidos para facilitar consultas sobre información almacenada en el sistema, proporcionar informes y, en resumen, facilitar la gestión independiente de la información por parte de los niveles intermedios de la organización.

Sistemas de Información

- Destacan entre ellos:
 - Sistemas Ofimáticos (OA).
 - Sistemas de Transmisión de Mensajería (E-mail y Fax Server).
 - Coordinación y control de tareas (Work Flow).
 - Tratamiento de documentos (Imagen, Trámite y Bases de Datos Documentarios).

Sistemas de Información

- **Sistemas Técnico-Operativos**, que cubren el núcleo de operaciones tradicionales de captura masiva de datos (Data Entry) y servicios básicos de tratamiento de datos, con tareas predefinidas (contabilidad, facturación, almacén, presupuesto, personal y otros sistemas administrativos).

Sistemas de Información

- **Sistemas Interinstitucionales**, este último nivel de sistemas de información recién está surgiendo, es consecuencia del desarrollo organizacional orientado a un mercado de carácter global, todo esto a partir de la generalización de las redes informáticas de alcance nacional y global (Internet), que se convierten en vehículo de comunicación entre la organización y el mercado.

Sistemas de Información

- Sin embargo, la tecnología data warehousing basa sus conceptos y diferencias entre dos tipos fundamentales de sistemas de información en todas las organizaciones:
 - *Sistemas técnico-operacionales.*
 - *Sistemas de soporte de decisiones.* Este último es la base de un data warehouse.

Datos Operacionales y Datos Informativos

- El proceso automatizado de un negocio utiliza *datos operacionales*, los que constituyen el conjunto de registros de las transacciones del negocio.
- Es difícil para los usuarios del negocio tener acceso a los datos operacionales debido a las limitaciones de performance y tecnología.

Datos Operacionales y Datos Informativos

- Lo que el usuario del negocio necesita como entrada a sus actividades de análisis son *datos informativos*.
- Estos son una combinación de datos operacionales que han sido modificados, depurados, transformados, consolidados y organizados desde diversas fuentes externas al proceso del negocio.

Datos Operacionales y Datos Informativos

- Los datos operacionales son específicos para cada aplicación y usualmente son almacenados de manera separada por otras aplicaciones.
- En cambio, para los datos informativos, el usuario necesita datos que crucen por varias aplicaciones, que estén reorganizados por temas de negocio, que se encuentren disponible para análisis durante períodos largos y que sea accesible de manera fácil y flexible.

Datos Operacionales y Datos Informativos

- Los datos operacionales son manejados, precisamente, por los sistemas operacionales o transaccionales (On Line Transactional Processing, OLTP):
 - Se concentran en la administración y la medición de indicadores empresariales (capital e inversión), indicadores financieros (márgenes de utilidades, rotación de inventarios), indicadores de ventas (identificación de clientes persistentes), etc.

Datos Operacionales y Datos Informativos

- Por su parte, los datos informativos son los que conforman un Data Warehouse, el cual tiene como fin comprender, medir y administrar parámetros empresariales estratégicos, como el crecimiento del ingreso y rentabilidad, la participación del mercado y los segmentos del cliente.

Sistemas de Soporte de Decisiones

Datos Operacionales	Datos Informativos
● Orientados a una aplicación	● Orientados a un tema
● Integración limitada	● Integrados
● Constantemente actualizados	● No volátiles
● Sólo valores actuales	● Valores a lo largo del tiempo
● Soportan operaciones diarias	● Soportan decisiones de administración

Características

Orientado a Temas:

- 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 nivel detallado.
- 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.

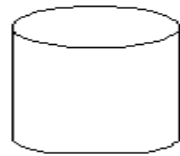
Características

- 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.

Características

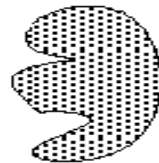
- En el ambiente data warehousing se organiza alrededor de sujetos tales como cliente, vendedor, producto y actividad. Por ejemplo, para un fabricante, éstos pueden ser clientes, productos, proveedores y vendedores.
- Para una universidad pueden ser estudiantes, clases y profesores. Para un hospital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

Características



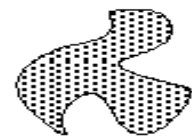
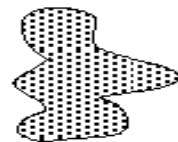
Operacional

préstamos



ahorros

tarjeta
bancaria

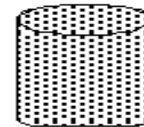


depósito

orientación a una aplicación

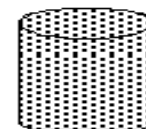
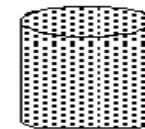


Data warehouse



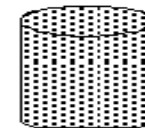
cliente

vendedor



producto

actividad



orientación al tema

Características

Integración:

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

Características

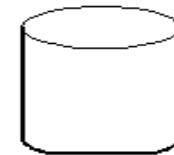
- La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros.

Características

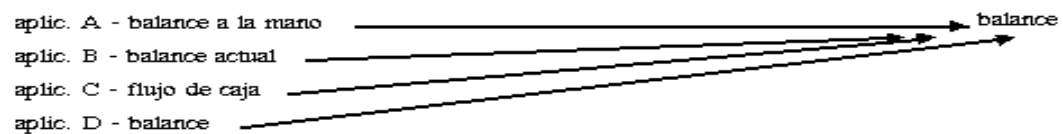
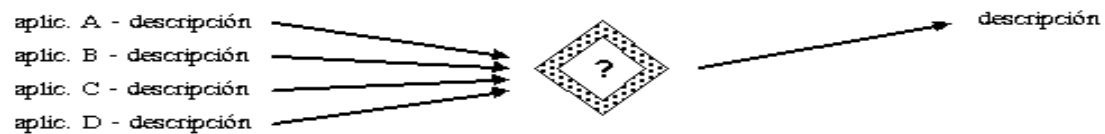
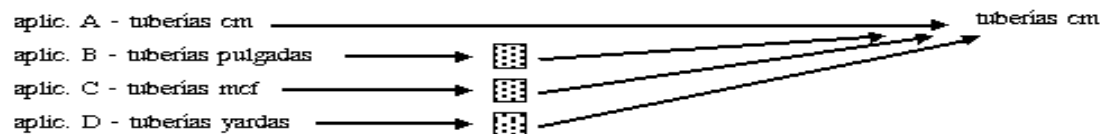
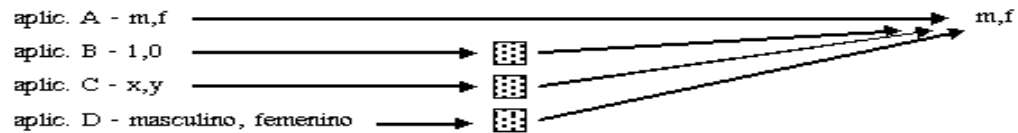


Operacional

Integración



Data warehouse



Características

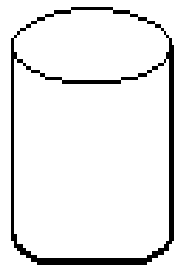
De Tiempo Variante:

- 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.

Características

- 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

Características

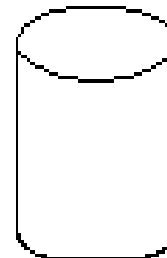


Operacional

Valor actual de los datos:

- Horizonte de tiempo: 60-90 días
- La clave puede, como no, tener un elemento de tiempo
- Los datos pueden ser actualizados

De tiempo variante



Data warehouse

Datos instantáneos:

- Horizonte de tiempo: 5-10 años
- La clave contiene un elemento de tiempo
- Una vez que el snapshot se realice, el registro no puede ser actualizado

Características

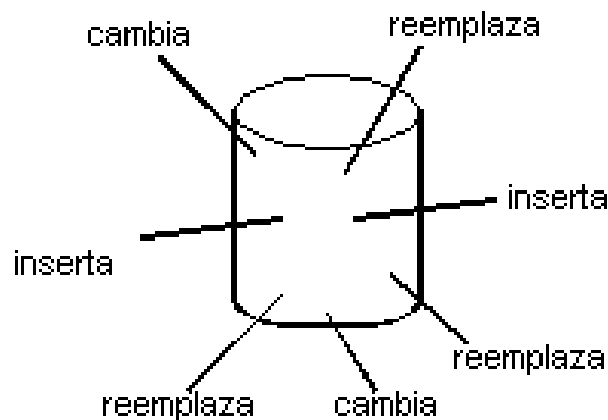
No Volátil:

- 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.

Características

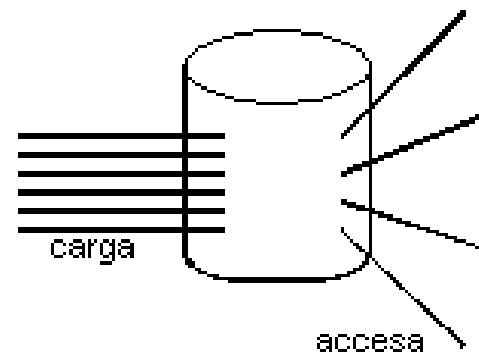
- La manipulación básica de los datos que ocurre en el data warehouse es mucho más simple.
- Hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No hay actualización de datos (en el sentido general de actualización) en el depósito, como una parte normal de procesamiento.

Características



Operacional

Normalmente, la data es actualizada registro por registro



Data warehouse

La data es cargada en el depósito de datos y es accesada allí, pero una vez que el snapshot está hecho, los datos en el depósito no cambian

Estructura de un Data Warehouse

- Detalle de datos actuales: En gran parte, el interés más importante radica en el detalle de los datos actuales, debido a que:
 - Refleja las ocurrencias más recientes, las cuales son de gran interés.
 - Es voluminoso, ya que se almacena al más bajo nivel de granularidad.
 - Casi siempre se almacena en disco, el cual es de fácil acceso, aunque su administración sea costosa y compleja.

Estructura de un Data Warehouse

- Detalle de datos antiguos: La data antigua es aquella que se almacena sobre alguna forma de almacenamiento masivo.
- No es frecuentemente accesada y se almacena a un nivel de detalle, consistente con los datos detallados actuales.
- Es poco usual utilizar el disco como medio de almacenamiento.

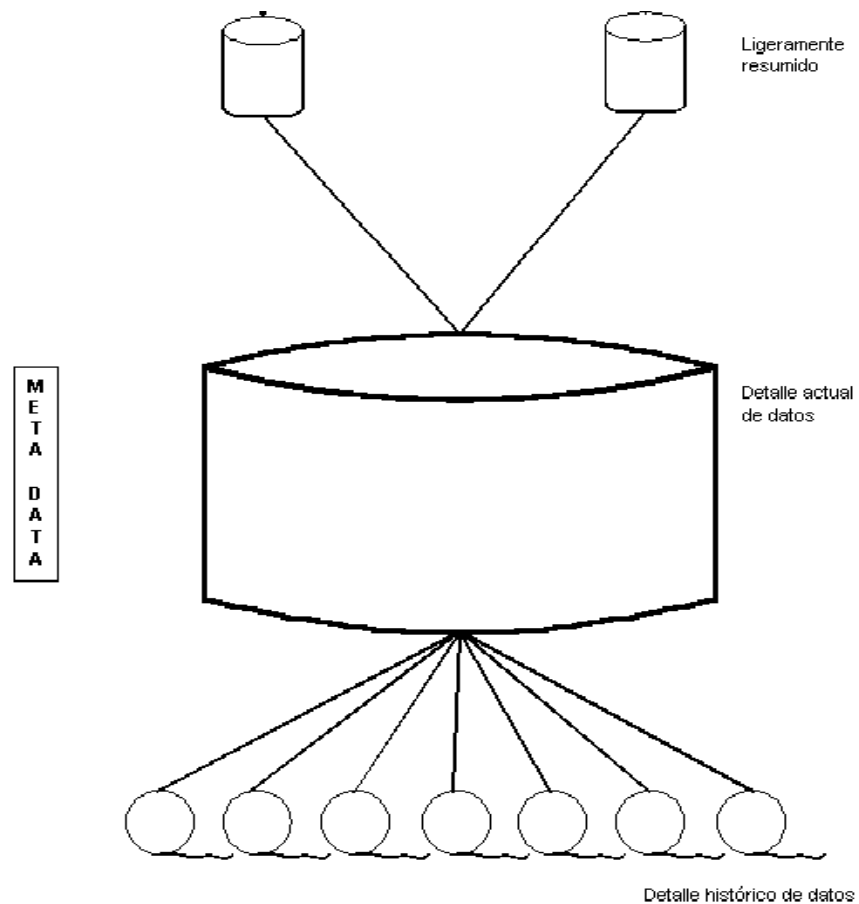
Estructura de un Data Warehouse

- Datos ligeramente resumidos: La data ligeramente resumida es aquella que proviene desde un bajo nivel de detalle encontrado al nivel de detalle actual.
- Este nivel del data warehouse casi siempre se almacena en disco. Los puntos en los que se basa el diseñador para construirlo son:
 - Que la unidad de tiempo se encuentre sobre la esquematización hecha.
 - Qué contenidos (atributos) tendrá la data ligeramente resumida.

Estructura de un Data Warehouse

- Estos datos son compactos y fácilmente accesibles.
- A veces se encuentra en el ambiente del data warehouse y en otros, fuera del límite de la tecnología que ampara al data warehouse (de todos modos, los datos completamente resumidos son parte del data warehouse sin considerar dónde se alojan los datos físicamente).

Estructura de un Data Warehouse



Estructura de un Data Warehouse

- Metadata: El componente final del data warehouse es el de la metadata.
- La metadata se sitúa en una dimensión diferente a la de otros datos del data warehouse, debido a que su contenido no es tomado directamente desde el ambiente operacional.

Estructura de un Data Warehouse

- La metadata es usada como:
 - Un directorio para ayudar al analista a ubicar los contenidos del data warehouse.
 - Una guía para el mapping de datos de cómo se transforma, del ambiente operacional al de data warehouse.
 - Una guía de los algoritmos usados para la esquematización entre el detalle de datos actual, con los datos ligeramente resumidos y éstos, con los datos completamente resumidos, etc.

Inteligencia de Negocio

- El término inteligencia empresarial se refiere al uso de los datos de una empresa para facilitar la toma de decisiones a las personas que deciden, es decir, la comprensión del funcionamiento actual y la anticipación de acciones para dar una dirección bien informada a la empresa.

Inteligencia de Negocio

- Las herramientas de inteligencia 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.

Inteligencia de Negocio

- Este conjunto de herramientas y metodologías tienen en común las siguientes características:
 - *Orientación al usuario final.*
 - *Apoyo en la toma de decisiones.*
 - *Accesibilidad a la información.*

Extracción, Transformación y Carga (ETL)

- Para poblar el Data Warehouse se deben mover bloques de datos, muchas veces desde diferentes sistemas operativos, estructuras de archivos y bases de datos, mediante procesos programados que se ejecutan frecuentemente fuera del horario de trabajo para no insumir tiempo de procesamiento del hardware de la empresa, entorpeciendo la operatoria de la misma.

Extracción, Transformación y Carga (ETL)

Extracción:

- El propósito principal de la fase de *extracción* es capturar y copiar los datos requeridos de uno o más sistemas operacionales o fuentes de datos.
- Los datos que se extraen son colocados en un archivo intermedio con un formato definido, que luego será utilizado por la siguiente fase del proceso.

Extracción, Transformación y Carga (ETL)

Transformación:

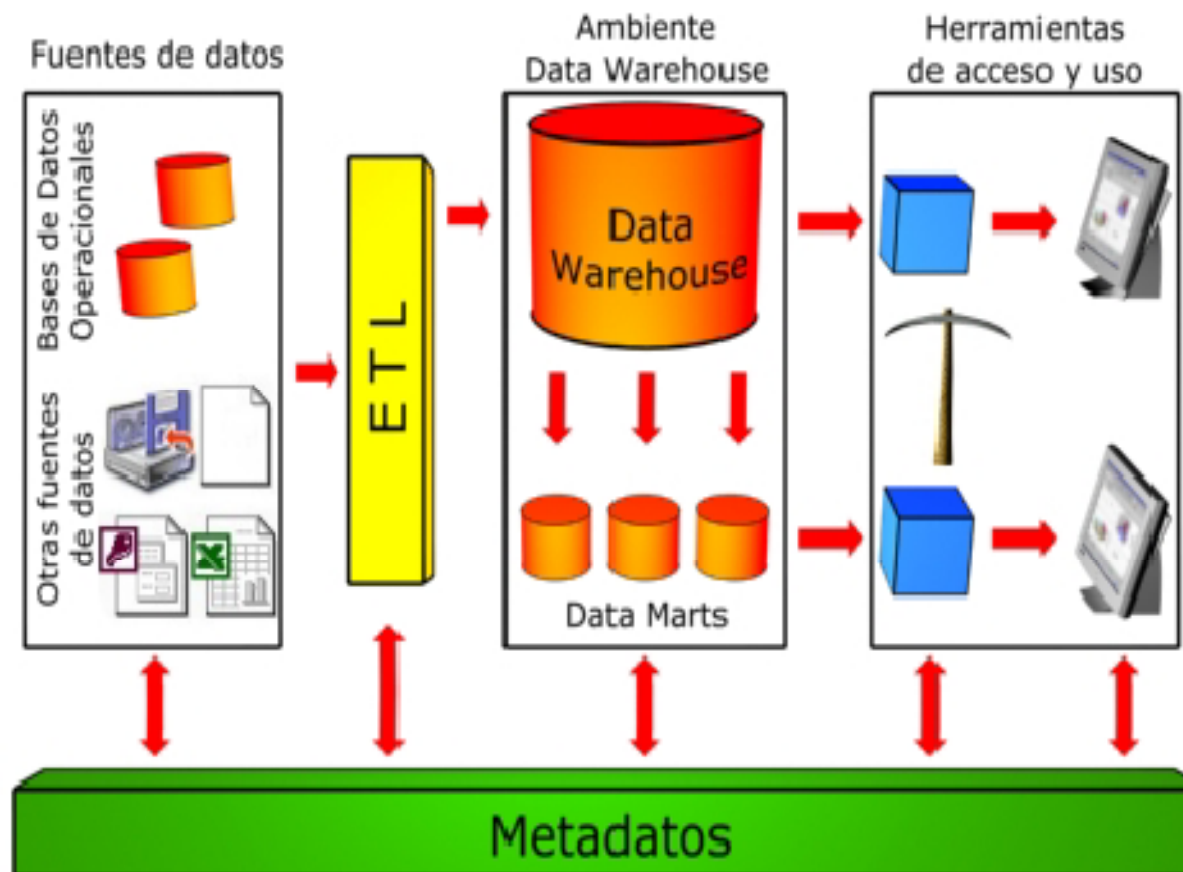
- Las funciones básicas a ser realizadas en esta fase consisten en leer los archivos intermedios generados por la fase de extracción, realizar las transformaciones necesarias, construir los registros en el formato del Data Warehouse y crear un archivo de salida conteniendo todos los registros nuevos a ser cargados en el Data Warehouse.

Extracción, Transformación y Carga (ETL)

Carga:

El objetivo de esta fase consiste en tomar los registros formateados por la fase de transformación y cargarlos en el Data Warehouse, que es el contenedor para todos los datos informativos (actuales e históricos) requeridos por las operaciones del Data Warehouse.

Extracción, Transformación y Carga (ETL)



Modelado de Datos

- Para comprender uno de los aspectos más relevantes de la arquitectura del Data Warehouse, como es el modelado de datos, es necesario establecer primero las diferencias sustanciales entre los dos mundos de modelado existentes: entidad-relación (ER) y dimensional.

El Modelo Relacional

- En la mayoría de los sistemas transaccionales el objetivo del modelo es garantizar la integridad de los datos, además de eliminar cualquier tipo de redundancia en los datos.
- Este enfoque es perfecto para los entornos de procesamiento transaccional, porque las transacciones son muy simples y deterministas.

El Modelo Dimensional

- El modelado dimensional es una técnica de diseño lógico que busca presentar la información en un marco estándar e intuitivo que permita un acceso de alto rendimiento.
- El modelado dimensional es esencialmente útil para resumir y organizar los datos y para soportar el análisis de la misma. Existen algunos conceptos básicos para comprender la filosofía de este tipo de modelado: áreas temas, medidas, dimensiones y hechos.

El Modelo Dimensional

- Las áreas tema en conjunto constituyen el ámbito de implementación del Data Warehouse.
- Una *medida* o *indicador* es un cuantificador del desempeño de un ítem o una actividad del negocio.
- La información que brinda una medida es usada por los usuarios en sus consultas para evaluar el desempeño de un área tema.

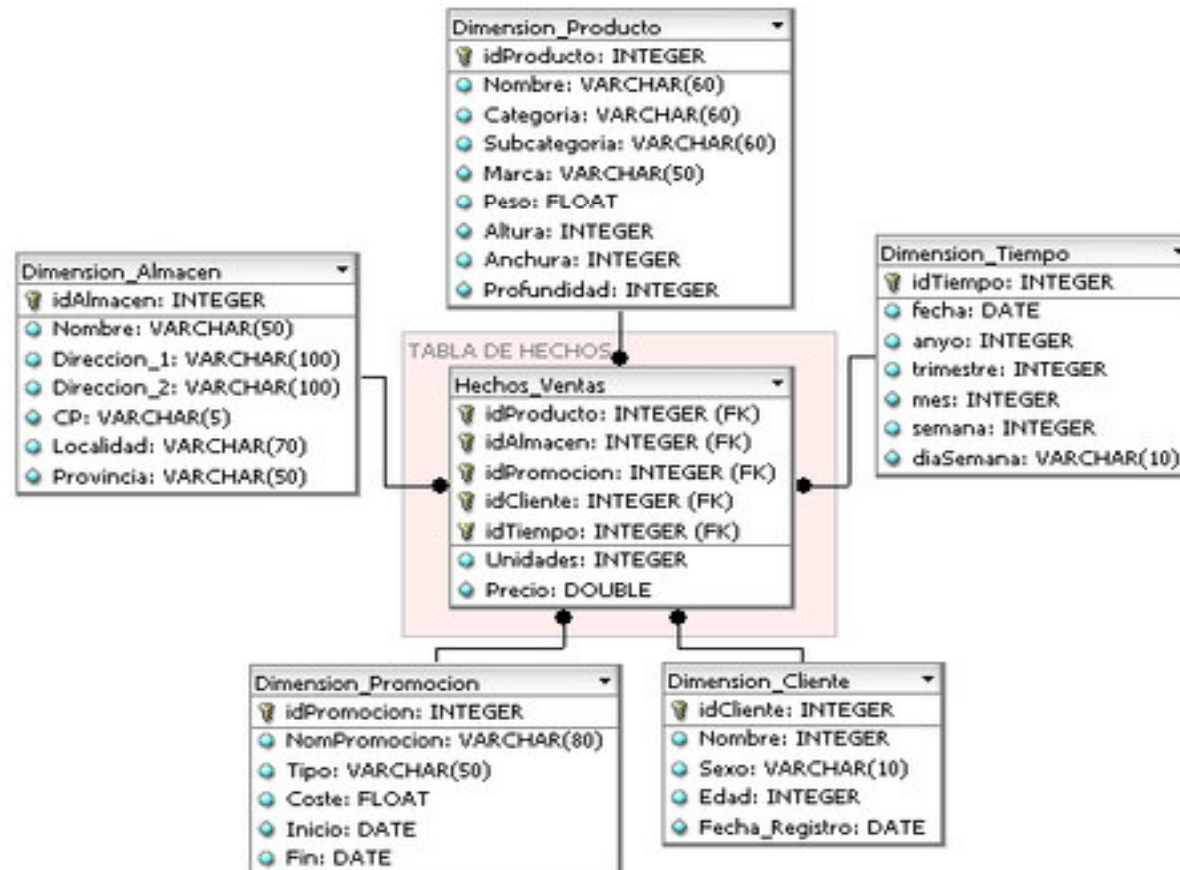
El Modelo Dimensional

- Una *dimensión* es una colección de miembros o entidades del mismo tipo y constituye un calificador conceptual que provee el contexto o significado para una medida.
- La forma de representar la organización de los datos en un modelo dimensional es a través de un *cubo* (el cual no necesariamente debe tener tres dimensiones).

El Modelo Dimensional

- Existen principalmente dos *esquemas para el modelo dimensional*: el *esquema estrella* (star), y el *esquema copo de nieve* (snowflake).
- En el *esquema estrella*, cada *modelo dimensional* está compuesto de una *tabla central* con una *clave primaria compuesta*, denominada *tabla de hechos*, y un conjunto de *tablas periféricas* denominadas *tablas de dimensiones*.

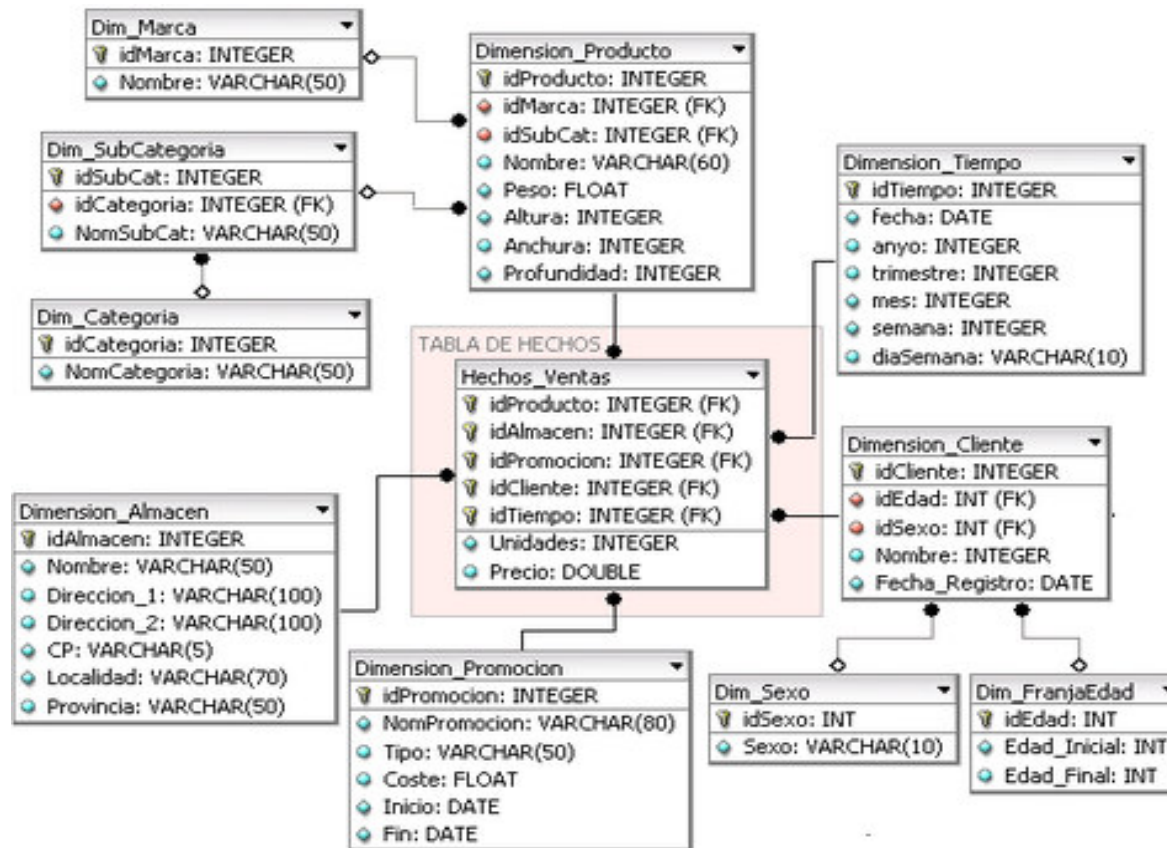
El Modelo Dimensional



El Modelo Dimensional

- Esquema en copo de nieve (bola de nieve) es una variedad más compleja del esquema estrella. El afinamiento está orientado a facilitar mantenimiento de dimensiones.

El Modelo Dimensional



Herramientas de Acceso y Uso

- Sin las herramientas adecuadas de acceso y análisis el Data Warehouse se puede convertir en una mezcla de datos sin ninguna utilidad. Es necesario poseer técnicas que capturen los datos importantes de manera rápida y puedan ser analizados desde diferentes puntos de vista.
- Actualmente a este tipo de herramientas se las conocen como Herramientas de Inteligencia de Negocio

Herramientas de Acceso y Uso

-OLAP (On Line Analytical Processing):

- En un Data Warehouse se depositan datos para consulta y análisis, a diferencia del Procesamiento Transaccional en Línea (OLTP), en donde los datos se almacenan para operación y control.
- El Procesamiento Analítico en Línea (OLAP) es una tecnología de análisis de datos que crea nueva información empresarial a partir de los datos existentes.

Herramientas de Acceso y Uso

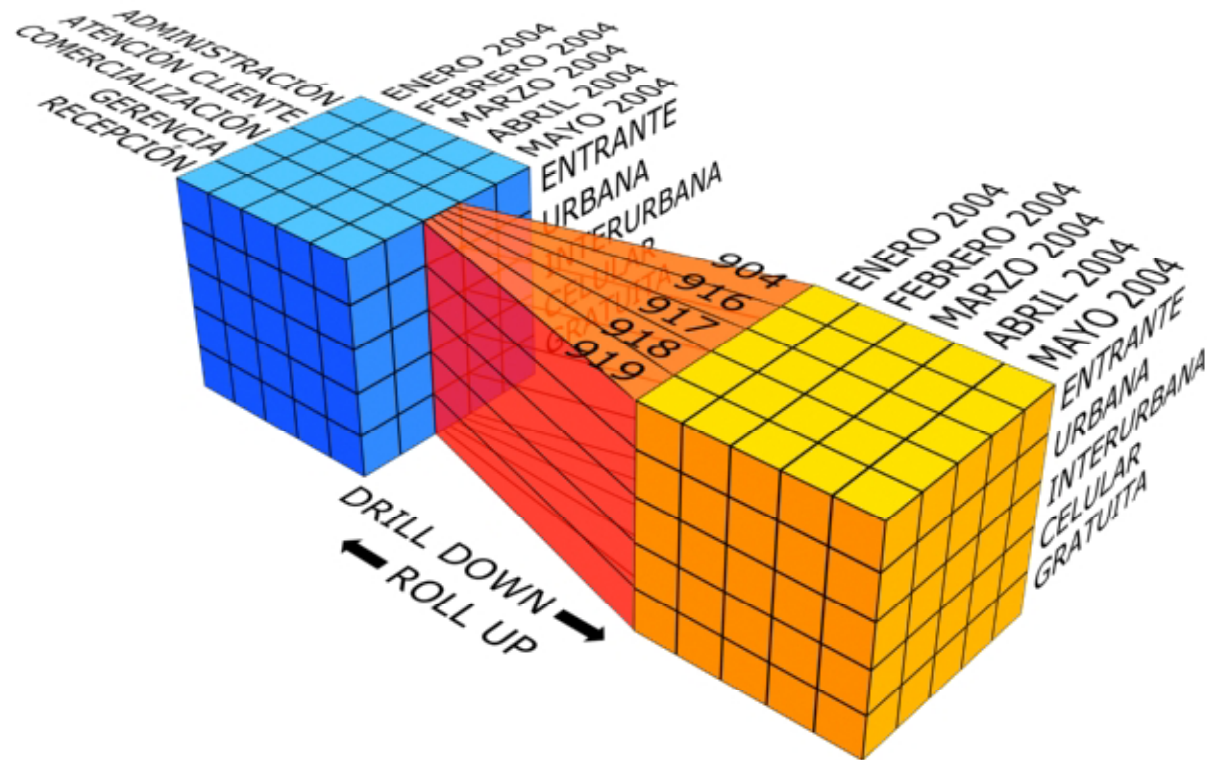
-Drill Down y Roll Up:

- Una de las características del acceso a la información es la posibilidad de representarla a diferentes niveles de agregación.
- Esto se logra mediante las características conocidas como Drill Down y Roll Up.

Herramientas de Acceso y Uso

- Drill Down explora los hechos hacia los niveles más detallados de la jerarquía de dimensiones, mientras que Roll Up explora los hechos iterativamente hacia el nivel más alto de agregación.

Herramientas de Acceso y Uso

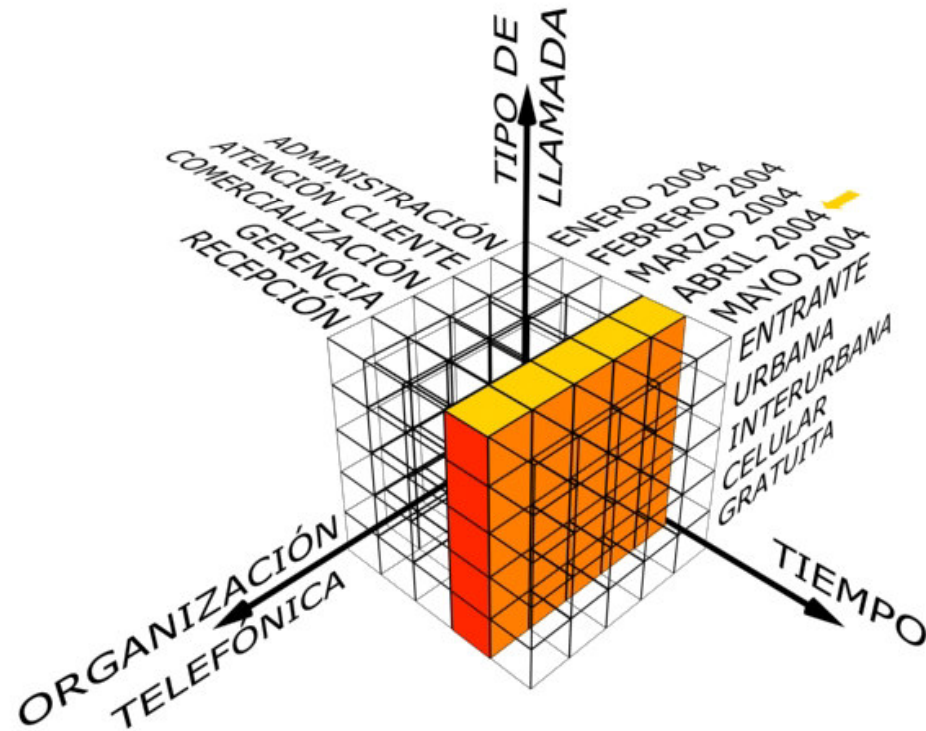


Herramientas de Acceso y Uso

-Slice y Dice:

- Slice (Rebanar) es la acción de conformar un subconjunto de los datos de un cubo aplicándole una única restricción a una sola dimensión, mediante la elección de un miembro en particular.

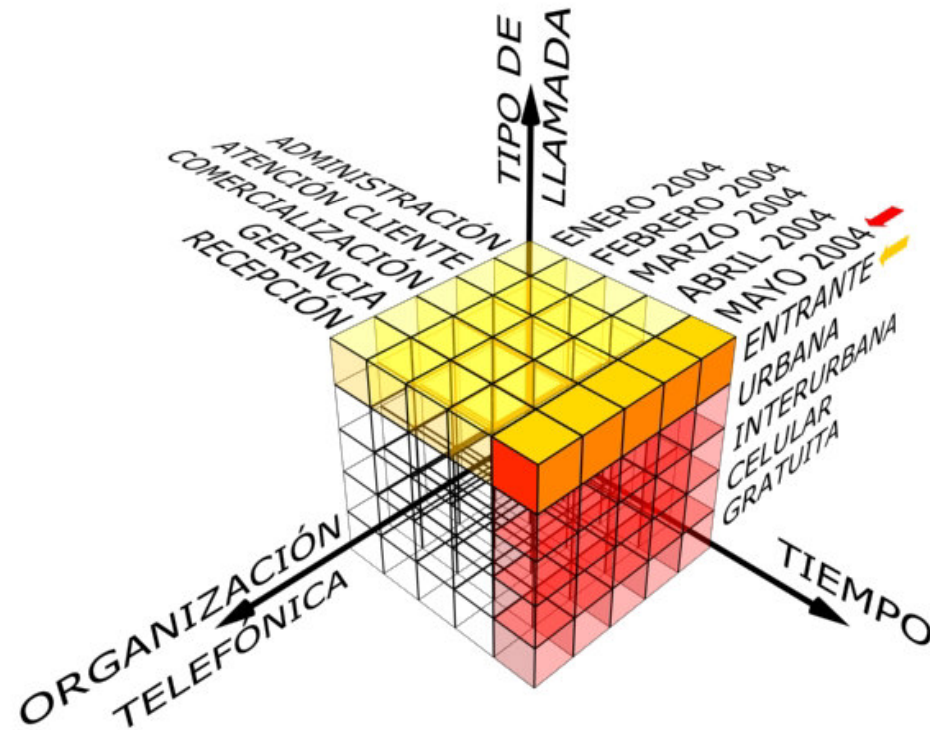
Herramientas de Acceso y Uso



Herramientas de Acceso y Uso

- Cuando se combinan varios Slice la acción efectuada es llamada *Dice* (Picar).
- Dice es equivalente a aplicar más de una restricción, ya sea para una misma dimensión o para varias dimensiones.

Herramientas de Acceso y Uso



Herramientas de Acceso y Uso

-Data Mining (Minería de Datos) :

- Las herramientas de Data Mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas. Estas herramientas exploran las bases de datos en busca de patrones ocultos.
- Las herramientas de Data Mining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información.

Conclusión

- La planificación y toma de decisiones por parte de los sectores gerenciales es un proceso que muchas veces se realiza en forma empírica.
- Esto a menudo puede generar problemas debido a la falta de datos cuantitativos que puedan respaldar el proceso, especialmente cuando la precisión es un factor crucial para el éxito en la implementación de las decisiones.

Conclusión

- Para transformar los tradicionales procesos empíricos en procesos heurísticos, que se basen en los datos cuantitativos del pasado para predecir racionalmente el futuro, hemos encontrado al Data Warehouse como una tecnología innovadora y flexible.

Sitios de Internet consultados

- [1] <http://www.lawebdelprogramador.com/cursos/mostrar.php?id=278&exto=Data+Warehouse>
- [2] <http://www.emagister.com/datawarehouse-cursos-1107904.htm>
- [3] <http://es.geocities.com/cibercero/mtd/foro/datawarehouse11.htm#data>
- [4] <http://www.sqlmax.com/dataw1.asp>
- [5] <http://www.ongei.gob.pe/publica/metodologias/Lib5084/14.HTM>
- [6] [http://es.wikipedia.org/wiki/BI_\(inform%C3%A1tica\)](http://es.wikipedia.org/wiki/BI_(inform%C3%A1tica))
- [7] http://es.wikipedia.org/wiki/Esquema_en_estrella
- [8] http://etl-tools.info/es/bi/almacenedatos_esquema-copo-de-nieve.htm