

1.La Tecnología Datawarehousing

1.1 Fundamento.

Los sistemas de Data Warehousing son el centro de la arquitectura de los Sistemas de Información de los 90's. Han surgido como respuesta a la problemática de *extraer información sintética a partir de datos atómicos almacenados en bases de datos de producción*. Uno de los objetivos principales de este tipo de sistemas es servir como base de información para la toma de decisiones. Los beneficios obtenidos por la utilización de este tipo de sistemas se basan en el acceso interactivo e inmediato a información estratégica de un área de negocios. Este acercamiento de la información al usuario final permite una toma de decisiones rápida y basada en datos objetivos obtenidos a partir de las bases de datos (eventualmente heterogéneas) de la empresa. Estos beneficios aumentan cuanto más importantes son las decisiones a tomar y cuanto más crítico es el factor tiempo.

Hoy en día las empresas cuentan en su mayoría con la automatización de sus procesos, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea. En esta información descansa el know-how de la empresa, constituyendo un recurso corporativo primario y parte importante de su patrimonio.

El nivel competitivo alcanzado en las empresas les ha exigido desarrollar nuevas estrategias de gestión. En el pasado, las organizaciones fueron típicamente estructuradas en forma piramidal con información generada en su base fluyendo hacia lo alto; y era en el estrato de la pirámide más alto donde se tomaban decisiones a partir de la información proporcionada por la base, con un bajo aprovechamiento del potencial de esta información. Estas empresas, han reestructurado y eliminado estratos de estas pirámides y han autorizado a los usuarios de todos los niveles a tomar mayores decisiones y responsabilidades. Sin embargo, sin información sólida para influenciar y apoyar las decisiones, la autorización no tiene sentido.

Esta necesidad de obtener información para una amplia variedad de individuos es la principal razón de negocios que conduce al concepto de Datawarehouse. El énfasis no está sólo en llevar la información hacia lo alto sino que a través de la organización, para que todos los empleados que la necesiten la tengan a su disposición.

Las Tecnologías de la Información (IT) han cambiado sustancialmente la forma de hacer negocios de las empresas. En un entorno donde la competitividad, la globalización, la consolidación de industrias, un ciclo de vida mas corto de los productos, saturación de mercados, etc. La información juega cada vez un papel más preponderante. La información referentes a mercados, competidores, clientes, incluso la relativa a los indicadores de rendimiento de la propia compañía, se ha convertido en un recurso clave. El problema radica en que las empresas disponen de una gran cantidad de datos, pero muy poca información. Varias razones motivan estos hechos: islas de información, carencia de arquitectura, gestión, responsabilidad, posesión de los datos, deficiencia en calidad, contenido, accesibilidad, fiabilidad de la información, múltiples y diversas aplicaciones operacionales, existencia de fuentes de información externa, etc. Gran parte del producto generado por tecnologías de información, no es información, sino solo datos brutos. Son generados por sistemas que fueron ideados para recogerlos, pero no para analizarlos. Los datos adquieren la categoría de información cuando disponen de una estructura inteligente. A su vez, esta información se convertirá en conocimiento si se le añade la ideas, intuición, capacidad del analista, es decir, conocimiento tácito. En puridad, según Nokata y Takeuchi, la información sería el conocimiento explícito, es decir algo susceptible de ser transmitido, pero solo la información no será capaz de aumentar y mejorar la base de conocimiento de una compañía. Es la inclusión del conocimiento tácito, la que promueve el ciclo virtuoso de la transformación de datos en información, información en conocimiento, y finalmente, conocimientos en acciones/decisiones mejor informadas y más afines a la realidad de la compañía. El ciclo se cierra al generar estas nuevas acciones/decisiones, más datos brutos que realimentaran el ciclo de la Gestión del Conocimiento.

Data Warehousing y Data Mining son dos procesos, entre otros, que posibilitarán la Gestión del Conocimiento en una compañía.

La aparición, relativamente reciente de estas tecnologías, que han surgido con una inusitada fuerza en el mercado de IT, han sorprendido a muchos profesionales del sector, y han generado una tremenda confusión en el mercado.

Un *Sistemas de Data Warehousing* incluye funcionalidades tales como:

- *Integración de bases de datos* heterogéneas (relacionales, documentales, geográficas, archivos, etc.).
- *Ejecución de consultas complejas no predefinidas* visualizando el resultado en forma de gráfica y en diferentes niveles de agrupamiento y totalización de datos.
- *Agrupamiento y desagrupamiento* de datos en forma interactiva.
- *Análisis de problema en términos de dimensiones*. Por ejemplo, permite analizar datos históricos a través de una dimensión tiempo.
- *Control de calidad de datos* para asegurar, no solo la consistencia de la base, sino también la relevancia de los datos en base a los cuales se toman las decisiones.

1.2 Qué es un Data Warehouse

Un Data Warehouse es una colección de datos

- orientada a sujetos
- integrada
- variante en el tiempo
- no volátil

que soporta el proceso de toma de decisiones. Un Data Warehouse soporta procesamiento informático, brindando una sólida plataforma de datos históricos, integrados, de los cuales hacer análisis.

1.3 Es orientado a sujetos:

Un primer aspecto de un Data Warehousing es que esta orientado a los mayores sujetos de la empresa. El mundo operacional esta diseñado alrededor de aplicaciones y funciones, como por ejemplo pagos, ventas, entregas de mercadería, para una institución comercial. Un Data Warehouse esta organizado alrededor de los mayores sujetos, como cliente, vendedor, producto y actividades El mundo operacional concierne al diseño de la base de datos y al diseño de procesos. Un Data Warehousing está enfocado en la modelización de los datos y el diseño de la base de datos, exclusivamente. El diseño de procesos (en su forma clásica) no es parte del Data Warehouse.

1.4 Los datos son integrados:

El aspecto más importante del ambiente de un Data Warehouse es que sus datos están integrados. Cuando los datos son movidos del ambiente operacional, son integrados antes de entrar en el Warehouse. Por ejemplo, un diseñador puede representar el sexo como "M" y "F", otro puede representarlo como "0" y "1", o "x" e "y", y otro usar las palabras completas "masculino" y "femenino". No importa la fuente de la cual el sexo llegue al Data Warehouse, debe ser guardado en forma consistente; los datos deben ser integrados.

1.5 Es variante en el tiempo

Los datos en el Warehouse son precisos para un cierto momento, no necesariamente ahora; por eso se dice que los datos en el Warehouse son variantes en el tiempo. La varianza en el tiempo de los datos de un Warehouse

se manifiestan de muchas maneras. El Data Warehouse contiene datos de un largo horizonte de tiempo. Las aplicaciones operacionales, sin embargo, contienen datos de intervalos de tiempo pequeños, por cuestiones de performance (tamaño chico de las tablas). Toda estructura clave en un Warehouse contiene implícita o explícitamente un elemento del tiempo. Esto no necesariamente pasa en el ambiente operacional. Los datos de un Warehouse, una vez almacenados, no pueden ser modificados (no se permiten updates). En el ambiente operacional, los datos, precisos al momento de acceso, pueden ser actualizados, según sea necesario.

1.6 Es simple de manejar

Updates, inserts y deletes son efectuados regularmente, en una base de record–por–record, a los datos operacionales. La manipulación de datos en un Warehouse, es mucho más sencilla. Solo ocurren dos operaciones, la carga inicial, y el acceso a los datos. No hay necesidad de updates (en su sentido general). Hay consecuencias muy importantes de esta diferencia de procesos con un sistema operacional: A nivel de diseño, en un Warehouse, no hay que controlar anomalías producidas por los updates, ya que no hay updates. Se pueden tomar libertades de diseño físico como optimizar el acceso a los datos, y denormalización física. Otra consecuencia es la simplicidad de la tecnología del Warehouse, en lo que respecta a backups, recuperación, locks, integridad, etc.

El DW (de ahora en adelante los términos Data Warehouse, Datawarehousing, Warehouse y DW serán utilizados en forma indistinta) convierte entonces los datos operacionales de una organización en una herramienta competitiva, por hacerlos disponibles a los empleados que lo necesiten para el análisis y toma de decisiones.

El objetivo del DW será el de satisfacer los requerimientos de información interna de la empresa para una mejor gestión. El contenido de los datos, la organización y estructura son dirigidos a satisfacer las necesidades de información de los analistas. El DW es el lugar donde la gente puede acceder sus datos.

No es atípico encontrar experimentos profesionales de IT, que no tienen claro cuál es la diferencia entre un Data Warehouse y el Data Ware housing, entre el Data Warehousing y el Data Mining, si la OLAP es Data Warehousing, o si hacer drill–down es Data Mining.

Según Bill H. Inmon, padre del Data Warehousing Un Data Warehouse es un conjunto integrado de bases de datos, con orientación temática, que están diseñados para el apoyo a la Toma de Decisiones, y donde cada unidad de datos es relevante en algún momento del tiempo.

Un Data Warehouse o Almacén de Datos, es una gran base de datos, normalmente medida en gigabytes (miles de millones de caracteres) o terabytes (billones de letras), que recoge información de múltiples sistemas fuentes u operacionales dispersos, y que su actividad se centra en la Toma de Decisiones, es decir, en el análisis de la información, en vez de su captura. Idealmente, toda pieza de información vertida en el Almacén será utilizada y no será redundante.

Podemos decir que un Almacén de Datos es un sistema que ofrece acceso fácil y rápido a información vital para la gestión y operación de una empresa, que permite acceso a datos históricos y detallados, que es un sistema donde la información es alimentada por múltiples sistemas fuentes, y que es un sistema de análisis de información destinado a todo tipo de usuario.

Data Warehousing, o Almacenamiento de Datos, es el proceso que facilita la creación y explotación de un Data Warehouse. Data Warehousing, es un proceso complejo, que hace real la Gestión del Conocimiento. Para conseguirlo será necesaria la aplicación de una Metodología, y la implantación de una Arquitectura Tecnológica de la Gestión del Conocimiento. Data Warehousing es un metaproceso compuesto por un conjunto de procesos. Por citar los más conocidos, formarían parte de este proceso la creación del Data Warehouse, su explotación analítica mediante técnicas de OLAP (On–line analytical processing) o técnicas de

Data Mining.

El concepto DataMart es una extensión natural del Data Warehouse, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de Finanzas o Marketing. Permitiendo así un mejor control de la información que se está abarcando.

Toda empresa puede ser vista en base al proceso productivo que la sustenta. El resultado de los costos y beneficios de este proceso productivo forman una cadena de valor, donde cada eslabón (proceso de negocios) adiciona valor a la empresa. De esta forma es claro, que las empresas deben buscar optimizar cada uno de sus eslabones sin perder de vista la cadena total.

Al manejar eficientemente la información de cada área de la empresa, se pueden tomar mejores decisiones y así efectuar acciones apropiadas y finalmente conseguir un mejor control sobre la producción empresarial.

En esta nueva tecnología cada eslabón de la cadena de valor será representado por una base de datos multidimensional, la cual permite potencialmente administrar la etapa productiva que representa. La cadena de valor total será representada entonces por el conjunto de bases de datos multidimensionales asociadas a cada eslabón.

OLAP o procesamiento analítico en línea, es el paradigma de análisis multidimensional de un Data Warehouse. Este concepto se contrapone al de OLTP o procesamiento transaccional en línea, que es el empleado por los sistemas fuentes para optimizar la recogida de información. En pocas palabras, mientras OLTP se encarga de procesar óptimamente multitud de pequeñas transacciones de captura de información (su entrada, modificación o eliminación), OLAP se dedica al análisis de enormes cantidades de información; por ende realizará pocas transacciones, pero éstas procesarán un volumen muy superior (cientos de miles de registros). Esto es viable, debido a que la información contenida en el Data Warehouse que va a ser analizada con OLAP, está resumida y agregada. Ambos paradigmas de acceso a la información, son contrapuestos, pero complementarios. OLTP se encargará de la recogida de la información en un modelo normalizado de base de datos, optimizado para procesar las transacciones en menos de un segundo, y la OLAP se dedicará al análisis de esa información, en un modelo multidimensional que facilitará la exploración y acceso a la información, invirtiendo minutos en obtener dichas vistas multidimensionales de los indicadores de rendimiento de la compañía.

1.7 Datawarehousing.

En primer lugar, DW no es un producto que pueda ser comprado en el mercado, sino más bien un concepto que debe ser construido. DW es una combinación de conceptos y tecnología que cambian significativamente la manera en que es entregada la información a la gente de negocios. El objetivo principal es satisfacer los requerimientos de información internos de la empresa para una mejor gestión, con eficiencia y facilidad de acceso.

La manera tradicional hasta ahora de entregar la información es a través de emisión de reportes impresos desde los sistemas operacionales, con consultas a nivel de cliente y extracción ocasional de datos para suplir actividades basadas en papel. Los problemas con la entrega de la información actual son muchos, incluyendo inconsistencia, inflexibilidad y carencia de integración a través de la empresa.

El DW puede verse como una bodega donde están almacenados todos los datos necesarios para realizar las funciones de gestión de la empresa, de manera que puedan utilizarse fácilmente según se necesiten. El contenido de los datos, la organización y estructura son dirigidos a satisfacer las necesidades de información de analistas. Los sistemas transaccionales son dinámicos, en el sentido que constantemente se encuentran

actualizando datos.

Analizar esta información puede presentar resultados distintos en cuestión de minutos, por lo que se deben extraer y almacenar fotografías de datos (snapshots), para estos efectos, con la implicancia de un consumo adicional de recursos de cómputo. Llevar a cabo un análisis complejo sobre un sistema transaccional, puede resultar en la degradación del sistema, con el consiguiente impacto en la operación del negocio.

El Data Warehouse intenta responder a la compleja necesidad de obtención de información útil sin el sacrificio del rendimiento de las aplicaciones operacionales, debido a lo cual se ha convertido actualmente en una de las tendencias tecnológicas más significativas en la administración de información.

Los almacenes de datos (o Datawarehouse) generan bases de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente. Estos datos se mantienen actualizados, pero no cambian al ritmo de los sistemas transaccionales. Muchos datawarehouses se diseñan para contener un nivel de detalle hasta el nivel de transacción, con la intención de hacer disponible todo tipo de datos y características, para reportar y analizar. Así un datawarehouse resulta ser un recipiente de datos transaccionales para proporcionar consultas operativas, y la información para poder llevar a

cabo análisis multidimensional. De esta forma, dentro de una almacén de datos existen dos tecnologías complementarias, una relacional para consultas y una multidimensional para análisis.

Existen muchas definiciones para el DW, la más conocida fue propuesta por Inmon (considerado el padre de las Bases de Datos) en 1992: Un DW es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales. En 1993, Susan Osterfeldt publica una definición que sin duda acierta en la clave del DW: Yo considero al DW como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. DW elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico.

Esta última definición refleja claramente el principal beneficio que el datawarehouse aporta a la empresa, eliminar aquellos datos que obstaculizan la labor de análisis de información y entregar la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión.

Puede considerarse que el modelo relacional en el cual se basa OLTP (Procesamiento Transaccional en Línea), tiene como objetivo guardar la integridad de la información necesaria para operar un negocio de la manera más eficiente. Sin embargo, este modelo no corresponde a la forma como el usuario percibe la operación de un negocio.

De hecho Codd, quien fue uno de los desarrolladores originales del concepto relacional, dijo: Aunque los RDBMS han sido tan beneficiosos para los usuarios, nunca han sido diseñados para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos.

DW se sustenta en un procesamiento distinto al utilizado por los sistemas operacionales, OLAP (Procesamiento Analítico En Línea), el cual surge como un proceso para ser usado en el análisis de negocios y otras aplicaciones que requieren una visión flexible del negocio.

1.8 Los requerimientos de un Data Warehouse

¿Cómo se empiezan a formular los requerimientos para un Data Warehouse? Existen muchas formas de considerar un Data Warehouse. Las siguientes son algunos ejemplos:

- Un Data Warehouse es simplemente un sistema de aplicación empresarial con su propia base de datos. Esta base de datos se genera a partir de otras bases de datos operacionales, no de información inicial que se introduce. El Data Warehouse ofrece una serie de características y funciones para implementar

procesos empresariales y enlazarlos con otros procesos fuera del ámbito del Data Warehouse. En forma muy similar a otros sistemas de aplicaciones empresariales, se requiere que el Data Warehouse proporcione al usuario final estos conjuntos prescritos de características y funciones del modo más eficiente posible.

- Un Data Warehouse es una capacidad latente. Almacena información resumida que se organiza de acuerdo con temas empresariales, tales como clientes y productos, para analizar la información con más facilidad. La carga de mostrar, organizar y reportar la información que guarda el Data Warehouse corresponde a las herramientas que deben incorporarse en el Data Warehouse. En esta visión, un Data Warehouse tiene una capacidad latente que sólo se vuelve útil cuando las herramientas de análisis y reporte se aplican con inteligencia a los datos que conserva el Data Warehouse. Se requiere que el Data Warehouse sustente un extenso rango de herramientas de acceso, operadas por un extensa gama de usuarios finales. El Data Warehouse también debe guardar y administrar un ámbito de información grande para servir a una extensa clientela.
- El Data Warehouse es una base de datos históricos, la cual es una acumulación de muchos años de información transaccional en línea, organizada para hacer eficiente el almacenamiento y facilitar la recuperación. Es necesario que el Data Warehouse organice grandes cantidades de información de manera compacta y eficiente. También se requiere que proporcione técnicas para resumir, a fin de que los usuarios finales comprendan las lecciones de los antecedentes con más facilidad.
- En ocasiones, el Data Warehouse es una tienda de datos operacionales. Entrega información operacional a un amplio rango de usuarios copiando información de los sistemas operacionales de bases de datos. En este caso, se requiere que el Data Warehouse distribuya información operacional de manera eficiente a un gran rango de usuarios. También se necesita que el Data Warehouse haga los cambios tecnológicos necesarios para mover la información de su base de datos operacional a la tecnología de almacenamiento que se emplea en el Data Warehouse.

Es evidente que los requerimientos de un Data Warehouse son tan variados y diversos como las clases de usuario que lo utilizan para obtener beneficios empresariales. Por lo tanto, es necesario clasificar los requerimientos del Data Warehouse utilizando técnicas clásicas.

2. OLTP v/s OLAP: Dos Mundos Diferentes

De acuerdo como se entiendan las diferencias entre estos dos tipos de sistemas uno gana un mejor entendimiento de OLAP. Esto es muy importante en especial para diseñadores, ya que ellos necesitan ver estas diferencias para poder llevar a cabo de mejor manera un proyecto de esta naturaleza.

2.1 Qué es OLAP

OLAP es un estándar para "On-Line Analytical Processing". En contraste al más familiar OLTP ("On-line Transaction Processing"), OLAP describe la tecnología asociada al acceso y análisis de datos en línea. Mientras que el procesamiento transaccional depende en general de bases de datos relacionales, OLAP se ha convertido en un sinónimo de bases de datos multidimensionales mediante las cuales se provee de una tecnología para el cálculo y análisis requerido por las aplicaciones analíticas para el "Business Intelligence".

2.2 Qué es una Base de Datos Multidimensional?

Las bases de datos relacionales están formadas por un conjunto de registros. Cada registro contiene la información organizada en campos. Un ejemplo de una base de datos relacional es:

Nombre Cliente	Cliente #	Teléfono	Dirección
Tiendas Walter	10556	350-7219	Ave. La Estancia, Chuao
Artefactos ABC	11204	266-4080	Calle el Recreo, Sabana Grande

En este caso se dice que la tabla relacional "Clientes" está basada en un formato de filas y columnas, donde las columnas representan los campos y las filas los registros.

En esencia esta tabla tiene una sola dimensión. Veamos un ejemplo de una base de datos relacional donde hay mas de una correspondencia entre los campos.. En el siguiente ejemplo se tienen las ventas de cada producto por región. Una compañía tiene tres productos (arandelas, tornillos, tuercas) que se venden en tres territorios (Este, Oeste, Central). A continuación se muestra la tabla relacional:

Producto	Region #	Ventas
Arandelas	Este	50000
Arandelas	Oeste	60000
Arandelas	Central	100000
Tornillos	Este	40000
Tornillos	Oeste	70000
Tornillos	Central	80000
Tuercas	Este	90000
Tuercas	Oeste	120000
Tuercas	Central	30000

Un camino para representar esta tabla en una forma mas óptima es a través de una matriz de dos dimensiones como lo muestra el próximo diagrama:

	Este	Oeste	Central
Arandelas	50000	60000	100000
Tornillos	40000	70000	80000
Tuercas	90000	120000	140000

De esta forma se pueden realizar preguntas como ¿Cuáles fueron las ventas de arandelas en el Este?, ¿Cuáles fueron las ventas de Tornillos en el Oeste?.

En casos simples no es necesario colocar la información en bases de datos multidimensionales, pero si nos hacemos preguntas como: ¿Cuál fue el total de ventas en el Este o en el Oeste? y tenemos un millón de productos la selección a través de un query nos tomaría mucho tiempo en una base de datos relacional mientras que usando la tecnología multidimensional OLAP nos tomaría escasos segundos.

Consolidación: la clave para las consultas rápidas y consistentes

Consolidación es la presumarización lógica de subtotales y totales.

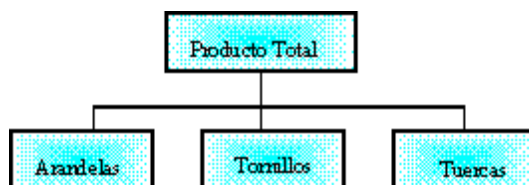
	Este	Oeste	Central	Total
Arandelas	50000	60000	100000	210000
Tornillos	40000	70000	80000	190000
Tuercas	90000	120000	140000	350000
Total	180000	250000	320000	750000

Las celdas que tiene la data original son llamadas *inputs*. Los totales calculados son llamados *ouputs*. Este, Oeste y Central son miembros de la *dimensión* Región. El total región es un *miembro output* de la *dimensión* región. Similarmente Arandelas, Tornillos y Tuercas y total son *miembros* de la *dimensión* Producto. Los valores que en este caso son cajas representan una *variable*. Para esta tabla la *variable* "cajas" está *dimensionada* por Producto y Región.

Jerarquías asociadas a las dimensiones

Las jerarquías representan los diferentes niveles de sumarización de la información

Simple Jerarquía



Múltiple Jerarquías

Drill Down: bajar a un nivel detallado de la información.

Drill Up: subir a un nivel agregado de la información.

Seguridad de las Bases de Datos

La seguridad es importante en el uso de cualquier base de datos que va a ser compartida por múltiples usuarios. La seguridad de la base de datos tiene dos propósitos principales:

- ◆ Guardar los usuarios que no están autorizados a consultar la información.
- ◆ Controlar el acceso a porciones de la base de datos por cada usuario.

El acceso a un subconjunto de la base de datos multidimensional se puede restringir por usuario o por grupo de usuarios.

Como conclusión podemos decir lo siguiente:

- Los servidores OLAP son una tecnología superior para las aplicaciones "Business Intelligence".
- Las aplicaciones OLAP permiten realizar consultas a un nivel agregado de la información. Como totales de ventas por línea de producto, región y vendedor.
- Las bases de datos OLAP son optimizadas para el análisis.
- Consultas rápidas y consistentes a cualquier nivel de sumariaización de la información.
- Las bases de datos OLAP son alimentadas de distintas fuentes de información.

Las diferencias entre ambos procesamientos se establecen en distintos ámbitos; el siguiente es un paralelo entre ambas filosofías:

1.–Orientación o Alineación de Datos.

2.– Integración

3.–Acceso y Manipulación de datos por parte de Usuarios finales.

4.– Administradores

5.– Transacción

6.–La dimensión Tiempo

2.3. Diferencia entre Data Warehouse y bases de datos operacionales (OLTP)

Un Data Warehouse es diferente de las bases de datos operacionales que soportan las aplicaciones de un Procesamiento de Transacción en Línea (OLTP, On-Line Transaction Processing). El Data Warehouse es lo siguiente:

- **Está orientado a una materia.** Organiza y orienta los datos desde la perspectiva del último usuario. 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 (la eficiencia de recuperación y actualización de la aplicación) al momento en que se implementó por primera vez.

- **Administra grandes cantidades de información.** La mayoría de los Data Warehouses contiene información histórica que se retira con frecuencia de los sistemas operativos porque ya no es necesaria para las aplicaciones operacionales 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 clasifican esta inmensa cantidad de datos. En resumen, los actuales usuarios de un Data Warehouse buscan árboles en el bosque. Por lo tanto, un Data Warehouse maneja información a diferentes niveles de granularidad. Por la necesidad de administrar toda la información histórica y además los datos actuales, un Data Warehouse es mucho mayor que las bases de datos operacionales.
- **Guarda información en diversos medios de almacenamiento.** Por los volúmenes de información que deben manejarse, un Data Warehouse frecuentemente guarda información en diferentes medios de almacenamiento.
- **Comprende múltiples versiones de un esquema de base de datos.** Debido a que el Data Warehouse tiene que guardar información histórica y administrarla, y como la 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 guardada por bases de datos operacionales para cualquier toma de decisiones sensata. Un Data Warehouse condensa y agrega la información para presentarla en forma comprensible a las personas. La condensación y adición es esencial para retroceder y entender la imagen global.
- **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, se requiere de Data Warehouse para recopilar y organizar en un solo lugar la información que estas aplicaciones han acumulado al paso de los años. Esta es una tarea desafiante por la diversidad de tecnologías de almacenamiento, de técnicas de administración de bases de datos y de la semántica de los datos.

2.4 Relación Entre Ambos Esquemas

Se definen dos fases en el Data Warehouse Dimensional (DDW): carga y consultas. En la primera se carga la snapshot para un tiempo dado, y en la segunda se pueden hacer consultas en la base de datos sin que haya cambios en ella durante el proceso.

Para visualizar fácilmente la relación entre ambos esquemas (OLTP y DDW), se muestra la siguiente figura:

FIGURA 1: ESQUEMA DE RELACIÓN ENTRE OLTP Y DDW.

2.5 Una máquina o dos?

Es mucho más habitual encontrar separado el DW del OLTP, debido a factores bien específicos y de considerable relevancia para su desempeño. Los argumentos que favorecen el tener en máquinas separadas el OLTP del DW son:

- El DW tiene una significativa y altamente variable demanda de recursos, por lo tanto puede entorpecer considerablemente el desempeño del OLTP.
- Los sistemas en cuestión son configurados muy diferentemente.

- A veces los datos del DW son integrados de múltiples sistemas OLTP remotos, y por lo tanto el DW puede verse como un conjunto de recursos centralizados. Es obvio entonces que estén físicamente separados ambos sistemas.

La razón para tenerlos en la misma máquina está en el hecho de que al ser la estructura básica del DW distinta a la del OLTP, el dato tiene que ser copiado y reestructurado por el DW. Para ahorrar envíos de datos entre máquinas, es mejor realizar este proceso dentro de una sola.

3 ARQUITECTURA DATAWAREHOUSE

3.1 Componentes y Estructuras

Antes de describir la arquitectura Datawarehouse vamos a señalar la siguiente consideración ya generalizada, presente en la literatura: el término Datawarehouse se utiliza indistintamente para hablar de la arquitectura en sí como también para uno de los componentes que la conforman, específicamente el que tiene relación con el almacenamiento físico de los datos. Ahora, con el propósito de facilitar el entendimiento por parte del lector, haremos especial énfasis en esta parte del capítulo sobre el contexto del cual se estará hablando al hacer referencia al término Datawarehouse.

La arquitectura lógica de un sistema de Data Warehousing es del tipo mostrado en la Figura 1. Un Sistema de Data Warehousing consta de tres niveles: (1) bases de datos fuentes (de producción e históricos), (2) una base de datos con datos resumidos extraídos de las bases de producción (el Data Warehouse), y (3) interfaces orientadas a usuarios que extraen información para la toma de decisiones. Las clásicas son: Análisis Multidimensional, consultas y reportes y Data Mining.

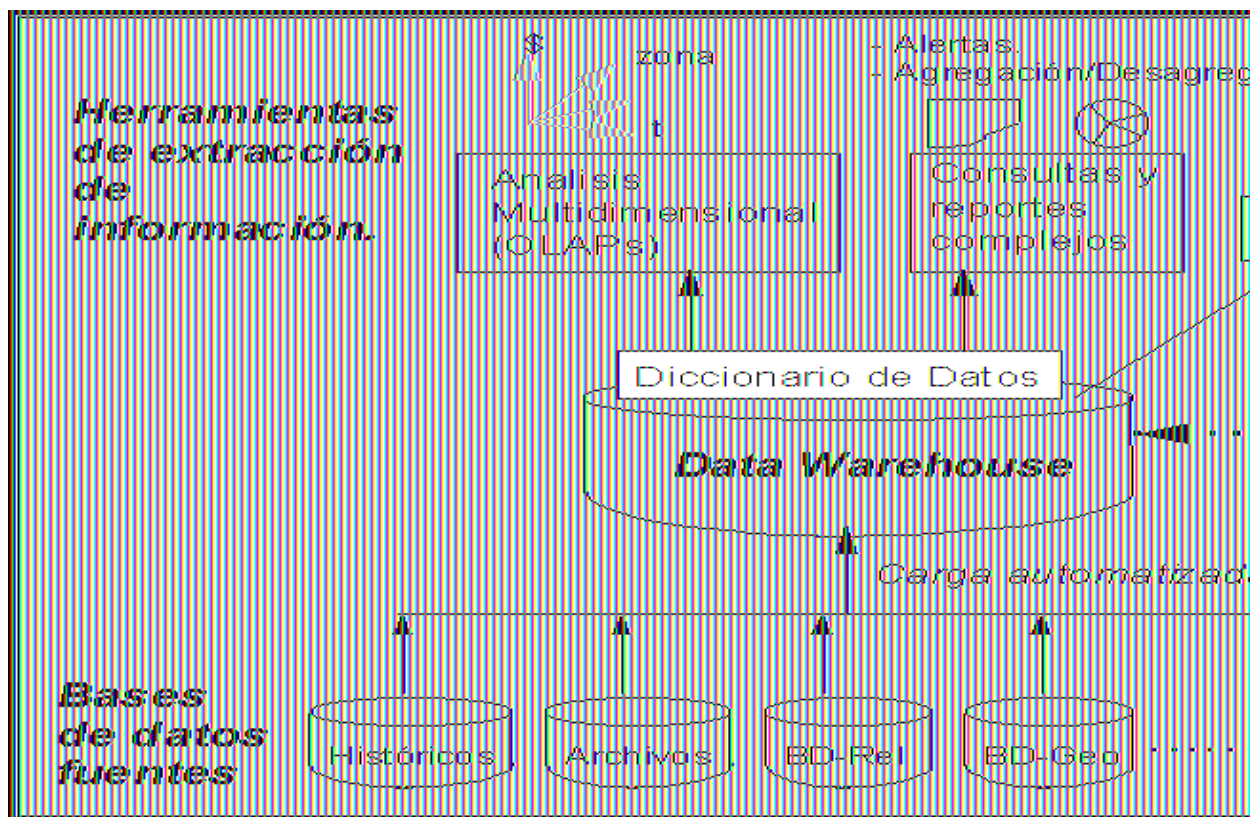


FIGURA 2. ARQUITECTURA LÓGICA DE UN SISTEMA DE DATA WAREHOUSING.

Las bases de datos fuentes consisten en bases de datos de producción así como en históricos de dichas bases.

Estas bases de datos pueden estar implementadas en diferentes tipos de sistemas: BD–Relacionales, BD–geográficas, BD–textos, archivos, etc. Una característica común es que almacenan ítems de datos atómicos, los cuales son relevantes como datos de producción, pero puede ser demasiado finos como base para la toma de decisiones. Además, la noción de calidad de los datos en estas bases se basa en la consistencia de dichos registros, independientemente de la relevancia que estos tengan dentro del problema.

La estructura básica de la arquitectura DW incluye:

1. Datos operacionales: un origen de datos para el componente de almacenamiento físico DW.
2. Extracción de Datos: selección sistemática de datos operacionales usados para poblar el componente de almacenamiento físico DW.
3. Transformación de datos: Procesos para sumarizar y realizar otros cambios en los datos operacionales para reunir los objetivos de orientación a temas e integración principalmente.
4. Carga de Datos: inserción sistemática de datos en el componente de almacenamiento físico DW.
5. Datawarehouse: almacenamiento físico de datos de la arquitectura DW.
6. Herramientas de Acceso al componente de almacenamiento físico DW: herramientas que proveen acceso a los datos.

FIGURA 3: ESTRUCTURA BÁSICA DW.

Los pasos 2, 3 y 4 considerados en la figura anterior, conforman el proceso conocido como ETT (Extracción, Transformación y Transporte).

3.2 Opciones de Implementación

La forma en la cual se estructure el almacenamiento de datos DW, genera una clasificación respecto a la forma de implementar una arquitectura DW. La estructura adoptada para el almacén de datos se debe realizar de la manera que mejor satisfaga las necesidades empresariales, siendo entonces dicha elección factor clave en la efectividad del DW. Las más básicas son:

- EL DW central: es una implementación de un solo nivel con un solo almacén para soportar los requerimientos de información de toda la empresa.
- El DW distribuido: es también una estructura de un nivel, pero particiona el almacén para distribuirlo a nivel departamental.
- El DW de dos niveles: combina ideas de los dos anteriores, siendo sus entregables tanto el almacén empresarial como los departamentales.

3.3 Costos v/s Valor De DW

En todo proyecto es importante e inevitable realizar un análisis desde la perspectiva Costo/Valor. A grandes rasgos, los costos asociados a un proyecto DW incluyen el costo de construcción y, la mantención y operación una vez que está construido. En cuanto al valor, éste considera, el valor de mejorar la entrega de información, el valor de mejorar el proceso de toma de decisiones y el valor agregado para los procesos empresariales.

4 Costos De Un DW

4.1 Costos De Construcción

Los costos de construir un DW son similares para cualquier proyecto de tecnología de información. Estos pueden ser clasificados en tres categorías:

RRHH: la gente necesita contar con un enfoque fuerte sobre el conocimiento del área de la empresa y de los procesos empresariales. Además es muy importante considerar las cualidades de la gente, ya que el desarrollo del DW requiere participación de la gente de negocios como de los especialistas tecnológicos; estos dos grupos de gente deben trabajar juntos, compartiendo su conocimiento y destrezas en un espíritu de equipo de trabajo, para enfrentar los desafíos de desarrollo del DW.

Tiempo: Se debe establecer el tiempo no tan solo para la construcción y entrega de resultados del DW, sino también para la planeación del proyecto y la definición de la arquitectura. La planeación y la arquitectura, establecen un marco de referencia y un conjunto de estándares que son críticos para la eficacia del DW.

Tecnología: Muchas tecnologías nuevas son introducidas por el DW. El costo de la nueva tecnología puede ser tan sólo la inversión inicial del proyecto.

4.2 Costos De Operación

Una vez que está construido y entregado un DW debe ser soportado para que tenga valor empresarial. Son justamente estas actividades de soporte, la fuente de continuos costos operacionales para un DW. Se pueden distinguir tres tipos de costos de operación:

Evolutivos: ajustes continuos del DW a través del tiempo, como cambios de expectativas y, cambios producto del aprendizaje del RRHH del proyecto mediante su experiencia usando el DW.

Crecimiento: Incrementos en el tiempo en volúmenes de datos, del número de usuarios del DW, lo cual conllevará a un incremento de los recursos necesarios como a la demanda de monitoreo, administración y sintonización del DW (evitando así, un incremento en los tiempos de respuesta y de recuperación de datos, principalmente).

Cambios: El DW requiere soportar cambios que ocurren tanto en el origen de datos que éste usa, como en las necesidades de la información que éste soporta.

Los dos primeros tipos de costos de operación, son básicos en la mantención de cualquier sistema de información, por lo cual no nos resultan ajenos; sin embargo, se debe tener especial cuidado con los costos de operación por cambios, ya que ellos consideran el impacto producto de la relación del OLTP y del Ambiente Empresarial, con el DW.

Resulta esencial para llevar a cabo un proyecto DW, tener claridad en la forma que éste se ve afectado por medio de cambios a nivel de OLTP como del Ambiente Empresarial; por ello entonces, a continuación se analiza más en detalle este tipo de costos de operación.

4.2.1 Cambios y el DW.

Cuando se implementa un DW, el impacto de cambios es compuesto. Dos orígenes primarios de cambios existen:

- Cambios en el ambiente empresarial: Un cambio en el ambiente empresarial puede cambiar las necesidades de información de los usuarios. Así, el contenido del DW se puede ver afectado y las aplicaciones DSS y EIS pueden requerir cambios.
- Cambios en la tecnología: Un cambio en la tecnología puede afectar la manera que los datos operacionales son almacenados, lo cual implicaría un ajuste en los procesos de Extracción, Transporte y Carga para adaptar las variaciones presentadas.

Un cambio de cualquiera de ellos impacta los sistemas operacionales. Un cambio en el ambiente operacional puede cambiar el formato, estructura o significado de los datos operacionales usados como origen para el DW. De esta forma serían impactados los procesos de Extracción, Transformación y Carga de datos.

4.3 Valor Del DW

El valor de un DW queda descrito en tres dimensiones:

1. Mejorar la Entrega de Información: información completa, correcta, consistente, oportuna y accesible. Información que la gente necesita, en el tiempo que la necesita y en el formato que la necesita.
2. Mejorar el Proceso de Toma de Decisiones: con un mayor soporte de información se obtienen decisiones más rápidas; así también, la gente de negocios adquiere mayor confianza en sus propias decisiones y las del resto, y logra un mayor entendimiento de los impactos de sus decisiones.
3. Impacto Positivo sobre los Procesos Empresariales: cuando a la gente se le da acceso a una mejor calidad de información, la empresa puede lograr por sí sola:
 - Eliminar los retardos de los procesos empresariales que resultan de información incorrecta, inconsistente y/o no existente.
 - Integrar y optimizar procesos empresariales a través del uso compartido e integrado de las fuentes de información.
 - Eliminar la producción y el procesamiento de datos que no son usados ni necesarios, producto de aplicaciones mal diseñados o ya no utilizados.

4.4 Balance de Costos v/s Valor.

Lograr una cuantificación económica de los factores de valor no es fácil ni natural a diferencia de los factores de costos, agregar valor económico a los factores de valor resulta ser en extremo complejo y subjetivo. Una alternativa a ello, es hacer una valoración desde la perspectiva de costos evitables, relacionados con los costos de no disponer en la organización de información apropiada, tanto a un nivel técnico como de procesos empresariales (en especial, para el proceso de Toma de Decisiones).

DW es una estrategia de largo plazo. Al querer implementar un DW, se debe evaluar el costo y el valor considerando un período de tiempo razonable para obtener beneficios. El retorno sobre la inversión de un DW, se comienza a percibir bastante más tarde del tiempo en el cual se realizó la inversión inicial. Si se calcula costo/valor desde una perspectiva de corto plazo, los costos serán significativamente más altos en proporción al valor.

5 Impactos DW

El éxito de DW no está en su construcción, sino en usarlo para mejorar procesos empresariales, operaciones

y decisiones. Posicionar un DW para que sea usado efectivamente, requiere entender los impactos de implementación en los siguientes ámbitos:

5.1 Impactos Humanos.

Efectos sobre la gente de la empresa:

- **Construcción del DW:** Construir un DW requiere la participación activa de quienes usarán el DW. A diferencia del desarrollo de aplicaciones, donde los requerimientos de la empresa logran ser relativamente bien definidos producto de la estabilidad de las reglas de negocio a través del tiempo, construir un DW depende de la realidad de la empresa como de las condiciones que en ese momento existan, las cuales determinan qué debe contener el DW. La gente de negocios debe participar activamente durante el desarrollo del DW, desde una perspectiva de construcción y creación.
- **Accesando el DW:** El DW intenta proveer los datos que posibilitan a los usuarios acceder su propia información cuando ellos la necesitan. Esta aproximación para entregar información tiene varias implicancias :

- a) La gente de la empresa puede necesitar aprender nuevas destrezas.
- b) Análisis extensos y demoras de programación para obtener información será eliminada. Como la información estará lista para ser accedida, las expectativas probablemente aumentarán.
- c) Nuevas oportunidades pueden existir en la comunidad empresarial para los especialistas de información.
- d) La gran cantidad de reportes en papel serán reducidas o eliminadas.
- e) La madurez del DW dependerá del uso activo y retroalimentación de sus usuarios.

- Usando aplicaciones DSS/EIS: usuarios de aplicaciones DSS y EIS necesitarán menos experiencia para construir su propia información y desarrollar nuevas destrezas.

5.2 Impactos Empresariales.

- **Procesos Empresariales Y Decisiones Empresariales.**

Se deben considerar los beneficios empresariales potenciales de los siguientes impactos:

- a) Los Procesos de Toma de Decisiones pueden ser mejorados mediante la disponibilidad de información. Decisiones empresariales se hacen más rápidas por gente más informada.
- b) Los procesos empresariales pueden ser optimizados. El tiempo perdido esperando por información que finalmente es incorrecta o no encontrada, es eliminada.
- c) Conexiones y dependencias entre procesos empresariales se vuelven más claros y entendibles. Secuencias de procesos empresariales pueden ser optimizados para ganar eficiencia y reducir costos.
- d) Procesos y datos de los sistemas operacionales, así como los datos en el DW, son usados y examinados. Cuando los datos son organizados y estructurados para tener significado empresarial, la gente aprende mucho de los sistemas de información. Pueden quedar expuestos posibles defectos en aplicaciones actuales, siendo posible entonces mejorar la calidad de nuevas aplicaciones.

- **Comunicación e Impactos Organizacionales.**

Apenas el DW comienza a ser fuente primaria de información empresarial consistente, los siguientes impactos pueden comenzar a presentarse:

- a) La gente tiene mayor confianza en las decisiones empresariales que se toman. Ambos, quienes toman las decisiones como los afectados conocen que está basada en buena información.
- b) Las organizaciones empresariales y la gente de la cual ella se compone queda determinada por el acceso a la información. De esta manera, la gente queda mejor habilitada para entender su propio rol y responsabilidades como también los efectos de sus contribuciones; a la vez, desarrollan un mejor entendimiento y apreciación con las contribuciones de otros.

- c) La información compartida conduce a un lenguaje común, conocimiento común, y mejoramiento de la comunicación en la empresa. Se mejora la confianza y cooperación entre distintos sectores de la empresa , viéndose reducida la sectorización de funciones.
- d) Visibilidad, accesibilidad, y conocimiento de los datos producen mayor confianza en los sistemas operacionales.

5.3 Impactos Técnicos De DW.

Considerando las etapas de construcción, soporte del DW y soporte de sistemas operacionales, se tienen los siguientes impactos técnicos:

- Nuevas destrezas de desarrollo: cuando se construye el DW, el impacto más grande sobre la gente técnica está dada por la curva de aprendizaje, muchas destrezas nuevas se deben aprender, incluyendo:

a)Conceptos y estructura DW.

b) El DW introduce muchas tecnologías nuevas (ETT, Carga, Acceso de Datos, Catálogo de Metadatos, Implementación de DSS/EIS), y cambia la manera que nosotros usamos la tecnología existente. Nuevas responsabilidades de soporte, nuevas demandas de recursos y nuevas expectativas, son los efectos de estos cambios.

c) Destrezas de diseño y análisis donde los requerimientos empresariales no son posibles de definir de una forma estable a través del tiempo.

d)Técnicas de desarrollo incremental y evolutivo.

e) Trabajo en equipo cooperativo con gente de negocios como participantes activos en el desarrollo del proyecto.

- Nuevas responsabilidades de operación: Cambios sobre los sistemas y datos operacionales deben ser examinados más cuidadosamente para determinar el impacto que estos cambios tienen sobre ellos, y sobre el DW.

5.4 La Importancia de la Performance

Los avances en procesamiento paralelo y data mining, de los Data Warehouses, se pueden aplicar también a los Data Marts. Varios vendedores ofrecen maneras específicas de mejorar la performance de los Data Marts. Hay que poner en la balanza conceptos como: performance para el usuario final, datos precalculados vs. sumariación a demanda, performance en la carga de los datos y el tamaño del Data Mart.

Red Brike Warehouse 5.0, de Red Brike Systems Inc., posee índices que continuamente se adaptan a los datos que están siendo procesados. Un nuevo tipo de Join: Hash híbrido, maneja más eficientemente algunas situaciones como recursión en subjoins. Consultas SQL pueden ser embebidas en la cláusula FROM de otra consulta, para evitar múltiples sentencias SQL.

Bases de Datos Multidimensionales (MDDB) como Essbase, de Arbol Software Corp.suportan actualización incremental, de manera que la estructura entera no necesita ser cambiada para cada actualización.

Pilot Decision Support Suite, de Pilot Software Inc., provee dimensiones dinámicas, que permiten que las agregaciones sean calculadas a demanda, en lugar de preagregadas y almacenadas en el cubo. Esto puede reducir significativamente el tamaño del cubo, al igual que el tiempo de consolidación requerido al generar el

cubo.

6 META DATOS

El conocimiento de los meta datos es tan esencial como el conocimiento de los datos del Data Warehouse.

Deben incluir dominio, reglas de validación, derivación y transformación de los datos extraídos. También describen las bases de datos del Warehouse, incluyendo reglas de distribución y control de la migración hacia los Data Marts. Los procesos que monitorean los procesos del Warehouse (como extracción, carga, y uso) crean meta datos que son usados para determinar que tan bien se comporta el sistema.

Los meta datos, deberían estar disponibles para los usuarios, para ser usados en sus análisis. Los administradores pueden manejar y proveer el acceso a través de los servicios del repositorio.

Las cinco funcionalidades del Warehouse proveen un marco de trabajo para controlar la arquitectura de los componentes. Este marco, describe las transformaciones de los datos desde un ambiente OLTP, a un ambiente OLAP.

6.1 Introducción

El uso efectivo de los Data Marts en un ambiente de Data Warehousing, es un factor importante para la efectividad del Warehouse, y puede también ser determinante en el éxito del proyecto de desarrollo. Los Data Marts son diseñados para satisfacer las necesidades específicas de grupos comunes de usuarios (divisiones geográficas, divisiones organizacionales, etc.). Los Data Marts son generalmente, subconjuntos del Data Warehouse, pero pueden también integrar un número de fuentes heterogéneas, e inclusive ser más grandes, en volumen de datos, que el propio Warehouse central. Como los Data Marts son un factor crítico para el éxito proyecto de Data Warehousing de mayor escala, también lo son su creación y mantenimiento.

Actualmente, las organizaciones se están convenciendo de que los Data Warehouse corporativos, son complejos tanto para construir como para usar. Implementar un Data Warehouse, requiere de un considerable equipo de desarrolladores, hardware, software, tiempo y dinero. Las necesidades de diferentes áreas de la empresa, a veces conflictivas, deben ser sobrellevadas en su conjunto. Los usuarios los encuentran difíciles de construir, y por lo tanto de navegar. En consecuencia, las empresas están construyendo Data Marts, en lugar de, o complementando a los Data Warehouses.

6.2 Crecimiento de los Data Marts

Los Data Marts están creciendo, llegando a tener tamaños semejantes a los Data Warehouse corporativos de menor escala.

Aunque hoy en día es difícil diferenciar a los Data Marts y Data Warehouses por su tamaño, algunas distinciones entre ellos son todavía importantes:

- ***Un Data Mart está enfocado a una sola área o grupo de usuarios***, mientras que un Data Warehouse contiene información de diferentes sujetos y áreas de la corporación.
- ***Una organización puede tener un sólo Data Warehouse, pero varios Data Marts.***
- ***Los Data Marts no contienen información almacenada como datos operacionales***, pero si la tienen los Data Warehouses.
- ***Como los Data Marts contienen menos información, son más fáciles de entender y navegar***, que los

Data Warehouses corporativos. Un Data Warehouse puede contener tanta información, que es difícil de manejar por los usuarios.

El reciente crecimiento de los Data Marts, ha generado también, muchos problemas a los usuarios, para acceder a la información de la organización.

- ***Se pierde performance a medida que aumenta el tamaño de los Data Marts.*** Los usuarios esperan mejor respuesta de los Data Marts, que de los Data Warehouses.
- ***Los usuarios requieren acceso a datos de muchos Data Marts.*** Los datos pueden ser replicados entre los Data Marts, pero se requieren mejores soluciones.
- ***Las compañías no pueden administrar fácilmente muchos Data Marts.*** Mientras sólo se tiene un Data Warehouse, se pueden tener muchísimos Data Marts.
- ***Las organizaciones tienen dificultades para construir los Data Marts.*** Aunque es aceptable que la construcción de un Data Warehouse lleve varios años, los Data Marts requieren un ciclo de desarrollo muy corto, para una inversión moderada.

Las soluciones de Data Marts, requieren una arquitectura de 3 capas: Los Data Warehouses son la primera capa (opcional), los Data Marts son la segunda capa, y las estaciones de trabajo de usuarios son la tercera.

6.3 Data Marts Virtuales y Meta Vistas

Los vendedores están desarrollando el concepto de Data Marts Virtuales para satisfacer la necesidad de los usuarios de acceder a muchos Data Marts, sin necesidad de excesivas replicaciones entre ellos. Los Data Marts Virtuales son vistas de varios Data Marts Físicos, o del Data Warehouse corporativo, brindadas a grupos específicos de usuarios.

Otros vendedores, como Sagent Data Mart Solution, de Sagent Technology Inc., proveen los conceptos de Vista Básica y Meta Vistas. Una Vista Básica es una representación gráfica de una base de datos que incluye tablas, columnas y joins. Una vez que una Vista Básica es creada, múltiples Meta Vistas se pueden derivar de ella. Una Meta Vista es una representación lógica de partes, de una o más Vistas Básicas. Inicialmente las tablas son desplegadas como categorías, y los campos como partes. Se pueden renombrar o remover categorías o partes de una Meta Vista. Esos cambios no afectan a las Vistas Básicas que la soportan. La Meta Vistas permiten usar una única Vista Básica para presentar diferentes partes de la información a diferentes grupos de usuarios.

La nueva tecnología de los Data Marts, está todavía en desarrollo.

6.4 Administración de los Data Marts

A medida que el número de Data Marts va creciendo, crece también la necesidad de administración y coordinación central, de actividades como manejar versiones, asegurar la consistencia e integridad de los datos, controlar la seguridad, y mantener la performance global. Sin la administración central, los datos se vuelven inconsistentes entre los diferentes departamentos, los usuarios no pueden acceder a la información de varios Data Marts a la vez, y eventualmente, los Data Marts, se vuelven tan desparejos que no pueden ser integrados en un Data Warehouse. La coordinación y administración de toda la colección de Data Marts, debe tener un enfoque centralizado, en lugar de distribuir las actividades de administración entre los diferentes usuarios.

La administración de los Data Marts, es un área con crecientes requerimientos, como la coordinación, la

extracción de los datos, la lectura, los procedimientos de replicación, los procedimientos de backup y recuperación, el manejo de metadatos, la seguridad, y la performance.

6.5 Paquetes de Data Marts

Muchos vendedores han reconocido la necesidad de hacer que los Data Marts sean más fáciles de instalar e implementar que un Data Warehouse corporativo. Los paquetes de Data Marts pueden proveer herramientas convenientes, y de relativamente bajo costo, que pueden ser el puntapié inicial para el desarrollo de los Data Marts. Aunque un Data Mart es relativamente fácil de instalar, hay que tener en cuenta otros aspectos como la lógica de los datos operacionales extraídos, la consistencia en la definición de los datos, y el diseño del Data Mart, para lograr una óptima performance.

6.6 Las promesas de Internet e Intranet

La tecnología de Internet e Intranet prometen dar a sus usuarios un acceso barato a los datos de los Data Warehouses y Data Marts, através de los Web Browsers.

Los productos, generalmente consisten en código situado entre los servidores Web y los productos OLAP. Los vendedores líderes, están empezando a incluir Java y/o ActiveX en sus productos, en comparación a las limitadas funcionalidades HTML de los productos iniciales.

Aunque los browsers no contienen toda la capacidad de acceso disponible en las aplicaciones Cliente-Servidor, Internet es una buena solución para dar acceso a los datos, cuando los usuarios son muchos o están geográficamente dispersos, especialmente, los usuarios que no necesitan una interacción muy sofisticada.

6.7 Que se debe hacer?

Hay varias maneras de dar a los usuarios un acceso a datos de soporte decisional.

- Una manera es construir un Data Warehouse corporativo, que puede ser usado directamente por los usuarios, o puede alimentar Data Marts.
- Otra manera es construir Data Marts planeados para eventualmente integrarlos en un Data Warehouse.
- Una técnica muy popular, es construir la infraestructura para un Data Warehouse, mientras al mismo tiempo, se construyen uno o más Data Marts para satisfacer las necesidades más inmediatas.

Los beneficios y costos dedicar tiempo y recursos a la construcción de un Data Warehouse corporativo, deben ser comparados contra los beneficios y costos de tener un Data Mart, corriendo ya, para satisfacer las oportunidades específicas de los negocios.

El enfoque de un Data Warehouse, tiene muchas ventajas: los requerimientos de todas las funciones de empresa pueden ser incluidos; las definiciones de los datos, y reglas de empresa son consistentes, y la redundancia de los datos, es eliminada. Las ventajas de una estrategia de Data Marts son: menor costo y esfuerzo en una implementación inicial, menor tiempo de desarrollo, la experiencia de los usuarios mejora la performance, y las funciones de cada área pueden controlar su propio Data Mart.

Aunque los Data Marts pueden proveer el éxito en solucionar muchos problemas de negocios, la proliferación de Data Marts no planeados, a través de la corporación puede llevar a inconsistencias en los datos, duplicación de éstos, y a que los usuarios no puedan acceder a todos los datos necesarios.

Los vendedores están llevando a cabo algunos de estos desafíos:

- Respuesta rápida, a medida que los Data Marts crecen en tamaño.
- Administración de los Data Marts de toda la organización, para asegurar consistencia en la definición de los datos, seguridad, y efectiva replicación de los datos.
- Implementación rápida y repetitiva, incluyendo el acceso a Internet, para proveer la capacidad de tomar decisiones de una manera más rápida, efectiva y menos costosa.

7 Porque construir un Data Warehouse?

La complejidad de los negocios actuales ha modificado la forma de administrar de las empresas. Los gerentes empresariales no sólo necesitan saber que está sucediendo en el negocio, sino además por qué (ver figura 14.1). En la fase de aplicar tecnología de la información para automatizar el procesamiento de datos, las empresas desarrollaron aplicaciones para medir con rapidez el factor ¿qué está sucediendo?. Ahora, en la fase de procesamiento de la información, las empresas requieren conocer el factor ¿por qué está sucediendo?; el ambiente competitivo y el ritmo de cambio lo demandan así. Las empresas desean pasar con rapidez a la siguiente fase. ¿qué debemos hacer y cuáles son los riesgos?

Necesidades

empresariales

Complejidad

FIGURA 4: Necesidades empresariales y administración de riesgos

Para crear recomendaciones factibles, los gerentes empresariales necesitan analizar el desempeño de sus negocios y el estado del mercado. El análisis siempre está libre de riesgos, ya que no se toma ninguna acción ni se establece un compromiso. La evaluación de riesgos es un ingrediente esencial de cualquier recomendación para equilibrar los costos potenciales no previstos con los beneficios. En el medio actual de ritmo acelerado, además de cada recomendación factible, la administración necesita también una orientación en el manejo de riesgos para la acción tomada. Las consecuencias de una acción precipitada pueden ser suicidas si la reacción y contrarreacción nos son oportunas. La tecnología del Data Warehouse tiene una función importante. Los Data Warehouses producen información factible, no sólo datos, consultas ad hoc y reportes muy resumidos.

7.1 El reto de la administración

En el ambiente empresarial de los años noventa, la administración ejecutiva debe examinar la naturaleza de su negocio debido a que fuerzas externas intensifican la competencia y presionan permanentemente los precios. La empresa debe esforzarse por continuar, ya sea sobresaliendo en el mercado, o bien sustentando un negocio de mercancías. La tecnología del Data Warehouse puede auxiliar en ambas tareas.

Los datos históricos de la empresa son una mina de oro de información sobre sus clientes, sus relaciones cliente/producto, los patrones de compra de los clientes, etcétera. Debido a que los Data Warehouses tienen el potencial de convertir esta información en conocimientos, las empresas con información profunda de sus clientes tienen el potencial de convertirlos en beneficios. Jonathan Berry declara en Business Week que: Los comercializadores están cada vez mas convencidos de que el comportamiento previo del cliente registrado en transacciones reales, es por mucho el mejor indicador de los patrones de compra futuros. Este conocimiento de sus clientes, extraído de las transacciones históricas de los mismos, es único para la empresa y, por lo tanto, se puede emplear para crear elementos que marquen una diferencia, tales como productos personalizados para cada segmento de micromercado, mayoreo de productos o servicios y paquetes especiales para clientes específicos. Analizar las relaciones con el cliente y su valor a través del tiempo (el ciclo del valor del cliente) ayuda a la empresa a evaluar el desempeño general de la comercialización.

Un negocio de mercancías se caracteriza por su pasión para reducir o eliminar costos del sistema empresarial. El análisis de la cadena de valor y la integración logística son dos formas en las que el Data Warehouse elimina costos de la empresa.

Por lo regular, un Data Warehouse con una buena arquitectura puede elevar las ventas en forma notable, incrementar la rotación de inventarios y ser un factor positivo en la productividad del personal, lo que da por resultado márgenes de operación muy mejorados.

7.2 Estado de los sistemas actuales

Es posible que la razón táctica más importante de construir un Data Warehouse sea lo inadecuado de los sistemas actuales y la falta de información empresarial, incluso cuando la empresa esta inundada en datos. Muchos sistemas de producción no satisfacen las necesidades del usuario empresarial. Por lo regular, los datos son inaccesibles e inconsistentes, tanto en forma como en significado. Por ejemplo, debido a la inconsistencia de datos, no coincide la información de ventas en diferentes reportes, la empresa carece de una imagen precisa de su ingreso. La falta de medidas comunes significa que los administradores no tienen una imagen clara del desempeño del negocio.

Los gerentes de comercialización y ventas requieren tener un acceso mas rápido a los datos, mas reportes y a mayor velocidad, análisis expeditos y reacciones oportunas para administrar el negocio y aumentar los ingresos. Incluso con costos sustanciales en tecnología de la información para crear y generar reportes, éstos son tardíos y, por lo tanto, la información pierde su novedad.

Distintos sistemas de producción mantienen información diferente sobre el mismo cliente en diversas bases de datos. Esta falta de una visión única y completa de cada cliente significa que, en el mejor de los casos, las oportunidades de ventas cruzadas, comercialización dirigida, mayoreo de productos, etc., representan un reto. No es posible atender a los clientes que desean hacer compras en un solo paso y no ventas con diferentes personas dentro de la misma compañía. Así, en vez de un crecimiento empresarial, existe una pérdida.

8 La Contribución de Microsoft a la Industria de Data Warehousing

Data warehousing se ha convertido en uno de los segmentos más dinámicos, interesantes y de más rápido crecimiento en la industria de la tecnología de la información. Data warehousing se refiere a los procesos y la tecnología involucrados en el almacenamiento de grandes volúmenes de datos con el fin de utilizarlos para análisis multidimensional. Microsoft Corporation se encuentra contribuyendo al rápido crecimiento de data warehousing para construir sistemas de soporte a las decisiones. Una combinación entre productos de Microsoft y los de una alianza de proveedores de servicios y de software independientes, les permite a los clientes operar sistemas de data warehouses que sean potentes y tengan un precio accesible. La base de datos del Microsoft® SQL Server!, implementada como un motor de acumulación de información para data warehouses, ofrece el precio/rendimiento, facilidad de administración, y la integración entre Windows NT y BackOffice que convierte al SQL Server en la plataforma de soluciones que se emplea preferentemente en muchos sistemas de data warehouse y de data mart.

Durante los últimos veinte años, Microsoft Corporation ha contribuido a que cada año la tecnología de la información esté al alcance de un mayor número de personas al reducir el costo y los retos asociados con la implementación de dicha tecnología. Microsoft se encuentra realizando en la actualidad una contribución similar en el campo de data warehousing y, por lo tanto, apoya de forma directa el rápido desarrollo de ese segmento de la industria de la tecnología de la información.

La base del concepto de data warehousing de Microsoft es el sistema de administración de bases de datos relacionales, Microsoft SQL Server (RDBMS), en el sistema operativo Windows NT. Microsoft SQL Server es un sistema de administración de bases de datos escalable y de alto rendimiento, diseñado específicamente

para la computación distribuida de cliente/servidor en Windows NT. Hay varias razones por las cuales los clientes han escogido a SQL Server para la implementación de sistemas de soporte a la toma de decisiones en el data warehouse. Entre éstas se encuentran:

- Liderazgo en precio/rendimiento y en costo total de la propiedad.
- Compatibilidad con muchos otros componentes y herramientas de software utilizadas

en la instalación de los data warehouses.

- Integración con Internet.
- Opciones de conectividad.
- Capacidades para la administración del sistema.
- La plataforma de mejor rendimiento en el sistema operativo Windows NT.
- Replicación de datos incorporada.
- Habilidad para incorporar fácilmente los datos obtenidos en las aplicaciones de oficina

del escritorio, entre las que se encuentran Microsoft Word, Microsoft Excel, Microsoft PowerPoint®, y el correo electrónico.

- La continua inversión de Microsoft en nuevas capacidades y productos de software.

Microsoft ofrece la plataforma de preferencia en muchas instalaciones de sistemas de data warehouse, y se encuentra invirtiendo en capacidades aún más novedosas en este importante mercado, como respuesta a las sugerencias de los clientes.

Toma de decisiones más eficaz gracias a un mejor acceso a los datos y análisis de los mismos.

Data warehousing es un proceso de soporte a la toma de decisiones que involucra a varios servicios y componentes de la tecnología de información.

Un Data Warehouse reúne información operacional desde una o múltiples fuentes, la transforma en objetos con significado y grupos de información, y entonces la guarda junto con información de tiempo e historia para permitir un soporte efectivo a la toma de decisiones. Éste ofrece incluso múltiples vistas de la información a una variedad de usuarios diferentes. El poder de este concepto radica en que les proporciona a los usuarios acceso a las tendencias en análisis y respuestas a las preguntas comerciales que antes se extraían de fuentes de datos que previamente no estaban relacionadas entre sí. Mientras que los sistemas de producción OLTP (Procesamiento de Transacciones En Línea) se configuran y optimizan para ofrecer tiempos de respuesta rápidos a las respuestas individuales, los sistemas de data warehouses almacenan los datos en forma estática, y se configuran y optimizan para dar soporte a decisiones complejas.

Las discusiones de la industria en general indican que, desde enero de 1997, aproximadamente un 70–80% de todos los data warehouses actualmente en producción son, de hecho, data marts.

El Microsoft SQL Server en Windows NT es una plataforma especialmente popular en la instalación de data marts. El precio/rendimiento y las capacidades de la plataforma de Microsoft se adecúan perfectamente a las prioridades de selección de los clientes que se encuentran implementando data marts.

Intelligent Solutions, Inc., compara los data warehouses con los data marts de la siguiente manera:

Data Warehouse	Data Mart
----------------	-----------

Construido para satisfacer las necesidades de información de toda la empresa.	Construido para satisfacer las necesidades de una función o unidad comercial específica.
Diseñado para optimizar la integración y la administración de los datos fuente.	Diseñado para optimizar la entrega de información de soporte a decisiones.
Administra grandes cantidades de historia a nivel atómico.	Primordialmente se concentra en administrar resúmenes y/o datos de muestreo.
Pertenece a, y se administra por, las organizaciones de Sistemas de Información (IS) de la empresa.	Puede ser propiedad de, y administrado por, el grupo de Sistema de Información (IS) en la Línea del Negocio.

8.1 Por qué ha crecido la popularidad de los Data Marts

Existe un número de sólidas razones detrás del aumento en popularidad de los data marts, en comparación con los sistemas de data warehouses a nivel de empresa.

Los data marts han reducido drásticamente el costo implícito en la creación y operación de un sistema de soporte a las decisiones. El concepto del data mart ha logrado situar la instalación de la tecnología de soporte a las decisiones dentro del rango de posibilidades económicas de un número mucho mayor de usuarios. Mientras que los presupuestos de instalación de data warehouses típicamente oscilan entre los \$2–5 millones de dólares, los data marts típicamente cuestan entre \$100.000 y 1 millón de dólares al presupuesto total del proyecto. El RDBMS (sistema de administración de base de datos relacionales) Microsoft SQL Server en Windows NT Server contribuye a reducir el costo de adquisición, instalación y la continua operación de los data marts. Entre los impulsores de los data marts se encuentra la instalación de servidores de bajo costo –con el sistema operativo Windows NT instalado de fábrica– que son poderosos, de bajo costo, y que ofrecen la suficiente memoria RAM y capacidad de almacenamiento de disco para satisfacer las necesidades de los data marts.

Wayne W. Eckerson, del Patricia Seybold Group, Inc., indicó en su informe, "Understanding Data Marts," ("Entendiendo a los Data marts"), Agosto de 1996, que los data marts son los preferidos por los departamentos autónomos y las pequeñas unidades comerciales que los emplean para crear sus propios sistemas de soporte a decisiones. Pero los data marts también se han convertido en los favoritos de la mayoría de los departamentos de Sistemas de Información (IS), para crear grandes almacenes centrales de datos. La idea consiste en crear un Data Warehouse paso a paso, añadiendo un data mart o área de estudio a la vez, adquiriendo gradualmente la experiencia y el soporte de administradores comerciales clave quienes ven beneficios concretos cada 3–6 meses.

Con los data marts, resulta mucho más fácil identificar un cliente o patrocinador comprometido dentro de una organización. En comparación con los data warehouses, los data marts son más limitados en cuanto a alcance, y se concentran más en un grupo específico de necesidades del usuario. La clave aquí radica en concentrarse en un reto y enfrentarlo con un grupo específicamente dedicado a esa tarea.

Los data marts permiten una prototificación más rápida para la captura de los requisitos del sistema de soporte a decisiones. Las encuestas realizadas entre los consumidores indican que los pilotos de los data marts se montan en 30–120 días. La completa instalación del sistema se logra en un período que oscila de 3 a 6 meses.

Resulta de por sí atractivo el poder comenzar un sistema de soporte a decisiones con un plan modesto, y crecer a partir de él, si fuera necesario, después de adquirir más conocimiento e información sobre los datos fuente y las necesidades de los usuarios finales. Esto les permite a las organizaciones basar su justificación para futuros data marts a medida que crecen.

Proyectos que comenzaron originalmente como data warehouses evolucionan hasta convertirse en data marts. Algunas organizaciones han acumulado una gran cantidad de datos históricos como soporte a decisiones que, al final, nunca se usó, o sólo raramente. Este descubrimiento ha permitido reducir el almacenamiento, o archivar información basándose en algunos criterios definidos (como la fecha, por ejemplo), y por lo tanto reducir sus data warehouses a data marts más concentrados. O la organización puede dividir su data warehouse en data marts separados para ofrecer:

- data marts más pequeños con tiempo de respuesta más rápido.
- acceso menos complejo para los usuarios a los data marts.
- data marts diseñados para grupos de usuarios específicos.

El resultado es una arquitectura de Data Warehouse que se visualiza de la siguiente manera:

FIGURA 5: Arquitectura

Los nuevos data warehouses a menudo se implementan con grandes cantidades de datos en detalle. Después de usarlo por un tiempo, es común que las solicitudes ad-hoc evolucionen hasta convertirse en informes de rutina. Estos informes pueden usar a menudo tablas de datos preresumidos. O el administrador del data warehouse puede observar patrones de acceso a consultas. Las respuestas a las consultas se pueden precomputarizar y almacenar en data marts de menor costo, quizás instalados en grupos de trabajo para simplificar el acceso a la red y mejorar todavía más los tiempos de respuesta. Esto mejora el rendimiento y reduce la complejidad y el tamaño de los data warehouses/data marts. A menudo, los data warehouses se convierten en data marts en este escenario.

Todas estas consideraciones hacen que, de por sí, las organizaciones encuentren atractiva la implementación de sus sistemas de soporte a decisiones en proyectos a escala de un data mart, más manejable. El META Group indicó, en su informe sobre la conferencia Data Warehouse World (El Mundo del Data Warehouse) del META Group/DCI 1997, efectuada en febrero de 1997 que: "La tecnología de la información ha cambiado de una etapa de justificación de costo de data warehouses empresariales a la entrega de aplicaciones de data marts internos".

8.2 Microsoft SQL Server en Windows NT para Data marts

En un informe con fecha de mayo de 1996, la encuesta a gran escala del Data Warehouse Institute – que incluyó a más de 1.500 organizaciones con proyectos de data warehouses en desarrollo – mostró que el sistema operativo Microsoft Windows NT se usó de alguna manera en un 46.8% de esos sitios. El Data Warehouse Institute identificó tres factores determinantes en el crecimiento de Windows NT en el terreno de data warehousing:

Primero, los planificadores de data warehousing han cambiado el énfasis en proyectos de almacenes empresariales por una reunión de sistemas departamentales con arquitectura propia, o data marts. Segundo, la reputación de "sistema pequeño" del Windows NT puede estar disminuyendo. (Windows) NT ya se soporta en procesadores Alpha de 64 bits de Digital, y los próximos sistemas basados en los nuevos procesadores P6 y P7 de Intel, combinados en sistemas de multiprocesador, prometen ofrecer una potencia enorme. La mayoría de estos sistemas de Intel usarán Windows NT Tercero, y lo más importante, el canal de distribución para data warehousing está cambiando. El data warehousing fue una vez el preservador exclusivo de expertos de bases de datos dentro de grandes organizaciones de IT; en la actualidad compañías de menor volumen y departamentos-usuarios finales se sienten cómodos a la hora de comprar tecnología para el data warehousing. Los canales de distribución que utilizan, la mayoría de las veces, también distribuyen sistemas basados en Intel, en los que Windows NT se considera un sistema operativo de "top end" que ofrece un alto nivel de confiabilidad, seguridad y facilidad de administración.

La mayoría de los proyectos de data warehousing en producción se encuentra bien situado dentro del rango de los sistemas de producción soportados por SQL Server y Windows NT. A principios de 1997, existen sistemas de producción SQL Server con bases de datos de 200 GB. Esto resulta suficiente para acomodar a una gran mayoría de los actuales proyectos de data warehouses, como se muestra en un estudio publicado por Forrester que mostró la siguiente distribución de volúmenes de los Data Warehouses de producción:

FIGURA 6

Microsoft continúa aumentando la escalabilidad de SQL Server para acomodar el crecimiento de los sistemas de data warehouses y data marts. Incrementos en la mejoras de la funcionalidad de hardware y software aumentan el número de usuarios concurrentes y el tamaño de la base de datos a la que se puede soportar. Además, la capacidad para agrupación de sistema ("clustering"), que Microsoft incluye con las versiones "Enterprise" de Windows NT y SQL Server, aumenta la capacidad del sistema. Un cluster o agrupación es un grupo de sistemas independientes que funcionan juntos en un mismo sistema. Un cliente interactúa con una agrupación o cluster como si ésta se tratara de un solo servidor. La agrupación también ofrece una sola entidad de administración. Las configuraciones en agrupación se utilizan para brindar una escalabilidad más amplia, mayor disponibilidad y una administración más fácil.

8.3 Mejor Base de Datos para Windows NT

SQL Server es la base de datos más rápida para Windows NT, lo que ya ha sido demostrado en las pruebas comparativas de TPC, cuyos resultados puede consultarse en www.microsoft.com/sql/tpc/ SQL Server forma parte de la familia de productos Microsoft BackOffice®. Esto garantiza que usted dispondrá de un RDBMS (sistema de administración de bases de datos relacionales) diseñado y probado con otras aplicaciones de Windows NT "mejores en su clase". Además del RDBMS SQL Server, entre dichas aplicaciones se incluyen: el Windows NT Server, el Internet Information Server, el SNA Server, el Systems Management Server, y el Mail Server. Microsoft BackOffice ofrece una sola arquitectura unificada en la que se comparten interfaces y herramientas de programación, modelos de administración, seguridad y redes, así como amplias opciones de hardware.

La tecnología de replicación heterogénea (heterogeneous replication) en SQL Server permite la publicación automática de los datos en otros sistemas que no sean SQL Server, entre los que se incluyen Oracle, DB2, Sybase, y Microsoft Access.

8.4 Liderazgo en Precio/Rendimiento

El liderazgo en precio/rendimiento que ofrece Microsoft con el RDBMS SQL Server en la plataforma Windows NT Server de Microsoft, ha contribuido a situar el costo de instalación de sistemas de data warehouse para soporte a decisiones dentro de un rango económicamente accesible para muchos nuevos clientes. Entre los factores que contribuyen a reducir el costo de propiedad se incluyen:

- procesos en hardware de bajo costo
- instalación rápida
- administración más fácil del sistema

Mientras que el rendimiento de Microsoft SQL Server se ha incrementado en casi un 275% en el pasado año, su costo por transacción se ha reducido en un 73%.

Si usted reúne el sistema completo del Microsoft SQL Server incluyendo hardware, soporte y mantenimiento sus costos serán un 26% menos que los del año pasado, y su sistema, sin embargo, tendrá un rendimiento dos y media veces más rápido.

8.5 Auge de Mercado y de Ventas

A medida que más bases de datos se instalen en el Microsoft SQL Server y el Windows NT, habrá más soluciones y experiencia disponible en dicha plataforma.

- Las ventas del Microsoft SQL Server aumentaron en más de un 100% en el año fiscal de 1996, y la tendencia continúa en 1997, lo que hace que la base instalada tenga ya más de 2 millones de usuarios.
- Según International Data Corp., el Microsoft SQL Server es el líder en cuanto a participación de mercado de bases de datos que corren en el Windows NT Server en participación por unidad.
- En agosto de 1996, cincuenta y siete por ciento de los participantes en una encuesta dirigida por Market Perspectives Inc., el META Group, y DCI identificaron al Microsoft Windows NT como el entorno para el data warehousing de sus organizaciones respectivas.
- El Microsoft SQL Server es la base de datos que más se vende, de acuerdo con los VARs recientemente encuestados en el número de junio de 1996 de la revista VAR Business Magazine. El Microsoft SQL Server se anotó un 38%, Oracle, 21%, IBM, 10%, Informix, 9%, y Sybase un 8%.
- Los desarrolladores han mostrado un gran interés, lo que se ha traducido en la venta de más de 25.000 copias de SQL Server 6.5 Training Kit (estuche de entrenamiento de SQL Server 6.5), durante los primeros dos meses de su aparición en el mercado. El estuche de entrenamiento consiste en un curso de Microsoft Press que ofrece una experiencia de primera mano y un aprendizaje que el usuario puede efectuar a su propio paso y conveniencia.
- Más de 3.200 Microsoft Solution Providers ofrecen productos y servicios para el Microsoft SQL Server un aumento de casi un 300% en los últimos dos años.
- El número de Profesionales Certificados por Microsoft (Microsoft Certified Professionals) que soportan y desarrollan aplicaciones para el Microsoft SQL Server se ha más que duplicado en el pasado año para incluir a más de 10.500 individuos.

8.6 Conectividad Superior por Dentro y Fuera

Microsoft SQL Server permite el acceso a datos internos a través de múltiples plataformas, entrega la información a sitios remotos de forma confiable, y publica información en páginas Web en Internet. El Microsoft SQL Server ofrece lo siguiente:

- Potente replicación de datos.
- Distribución segura de los datos a través de servidores y mainframes remotos.
- Fluída integración entre Internet e intranet.
- El Web Assistant (Asistente del Web) genera páginas HTML a partir de los datos de SQL Server en base a un disparador programado.
- Estrecha integración de correo.
- Garantiza que la información más actual esté al alcance de todo el que la necesite, sin demoras.
- El Microsoft SNA Server ofrece conectividad vía LU 6.2 a muchos entornos IBM, y con el nuevo Host
- Data Replicator permite replicación bi-direccional de datos entre DB2 y SQL Server.

8.7 Integración con Productos de Software para Escritorio

Algunos de los más populares productos de acceso para el usuario del escritorio de Microsoft, utilizados en sistemas de data warehousing, son el Microsoft Access, el Microsoft Excel, el Microsoft Query, y el Microsoft Internet Explorer. En septiembre de 1996, un setenta y cuatro por ciento de los participantes en una encuesta realizada por Market Perspectives Inc. y por el META Group para DCI, indicó que utilizaban estos productos de Microsoft de acceso a la información en sus respectivas organizaciones.

Los productos de software para escritorio de Microsoft tienen fuertes vínculos con el Microsoft SQL Server.

Aunque funcionan con cualquier sistema de base de datos que se rige por ODBC, están desarrollados y probados con el Microsoft SQL Server. Dichos productos utilizan ODBC nativo que no se coloca como una capa por encima de otros manipuladores (drivers), lo que mejora el rendimiento. Los clientes aprecian el poder usar las herramientas de acceso a la información que ya tienen y saben usar, cuando estas herramientas satisfacen adecuadamente sus necesidades.

El Microsoft Internet Explorer está ganando rápidamente popularidad como explorador del Web de acceso a información. Un producto lanzado en 1997 llamado Microsoft Visual InterDev! ayuda a convertir SQL Server en una de las mejores bases de datos para el desarrollo en el Web. Proporciona conectividad dinámica del Web a la base de datos, del tipo apunta-clic, incluyendo expertos y controles que automáticamente generan la lógica para crear informes personalizados o formularios HTML ligados a los datos. Visual InterDev les permite a los usuarios construir gráficamente consultas SQL para crear nuevas bases de datos SQL Server. Y ofrece una interfaz gráfica del usuario para administrar bases de datos SQL Server con capacidades que incluyen desde la adición de nuevas tablas, cambio de estructuras de la tabla "al vuelo", y la creación/alteración de índices y relaciones.

Debido a que OLE es común en todo Microsoft Office, los resultados de los datos pueden incorporarse rápidamente a un documento de Microsoft Word, enviarse por correo a otro lugar en la compañía, o incorporarse en una presentación en PowerPoint.

Además, muchas herramientas de acceso para el usuario de terceras partes, que ofrecen una variedad de capacidades, se han conectado para funcionar con el Microsoft SQL Server en Windows NT. Más información sobre algunas de éstas se encuentra en el sitio Web de Microsoft: www.microsoft.com/sql/.

8.8 Implementación por parte de muchos Solution Providers Independientes

Muchos Integradores de Sistemas y otros proveedores de soluciones implementan ahora soluciones de data warehousing para clientes que emplean el Microsoft SQL Server como motor de acumulación de información del Data Warehouse. Ellos encuentran que la facilidad de implementación y administración del Microsoft SQL Server en Windows NT los ayuda a ofrecer una solución al cliente de forma más eficiente. Y además se reduce el esfuerzo asociado con capacitar a los clientes para administrar sus propios sistemas después que ya han sido implementados.

8.9 Otras Capacidades de SQL Server para el Data Warehousing

- Cursores del servidor, que resultan útiles para las aplicaciones del explorador del escritorio. Usados en el tipo de soporte a las decisiones que ocurre cuando los usuarios no saben cuál información quieren hasta que la ven, mientras recorren los datos existentes en la base de datos.
- Un optimizador basado en el costo, con análisis de costo de consultas basado en estadísticas para una respuesta mejorada en el procesamiento de consultas mixtas contra la base de datos.
- Barrido paralelo de datos para un rendimiento mejorado de las consultas ad-hoc de soporte a decisiones.
- Extensiones de consultas OLAP, CUBE y ROLLUP, permiten que una consulta devuelva datos detallados y agregados a través de múltiples dimensiones, simplificando la recuperación de información que se necesita para efectuar análisis.
- Capacidad de Conducto de Datos (Data Pipes) que permite al SQL Server recuperar información programáticamente desde múltiples fuentes y popular las tablas SQL Server con esos resultados.
- Insert Exec permite almacenar en tablas los resultados de procedimientos guardados de forma remota o extendida.
- Respaldo y restauración de tablas individuales.
- Reconstrucción de índices sin caída y recreación del índice.
- Entrada/Salida asincrónica soporta acceso paralelo a múltiples dispositivos de disco para un mayor

caudal de procesamiento (throughput) cuando se procesan consultas grandes que necesitan leer datos de múltiples unidades de disco.

- Seguridad integrada para proteger sus datos de soporte a decisiones de posibles intrusiones.
- Administración de bases de datos centralizada con programación y manejo de excepciones.
- Soporte incorporado a multiprocesamiento simétrico (SMP) con balance automático de la carga de trabajo entre múltiples procesadores para aumentar el caudal de procesamiento total del CPU.

8.10 Las capacidades de base de datos multidimensional de Microsoft

En 1998, Microsoft está planeando lanzar una solución OLAP (Online Analytical Processing o Procesamiento Analítico en Línea) que administre datos multidimensionales. El desarrollo se llama proyecto "Plato." Éste incluirá capacidades para OLAP multidimensional (MOLAP) y OLAP relacional (ROLAP). Muchos clientes que instalan sistemas de data warehouses de soporte a decisiones descubren que una porción de sus necesidades de soporte a decisiones está limitada por la estructura de datos bidimensionales de los sistemas relacionales de administración de bases de datos. El producto OLAP de Microsoft está diseñado con las siguientes características:

- La solución OLAP de Microsoft funcionará con cualquier proveedor de datos que exponga ODBC u OLE DB.
- Microsoft ofrecerá una atractiva y única integración de sus servidores OLAP y SQL.
- La solución OLAP de Microsoft expondrá un modelo del objeto que será programable vía los lenguajes de programación Microsoft Visual Basic®, Java!, y C++®. Microsoft también espera ver a muchas aplicaciones de terceras partes desarrollar productos de "front end" especializados.
- La implementación de Microsoft no se trata simplemente de un concepto OLAP híbrido, sino que más bien ofrecerá a los clientes la más flexible arquitectura MOLAP/ROLAP/híbrida posible. ROLAP permite acceso a mayores acumulaciones de datos en una base de datos relacional. MOLAP brinda respuesta más rápida a consultas que pueden contestarse con preagregados y guardarse en una distribución multidimensional. Esta tecnología híbrida permite que las consultas tengan acceso a datos MOLAP o ROLAP, separada o concurrentemente, y el tipo de acceso pueda ser transparente para el usuario final.
- El producto OLAP de Microsoft resolverá dos de las preocupaciones más comunes en cuanto a análisis multidimensional – específicamente, alto precio y costos de implementación– por lo tanto hará que OLAP sea factible para clientes que no hayan podido, por razones de economía, instalar soluciones OLAP disponibles anteriormente.

8.11 Resumen

El Microsoft SQL Server en el sistema operativo Windows NT Server resulta popular en la instalación de sistemas de data warehousing debido a que reduce el costo de adquisición, de instalación y de administración. La solución de Microsoft resulta especialmente atractiva a aquellos clientes que anteriormente prefirieron complejos sistemas de almacenamiento de datos empresariales y que ahora favorecen proyectos de data marts más manejables. Los data marts en el Microsoft SQL Server y el Windows NT Server ofrecen una entrega y devolución más rápidas de la funcionalidad que necesitan las empresas para soporte a decisiones. Las capacidades desarrolladas en Microsoft para los productos de software para escritorio se extienden ahora a productos de múltiples usuarios, incluyendo integración del escritorio con el servidor de la base de datos. Y Microsoft continúa invirtiendo en próximos desarrollos y mejoras para garantizar que exista una amplia y atractiva selección de productos de data warehousing en el futuro.

9 Consideraciones finales del análisis

Por último se puede decir que un proyecto Data Warehousing se considera exitoso, cuando su objetivo final comienza a concretarse, es decir que la gente de la empresa use el DW para satisfacer sus necesidades empresariales.

Como ya hemos visto, son variados los cambios que comenzarán a producirse al implementar un DW. Es importante entonces anticiparse a estos cambios, considerar sus implicancias y planificarlos en la empresa. Las siguientes situaciones, gatillan el comienzo de estos cambios:

- La gente de la empresa depende del DW como un recurso primario de información.
- La gente de empresa se vuelve menos dependiente de los sistemas operacionales y de sus

bases de datos para sus necesidades de información.

- Se ve reducida o eliminada la demanda por programación especializada para encontrar la

información necesaria.

- Los usuarios y uso del DW crecen, con un correspondiente incremento en la demanda de

soporte.

- La complejidad de cambios en los sistemas operacionales se incrementa, y su efecto sobre el DW debe ser considerado.

- La finalidad de un Data Warehouse consiste en asistir a la administración a comprender el

pasado y planear el futuro. Aunque el Data Warehouse contiene información estratégica crucial, extraer un valor empresarial del Data Warehouse es un esfuerzo complejo. Obtener este valor requiere de una mezcla de aptitudes empresariales, habilidad técnica, intuición y experiencia.

- Un Data Warehouse puede producir la información correcta para los ejecutivos empresariales, gerentes y analistas. Conectarse al potencial del Data Warehouse sólo está limitado por la capacidad, habilidad y creatividad de los usuarios empresariales. Las experiencias anteriores apuntan hacia un amplio rango de usos: análisis de rentabilidad y crecimiento, administración estratégica, administración de relaciones con el cliente, administración de activos corporativos y manejo de la reingeniería del proceso empresarial (BPR) y la administración de calidad total (TQM). Los Data Warehouses se han aplicado con éxito en los sectores de manufactura, bienes de consumo y distribución, y en banca y finanzas.
- El uso del Data Warehouse en el curso diario de una empresa consiste en dos grandes actividades: acceso y recuperación, y análisis y reportes. Estas actividades se manejan a través del bloque de Acceso y Uso de la arquitectura de referencia. Existen dos modelos para aprovechar el Data Warehouse: el modelo de verificación, que utiliza técnicas de procesamiento analítico e informático, y el modelo de descubrimiento, que emplea técnicas de minería de datos.

10 Referencias y Bibliografía

- Harjinder S. Gill y Prakash C. Rao, Data Warehousing, Prentice Hall Hispanoamericana, 1996. Capítulos 1,2, 4. Pág.5, 6, 21, 305,306
- RPP Revista Profesional para programadores N° 45, Noviembre 1998

- Computer World N°145, 12 Noviembre 1997.
- Computer World N°151 18 Febrero 1998.
- Oracle Data Warehousing, Michael J.Corey & Michael Abbey, Computer World 1993 – pág. 218.
- Revista Informática Volumen 18, 8 Septiembre 1996.
- Designing The Perfect Datawarehouse, Frank McGuff, <http://www.techguide.com/>
- Datawarehouse Modeling, Frank McGuff, <http://www.techguide.com/>
- "Managing The Warehouse Throughout Its Lifecycle, <http://www.techguide.com/>
- Building A Decision Support Architecture For Datawarehousing, <http://www.techguide.com/>
- Putting Metadata To Work In The Warehouse , <http://www.techguide.com/>
- A Practical Guide To Getting Started With Data Warehousing, <http://www.techguide.com/>
- Guía de la Computación <http://www.guia.com.uy>

Direcciones de Internet

- <http://sol.facea.puc.cl/a352m/datawarehouse/index.htm>
- <http://www.gcc.com.mx/soluciones/dwhouse.htm>
- http://www.consisa.com/herramientas/pag_dw.htm
- <http://www.sybase.com/products/dataware/studio.html>
- http://www.microsoft.com/Spain/PRENSA/11_98/NOTA3.HTM

41

Ingeniería de la Información

Qué

Por qué

Acción