

Existen dos formas de tomar decisiones:



Decisiones *lentas*, basadas sólo en la *experiencia*, prediciendo el futuro en base a la *intuición*.



Decisiones *rápidas y oportunas*, tomadas sobre *cimientos firmes*, prediciendo el futuro en base a *información* del pasado.

¿Usted cómo lo haría en su empresa?



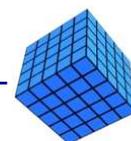
DATA WAREHOUSE

SISTEMAS
DE GESTIÓN II

DOCENTES: Ing. Bigatti, Cristian G.
Est. Grasso, Mónica

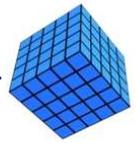
COLABORADORES: Alvarez, Ezequiel
Gorza, Luciano
Leala, Guillermo





Índice

| | |
|---|----|
| ÍNDICE | 2 |
| RESUMEN | 3 |
| INTRODUCCIÓN | 5 |
| DATOS OPERACIONALES Y DATOS INFORMATIVOS | 6 |
| DATA WAREHOUSE | 7 |
| SISTEMAS DE SOPORTE DE DECISIONES | 8 |
| INTELIGENCIA DE NEGOCIO | 8 |
| PROBLEMAS QUE DAN ORIGEN A UN DATA WAREHOUSE | 9 |
| BENEFICIOS ASOCIADOS AL DATA WAREHOUSE | 10 |
| ARQUITECTURA DATA WAREHOUSE | 12 |
| <i>Extracción (Extraction)</i> | 13 |
| <i>Transformación (Transformation)</i> | 13 |
| <i>Carga (Loading)</i> | 14 |
| <i>Data Mart</i> | 14 |
| MODELADO DE DATOS | 17 |
| EL MODELO RELACIONAL | 17 |
| EL MODELO DIMENSIONAL | 18 |
| <i>Ventajas del modelo dimensional</i> | 22 |
| OLAP (ON LINE ANALYTICAL PROCESSING) | 24 |
| DRILL DOWN Y ROLL UP..... | 25 |
| SLICE Y DICE | 26 |
| DATA MINING (MINERÍA DE DATOS) | 28 |
| CICLO DE VIDA | 29 |
| PLANIFICACIÓN DEL PROYECTO | 29 |
| DEFINICIÓN DE REQUERIMIENTOS DE NEGOCIO | 33 |
| PISTA DE TECNOLOGÍA..... | 34 |
| <i>Diseño de la Arquitectura Técnica</i> | 34 |
| <i>Selección e Instalación del Producto</i> | 35 |
| PISTA DE DATOS | 36 |
| <i>Modelado Dimensional</i> | 36 |
| <i>Diseño Físico</i> | 39 |
| <i>Diseño ETL</i> | 39 |
| PISTA DE APLICACIÓN ANALÍTICA | 39 |
| <i>Especificación de Aplicación Analítica</i> | 39 |
| <i>Desarrollo de Aplicación Analítica</i> | 39 |
| IMPLEMENTACIÓN | 40 |
| MANTENIMIENTO Y CRECIMIENTO | 40 |
| HERRAMIENTA OLAP - O3 | 41 |



Resumen

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. Dicha información debe estar estructurada y elaborada de acuerdo a parámetros de calidad, a fin de posibilitar una adaptación ágil y precisa a las fluctuaciones del ambiente externo.

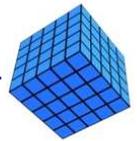
Las empresas disponen, para la gestión de sus procesos de negocio, de sistemas transaccionales corporativos que manejan enormes cantidades de datos, organizados de forma tal que puedan ser utilizados por las aplicaciones operacionales existentes. Los niveles gerenciales necesitan a menudo tomar decisiones de alto nivel, cruciales para el funcionamiento de la empresa. 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.

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.

La filosofía de trabajo del Data Warehouse es diferente a la de los sistemas transaccionales. Se modelan los datos a partir de dimensiones, en lugar del tradicional modelado relacional, y las herramientas de acceso a los datos se basan en una tecnología de procesamiento analítico, distinta al procesamiento transaccional de los sistemas operacionales.

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. 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.

Los datos dentro de un ambiente Data Warehouse pueden organizarse en un Data Warehouse corporativo, o dividirse por área o sector departamental, almacenándose en Data Marts, que son similares a una versión reducida de un Data Warehouse. Otra de las

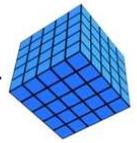


características particulares de un Data Warehouse es la importancia que presentan los metadatos, o datos acerca de los datos, considerados en todas las etapas de su desarrollo.

Las aplicaciones de usuario final que acceden al Data Warehouse brindan a los gerentes la posibilidad de ver la información a diferentes niveles de agregación (detallados o resumidos) y filtrar las consultas por distintas variables (“rebanar” y “picar” la información).

Finalmente, el Data Warehouse permite aplicar herramientas como el Data Mining, para encontrar relaciones entre los datos a fin de comprender las causas de variabilidad presentes y realizar pronósticos con el apoyo de modelos estadísticos.

En la sociedad actual, la información constituye un activo esencial de cualquier organización, proporcionando beneficios significativos, tangibles y cuantificables. Como consecuencia, la integración de un Data Warehouse a la empresa representa una ventaja competitiva en el mundo de los negocios.



Introducción

Los continuos cambios en los mercados crearán, y de hecho ya están creando, nuevas oportunidades de negocio, así como una competencia implacable. A medida que esta pugna se intensifica, las empresas se enfrentan con la necesidad de reducir los costos, añadir valor a sus productos y servicios y asegurar la diferenciación como señal de identidad. En un mercado en el que el cliente puede escoger entre distintas empresas, la inteligencia de negocios se convierte en una función estratégica para la retención e incremento del segmento de mercado de la empresa.

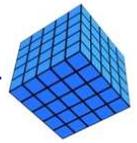
Todos estos cambios exigen de las empresas una infraestructura receptiva y flexible, en una palabra: *ágil*. Ágil en su enfoque del mercado, en la manera en que responde a sus clientes y en la puesta en marcha de los nuevos servicios o el desarrollo de los nuevos productos. El desafío no es ser el más grande, sino el más ágil en adaptarse al cambio.

Hacer accesible la información de ayuda a la toma de decisiones, de manera instantánea a los responsables de la gestión de la empresa, constituye en definitiva el objetivo intrínseco del *Data Warehouse*. Cuando se construye adecuadamente el Data Warehouse, en unión con un conjunto adecuado de aplicaciones de ayuda a la toma de decisiones, proporciona de manera natural un mejor conocimiento de la rentabilidad y la orientación adecuada al esfuerzo de los objetivos empresariales.

El material básico necesario (los datos operacionales) para el desarrollo de un Data Warehouse corporativo es abundante. La información sobre los clientes y los productos, por ejemplo, son a la vez recursos disponibles, profusos e infrautilizados. Por ende, los datos operacionales por sí solos no generan información significativa y conocimiento, a menos que los objetivos empresariales de la compañía y la tecnología de la información estén alineados.

Para que la tecnología se convierta en un elemento que facilite la creación de nuevas estrategias, debe primero integrarse en las funciones de negocio de la empresa. El futuro éxito o fracaso de las compañías dependerá en gran medida del conocimiento de la propia empresa, el cual a su vez se basará en la construcción exitosa de herramientas como el Data Warehouse.

Dada la gran importancia del Data Warehouse para el crecimiento de la empresa, se hace necesario conocer los conceptos relacionados a su filosofía de trabajo, a su proceso y la forma de utilizarlo. En este trabajo expondremos las características principales, su estructura y su funcionamiento, así como los resultados que puede esperar el usuario final al interactuar con las aplicaciones que utilizan esa herramienta. Para ello, junto con el desarrollo teórico hemos



tomado un caso práctico de aplicación que permitirá ver la implementación de la tecnología de Data Warehouse. Este caso se describe a continuación.

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. Estos datos son creados durante la ejecución de estos procesos y son almacenados en un archivo o en una base de datos. Frecuentemente contienen valores incorrectos, son muy detallados y son de mínimo uso en los negocios debido a su gran volumen, ubicación y formatos. En conclusión, es difícil para los usuarios del negocio tener acceso a los datos operacionales debido a las limitaciones de performance y tecnología.

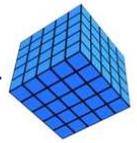
Lo que el usuario del negocio necesita como entrada a sus actividades de análisis son *datos informativos*. Estos son una combinación de datos operacionales que han sido modificados, depurados, transformados, consolidados y organizados desde diversas fuentes externas al proceso del negocio. Este tipo de información generalmente es específico para un conjunto de usuarios del negocio que lo hacen significativo y útil para su análisis.

Ambos tipos de datos y ambos tipos de uso son muy importantes, pero es difícil cumplir con ambos propósitos en el mismo sistema.

Los datos operacionales son específicos para cada aplicación y usualmente son almacenados de manera separada por otras aplicaciones. Estos datos son útiles en la medida en que se aprovechen para satisfacer el proceso de las aplicaciones predefinidas. Mayormente se requieren sólo datos actuales y estos deben ser mantenidos al día haciendo actualizaciones frecuentes en la base de datos. En cambio, para los datos informativos, el usuario necesita datos que crucen por varias aplicaciones, que estén reorganizados por temas de negocio, que contengan valores históricos, que se encuentren disponible para análisis durante períodos largos y que sea accesible de manera fácil y flexible.

Los datos operacionales son manejados, precisamente, por los sistemas operacionales o transaccionales (On Line Transactional Processing, OLTP), los cuales 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.

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. En el siguiente cuadro se muestran las diferencias entre los datos operacionales y los datos informativos.

Datos Operacionales

- Orientados a una aplicación
- Integración limitada
- Constantemente actualizados
- Sólo valores actuales
- Soportan operaciones diarias

Datos Informativos

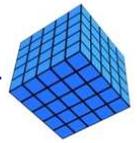
- Orientados a un tema
- Integrados
- No volátiles
- Valores a lo largo del tiempo
- Soportan decisiones de administración

Data Warehouse

Un *Data Warehouse* es un conjunto de datos integrados orientados a una materia, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una organización.

A continuación describiremos las características distintivas de un Data Warehouse, las cuales nos ayudarán a comprender mejor su concepto.

- *Integra y asocia información de muchas fuentes de información.* Debido a que las organizaciones han administrado históricamente sus operaciones utilizando numerosas aplicaciones de software y múltiples bases de datos (datos operacionales), se requiere de Data Warehouse para recopilar y organizar en un solo lugar (datos informativos) la información que estas aplicaciones han acumulado al paso de los años.
- *Está orientado a una materia.* Muchos sistemas operativos organizan sus datos desde la perspectiva de la aplicación, de modo que el acceso de la aplicación a los datos tenga la mayor eficiencia posible. Con frecuencia, la información que está organizada para que una aplicación del negocio la recupere y actualice con facilidad no está organizada necesariamente de modo que un analista con herramientas gráficas inteligentes de consulta pueda formular las preguntas empresariales correctas. Esto se debe al enfoque del diseño de la base de datos. El Data Warehouse resuelve este problema organizando y orientando los datos a una materia desde la perspectiva del usuario final.
- *Administra grandes cantidades de información.* La mayoría de los Data Warehouse contienen información histórica que se retira con frecuencia de los sistemas operacionales porque ya no es necesaria para las aplicaciones operativas y de producción. Por el volumen de información que un Data Warehouse debe manejar, también debe ofrecer opciones para la adición y la condensación que clasifiquen esta



inmensa cantidad de datos, es decir, un Data Warehouse maneja información a diferentes niveles de detalle. Por lo tanto, el volumen de los datos que debe almacenar es generalmente mayor al de los datos de las bases de datos operacionales.

- *Gestiona múltiples versiones de un esquema de base de datos.* Debido a que el Data Warehouse tiene que guardar la información histórica y administrarla, y como esta información histórica ha sido manejada en distintos momentos por diferentes versiones de esquemas de bases de datos, en ocasiones el Data Warehouse tiene que controlar información originada en organizaciones de bases de datos diferentes.
- *Condensa y agrega información.* Con frecuencia, es muy alto el nivel de detalle de la información almacenada por bases de datos operacionales para cualquier toma de decisiones sensata. Un Data Warehouse condensa y agrega información para presentarla en forma comprensible a los usuarios finales.

Sistemas de Soporte de Decisiones

En el proceso de tomar decisiones hay dos partes en general: tener los datos y obtener respuestas de los datos.

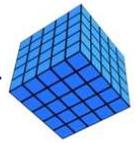
Un *Sistema de Soporte de Decisiones (SSD)* puede ser definido de una manera general como un sistema de computación diseñado para apoyar al proceso de toma de decisiones (de planeamiento, de administración y operacionales) en un negocio.

Este tipo de Sistema contiene todos los servicios o procesos para elegir, manipular y analizar datos informativos y presentar los resultados. Debería brindar acceso transparente a los datos en varias partes del Data Warehouse y proveer de interfaces comunes a un grupo de usuarios de la organización. Fundamentalmente, un Sistema de Soporte de Decisiones constituye la ventana del usuario a los datos informativos almacenados en el Data Warehouse.

Inteligencia de Negocio

La *Inteligencia de Negocios* (Business Intelligence, BI) es una disciplina que combina consultas, servicios, aplicaciones y tecnologías para obtener, administrar, analizar y transformar datos en información útil para desarrollar el entendimiento necesario para tomar decisiones adecuadas.

Esta definición se basa en el concepto de *Inteligencia*, que consiste en la aplicación de la información, habilidades, experiencias y razonamiento para resolver un problema de negocios.



Las actividades primarias consisten en obtener, preparar y analizar datos. El resultado de combinar Data Warehouse con Sistemas de Soporte de Decisiones es la Inteligencia de Negocios.

Los datos por sí solos deben ser de alta calidad, por lo que se necesita un proceso para alcanzar ese nivel. La información almacenada en el Data Warehouse tiene que estar disponible para los usuarios, de aquí la necesidad de un Sistema de Soporte de Decisiones para el acceso a esa información.

Los datos y la información que están en un Data Warehouse deben ser administrados con cuidado, ya que son el material básico para obtener el conocimiento. Los Sistemas de Soporte de Decisiones proveen la facilidad para que los usuarios de los procesos puedan obtener y procesar ese material básico para adquirir el conocimiento que necesitan para desarrollar las estrategias de negocio adecuadas.

El Data Warehouse contiene hechos históricos y datos derivados que son analizados con un Sistema de Soporte de Decisiones para entender lo que sucedió en el pasado y lo que ocasionó las condiciones actuales. Estos datos son también analizados de diferentes formas para desarrollar una visión de lo que podría ocurrir en el futuro si se toman ciertas decisiones. Después de la extracción y análisis de datos, los resultados combinados pueden ser presentados utilizando herramientas de Soporte de Decisiones, de una manera que facilite el entendimiento de la información, lo cual puede llevar a decisiones mejor respaldadas.

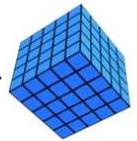
La meta es utilizar el Sistema de Inteligencia de Negocios para entender el pasado en orden de optimizar las condiciones del negocio en el futuro.

Por todo lo expresado, podemos definir a la Inteligencia de Negocios como el proceso de convertir datos en conocimiento y conocimiento en acción para el beneficio del negocio.

Problemas que dan origen a un Data warehouse

Los Data Warehouse surgen por los siguientes problemas:

- *Falta de integración de la información.* Muchos de los datos de que disponen las empresas se encuentra en diversos almacenes de datos porque han evolucionado como sistemas independientes, en diferentes lenguajes, desarrollados por diferentes personas y en diferentes momentos. Estos almacenes de datos son incapaces de ofrecer a la empresa una visión consolidada ya que no poseen vinculación entre ellos.
- *Sobrecarga de información.* En las últimas décadas, los sistemas operacionales han producido una gran cantidad de datos, los que son difíciles de analizar para tomar



decisiones de negocio. A pesar de las grandes sumas invertidas en tecnología de la información, los distintos departamentos dentro de las empresas todavía carecen de herramientas para explorar esos datos de una manera ágil y eficaz.

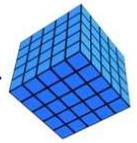
- *Exceso de información genérica y despersonalizada.* Por lo general, los informes producidos por un sistema operacional contienen una gran cantidad de datos. Sin embargo, ofrecen muy poca información que ayude a la toma de decisiones. Estos informes no tienen la capacidad de adaptarse a las necesidades actuales de los usuarios, y carecen de información personalizada y/o relevante para los distintos perfiles que existen en un negocio.
- *Falta de retroalimentación oportuna para la mejora de los negocios.* En las organizaciones, los ejecutivos trabajan haciéndose preguntas e imaginando posibles situaciones futuras. Los resultados obtenidos de esa labor generarán a su vez más preguntas y el planteo de nuevas situaciones repitiendo el ciclo hasta conseguir los fundamentos para tomar una decisión. Con el uso de los sistemas operacionales, este proceso se hace dificultoso, sino imposible, al no tener posibilidad de generar informes adaptados a las necesidades del momento.

Beneficios Asociados al Data Warehouse

Un Data Warehouse puede dar lugar a una serie de importantes beneficios para la organización. En cualquier caso, su utilización permitirá que la información de gestión sea: accesible, correcta, uniforme y actualizada.

Estas características asociadas a la información contenida en un Data Warehouse, junto con otra serie de aspectos inherentes al mismo dan lugar a la obtención de un conjunto de ventajas, que podemos resumir del siguiente modo:

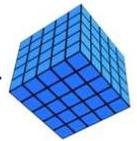
- Menor costo en la toma de decisiones. Se suprime el desperdicio de tiempo que se podía producir al intentar ejecutar consultas de datos largas y complejas con bases de datos que estaban diseñadas específicamente para transacciones más cortas y sencillas.
- Posibilidad de encontrar relaciones ocultas. Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos, obteniendo un valor añadido para el negocio.
- Aprendizaje del pasado. Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.



- Mayor flexibilidad ante el entorno. El Data Warehouse convierte los datos operacionales en información relacionada y estructurada, que genera el conocimiento necesario para la toma de decisiones. Esto permite establecer una base única del modelo de información de la organización, que puede dar lugar a una visión global de la información en base a los conceptos de negocio que tratan los usuarios. Además, aporta una mejor calidad y flexibilidad en el análisis del mercado, y del entorno en general.

Esta visión global puede conllevar también la obtención de otras ventajas competitivas, al permitir el desempeño de tareas que en los sistemas tradicionales sufrirían un costo adicional, por ejemplo:

- *Mejor servicio al cliente.* Lo dicho en el punto anterior implica una importante mejora en la calidad de gestión, lo que también repercute en la relación con el cliente. De hecho, el que un Data Warehouse implique una mayor flexibilidad ante el entorno tiene una consecuencia directa en una mayor capacidad para responder a las necesidades de los clientes.
- *Rediseño de procesos.* Ofrecer a los usuarios una capacidad de análisis de la información de su negocio que tiende a ser ilimitada y permite con frecuencia obtener una visión más profunda y clara de los procesos de negocio propiamente dichos, lo que a su vez permite obtener ideas renovadoras para el rediseño de los mismos.



Arquitectura Data Warehouse

Poder transformar los datos en conocimiento es un proceso complejo. Un Data Warehouse es mucho más que simplemente copiar datos de un lugar a otro, de los sistemas operacionales a una base de datos informativa independiente. Un Data Warehouse es, en primer lugar, una *arquitectura* que debe servir como infraestructura para proporcionar una solución completa a los problemas anteriormente mencionados. La arquitectura de un Data Warehouse se representa en la Figura 4.

Fuentes de datos

Este componente es el que normalmente está presente originariamente en las organizaciones, y a partir del cual se realiza la captura de datos que se contemplará en el Data Warehouse. Estas fuentes de datos pueden ser sistemas operacionales corporativos (representan el entorno del que se obtienen la mayor parte de los datos significativos de la operativa diaria de la compañía), sistemas operacionales departamentales y fuentes externas como copias de respaldo, planillas de cálculo, archivos planos, reportes de la empresa, resultados de investigaciones de mercado, etc.

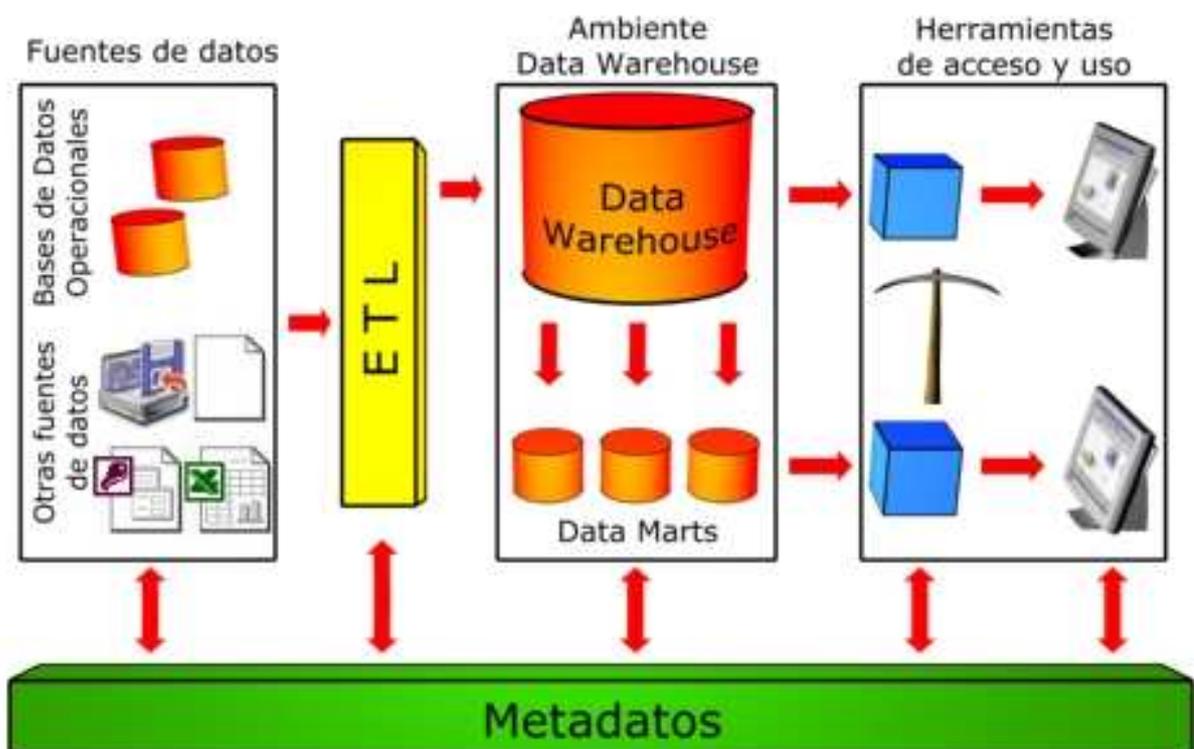
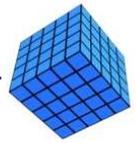


Figura 4 – Componentes de la arquitectura Data Warehouse.



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.

Los subsistemas para poblar el Data Warehouse se pueden construir utilizando herramientas y productos disponibles en el mercado, programas y procesos codificados desde cero, o combinaciones de estos elementos.

Al construir los sistemas para poblar el Data Warehouse, se debe considerar la posibilidad de que estos permitan regular el crecimiento evolutivo del Data Warehouse, brindando escalabilidad y soporte para grandes cantidades de datos y consultas complejas. Se pueden encontrar dificultades adicionales dependiendo de las fuentes de datos que se tengan disponibles, que implican el uso de diferentes herramientas y tecnologías para acceder a cada uno de ellos.

Extracción (Extraction)

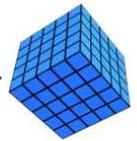
El propósito principal de la fase de *extracción* es capturar y copiar los datos requeridos de uno o más sistemas operacionales o fuentes de datos. Los datos que se extraen son colocados en un archivo intermedio con un formato definido, que luego será utilizado por la siguiente fase del proceso.

Los registros que sean rechazados en el proceso deben ser registrados en un archivo o *log* de rechazos para que puedan ser analizados posteriormente y así tener la posibilidad de cargarlos en el Data Warehouse correctamente. Además, esto permite descubrir los errores que han ocurrido en los procesos de creación de los datos operacionales. Ejemplos de estos errores son violaciones de integridad, claves duplicadas, formatos de datos incorrectos y datos inválidos como campos vacíos, fechas futuras e importes negativos cuando estos no correspondan. En nuestro Caso: números telefónicos inválidos, estándares no respetados, llamadas duplicadas y campos nulos.

Hay que tener en cuenta que después de la fase inicial de carga del Data Warehouse sólo es necesario cargar los datos nuevos y que han sido modificados.

Transformación (Transformation)

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. La mayor parte del trabajo en esta fase involucra el efectuar las transformaciones necesarias. Estas transformaciones incluyen:

- Combinar campos múltiples de nombres y apellidos en un solo campo.
- Fusionar campos o datos homónimos.
- Separar un campo de fecha en campos de año, mes y día.
- Cambiar la representación de los datos, como TRUE (verdadero) a 1, y FALSE (falso) a 0, o códigos postales numéricos a alfanuméricos, respetando los estándares de la empresa.
- Cambiar un dato que tiene múltiples representaciones a una sola representación, como por ejemplo definir un formato común para números telefónicos, o establecer un término común para los nombres de los campos o los valores de los datos que sean sinónimos.

Carga (Loading)

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. Generalmente los datos son insertados en el Data Warehouse, rara vez son actualizados o eliminados.

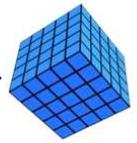
Ambiente Data Warehouse

Es el bloque donde se almacenan los datos informativos, utilizado principalmente para usos estratégicos. No obstante, existen herramientas que no hacen uso de este bloque, realizando las consultas multidimensionales directamente sobre la base operacional. En este caso se puede pensar en el Data Warehouse simplemente como una vista lógica o virtual de datos.

Data Mart

Un *Data Mart* es una implementación de un Data Warehouse con un determinado alcance de información y un soporte limitado para procesos analíticos, que sirve a un sólo departamento de una organización o para el análisis de problemas de un tema particular.

El Data Mart es un subconjunto de información corporativa con formato adicional a la medida de un usuario específico del negocio. Un Data Mart será siempre menor en complejidad



y alcance de los datos. Un Data Warehouse tiene más usuarios y más temas que un Data Mart, brindando una vista más amplia entre múltiples áreas.

Existen dos grandes filosofías con respecto a la relación entre los conceptos de Data Warehouse y Data Mart.

Bill Inmon, quien es considerado el padre del Data Warehouse, propuso la idea de que los Data Marts se sirven del Data Warehouse para extraer información. La misma está almacenada en tercera forma normal, en un modelo relacional.

Por su parte, Ralph Kimball, el principal propulsor del enfoque dimensional para el diseño del Data Warehouse, sostiene que el Data Warehouse es el resultado de la unión de los Data Marts de la empresa.

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. También deben transformar los datos capturados en información útil para el negocio. Actualmente a este tipo de herramientas se las conocen como Herramientas de Inteligencia de Negocio (Business Intelligence Tools, BIT) y están situadas conceptualmente sobre el Data Warehouse. Cada usuario final debe seleccionar la herramienta que mejor se ajusta a sus necesidades y a su Data Warehouse. Entre ellas podemos citar las Consultas SQL (Structured Query Language), las Herramientas MDA (Multidimensional Analysis), OLAP (On-Line Analytical Processing) y las herramientas Data Mining.

Este bloque también incluye el hardware y software involucrados en mostrar la información en pantalla y emitir reportes de impresión, hojas de cálculo, gráficos y diagramas para el análisis y presentación.

Metadatos

Los metadatos son datos acerca de los datos. En una base de datos los metadatos son la representación de los diversos objetos que definen una base de datos, por ejemplo, ubicación y descripción de base de datos, tablas, nombres y resúmenes. También podemos mencionar las descripciones lógicas y físicas de tablas, columnas y atributos.



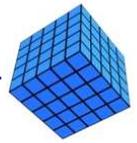
Uno de los problemas con el que pueden encontrarse los usuarios de un Data Warehouse es saber lo que hay en él y cómo pueden acceder a lo que quieren. A fin de proveer el acceso a los datos universales, es absolutamente necesario mantener los metadatos. Un componente llamado *repositorio* les ayuda a conseguirlo. Los metadatos son sólo una de las utilidades del repositorio, pero éste tiene muchas funcionalidades: catalogar y describir la información disponible, especificar el propósito de la misma, indicar las relaciones entre los distintos datos, establecer quién es el propietario de la información, relacionar las estructuras técnicas de datos con la información de negocio, establecer las relaciones con los datos operacionales y las reglas de transformación, y limitar la validez de la información.

Metadato para Entidad

| | |
|----------------------------|---|
| Nombre | Cliente |
| Nombre alternativo | Cuenta |
| Definición | Un cliente es una persona o empresa que ha comprado a la corporación bienes o servicios por lo menos en una ocasión |
| Fecha creación | 15 de enero de 2000 |
| Fecha última actualización | 01/03/2005 |
| Claves | Identificación del cliente |
| Ciclo de actualizac. | Extraído cada mes |
| Patrones de acceso | Fecha de último acceso 20/03/2005 |

Metadato para Atributo

| | |
|----------------------------------|---|
| Nombre | Nombre de cliente |
| Nombre alternativo | Nombre de cuenta |
| Definición | Es el nombre oficial que usa la persona o empresa cliente |
| Fecha creación | 15 de enero de 2000 |
| Fecha última actualización | 01/03/2003 |
| Clave | N |
| Fuente de dato | Sistema de ingreso de pedidos. El atributo proviene de la tabla clientes y se denomina nom_cuenta |
| Inf. de transf./ conversión | 30 caracteres |
| Algoritmo de resumen /derivación | Ninguno |
| Valor predeterminado | Ninguno |
| Fuentes operacionales múltiples | No |



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.

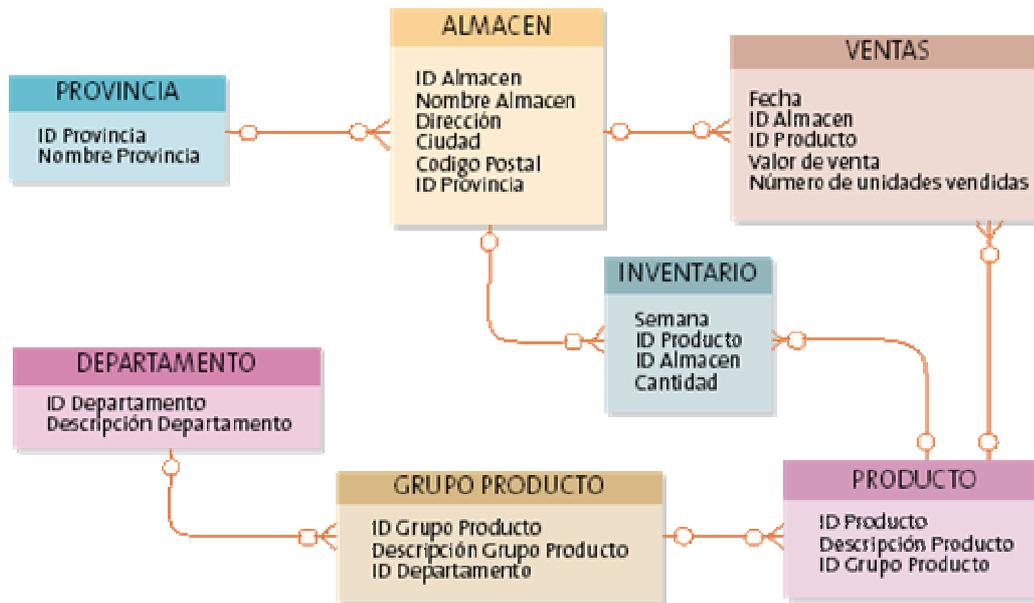
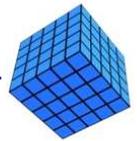
El modelado entidad-relación se utiliza habitualmente para crear un único modelo complejo de todos los procesos de una organización. Este enfoque ha demostrado ser efectivo para crear sistemas eficientes de procesamiento transaccional en línea (OLTP). Por otra parte, el *modelado dimensional* crea modelos individuales para reflejar procesos discretos de negocio. Este modelado organiza la información en estructuras que usualmente corresponden a la forma en que los analistas realizan sus consultas de los datos del Data Warehouse.

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 éxito del procesamiento transaccional en un entorno de bases de datos relacionales se debe básicamente a esta disciplina de modelado.

Sin embargo, para el propósito de un Data Warehouse, el modelo relacional (ER) presenta los siguientes problemas:

- *Legibilidad limitada.* Los usuarios finales no son capaces de entender el modelo ER. Por tanto, no pueden “navegar” por dicho modelo en busca de información.
- *Dificultad para las herramientas de consulta en el acceso a un modelo ER general.* Las herramientas de consulta a menudo poseen prestaciones mediocres o inaceptables cuando se trabaja en entornos relacionales de grandes volúmenes de información.
- *La utilización de la técnica de modelado ER frustra el principal atractivo del Data Warehouse.* Al utilizar el modelado ER queda frustrada la recuperación de información intuitiva y con alto rendimiento.



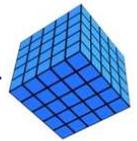
El modelo dimensional

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. Este modelado se vale de los principios de la disciplina que emplea el modelo relacional con algunas importantes restricciones. El modelado dimensional es esencialmente útil para resumir y organizar los datos y la presentación de información 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*.

Un área tema como es una cuestión de interés de una función empresarial. Las áreas tema en conjunto constituyen el ámbito de implementación del Data Warehouse. Por ejemplo, el departamento de Comercialización de una empresa puede estar interesado en las áreas tema de Pedidos, Promociones, Mercados y Ventas.

Para especificar las áreas tema se deben identificar las medidas. 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. Ejemplos de medidas son duración de las llamadas y cantidad de llamadas.

El Data Warehouse organiza un gran conjunto de datos operacionales mediante múltiples dimensiones. 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. Ejemplos de dimensiones son Tipo de Llamada, Tiempo, Líneas y Organización Telefónica (Áreas e Internos).



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). Por ejemplo, podemos pensar en un cubo que posea como medida Duración de las Llamadas y como dimensiones Tiempo, Tipo de Llamada y Organización Telefónica. La representación de este modelo se muestra en la Figura 1.

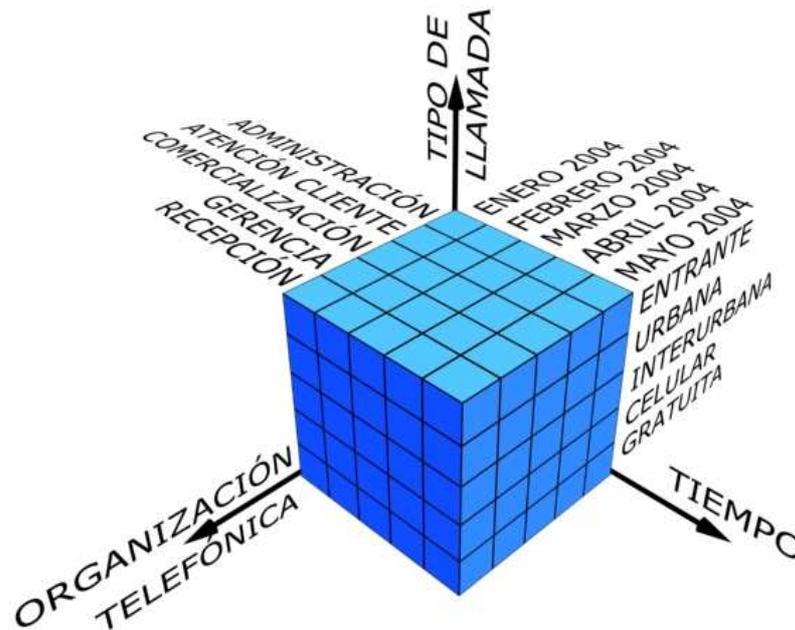
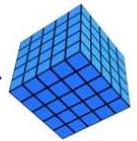


Figura 1 – Cubo con dimensiones Tiempo, Tipo de Llamada y Organización Telefónica.

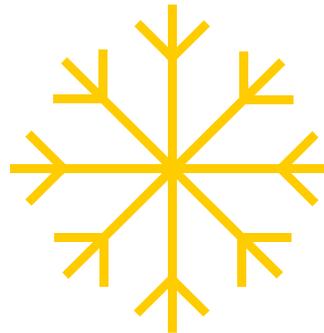
Cada porción del cubo de la Figura 1 es la medida a la que hacemos referencia, y expresa la duración de las Llamadas de un tipo determinado efectuadas en un Área en un mes. Las dimensiones están representadas por los ejes. Una consulta para el cubo de la Figura 1 podría ser la duración de las llamadas salientes del mes de enero de 2004 discriminadas por Área.

Los miembros de una dimensión pueden estar organizados en una o más *jerarquías*. Una jerarquía es un conjunto de miembros de una dimensión, los cuales se definen por su posición relativa con respecto a los otros miembros de la misma dimensión, y forman en su totalidad una estructura de árbol. Partiendo de la raíz del árbol, los miembros son progresivamente más detallados hasta llegar a las hojas, donde obtenemos el mayor nivel de detalle. Por ejemplo, para la dimensión de Organización Telefónica podemos establecer Área como raíz, luego, dentro de cada Área existen muchos Internos, los que constituyen las hojas. Puede darse el caso en que una dimensión no necesite jerarquizarse debido a que ninguno de sus miembros posee una posición relativa con respecto a los otros miembros. Por ejemplo, una dimensión Cliente que tiene como miembros nombre, sexo y fecha de nacimiento, no necesita organizar



estos miembros porque todos están al mismo nivel de detalle, a menos que desee agruparlos por alguno de ellos para visualizar los datos.

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*.

Cada una de las tablas de dimensiones tiene una clave primaria que corresponde exactamente con uno de los componentes de la clave compuesta de la tabla de hechos. Las tablas de hechos, además de sus campos clave, contienen una o más medidas, indicadores o “hechos”. Las medidas más útiles en una tabla de hechos son *numéricas* y *aditivas*. La aditividad es crucial porque las aplicaciones Data Warehouse casi nunca recuperan un solo registro de la tabla de hechos, sino que acceden a cientos, miles o incluso millones de registros a la vez.

Las tablas de dimensiones, por el contrario, contienen información *textual descriptiva*. Los atributos de las dimensiones se emplean como fuente de las restricciones en las consultas al Data Warehouse.

En el modelo estrella las dimensiones no se normalizan. Con ello se logra minimizar el número de uniones y, por consiguiente, incrementar el rendimiento de las consultas (una tabla de hechos está relacionada con numerosas tablas de dimensiones). Ej. en la Fig. 2.

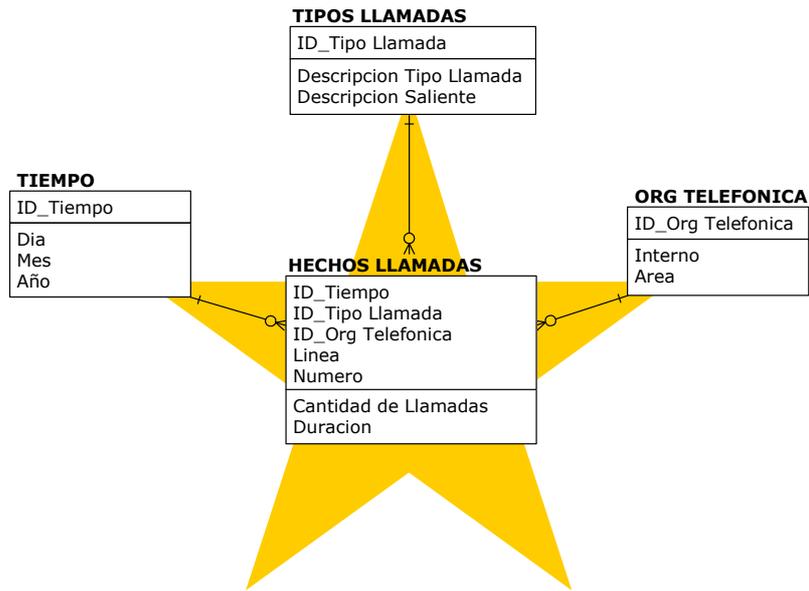


Figura 2 - Modelo estrella.

Una variante del modelo en estrella es el modelo copo de nieve o *snowflake*.

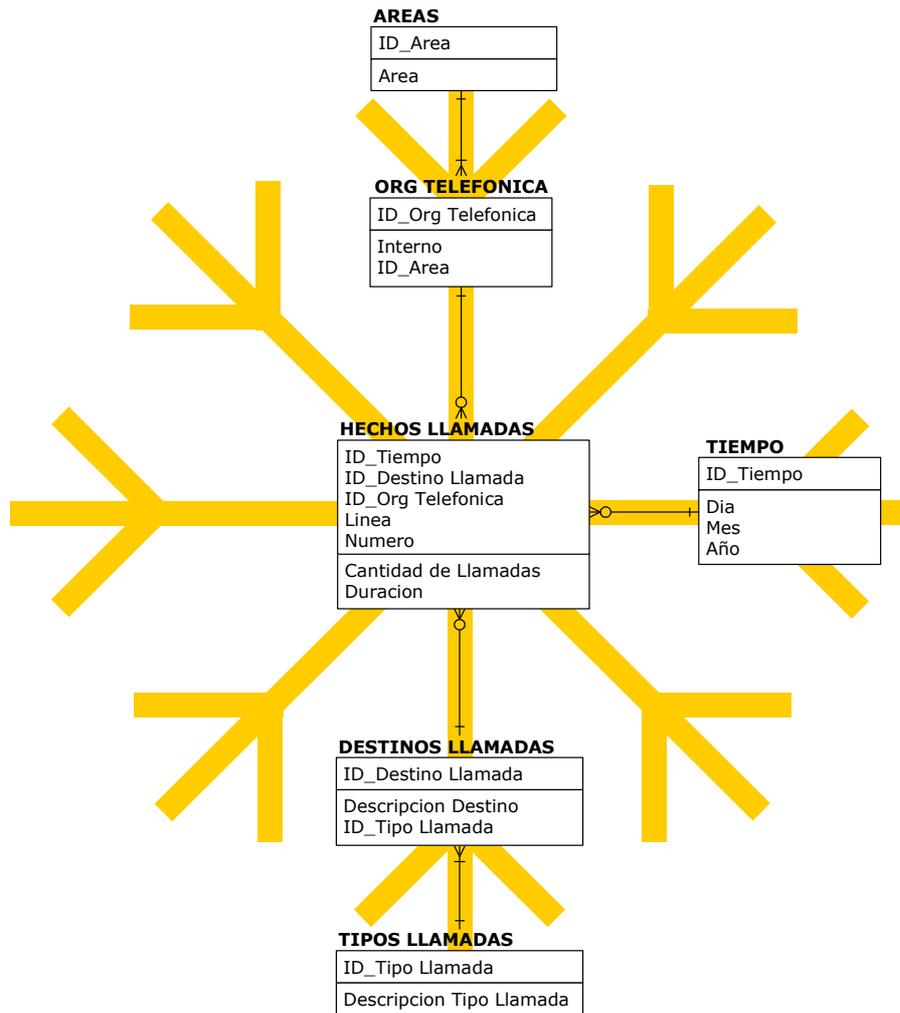
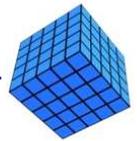


Figura 3 - Modelo copo de nieve.



En este modelado se normalizan las dimensiones reflejando las jerarquías en las mismas y conservando lo esencial del modelo en estrella: las tablas de hechos. La ventaja del modelo copo de nieve es eliminar la redundancia de datos y por lo tanto ocupar menos espacio en disco. En la Figura 3 se da un ejemplo de un modelo copo de nieve.

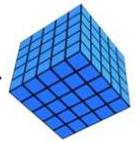
Ventajas del modelo dimensional

El modelo dimensional presenta importantes ventajas de las que carece el modelo relacional. Uno de los puntos fuertes del modelo dimensional es que el marco predecible del esquema estrella resiste a los cambios inesperados en el comportamiento del usuario. Cada dimensión es equivalente a las demás y todas las dimensiones pueden ser concebidas como puntos de entrada hacia la tabla de hechos. El diseño lógico puede realizarse independientemente de los patrones de consulta esperados, siendo consideradas de la misma forma tanto las interfaces de usuario como las estrategias de consulta, así como el lenguaje de consulta generado contra el modelo dimensional.

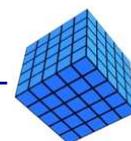
Otra cualidad del modelo dimensional es la flexibilidad. Los nuevos elementos de datos y las nuevas decisiones de diseño son fácilmente adaptables. Todas las tablas pueden modificarse simplemente agregando nuevos registros de datos o se pueden incluir nuevas dimensiones al modelo sin necesidad de volver a cargar los datos posteriormente. Además no es necesario volver a programar las herramientas de consulta o de informes para adaptarse a los cambios, y las aplicaciones existentes pueden continuar su ejecución brindando los mismos resultados. Las modificaciones ante las cuales el modelo dimensional es flexible incluyen:

- Agregar medidas a la tabla de hechos, siempre que sean aditivas y consistentes con el mayor nivel de detalle de las dimensiones.
- Agregar atributos a las dimensiones.
- Agregar nuevas dimensiones, siempre que exista un único valor de dicha dimensión definido para cada registro de la tabla de hechos.
- Particionar los registros de una dimensión a un mayor nivel de detalle a partir de un determinado punto en el tiempo. Los registros anteriores permanecerán sin cambios mientras que los futuros registros se almacenarán de acuerdo al nuevo modelo.

Una ventaja adicional del modelo dimensional es el creciente número de utilidades administrativas y aplicaciones que gestionan y utilizan los *agregados*. Los agregados son registros resumidos que son lógicamente redundantes con la información ya existente en el Data Warehouse y son empleados para mejorar el rendimiento de las consultas. Cualquier



implementación de tamaño mediano o grande del Data Warehouse requiere la creación de una estrategia de agregados. Todas las aplicaciones software de gestión de agregados, así como las utilidades de navegación de agregados, dependen de una estructura específica de las tablas de hechos y dimensiones que es absolutamente dependiente del modelo dimensional. Si no se emplea el enfoque del modelo dimensional no es posible beneficiarse de tales aplicaciones. Por ejemplo, en el cubo de la Figura 1 se muestran los datos discriminados por mes, pero en la tabla de la dimensión Tiempo podrían estar almacenados por día. Con una determinada estrategia de agregados, los agregados de los meses podrían almacenarse para acceder directamente a los datos sin necesidad de invertir tiempo en cálculos, optimizando así el rendimiento en las consultas.



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.

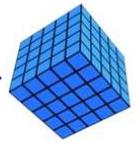
El Procesamiento Analítico en Línea (OLAP) es una tecnología de análisis de datos que crea nueva información empresarial a partir de los datos existentes. Posee las siguientes características:

- Presenta una visión multidimensional lógica de los datos del Data Warehouse, independiente de su forma de almacenamiento.
- Crea resúmenes, adiciones y jerarquías.
- Comprende consultas interactivas y análisis de los datos. Permite una profundización hacia niveles más detallados o un ascenso a niveles superiores de resumen y adición.
- Ofrece opciones de modelado analítico, incluyendo un motor de cálculo para medir datos numéricos a través de muchas dimensiones, así como también provee modelos estadísticos básicos.
- Responde con rapidez a las consultas, de modo que el proceso de análisis no se interrumpe y la información no se desactualiza.
- Recupera y exhibe datos tabulares en dos o tres dimensiones, cuadros o gráficos, con un fácil pivoteo de los ejes.

Esta tecnología es independiente de la implementación y permite el empleo de cualquier base de datos, ya sea relacional (ROLAP, cuando se aplica el modelo dimensional a una base de datos relacional), dimensional (MOLAP, modelo dimensional sobre base de datos dimensional), de objetos, etc.

En el siguiente cuadro se exponen las principales diferencias entre OLAP y OLTP.

| | OLAP | OLTP |
|-----------------------------------|---|---|
| Usuario | Gerente, Ejecutivo | Empleado, Profesional de Sistemas |
| Función | Soporte a las Decisiones | Operación diaria |
| Diseño de la Base de Datos | Orientado a una materia | Orientado a una aplicación |
| Datos | Históricos, resumidos, multidimensionales, integrados, consolidados, opcionalmente detallados | Actuales, detallados, relacionales, aislados |
| Uso | Ad hoc, requerimientos del momento | Repetitivo |
| Acceso | Exploración | Lectura/Escritura, Índices/Algoritmos de dispersión |
| Unidad de Trabajo | Consultas complejas | Transacciones simples |



| | OLAP | OLTP |
|--|--------------------------------|------------------------------------|
| Cantidad de registros accedidos | Millones | Decenas |
| Cantidad de usuarios | Cientos | Miles |
| Volumen de la Base de Datos | Mayor | Menor |
| Métrica | Consultas por unidad de tiempo | Transacciones por unidad de tiempo |

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. Estas son técnicas para navegar a través de distintos niveles de detalle de una jerarquía de datos, desde los de mayor nivel de agregación (también llamados datos sumariados) hasta los más detallados. 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.

En la Figura 5, observamos los datos que surgen a partir de aplicar Drill Down sobre el miembro Comercialización de la dimensión Organización Telefónica del cubo de la Figura 1. Teniendo la información detallada podemos efectuar la operación Roll Up para llegar a la información resumida.

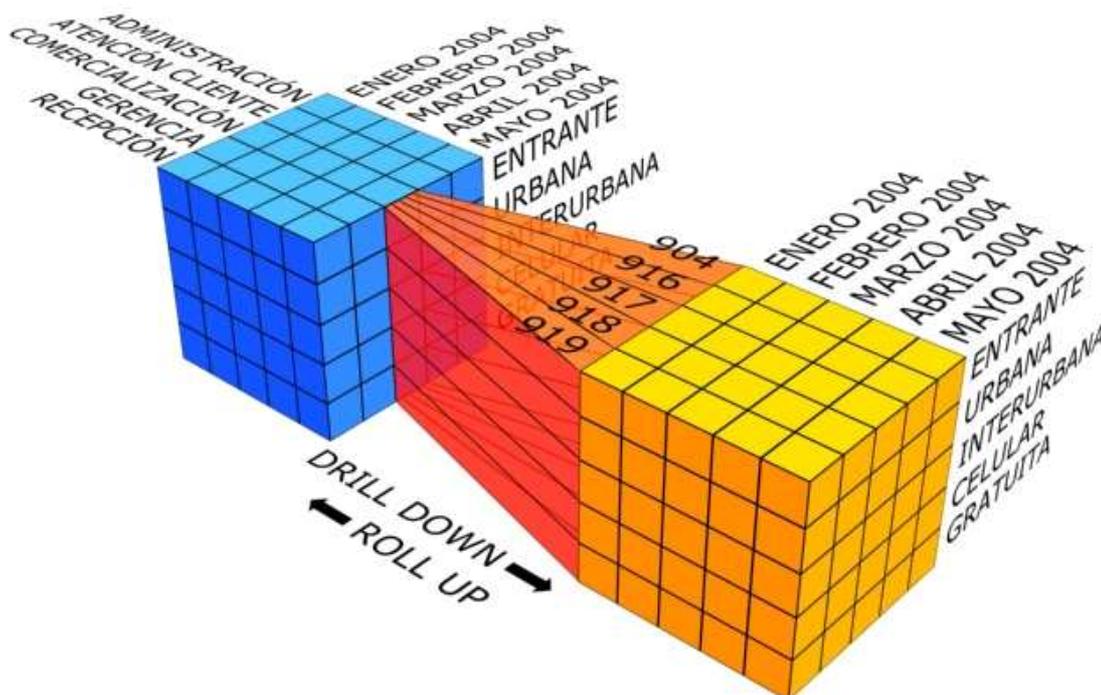
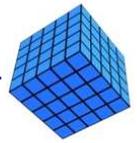


Figura 5 – Drill Down y Roll Up.



Slice y Dice

Estos términos son utilizados para describir como los datos multidimensionales pueden ser mostrados aplicando diferentes filtros a los cubos.

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. Haciendo referencia al cubo de la Figura 1, se muestra un ejemplo de Slice en la Figura 6. En este cubo se forma el subconjunto filtrando los datos correspondientes al mes de Abril de 2004, para todas las Áreas y todos los Tipos de Llamada.

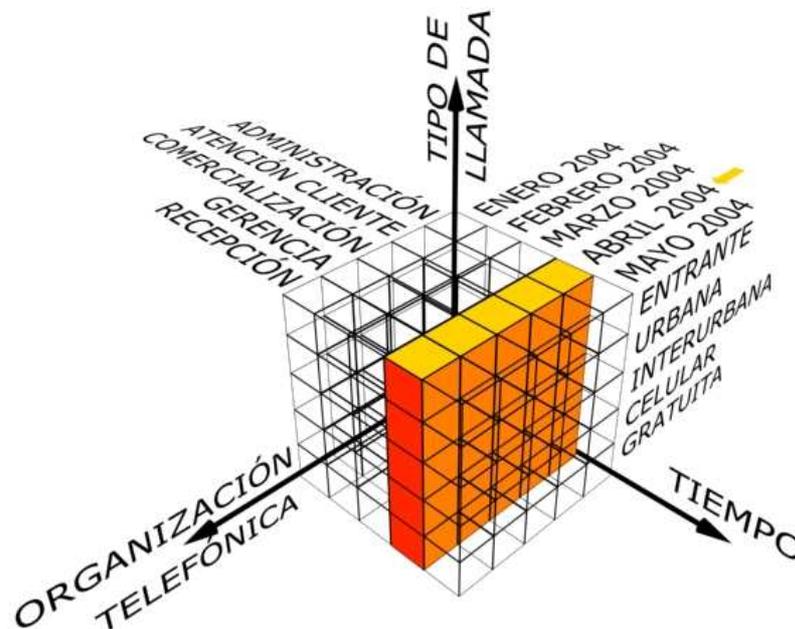


Figura 6 – Slice de un cubo para el mes de Abril de 2004.

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. Un tipo de Dice sería equivalente a efectuar dos Slice, cada uno por un diferente mes de 2004. Otros casos de Dice más complejos son mostrados en las Figuras 7 y 8. En la Figura 7 hacemos Dice por el mes de Mayo de 2004 y el Tipo de Llamada Entrante, filtrando así por dos dimensiones. El resultado está representado por la sección en color anaranjado.

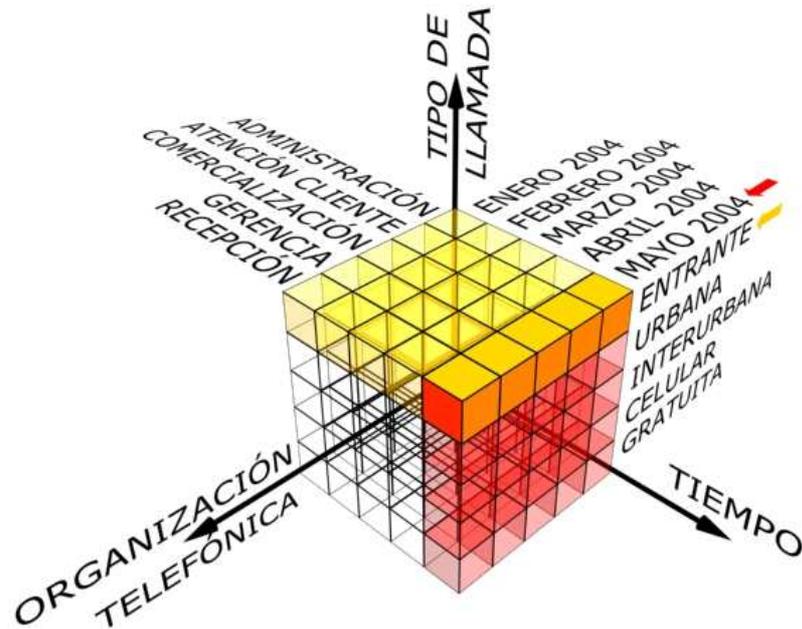
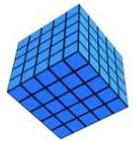


Figura 7 – Dice de un cubo para el mes de Mayo de 2004 y el Tipo de Llamada Entrante.

En la Figura 8 realizamos Dice agregando la dimensión Área al filtro, específicamente el Área Recepción. El resultado es el pequeño cubo en color gris.

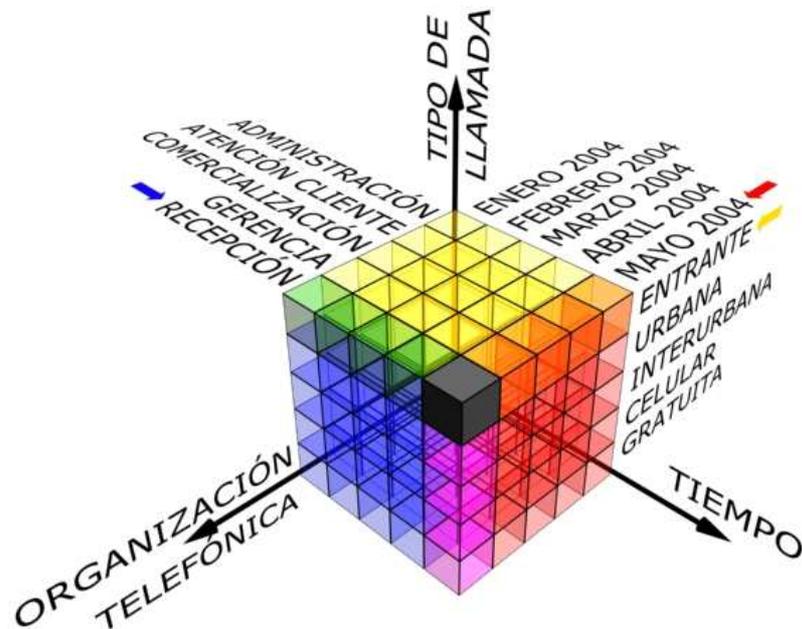
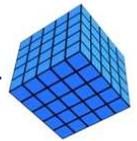
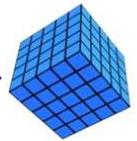


Figura 8 – Dice de un cubo para el mes de Mayo de 2004, el Tipo de Llamada Entrante y el Área Recepción.



Data Mining (Minería de Datos)

El Data Mining consiste en el análisis y modelización estadística de datos. Es una poderosa tecnología con gran potencial para ayudar a las compañías a concentrarse en la información más importante de su Data Warehouse. 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. Los análisis prospectivos automatizados ofrecidos por un producto así van más allá de los eventos pasados provistos por herramientas retrospectivas típicas de Sistemas de Soporte a las Decisiones. 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, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas.

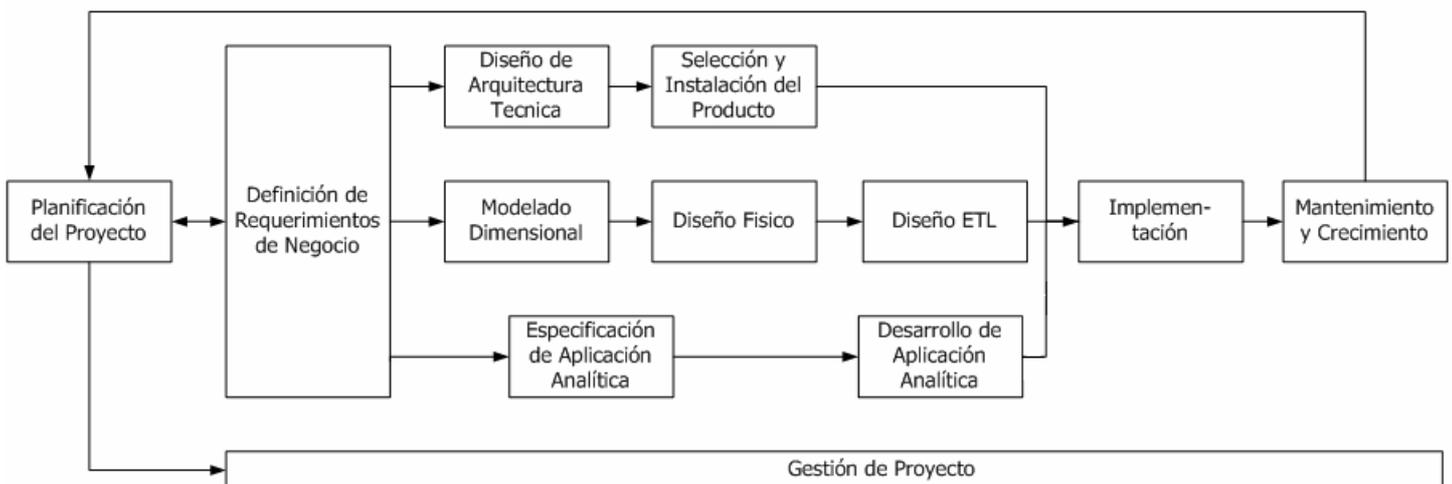


ciclo de vida

Para conducirnos a un lugar al que nunca fuimos debemos tener un mapa de ruta. Lo mismo pasa con un proyecto de Data Warehouse, es necesario contar con un ciclo de vida.

Pero antes de empezar con la definición del ciclo de vida es necesario tener en cuenta ciertas observaciones. En primer punto, un proyecto de Data Warehouse debe mantener su foco en las necesidades de negocio. En segundo lugar, los datos deben ser presentados en forma dimensional.

Vamos a usar el siguiente diagrama para encapsular las etapas del ciclo de vida. El diagrama ilustra la secuencia, dependencia y concurrencia de etapas. Este sirve como mapa de ruta para ayudar al equipo a realizar su trabajo en forma correcta en el momento correcto. El diagrama no refleja una línea de tiempo absoluta.

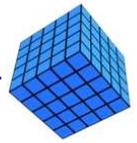


Planificación del Proyecto

El ciclo de vida del Data Warehouse comienza con el planeamiento del proyecto. Durante esta etapa se realizarán los siguientes pasos:

1. Evaluar la disposición de la organización a la iniciativa del Data Warehouse.
2. Definir el contexto del proyecto

Definir el contexto del proyecto permite precisar cual es el alcance del desarrollo, es decir, que vamos a realizar y que no. Es necesario que el contexto sea significativo en términos del valor para la organización, como también manejable.



En la mayoría de las organizaciones, se empieza con una implementación que produzca beneficios rápidos a un grupo de usuarios, después de definir un rumbo general y un conjunto general de objetivos.

El ámbito se puede restringir teniendo en cuenta:

- Presupuesto
- N° de fuentes seleccionadas
- N° de departamentos atendidos
- Tiempo asignado a un proyecto.

3. Definir la justificación del proyecto

La justificación requiere una estimación de los beneficios y costos asociados del Data Warehouse; por suerte, los beneficios anticipados pesan bastante más que los costos.

4. Definir el equipo de trabajo y los recursos

5. Selección de la estrategia de implementación

Enfoque de arriba hacia abajo:

De lo general a lo particular. Se identifican primero los requerimientos empresariales que debe cumplir el DW.

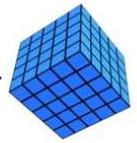
Se debe establecer el ámbito del DW, o sea las fronteras de los datos que definirán el territorio del mismo.

Ventajas:

- los requerimientos empresariales delimitan con claridad las fronteras de implementación.
- La tecnología es conducida por el negocio y no a la inversa.
- Resulta fácil comunicar los beneficios del DW a quienes toman decisiones y a los inversionistas.

Desventajas:

- En ocasiones pueden quedar oportunidades fuera del horizonte empresarial, resultado de demasiada concentración.
- La tecnología puede impulsar al negocio y ofrecer una ventaja competitiva que en un principio, no sea muy obvia para las actividades.



- Las expectativas iniciales pueden restringir la persecución de objetivos con recompensas potencialmente mayores una vez implementado.

Se recomienda cuando:

- La organización esta familiarizada con la tecnología
- Los ejecutivos y los que toman decisiones prevén un conjunto de objetivos claros para el DW
- Tienen una idea clara de dónde encaja el DW en su estructura organizacional como herramienta de apoyo a las decisiones.
- Tienen claro que el DW es un subproceso dentro de un proceso empresarial ya establecido.

Este enfoque es útil cuando la tecnología esta madura y se comprende bien y cuando están claros los problemas empresariales que deben resolverse.

Enfoque de abajo hacia arriba:

De lo particular a lo general. Comienza con experimentos y prototipos. Se selecciona un subconjunto específico, bien entendido, de la problemática empresarial y se formula una solución para ese subconjunto. Es más rápido. Es útil en las etapas de madurez de la tecnología.

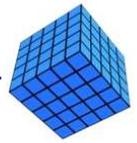
Permite con un gasto considerablemente menor evaluar los beneficios de la tecnología. Se toma este enfoque para implementar un Mercado de Datos que esta orientado a responder unas cuantas consultas bien definidas en un área determinada como contabilidad, análisis de mercado, administración de productos.

Ventajas:

- En ocasiones es importante la necesidad de implementación y de comenzar a usar el DW
- En las primeras etapas de la tecnología, permite evaluar beneficios sin grandes compromisos.
- La participación de poca gente trabajando en el proyecto.

Desventajas:

- Después de la implementación inicial hay que volver atrás y observar como se puede ampliar la solución para dar servicio a toda la empresa.
- La falla de un solo proyecto puede perjudicar la implementación de la tecnología



- El equipo de trabajo debe integrarse a uno mayor.

Se recomienda cuando:

- La tecnología se encuentra en el extremo izquierdo de la curva de madurez.
- La organización todavía no está comprometida con la tecnología DW, pero busca una evaluación tecnológica para determinar cómo, cuándo y dónde desplegar esta tecnología.
- La organización trata de obtener una idea de costos adicionales para implementarlo
- No están claros los objetivos empresariales.

6. Selección de un ámbito inicial de implementación:

En la mayoría de las organizaciones, se empieza con una implementación que produzca beneficios rápidos a un grupo de usuarios, después de definir un rumbo general y un conjunto general de objetivos.

El ámbito se puede restringir teniendo en cuenta:

- Presupuesto
- Nº de fuentes seleccionadas
- Nº de departamentos atendidos
- Tiempo asignado a un proyecto.

7. Enfoque de selección de una arquitectura:

- Almacenamiento operacional vs. uso de copias de datos operacionales
- Solo DW
- Solo Mercado de datos
- DW y mercado de datos

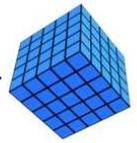
8. Selección de la metodología de desarrollo:

Método de desarrollo en espiral. El énfasis está en la velocidad y la culminación, con un reconocimiento de que los requerimientos no se pueden identificar con claridad al inicio.

Este método es partidario de la rápida generación de sistemas cada vez más funcionales con intervalos cortos entre versiones sucesivas.

Se usa cuando

- No se puede predecir con claridad ni anticipación los requerimientos.
- El tiempo de colocación es un ingrediente importante.



- Es necesaria una mejora iterativa para hacer correcciones de mercado.
- A la organización le toma un tiempo absorber las versiones sucesivas.
- La ventaja competitiva sostenida proviene de mejoras súbitas en forma continua.

9. Desarrollo y Mantenimiento del Plan de Proyecto

En este punto es necesario definir todas las tareas necesarias para el desarrollo del Data Warehouse, junto con sus tiempos y precedencias. Para ello suele ser útil desarrollar herramientas gráficas como Gantt o Perts.

Definición de Requerimientos de Negocio

Es una especificación de las funciones que se obtendrán del DW.

Además de las características y funciones, describirán el ambiente operativo en el que se entregará el DW.

El volumen de acopio inicial de requerimientos depende del enfoque de implementación (de abajo hacia arriba o viceversa).

Requerimientos del propietario:

La recolección de requerimientos debe dar respuesta a estas preguntas:

¿Por qué construir un DW o un Mercado de datos?

¿Qué problema empresarial abordará?

¿Cuáles son los objetivos empresariales?

¿Quién es el cliente?

¿Cuánto costará?

¿Cuándo estará listo?

¿Cuál es el impacto sobre la gente?, ¿Sobre las habilidades?, ¿Sobre la organización?

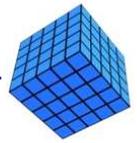
¿Cómo afecta nuestras inversiones actuales en cómputo?

¿Tenemos capacidad para hacerlo?

¿Cuáles son los riesgos?

Áreas de recolección:

- Objetivos empresariales
- Ambito y objetivos del DW / MD y clientes
- Fuentes de datos
- Un plan, Presupuesto, calendario y recursos

**Requerimientos de usuario final:**

Este ve al DW como una caja negra cuyo acceso es a través de aplicaciones y herramientas de consulta y reportes.

Para relevar los requerimientos se puede ayudar con los escenarios de uso empresarial. Se captan las consultas empresariales expresadas en terminología de usuario final:

- Departamento de ventas:

Selección de personal: ¿Cuál representante de ventas tiene menos clientes y mejores contratos?

Análisis de pedidos: ¿Cuál es nuestra fuente más grande de pedidos en navidad?

Administración de clientes: ¿Cuáles son nuestros mejores clientes?, ¿en este año?, ¿en el año anterior?, ¿Por qué algunos clientes salieron de la lista este año?

Evaluación corporativa: ¿Qué región ha sido consistentemente la mejor durante los últimos 6 trimestres?

Éxito de producto: ¿Cuáles son los 10 primeros productos en mi región para las ventas en los últimos 4 trimestres?

- Departamento de comercialización:

Mezcla de productos: ¿Cuáles son los 10 mejores productos del año pasado?, ¿De éste año?

Análisis de promoción: ¿Qué tan buenas fueron las promociones de este año comparadas con las del año anterior?, ¿Qué productos responden bien a las promociones?, ¿Qué promociones deberíamos discontinuar?

Desviaciones regionales: ¿Cuáles fueron los 2 principales productos por región el año anterior? ¿En los últimos dos años?

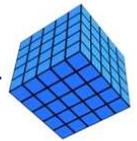
Cambios demográficos: ¿Cómo se correlacionan los cambios demográficos en una región con los cambios en el éxito del producto?

Posición del mercado: ¿Qué mercados dominamos este año?, ¿El año pasado?

¿Quiénes son los competidores en estos mercados?

Pista de Tecnología***Diseño de la Arquitectura Técnica***

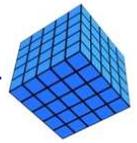
Como los planos para una nueva casa, la arquitectura técnica es el plano para los elementos y servicios técnicos del Data Warehouse. La arquitectura sirve como un marco de trabajo para



soportar la integración de tecnologías. Como un plano, la arquitectura técnica consiste en una serie de modelos que profundizan en un mayor detalle, que considera cada uno de los componentes principales. En ambas situaciones, la arquitectura nos permite descubrir los problemas en el papel (como tener el lavaplatos muy lejos del fregadero) y minimizar las sorpresas a mitad del proyecto. La arquitectura identifica los componentes inmediatamente requeridos contra aquéllos que se incorporarán más tarde. Más importante, es que la arquitectura sirve como una herramienta de comunicación. Los planos de construcción de una casa permiten al arquitecto, contratista general, subcontratantes, y propietario comunicarse mediante un documento común.

Selección e Instalación del Producto

Usando las capacidades identificadas en el diseño de la arquitectura como una lista de compras, vamos a evaluar y seleccionar los productos específicos que se amolden a nuestras necesidades de funcionalidad, que encajen en nuestro marco de trabajo definido en la etapa anterior, y produzcan el menor costo.



Pista de Datos

Modelado Dimensional

En esta etapa vamos a transformar los requerimientos en un modelo dimensional. Para esto se deben seguir cuatro pasos:

1. Seleccionando el área tema a modelar

Un área tema son temas de interés de diversas funciones empresariales. Ej.: en área comercialización: investigación de mercados, análisis de competitividad, comportamiento del comprador, segmentación de mercado, decisiones sobre precios y presupuesto, decisiones sobre productos, sobre promoción, sobre canales, pronóstico de tendencias, pruebas de calidad. Estos temas producen las siguientes áreas tema:

Pedidos

Promociones

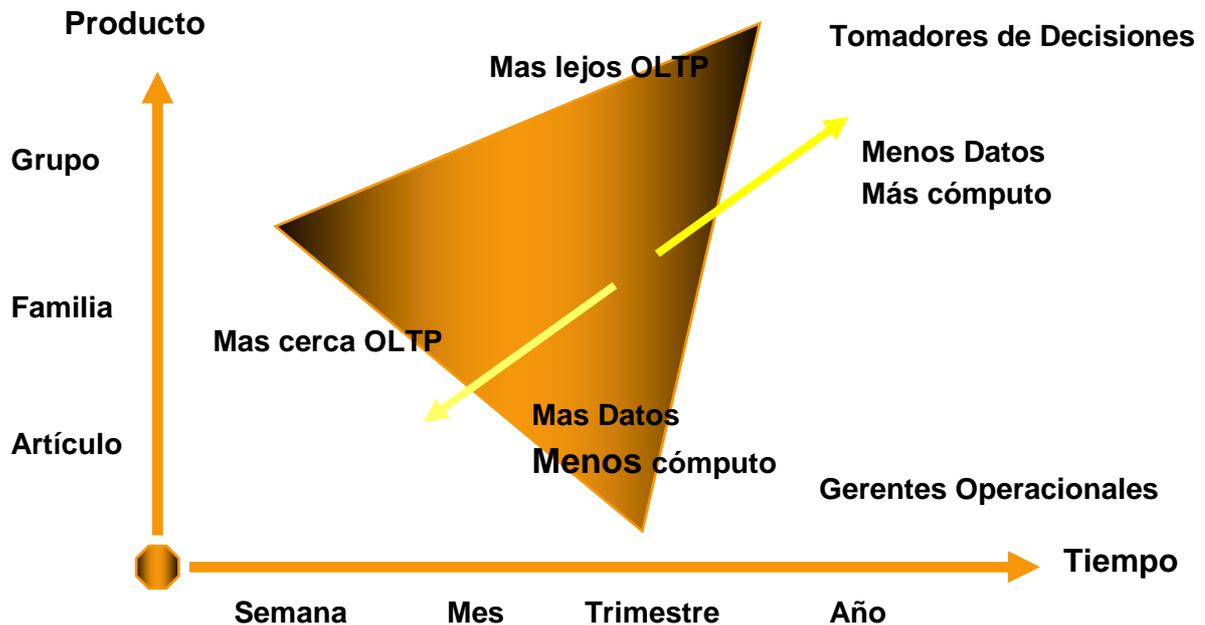
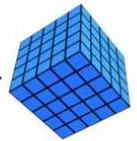
Mercados

Ventas

Ciclo de tiempo

2. Declarando la granularidad del área tema

La granularidad esta referido al nivel de detalle de la información requerida. Tiene relación directa con las actividades de resumen y adición. A menor nivel de granularidad, mayor nivel de detalle. Los datos operacionales se consideran con el nivel mas bajo de granularidad. A mayor granularidad, mayor cantidad de procesamiento para convertir y resumir datos operacionales y menor volumen de almacenamiento y se pueden consultar con mayor rapidez. Declarar la granularidad significa especificar lo que un registro de la tabla de hecho representa exactamente. La granularidad comunica el nivel de detalle asociado con las medidas de de la tabla de hechos. Proporciona la respuesta a la pregunta, "¿Cómo usted describe un solo registro la tabla de hechos?"



Ejemplos:

- ¿Cuál es el nº típico de clientes por representante de ventas? Esto ayuda a determinar si los clientes estan bien atendidos o si los vendedores estan sobrecargados.
- ¿Quién es el representante de ventas con más alto ingreso?
- ¿Cuál es el ingreso por región en los últimos 6 meses?
- ¿Cuáles son los nombres de los 10 primeros productos por región para ventas en volumen?

3. Eligiendo las dimensiones

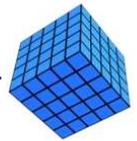
Las dimensiones responde a la pregunta, "¿Cómo la gente de negocio describen los datos que son el resultado del área tema?" Nosotros queremos dar contexto a nuestras tablas de hecho con un conjunto robusto de dimensiones que representen todas las posibles descripciones que se necesiten asumir en el contexto de cada medida.

Una dimensión importante es el tiempo. Otros ejemplos son: grupo de clientes, familia de productos, geografía, estructura de la organización, etc.

4. Identificar las medidas que van a poblar la tabla de hechos

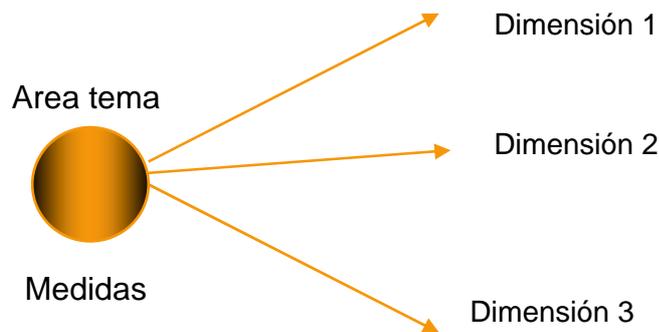
Las medidas van a ser determinadas respondiendo a la pregunta, "¿Qué quiero medir?". Todas las medidas candidatas deben estar validadas por la granularidad definida en el paso 2

Una vez finalizado estos pasos debemos definir el modelo starnet.



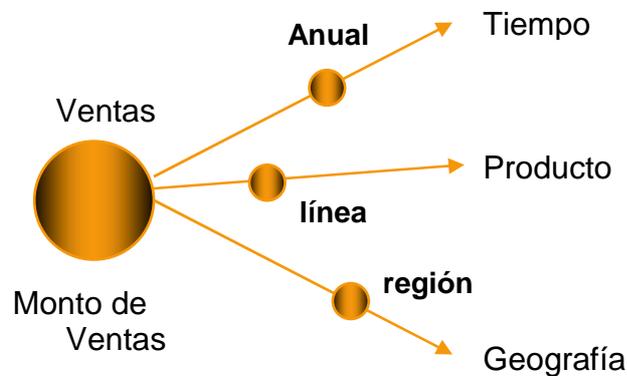
La técnica de modelado para entender, modelar y analizar consultas empresariales consiste en construir moldes de consultas.

Permite buscar candidatos potenciales de tablas de hecho y dimensión.

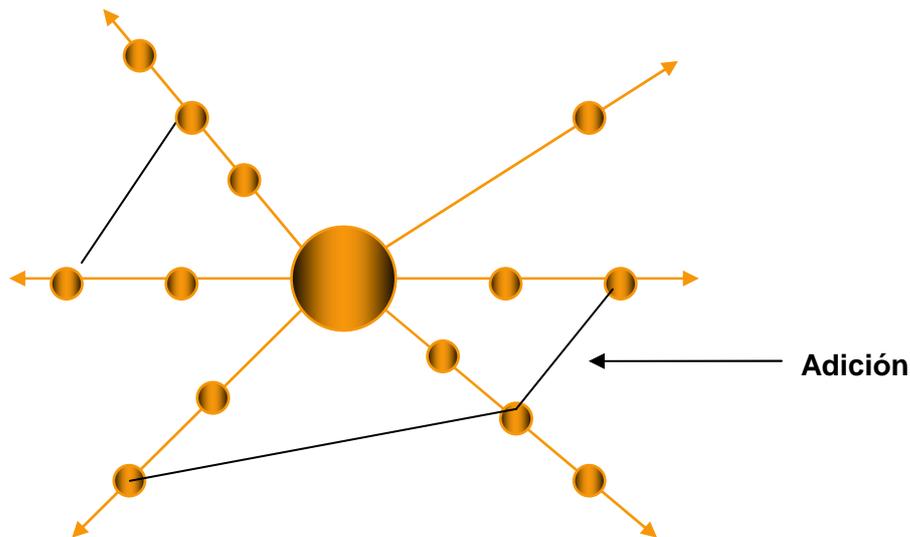
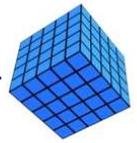


Ejemplo:

Consulta empresarial: Ventas anuales por línea de producto y región



Una vez identificados los moldes de consulta para cada consulta empresarial, las consultas y sus moldes se consolidan. El diagrama consolidado resultante para cada área tema es el modelo starnet (red estrella). Es la representación combinada de todas las consultas empresariales.



Diseño Físico

Los modelos dimensionales desarrollados en la sección precedente necesitan ser traducidos en un diseño. El modelo Físico difiere del modelo lógico en cuanto a los detalles especificados para la base de datos física, incluyendo los nombres de columna físicos, los tipos de datos, las claves, y los permisos de valores nulos. A estas alturas el diseño físico también contiene cosas que mejoren el desempeño del Data Warehouse, como ser índices o atributos calculados.

Diseño ETL

En esta etapa debemos diseñar todo el proceso ETL, que se encarga de pasar los datos de los sistemas operacionales al Data Warehouse, teniendo muy en cuenta el diseño de la staging area y los valores erróneos.

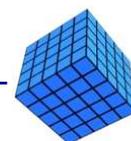
Pista de Aplicación Analítica

Especificación de Aplicación Analítica

En esta etapa nos encargamos de definir las aplicaciones necesarias que cumplan con los requerimientos del cliente, como pueden ser reportes, tableros de control, gráficos, etc.

Desarrollo de Aplicación Analítica

En este punto debemos realizar el desarrollo de las aplicaciones analíticas definidas en el punto anterior. Es importante tener en cuenta que esta etapa puede comenzar una vez que el



diseño físico se ha realizado. Aquí es cuando la herramienta de acceso elegida toma un papel importante ya que define la simplicidad o complejidad de este desarrollo.

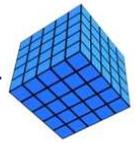
Implementación

Una vez finalizadas las tres pistas podemos comenzar con la implementación del Data Warehouse. Este es el punto que más dificultad estima el tiempo de desarrollo.

Como todo desarrollo de software, el Data Warehouse, conlleva una serie de versiones previas antes de su versión final.

Mantenimiento y Crecimiento

En esta etapa como su nombre lo indica nos encargamos de realizar el soporte a los usuarios, controlando que estos lo utilicen de la manera correcta; ofrecer un programa de educación que explique el funcionamiento del Data Warehouse junto con sus capacidades de uso; ofrecer el mantenimiento de errores que lleguen a existir; como también controlar su crecimiento, para que el sistema no colapse ante las nuevas necesidades.



Herramienta OLAP - 03

O3 es un producto de Business Intelligence con multi-propósito que puede ser utilizado por cualquier empresa o persona que necesite analizar información, consolidarla a partir de varias fuentes de datos, suministrarla (gratuitamente o no) o simplemente presentarla como en un reporte, pero diseñado por el propio usuario final de acuerdo a sus necesidades y al instante.

La tecnología OLAP asiste al usuario en el análisis de la información de la empresa, a través de una mirada personalizada y comparativa, que incluye el análisis histórico y proyectado de los datos en distintos escenarios.

O3 posee un conjunto de funcionalidades que le convierten en una herramienta OLAP potente, flexible y de fácil manejo, entre las que citaremos:

- Posibilita la realización de un análisis multidimensional y dinámico de los datos consolidados, dándole al usuario la capacidad de navegarlos estableciendo sus propios caminos de análisis.
- Permite el análisis gráfico y en forma tabular de información proveniente de una diversidad de base de datos (ya sean relacionales, archivos ASCII, DataWarehouses, o legacy systems) en forma simultánea, integrándolos en un modelo multidimensional.
- O3 da al usuario la oportunidad de comprender no solo lo que está sucediendo, sino cuándo, dónde, por qué, y cómo. Adicionalmente O3 permite responder a interrogantes más estratégicas del tipo What if.

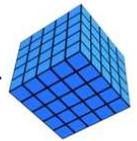
O3 cuenta con diversas funcionalidades que posibilitan analizar profundamente la información disponible desde diferentes ángulos, con una exploración y navegación flexible:

- **Drag & Drop:** Es la acción de arrastrar las diferentes dimensiones del cubo sobre los distintos ejes del gráfico o eventualmente en el formato de tabla, llevarlas hasta ubicarlas en las filas o columnas deseadas. Permite cambiar el contexto de análisis de los datos.
- **Drill Down & Roll Up:** Dado que las dimensiones se organizan de forma jerárquica, es posible analizar la información en distintos niveles de detalle, pudiendo fácilmente navegar hacia el nivel más global (Roll Up) o al nivel de mayor detalle. (Drill Down).
- **Filtros:** Los modelos multidimensionales, data-marts o cubos, tienen varias dimensiones, las cuales pueden guiar el análisis de la información, directamente poniéndolas en los ejes del gráfico o filtrándolas desde los combos de dimensiones.



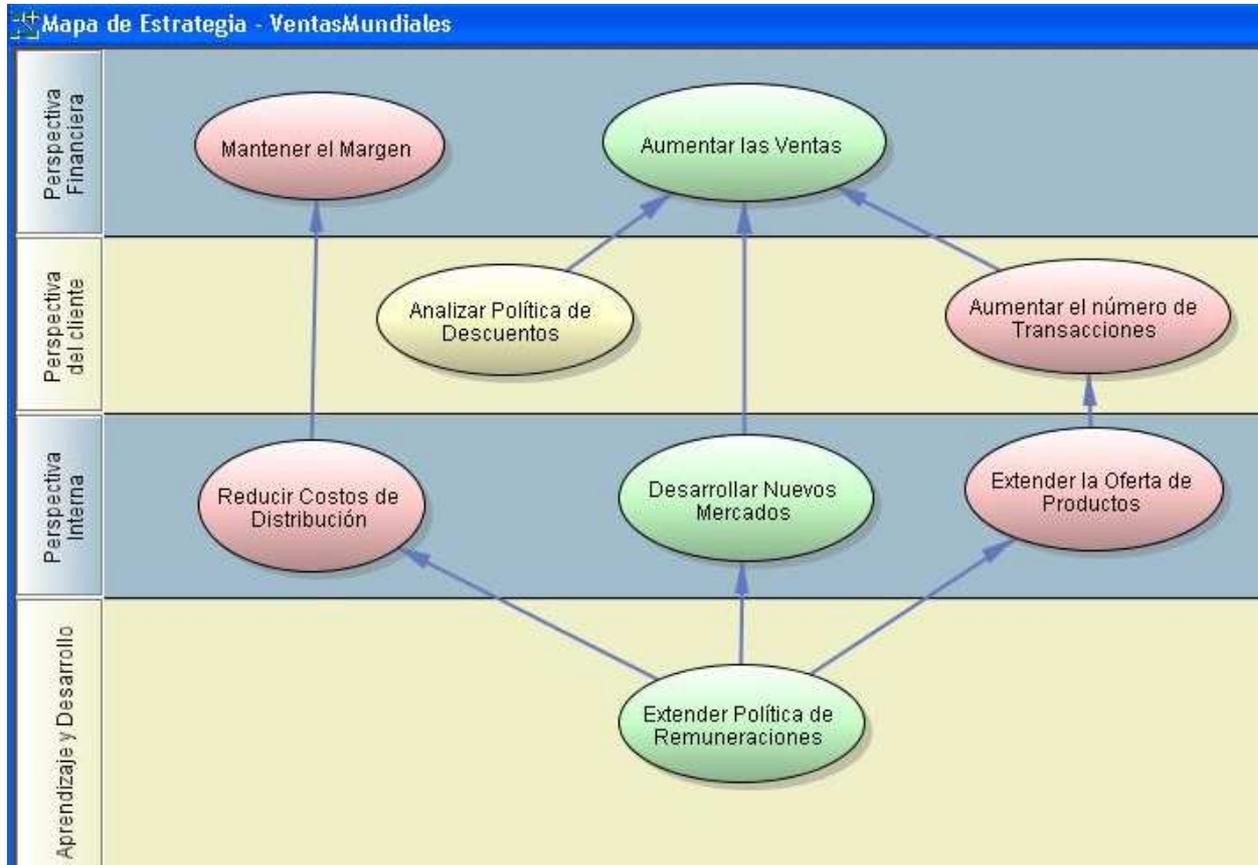
De este modo se analiza la información en hipercubos con la cantidad de dimensiones requeridas y no en un plano cartesiano con únicamente dos dimensiones.

- **Ocultar-Visualizar:** Es posible ocultar y visualizar elementos de las dimensiones que estén en los ejes del gráfico o en las filas o columnas de la tabla permitiendo la selección de los elementos en los cuales concentrar nuestra atención.
- **Ranking y Selección:** Esta funcionalidad permite ordenar los elementos de una dimensión, ya sea según el valor total, por orden alfabético de sus etiquetas o por algún elemento de dimensiones establecidas en alguno de los ejes de análisis. Además de ordenar los valores o etiquetas, O3 permite seleccionar un conjunto de elementos según algún criterio definido por el usuario en el momento de realizar la selección (por ejemplo seleccionar los 5 productos más vendidos, o los productos que reúnan el 60% de las ventas, o inclusive los primeros elementos que sumen 100 unidades vendidas).
- **Mostrar Totales:** Es posible mostrar totales y subtotales en los ejes del gráfico, así como totalizar las filas y columnas de la grilla, a través del uso del menú que se despliega con un simple click del botón derecho.
- **Desplegar Porcentajes:** Es posible mirar los valores en porcentajes, en función de uno de los ejes o de los dos, ó en la tabla según filas y columnas.
- **Tipos Gráficos:** O3 tiene ofrece diferentes tipos de gráficos. Como ejemplo citamos: Barras apiladas, Puntos, Líneas, Sectores, Cintas, Pirámides, gráficos de sectores.
- **Tabla:** Además de las funcionalidades ya descritas, el formato tabla permite agrupar dimensiones en las filas y columnas. Esto significa que podemos agrupar los elementos de una dimensión (por ejemplo clientes) según otra (por ejemplo países), pudiendo analizar las ventas realizadas a los clientes según cada país, y de manera muy sencilla agregar subtotales por país. O3 ofrece un gran poder de cálculo posibilitando la definición de fórmulas que pueden agregarse a los cubos mediante un sencillo lenguaje de expresiones. Tales fórmulas pueden ser estándares, predefinidas por el producto ó definidas por el usuario.
- **Cálculos en la Tabla:** La tabla tiene un gran potencial en cuanto a cálculos se refiere. Por un lado existe la posibilidad de realizar cálculos a distintos niveles:
 - global
 - a nivel de dimensión (en caso de tener anidaciones para esa dimensión)



- o a nivel de elemento (en caso de tener anidaciones para ese elemento).

Las fórmulas que se pueden agregar pueden ser estándares, predefinidas por el producto (por ejemplo suma, mínimo, promedios) o definidas por el usuario.



Archivo Escritorio Editar Ver Explorar Herramientas Ventanas Apariencia Ayuda

Sales Analysis

Viewer - file:C:\03\files\VentasMundiales\VentasMundiales.bsDat

Evolución de las ventas

Estado

Fecha: Trimestre 1 / 2001

| Ubicación | Valor | Meta | Tolerancia |
|----------------|-----------|----------|------------|
| Estados Unidos | 699565,00 | 40000,00 | 3000,00 |
| Mexico | 25495,00 | 40000,00 | 3000,00 |
| Inglaterra | 127421,00 | 40000,00 | 3000,00 |
| Francia | 12000,00 | 40000,00 | 3000,00 |
| España | 157540,00 | 40000,00 | 3000,00 |
| Argentina | 56100,00 | 40000,00 | 3000,00 |
| Brasil | 34209,00 | 40000,00 | 3000,00 |

Cuadro de Mando Integral
 - Perspectiva Financiera
 - Mantener el margen
 - Margen por país
 - Aumentar las ventas
 - Evolución de las ventas
 - Perspectiva del Cliente
 - Analizar política de descuentos a clientes
 - Análisis de descuentos
 - Aumentar el número de transacciones
 - Número de transacciones
 - Perspectiva Interna
 - Reducir los costos de distribución
 - Costos de distribución
 - Desarrollar nuevos mercados
 - Introducir la distribución en Italia
 - Introducir la distribución en Brasil
 - Extender la oferta de productos
 - Introducción de bananas
 - Introducción de Kiwis
 - Perspectiva de Aprendizaje y Desarrollo
 - Política de remuneraciones
 - Porcentaje destinado a comisiones

