

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA – LEÓN

FACULTAD DE CIENCIAS

DEPARTAMENTO DE COMPUTACIÓN



MONOGRAFÍA

**“DATAWAREHOUSE, CONCEPTOS Y APLICACIÓN EN SQL
SERVER 7 BAJO PLATAFORMA WINDOWS 2000 SERVER”**

INTEGRANTES

**BR. URIEL OCTAVIO CASCO LEÓN
BR. DOUGLAS VIRGILIO BORDA GARCÍA**

TUTOR

MSC. MARTÍN IBARRA

**PREVIO REQUISITO PARA OPTAR AL TITULO DE LICENCIADO
EN COMPUTACIÓN**

LEÓN, 2004

DEDICATORIA

Queremos dedicar el presente trabajo monográfico a Dios nuestro creador, él ha hecho posible que este documento se haya realizado y se encuentre ahora en sus manos, nos ha encaminado siempre por el buen camino y ha hecho que este momento especial se haga realidad.

También dedicamos este trabajo a nuestros padres por todo su sacrificio y apoyo que nos han brindado siempre, ellos han sabido guiarnos en el transcurso de nuestros estudios y con su amor nos han formado hombres de bien.

Por último pero sin menor importancia se la dedicamos a todas las personas que de alguna u otra forma nos han brindado todo el calor, confianza y solidaridad durante el transcurso de nuestra carrera.

A todos ustedes... es la dedicatoria de esta Monografía.

AGRADECIMIENTOS

Queremos aprovechar la oportunidad que nos presenta este estudio monográfico para agradecer infinitamente a Dios, por habernos permitido culminar nuestros estudios y ser el principal impulsor en nuestras vidas a ser mejores cada día.

Les agradecemos con todo nuestro corazón a:

Nuestros padres: los cuales nos han brindado toda la seguridad económica y emocional que necesitamos, su apoyo incondicional y su amor tienen sus frutos en este día.

Nuestras familias: su cariño nos ha dado fuerzas para alcanzar cada una de nuestras metas.

Nuestros amigos: porque han estado siempre a nuestro lado, por su apoyo fiel y constante en todos los momentos. Gracias.

Al Personal Docente que de uno u otro modo nos ha ayudado a adquirir conocimientos en el transcurso de nuestra carrera.

A nuestro tutor: por su dedicación y apoyo científico.

ÍNDICE

| | Páginas |
|---|----------------|
| I. Introducción | 4 |
| II. Objetivos | 8 |
| III. Marco Teórico | 9 |
| A. Atributos y Conceptos | 9 |
| B. Sistemas de Datawarehousing | 17 |
| C. Control de Calidad de los Datos | 18 |
| D. Características | 19 |
| E. Operaciones en un Datawarehousing | 22 |
| F. Usos del Datawarehousing | 23 |
| G. Arquitectura | 24 |
| H. Estructura | 28 |
| I. Beneficios en el Uso del Datawarehouse | 29 |
| J. Ámbito de la Aplicación | 30 |
| IV. Aplicación | 32 |
| A. Escenario | 32 |
| B. Guía Práctica | 33 |
| V. Conclusiones | 46 |
| VI. Recomendaciones | 47 |
| VII. Bibliografía | 48 |
| VIII. Anexos | 49 |

I- INTRODUCCIÓN

Desde que se inició la era de la computadora, las organizaciones han generado y procesado sus datos haciendo uso de sus sistemas operacionales para atender sus necesidades de información. Algunas proporcionan acceso directo a la información contenida dentro de las aplicaciones operacionales. Otras, han extraído los datos desde sus bases de datos operacionales para combinarlos de varias formas no estructuradas, en su intento por atender a los usuarios en sus necesidades de información.

Ambos métodos han evolucionado a través del tiempo y ahora las organizaciones pueden estar manejando un banco de datos inseguro e inconsistente, sobre en la mayoría de las veces, se toman decisiones de vital importancia para dicha organización.

La gestión administrativa reconoce que una manera de elevar su eficiencia está en hacer el mejor uso de los recursos de información que ya existen dentro de la organización y poder conjugar dichos datos con nuevos sistemas automatizados basados en nuevas tecnologías.

La información es muy importante y útil para organizar las actividades de la empresa. Contar con una buena organización de la información y con datos adecuadamente organizados es vital para la toma de decisiones y para la generación de estadística que ayudan a la empresa a analizar su desempeño y los factores que lo influyen para resolver problemas de productividad.

Frecuentemente la información no está digitalizada, y si lo esta, se encuentra en archivos o fuentes con formato que no facilitan su análisis. Además, en muchas empresas no se ha hecho el esfuerzo de identificar los criterios de observación adaptados a sus características organizacionales, a su estructura y a su funcionamiento, que les ayudarán a analizar las situaciones que caracterizan su desempeño y que determinan su productividad.

El datawarehousing, es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

Reunir los elementos de datos apropiados desde diversas fuentes de aplicación en un ambiente integral centralizado, simplifica el problema de acceso a la información y en consecuencia, acelera el proceso de análisis, de consultas y el menor tiempo de acceso a la información.

Nuestro trabajo consiste en:

- Recopilar bibliografía.
- Investigar sobre datawarehouse.
- Organizar la documentación.
- Presentar resultados.

Ya que consideramos que con esto aportamos a la demostración del manejo del datawarehouse. En nuestro caso particular, nuestro estudio se limita a:

- Ambiente.
- Sistema Operativo.
- Gestión de Base de Datos.

Se desarrollo un ejemplo completo del diseño del datawarehouse y se proporciona una guía detallada del proceso.

Sobre el Datawarehouse existe una serie de antecedentes a destacar, por un lado tenemos el legado de sistemas de información que fueron desarrollados desde los años 70 sobre la plataforma IBM utilizando herramientas tales como Cobol, CICS, IMS, DB2, etc.; durante los años 80 utilizando AS/400 y VAX/VMS, y a finales de los 80 utilizando UNIX, plataforma que introduce la arquitectura cliente / servidor.

Por otro lado, otra categoría popular de sistemas de análisis es la formada por los sistemas de apoyo en la toma de decisiones y los sistemas de información ejecutiva. Los sistemas de apoyo en la toma de decisiones tienden a enfocar más los detalles de los datos y están enfocados a los encargados de las tomas de decisiones de nivel medio. Los sistemas de información ejecutiva generalmente proveen un alto nivel de consolidación y una vista

multidimensional de los datos necesarios para los altos ejecutivos que dividen y juegan con la información sin tener que detenerse en los detalles de los datos.

Estos dos tipos de sistemas son los más cercanos antepasados del datawarehouse. Hoy en día los sistemas basados en datawarehouse proveen todas las herramientas analíticas que exponen sus precursores.

Existen también otros antecedentes que favorecieron al desarrollo de sistemas basados en datawarehouse. Primero, el incremento en las capacidades de procesamiento y almacenamiento del hardware moderno a la par del decremento en los precios de los mismos. Segundo, durante la década pasada se dio el incremento en la popularidad de las computadoras personales las que han introducido un sin número de oportunidades y nuevas opciones para el análisis de los negocios. El puente entre los programadores y los usuarios finales ha iniciado a cerrarse a medida que los analistas de negocios tienen a sus posibilidades muchas de las herramientas requeridas para mejorar la eficiencia en el uso de hojas de cálculo para la representación y análisis gráfico de la información. Las capacidades modernas de las computadoras personales han permitido desarrollar sistemas basados en la arquitectura cliente / servidor de tal forma que la mayoría de datawarehouses son accedidas mediante las herramientas para computadoras personales. Tercero, desarrollo de sistemas operativos servidores tales como WindowsNT y Unix han sido de vital importancia en el desarrollo de sistemas distribuidos cuyos costos han permanecido estables o en disminución.

Todo esto combinado con los conceptos de memoria virtual, multitarea y multiprocesamiento simétrico han ayudado a obtener plataformas de desarrollo para datawarehouses relativamente a bajo costo. Finalmente, el fenómeno más importante en el desarrollo de la computación ha sido la explosión en el desarrollo de sistemas basados en Internet y la Web. Uno de los campos de mayor desarrollo en las industrias son las llamadas Intranets que no son más que redes privadas basadas en los estándares de Internet y que son diseñadas para uso interno de una empresa o negocio.

También se dan factores tales como la globalización basada en la demanda de Información, el cambio en la naturaleza de los negocios, la caída del comunismo y la liberación de las economías de Asia y América del Sur y el surgimiento de aplicaciones estandarizadas para negocios han tenido un efecto directo en el uso de datawarehouses como mecanismo para reemplazar los sistemas anteriores y para facilitar las transacciones sobre grandes bancos de datos que son accedidos a través de Intranet o de Internet a grandes servidores multitareas.

La mayoría de las organizaciones hacen lo posible por conseguir buena información, pero el logro de ese objetivo depende fundamentalmente de su arquitectura actual, tanto de hardware como de software, producto de esto hemos orientado nuestro trabajo al uso del DataWareHouse por que consideramos vital la implementación de esta nueva tecnología en las empresas, que día a día necesitan y deben hacer más práctica y fácil la explotación de datos para una mayor eficacia del negocio. Con este nuevo enfoque se pretende brindar a las empresas una nueva alternativa que le permita llevar una mejor organización y control de su información para toma de decisiones más acertadas y eficaces, asegurando de este modo el crecimiento mismo de la empresa.

El datawarehouse no es una aplicación como otra (Visual Basic, FoxPro, etc), dado que esta se encuentra en un nivel mas alto ya que proporciona una mayor explotación (comprensión) de la información que maneja la empresa o la institución. Se puede considerar para la empresa la piedra angular en el tratamiento de la información que posee, sustentadas en las bases de datos existentes y algún otro mecanismo que sirva de almacenamiento de información en la misma.

II - OBJETIVOS

OBJETIVOS GENERALES

1. Recopilar y estudiar la tecnología del DataWareHouse.
2. Realizar la construcción e implementación de un DataWareHouse para el análisis de la información empresarial.

OBJETIVOS ESPECIFICOS

1. Explicar qué es un DataWareHouse, sus utilidades y funcionamiento.
2. Mostrar las ventajas existentes en el uso del DataWareHouse.
3. Dar a conocer las herramientas que conforman el DataWareHouse.
4. Construir un DataWareHouse.
5. Instrumentar una aplicación que valide el uso del DataWareHouse.
6. Adquirir conocimientos y experiencia en el manejo de un DataWareHouse.

III - MARCO TEORICO

A - ATRIBUTOS Y CONCEPTOS.

DataWareHouse

- ✚ Es una colección de información corporativa derivada directamente de los sistemas operacionales y de algunos orígenes de datos externos.
- ✚ Un datawarehousing es una colección de datos en la cual se encuentra integrada la información de la Institución y que se usa como soporte para el proceso de toma de decisiones gerenciales.
- ✚ Es una copia de datos de transacciones específicamente estructurada para consultas y reportes. También se podría decir que un datawarehouse es la manipulación de datos estructurados después y fuera de un sistema operacional.

Aunque diversas organizaciones y personas individuales logran comprender el enfoque de un Warehousing, la experiencia ha demostrado que existen muchas dificultades potenciales.

Su propósito específico es soportar la toma de decisiones en un negocio, no las operaciones de un negocio.

A estos conceptos aclararemos dos aspectos: 1) a veces se guardan datos no transaccionales en un almacén de datos - aunque probablemente 95-99% de los datos normalmente son datos de la transacción. 2) Las preguntas e informes generados de datos guardados en un almacén de datos pueden o no pueden usarse para el análisis.

Algo que no se dice en las definiciones expuestas anteriormente es que la forma en que los datos son almacenados no tiene nada que ver con el hecho de que esos datos constituyan un datawarehouse.

Un datawarehouse puede normalizarse o desnormalizarse. Puede ser un banco de datos correlativo, un banco de datos multidimensional, un archivo llano, un banco de datos jerárquico, un banco de datos de objetos, etc. Los datawarehouses se enfocan a menudo en una actividad específica o entidad.

El datawarehouse no es precisamente para llenar las necesidades de los "tomadores de decisiones" o para ser utilizado en el proceso de hacer decisiones. Por supuesto si usted quiere definir a cada usuario como tomador de decisión y todas las actividades como procesos de tomas de decisión, entonces lo antes dicho es falso. Pero por experiencia de otros investigadores, los usos masivos de datawarehousing son para cosas rutinarias, de hecho, no se quiere decir que hemos concluido diciendo que el datawarehousing en el proceso de hacer decisiones no es un esfuerzo maravilloso y potencialmente de alto retorno.

Pero la cautela es que aunque existe presión de los fabricantes, vendedores, y muchos expertos de la industria que alaban el papel del datawarehousing versus el haciendo decisión, ésta es una área en realidad que nosotros no tenemos una comprensión clara hasta el momento.

Las aplicaciones para soporte de decisiones basadas en un dataWarehousing, pueden hacer más práctica y fácil la explotación de datos para una mayor eficacia del negocio, que no se logra cuando se usan sólo los datos que provienen de las aplicaciones operacionales, en los que la información se obtiene realizando procesos independientes y muchas veces complejos.

Un datawarehousing se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales. La data extraída es transformada para eliminar inconsistencias y resumir si es necesario y luego, cargadas en el datawarehousing. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Institucional. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad.

La innovación de la Tecnología de Información dentro de un ambiente dataWarehousing, puede permitir a cualquier organización hacer un uso más óptimo de los datos, como un ingrediente clave para un proceso de toma de decisiones más efectivo. Las organizaciones tienen que aprovechar sus datos para crear la información de la operación del negocio, pero deben considerarse las estrategias tecnológicas necesarias para la implementación de una arquitectura completa de datawarehousing.

DataWarehousing es el centro de la arquitectura para los sistemas de información en la década de los '90. Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico, informático sobre una amplia perspectiva de tiempo.

Se puede caracterizar un dataWarehousing haciendo un contraste de cómo los datos de un negocio almacenados en un dataWarehousing, difieren de los datos operacionales usados por las aplicaciones de producción.

| Base de Datos Operacional | Data Warehousing |
|----------------------------------|------------------------------------|
| Datos Operacionales | Datos del negocio para Información |
| Orientado a la aplicación | Orientado al sujeto |
| Actual | Actual + histórico |
| Detallada | Detallada + más resumida |

Dentro de una empresa el uso de un Sistema de DataWarehousing se aplica no solo en el soporte de la gerencia de alto nivel que debe tomar decisiones de objetivos y planificación estratégica, sino también en niveles más técnicos de toma de decisión.

Olap

 Es un tipo de tecnología que permite a los usuarios mejorar la visión que tienen de sus datos de una manera rápida, interactiva y fácil de usar.

Los servicios OLAP de SQL Server 7.0 ofrecen potentes herramientas que ayudan a las empresas a utilizar los datos para analizar la actividad empresarial. Mediante un servidor SQL Server OLAP y un lado cliente motor de cálculo y almacenamiento en caché denominado Microsoft PívoTable Service, los servicios OLAP permiten a las empresas descubrir tendencias de mercado, por ejemplo, que pueden servir para respaldar las decisiones de optimización de los procesos empresariales.

Microsoft proporciona una potencia de análisis de datos y creación de informes sin precedentes a las organizaciones que antes no podían justificar el coste de una herramienta de inteligencia empresarial tan sofisticada.

Cubos

- Proporcionan un mecanismo de consulta de datos con un tiempo corto de respuesta.
- Se organizan y resumen en estructuras multidimensionales.

Los Cubos están constituidos por:

- Origen de Datos
- Dimensiones
- Medidas
- Particiones

Esquemas relacionales y cubos

La manera más común de administrar datos relacionales para su empleo multidimensional es un esquema de estrella. Un esquema de estrella consiste en una única tabla de hechos que se combina con varias tablas de dimensiones. La tabla de hechos contiene los datos numéricos que se corresponden con las medidas de un cubo. Las columnas de la tabla de dimensiones, tal como implica su nombre, asignan los niveles jerárquicos de una dimensión.

No se necesita un esquema de estrella para crear un cubo. También podrá utilizar un esquema radial ramificado o, incluso, un esquema sencillo de tabla.

Dimensiones de un cubo

Las dimensiones de un cubo representan distintas categorías para analizar datos empresariales. Categorías tales como fecha, geografía o línea de productos son dimensiones típicas de cubos.

Los cubos no están limitados a tres dimensiones. Pueden contener hasta 64 dimensiones.

Dimensiones y jerarquías

Las dimensiones se suelen organizar en jerarquías de información que se asignan a columnas en una base de datos relacional. Las jerarquías de dimensiones están agrupadas en niveles que constan de los miembros de una dimensión. Podrá unir los niveles de una dimensión para formar los valores de los que constará el siguiente nivel superior. Por ejemplo, en una dimensión temporal, los días se unen en meses y los meses forman trimestres. Ciertas columnas podrán especificar una jerarquía de niveles de una dimensión.

Una dimensión se puede clasificar en:

- privada
- compartida

Tabla de Hechos

Contiene las medidas a incluir en el cubo. Se debe conservar el nivel de granularidad más fino. Posee claves externas que se combinan con las Claves principales en las tablas de dimensiones.

Es la tabla central de un esquema:

- esquema estrella
- esquema radial ramificado

Medidas de un cubo

Constituyen los datos numéricos que los usuarios ven y analizan. Las medidas son los valores cuantitativos contenidos en la base de datos que desea analizar. Cada medida se corresponde con una columna de la tabla de hechos. Las medidas típicas son ventas, costo y datos presupuestarios. Las medidas se analizan contra las distintas categorías de dimensiones de un cubo. Por ejemplo, tal vez desee analizar datos de ventas y de presupuestos (sus medidas) para un determinado producto (una dimensión) correspondientes a varios países (niveles específicos de una dimensión geográfica) durante dos años concretos (niveles de una dimensión temporal).

Cubos virtuales

Los cubos virtuales le permitirán ampliar los cubos que haya definido sin necesidad de aumentar los requisitos de almacenamiento de la base de datos. En este aspecto, los cubos virtuales son similares a las vistas de una base de datos relacional.

Combinación de varios cubos

Cuando cree un cubo virtual, incluirá medidas y dimensiones de varios cubos para proporcionar una vista más amplia de los datos. Por ejemplo, se podrían combinar los datos de un cubo de ventas y de un cubo de mercadotecnia para proporcionar un método de comparación lado a lado que permita ver cómo afectan las promociones de mercadotecnia a las cantidades vendidas.

Cubos virtuales y almacenamiento de datos

Un cubo virtual utiliza las opciones de rendimiento de la consulta y los modelos de almacenamiento de los cubos que lo definen, pero no necesita un espacio adicional para almacenar datos. Un cubo que utiliza el almacenamiento MOLAP se podrá combinar con otros cubos que utilicen almacenamiento ROLAP y HOLAP para crear un cubo virtual.

Almacenamiento de Datos

El Administrador de OLAP proporciona tres maneras distintas de almacenar los datos en un cubo:

- OLAP multidimensional (MOLAP)
- OLAP relacional (ROLAP)
- OLAP híbrido (HOLAP)

Cada una de estas opciones proporciona ciertas ventajas en función del tamaño de la base de datos y de cómo se van a utilizar los datos.

Almacenamiento MOLAP

MOLAP es un formato de almacenamiento de datos multidimensional de alto rendimiento. Con MOLAP, los datos se almacenan en el servidor OLAP. MOLAP proporciona el mejor rendimiento de consulta, porque está específicamente optimizado para consultas de datos multidimensionales.

El almacenamiento MOLAP es apropiado para conjuntos de datos de tamaño pequeño a medio en los que la copia de todos los datos al formato multidimensional no exigiría un tiempo de carga significativo o utilizar grandes cantidades de espacio en disco.

Almacenamiento ROLAP

Con ROLAP, los datos permanecen en las tablas relacionales originales. Se utiliza un conjunto separado de tablas relacionales para almacenar y hacer referencia a los datos de agregados. ROLAP es ideal para bases de datos de gran tamaño o datos antiguos que se consultan con poca frecuencia.

Almacenamiento HOLAP

HOLAP combina elementos de MOLAP y ROLAP. HOLAP mantiene los datos originales en tablas relacionales pero almacena los agregados en formato multidimensional. HOLAP proporciona conectividad a grandes conjuntos de datos en tablas relacionales a la vez que aprovecha el mayor rendimiento del almacenamiento multidimensional de agregados.

Creación de particiones en un cubo

El Administrador de OLAP le permitirá almacenar, administrar y distribuir los datos de un cubo mediante particiones. Las particiones dividen un cubo en segmentos independientes que se pueden optimizar de manera individual, a la vez que se podrán consultar como si se tratara de un único cubo.

Opciones de almacenamiento en particiones

Cada cubo consta, al menos, de una partición; sin embargo, un cubo puede dividirse en varias particiones. Las distintas particiones podrán tener distintas opciones de almacenamiento de datos. Por ejemplo, un cubo puede tener tres particiones, una que utilice ROLAP, otra con HOLAP y la tercera con MOLAP.

Distribución de datos

Las particiones le permitirán separar los datos del cubo en un clúster de servidores. Por ejemplo, puede optar por almacenar los datos más antiguos y menos consultados en los servidores más lentos.

Los datos más recientes y consultados con mayor frecuencia se podrán almacenar en un servidor de alta velocidad para aumentar el rendimiento de las consultas.

Rebanadas de datos

Una rebanada de datos representa un subconjunto de los datos almacenados en una partición. Por ejemplo, podría crear una rebanada si desea analizar los datos de ventas de un producto específico durante varios años.

B - SISTEMAS DE DATA WAREHOUSING

En los últimos años se ha acrecentado el interés por disponer de información altamente sintética como base para la toma de decisiones. Dicha información debe ser extraíble a partir de datos atómicos almacenados en Bases de Datos de producción. El interés por obtener información a partir de datos de producción se manifiesta bajo la forma de comentarios tales como:

Los Sistemas de DataWarehousing surgen con el objetivo de resolver estas necesidades. Un Sistema de DataWarehousing incluye funcionalidades tales como: (a) Integración de bases de datos heterogéneas (relacionales, documentales, geográficas, archivos, etc.). (b) Ejecución de consultas complejas no predefinidas visualizando el resultado en forma de gráfica y en diferentes niveles de agrupamiento y totalización de datos. (c) Agrupamiento y desagrupamiento de datos en forma interactiva. (d) Análisis de problema en términos de dimensiones. Por ejemplo, permite analizar datos históricos a través de una dimensión tiempo. (e) 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.

La arquitectura lógica de un sistema de DataWarehousing es del tipo mostrado a continuación.



C - CONTROL DE CALIDAD DE LOS DATOS.

En el DataWarehousing es un tema fundamental en el éxito de un proyecto de este tipo: Si quien debe tomar decisiones duda de la calidad de la información, entonces no utilizará el sistema de DataWarehousing. La calidad de datos en el DataWarehousing involucra, no solo la consistencia clásica en bases de datos, sino también las nociones de pertinencia y relevancia de los datos almacenados. Finalmente, un sistema de DataWarehousing no es un sistema monolítico, sino una combinación flexible de múltiples módulos. Dichos módulos pueden ser productos o soluciones específicamente desarrolladas para la aplicación a resolver.

D - CARACTERISTICAS

Orientado a temas.

Una primera característica del datawarehouse es que la información se clasifica en base a los aspectos que son de interés para la empresa. Siendo así, los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones. El ambiente operacional se diseña alrededor de las aplicaciones y funciones tales como préstamos, ahorros, tarjeta bancaria y depósitos para una institución financiera. Por ejemplo, una aplicación de ingreso de órdenes puede acceder a los datos sobre clientes, productos y cuentas. La base de datos combina estos elementos en una estructura que acomoda las necesidades de la aplicación.

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

Integración.

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

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

El contraste de la integración encontrada en el datawarehousing con la carencia de integración del ambiente de aplicaciones.

Funciones Múltiples.

El mismo elemento puede derivarse desde fuentes múltiples. En este caso, el proceso de transformación debe asegurar que la fuente apropiada sea usada, documentada y movida al depósito.

Tal como se muestra en la figura, los puntos de integración afectan casi todos los aspectos de diseño - las características físicas de los datos, la disyuntiva de tener más de una de fuente de datos, el problema de estándares de denominación inconsistentes, formatos de fecha también inconsistentes y otros.

Cualquiera que sea la forma del diseño, el resultado es el mismo - la información necesita ser almacenada en el datawarehousing en un modelo globalmente aceptable y singular, aun cuando los sistemas operacionales subyacentes almacenen los datos de manera diferente.

De tiempo volátil

Toda la información del datawarehousing es requerida en algún momento. Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En éstos, la información se requiere al momento de acceder. En otras palabras, en el ambiente operacional, cuando usted accesa a una unidad de información, usted espera que los valores requeridos se obtengan a partir del momento de acceso.

Como la información en el datawarehousing es solicitada en cualquier momento (es decir, no "ahora mismo"), los datos encontrados en el depósito se llaman de "tiempo variante".

El tiempo variante se muestra de varias maneras:

1° La más simple es que la información representa los datos sobre un horizonte largo de tiempo - desde cinco a diez años. El horizonte de tiempo representado para el ambiente operacional es mucho más corto - desde valores actuales hasta sesenta a noventa días.

Las aplicaciones que tienen un buen rendimiento y están disponibles para el procesamiento de transacciones, deben llevar una cantidad mínima de datos si tienen cualquier grado de flexibilidad. Por ello, las aplicaciones operacionales tienen un corto horizonte de tiempo, debido al diseño de aplicaciones rígidas.

2° La segunda manera en la que se muestra el tiempo variante en el datawarehouse está en la estructura clave. Cada estructura clave en el datawarehouse contiene, implícita o explícitamente, un elemento de tiempo como día, semana, mes, etc.

El elemento de tiempo está casi siempre al pie de la clave concatenada, encontrada en el datawarehousing. En ocasiones, el elemento de tiempo existirá implícitamente, como el caso en que un archivo completo se duplica al final del mes, o al cuarto.

3° La tercera manera en que aparece el tiempo variante es cuando la información del datawarehouse, una vez registrada correctamente, no puede ser actualizada. La información del datawarehousing es, para todos los propósitos prácticos, una serie larga de "snapshots" (vistas instantáneas).

Por supuesto, si los snapshots de los datos se han tomado incorrectamente, entonces pueden ser cambiados. Asumiendo que los snapshots se han tomado adecuadamente, ellos no son alterados una vez hechos. En algunos casos puede ser no ético, e incluso ilegal, alterar los snapshots en el datawarehouse. Los datos operacionales, siendo requeridos a partir del momento de acceso, pueden actualizarse de acuerdo a la necesidad.

No volátil.

La información es útil sólo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable.

La actualización (insertar, borrar y modificar), se hace regularmente en el ambiente operacional sobre una base de registro por registro. Pero la manipulación básica de los datos que ocurre en el datawarehousing es mucho más simple. Hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No hay actualización de datos (en el sentido general de actualización) en el depósito, como una parte normal de procesamiento.

Hay algunas consecuencias muy importantes de esta diferencia básica, entre el procesamiento operacional y del datawarehousing. En el nivel de diseño, la necesidad de ser precavido para actualizar las anomalías no es un factor en el datawarehousing, ya que

no se hace la actualización de datos. Esto significa que en el nivel físico de diseño, se pueden tomar libertades para optimizar el acceso a los datos, particularmente al usar la normalización y de normalización física.

E - OPERACIONES EN UN DATA WAREHOUSING

Entre algunos de los tipos de operaciones que se efectúan dentro de un ambiente dataWarehousing tenemos:

A) Sistemas Operacionales

Los datos administrados por los sistemas de aplicación operacionales son la fuente principal de datos para el datawarehousing.

Las bases de datos operacionales se organizan como archivos indexados (UFAS, VSAM), bases de datos de redes/jerárquicas (I-D-S/II, IMS, IDMS) o sistemas de base de datos relacionales (DB2, Oracle, Informix, etc.). Según las encuestas, aproximadamente del 70% a 80% de las bases de datos de las empresas se organizan usando DBMSs no relacional.

B) Extracción, Transformación Y Carga De Los Datos

Se requieren herramientas de gestión de datos para extraer datos desde bases de datos y/o archivos operacionales, luego es necesario manipular o transformar los datos antes de cargar los resultados en el datawarehousing.

Tomar los datos desde varias bases de datos operacionales y transformarlos en datos requeridos para el depósito, se refiere a la transformación o a la integración de datos. Las bases de datos operacionales, diseñadas para el soporte de varias aplicaciones de producción, frecuentemente difieren en el formato. Los mismos elementos de datos, si son usados por aplicaciones diferentes o administrados por diferentes software DBMS, pueden definirse al usar nombres de elementos inconsistentes, que tienen formatos inconsistentes y/o ser codificados de manera diferente. Todas estas inconsistencias deben resolverse antes que los elementos de datos sean almacenados en el datawarehousing.

C) Meta Data

Otro paso necesario es crear la meta data. La meta data (es decir, datos acerca de datos) describe los contenidos del datawarehousing. La meta data consiste de definiciones de los elementos de datos en el depósito, sistema(s) del (os) elemento(s) fuente.

D) Acceso De Usuario Final

Los usuarios accesan al data warehousing por medio de herramientas de productividad basadas en GUI (Graphical User Interface - Interfase gráfica de usuario). Pueden proveerse a los usuarios del data warehousing muchos de estos tipos de herramientas. Estos pueden incluir software de consultas, generadores de reportes, procesamiento analítico en línea, herramientas data/visual mining, etc., dependiendo de los tipos de usuarios y sus requerimientos particulares. Sin embargo, una sola herramienta no satisface todos los requerimientos, por lo que es necesaria la integración de una serie de herramientas.

F - USOS DEL DATA WAREHOUSING

Los datos operacionales y los datos del data warehousing son accesados por usuarios que usan los datos de maneras diferentes.

| Uso de Base de Datos Operacionales | Uso de Data Warehousing |
|---|---|
| Muchos usuarios concurrentes | Pocos usuarios concurrentes |
| Consultas predefinidas y actualizables | Consultas complejas, frecuentemente no anticipadas. |
| Cantidades pequeñas de datos detallados | Cantidades grandes de datos detallados |

Los usuarios que accedan a los datos operacionales, comúnmente efectúan tareas predefinidas que, generalmente requieren acceso a una sola base de datos de una aplicación. Por el contrario, los usuarios que accedan al datawarehouse, efectúan tareas que requieren acceso a un conjunto de datos desde fuentes múltiples y frecuentemente no son predecibles. Lo único que se conoce (si es modelada correctamente) es el conjunto inicial de datos que se han establecido en el depósito.

Por ejemplo, un especialista en el cuidado de la salud podría necesitar acceder a los datos actuales e históricos para analizar las tendencias de costos, usando un conjunto de consultas predefinidas. Por el contrario, un representante de ventas podría necesitar acceder a los datos de cliente y producto para evaluar la eficacia de una campaña de marketing, creando consultas base o ad-hoc para encontrar nuevamente necesidades definidas.

G - ARQUITECTURA

Una de las razones por las que el desarrollo de un data warehousing crece rápidamente, es que realmente es una tecnología muy entendible. De hecho, data Warehousing puede representar mejor la estructura amplia de una empresa para administrar los datos informacionales dentro de la organización. A fin de comprender cómo se relacionan todos los componentes involucrados en una estrategia data Warehousing.

✓ *Elementos Constituyentes De Una Arquitectura DataWarehousing.*

Una Arquitectura Data Warehousing (Data Warehousing Architecture - DWA) es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa.

La arquitectura se constituye de un número de partes interconectadas:

✓ *Base De Datos Operacional / Nivel De Base De Datos Externo*

Los sistemas operacionales procesan datos para apoyar las necesidades operacionales críticas. Para hacer eso, se han creado las bases de datos operacionales históricas que proveen una estructura de procesamiento eficiente, para un número relativamente pequeño de transacciones comerciales bien definidas.

Sin embargo, a causa del enfoque limitado de los sistemas operacionales, las bases de datos diseñadas para soportar estos sistemas, tienen dificultad al acceder a los datos para otra gestión o propósitos informáticos.

Esta dificultad en acceder a los datos operacionales es amplificada por el hecho que muchos de estos sistemas tienen de 10 a 15 años de antigüedad. El tiempo de algunos de estos sistemas significa que la tecnología de acceso a los datos disponible para obtener los datos operacionales, es así mismo antigua.

Ciertamente, la meta del data Warehousing es liberar la información que es almacenada en bases de datos operacionales y combinarla con la información desde otra fuente de datos, generalmente externa.

Cada vez más, las organizaciones grandes adquieren datos adicionales desde bases de datos externas. Esta información incluye tendencias demográficas, econométricas, adquisitivas y competitivas (que pueden ser proporcionadas por Instituciones Oficiales - INEI). Internet o también llamada "information superhighway" (supercarretera de la información) provee el acceso a más recursos de datos todos los días.

✓ ***Nivel De Acceso A La Información***

El nivel de acceso a la información de la arquitectura data warehousing, es el nivel del que el usuario final se encarga directamente. En particular, representa las herramientas que el usuario final normalmente usa día a día. Por ejemplo: Excel, Lotus 1-2-3, Focus, Access, SAS, etc.

Este nivel también incluye el hardware y software involucrados en mostrar información en pantalla y emitir reportes de impresión, hojas de cálculo, gráficos y diagramas para el análisis y presentación. Hace dos décadas que el nivel de acceso a la información se ha expandido enormemente, especialmente a los usuarios finales quienes se han volcado a los PCs monousuarios y los PCs en redes.

Actualmente, existen herramientas más y más sofisticadas para manipular, analizar y presentar los datos, sin embargo, hay problemas significativos al tratar de convertir los datos tal como han sido recolectados y que se encuentran contenidos en los sistemas operacionales en información fácil y transparente para las herramientas de los usuarios finales. Una de las claves para esto es encontrar un lenguaje de datos común que puede usarse a través de toda la empresa.

✓ *Nivel De Acceso A Los Datos*

El nivel de acceso a los datos de la arquitectura data warehousing está involucrado con el nivel de acceso a la información para conversar en el nivel operacional. En la red mundial de hoy, el lenguaje de datos común que ha surgido es SQL. Originalmente, SQL fue desarrollado por IBM como un lenguaje de consulta, pero en los últimos veinte años ha llegado a ser el estándar para el intercambio de datos.

Uno de los adelantos claves de los últimos años ha sido el desarrollo de una serie de "filtros" de acceso a datos, tales como EDA/SQL para acceder a casi todo los Sistemas de Gestión de Base de Datos (Data Base Management Systems - DBMSs) y sistemas de archivos de datos, relacionales o no. Estos filtros permiten a las herramientas de acceso a la información, acceder también a la data almacenada en sistemas de gestión de base de datos que tienen veinte años de antigüedad.

El nivel de acceso a los datos no solamente conecta DBMSs diferentes y sistemas de archivos sobre el mismo hardware, sino también a los fabricantes y protocolos de red. Una de las claves de una estrategia data Warehousing es proveer a los usuarios finales con "acceso a datos universales".

El acceso a los datos universales significa que, teóricamente por lo menos, los usuarios finales sin tener en cuenta la herramienta de acceso a la información o ubicación, deberían ser capaces de acceder a cualquier o todos los datos en la empresa que es necesaria para ellos, para hacer su trabajo.

El nivel de acceso a los datos entonces es responsable de la interfaces entre las herramientas de acceso a la información y las bases de datos operacionales. En algunos casos, esto es todo lo que un usuario final necesita. Sin embargo, en general, las organizaciones desarrollan un plan mucho más sofisticado para el soporte del data Warehousing.

✓ ***Nivel De Directorio De Datos (Meta Data)***

A fin de proveer el acceso a los datos universales, es absolutamente necesario mantener alguna forma de directorio de datos o repositorio de la información meta data. La meta data es la información alrededor de los datos dentro de la empresa.

Las descripciones de registro en un programa COBOL son meta data. También lo son las sentencias DIMENSION en un programa FORTRAN o las sentencias a crear en SQL.

A fin de tener un depósito totalmente funcional, es necesario tener una variedad de meta data disponible, información sobre las vistas de datos de los usuarios finales e información sobre las bases de datos operacionales. Idealmente, los usuarios finales deberían de acceder a los datos desde el data warehousing (o desde las bases de datos operacionales), sin tener que conocer dónde residen los datos o la forma en que se han almacenados.

✓ ***Nivel De Gestión De Procesos***

El nivel de gestión de procesos tiene que ver con la programación de diversas tareas que deben realizarse para construir y mantener el data warehousing y la información del directorio de datos. Este nivel puede depender del alto nivel de control de trabajo para muchos procesos (procedimientos) que deben ocurrir para mantener el data warehousing actualizado.

✓ ***Nivel De Mensaje De La Aplicación***

El nivel de mensaje de la aplicación tiene que ver con el transporte de información alrededor de la red de la empresa. El mensaje de aplicación se refiere también como "subproducto", pero puede involucrar sólo protocolos de red. Puede usarse por ejemplo, para aislar aplicaciones operacionales o estratégicas a partir del formato de datos exacto, recolectar transacciones o los mensajes y entregarlos a una ubicación segura en un tiempo seguro.

✓ ***Nivel Data Warehousing (Físico)***

En el data warehousing (núcleo) es donde ocurre la data actual, usada principalmente para usos estratégicos. En algunos casos, uno puede pensar del datawarehouse simplemente como una vista lógica o virtual de datos. En muchos ejemplos, el data warehousing puede no involucrar almacenamiento de datos.

En un data warehousing físico, copias, en algunos casos, muchas copias de datos operacionales y/o externos, son almacenados realmente en una forma que es fácil de acceder y es altamente flexible. Cada vez más, los data warehousing son almacenados sobre plataformas cliente/servidor, pero por lo general se almacenan sobre mainframes.

H- ESTRUCTURA.

Los data warehousing tienen una estructura distinta. Hay niveles diferentes de esquematización y detalle que delimitan el data warehousing.

✓ ***Detalle De Datos Actuales***

En gran parte, el interés más importante radica en el detalle de los datos actuales, debido a que:

Refleja las ocurrencias más recientes, las cuales son de gran interés

Es voluminoso, ya que se almacena al más bajo nivel de granularidad.

Casi siempre se almacena en disco, el cual es de fácil acceso, aunque su administración sea costosa y compleja.

✓ ***Detalle De Datos Antiguos***

La data antigua es aquella que se almacena sobre alguna forma de almacenamiento masivo. No es frecuentemente accesada y se almacena a un nivel de detalle, consistente con los datos detallados actuales. Mientras no sea prioritario el almacenamiento en un medio de

almacenaje alternativo, a causa del gran volumen de datos unido al acceso no frecuente de los mismos, es poco usual utilizar el disco como medio de almacenamiento.

✓ ***Datos Ligeramente Resumidos***

La data ligeramente resumida es aquella que proviene desde un bajo nivel de detalle encontrado al nivel de detalle actual. Este nivel del data warehousing casi siempre se almacena en disco. Los puntos en los que se basa el diseñador para construirlo son:

1. Que la unidad de tiempo se encuentre sobre la esquematización hecha.
2. Que contenidos (atributos) tendrá la data ligeramente resumida.

✓ ***Datos Completamente Resumidos***

El siguiente nivel de datos encontrado en el data warehousing es el de los datos completamente resumidos. Estos datos son compactos y fácilmente accesibles.

I - BENEFICIOS EN EL USO DEL DATAWAREHOUSE

Los beneficios que aporta el datawarehouse pueden resumirse:

- ✓ Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- ✓ Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén, obteniendo un valor añadido para el negocio de dicha información.
- ✓ Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- ✓ Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

- ✓ Supone una optimización tecnológica y económica en entornos de Centros de información, estadística, o de generación de informes con retornos de la inversión (espectaculares).

J - AMBITO DE APLICACIÓN.

A continuación se listan algunos requerimientos típicos resolubles por Sistemas de Data Warehousing en diferentes áreas.

Bancos:

- El cliente X tiene una cuenta de ahorro, un deposito compartido, una hipoteca, etc. ¿Es un cliente beneficioso para el banco?
- Basándose en sus perfiles de inversiones, ¿qué clientes deberían ser focalizados en nuestra próxima campaña?
- ¿Qué tan beneficiosa es una inversión en negocios financieros en la región Z?

Seguros

- Se espera lanzar un paquete completo de seguros. ¿A qué clientes se les debería ofrecer una actualización de sus coberturas actuales?
- Se ha dado a conocer una previsión de cambios en los perfiles de edad de la población. ¿Qué implicaciones tienen estos cambios en las ventas de diferentes tipos de pólizas?

Producción

- El tiempo de orden-a-envió para el producto X ha sido disminuido en un 20%. ¿Esto tiene consecuencias (negativas o positivas) en los tiempos de entrega de otros productos?
- Las ventas del producto A estuvieron un 10% por debajo de lo esperado en este último periodo. ¿Qué efecto tuvo esto en las ganancias?

Ventas.

- Un competidor apareció cerca de mi sucursal A. ¿Cuáles han sido las consecuencias en la marcha de los negocios en la zona para los diferentes tipos de productos?
- Dentro del análisis de una canasta de mercado, especificar que promociones de productos específicos se pueden focalizar a ciertos grupos de clientes.

Servicios públicos.

- ¿De que forma una nueva tarifa telefónica nocturna afectaría la carga de la red y los ingresos globales?
- ¿Cómo se compara el mercado del teléfono celular en comparación al mercado telefónico global?
- Se han implementado nuevas líneas de transporte público colectivo. ¿Cómo ha afectado esto a la movilidad de la población en diferentes zonas? ¿Pueden identificarse grupos de usuarios con comportamiento homogéneo? ¿Ha tenido repercusiones en otros aspectos del tránsito?

Capacitación

- Es vital para una exitosa puesta en funcionamiento de los Sistemas de Data Warehousing, que los usuarios y administradores del sistema tengan conocimiento de las herramientas y su aplicación al negocio en particular.

IV - APLICACIÓN

Entre las funciones que se llevarán a cabo en nuestro proyecto tenemos:

- ✓ Configurar la conexión con el origen de datos del sistema.
- ✓ Iniciar el administrador OLAP.
- ✓ Configurar la base de datos y el origen de datos.
- ✓ Generar el CUBO.
- ✓ Agregar funciones al CUBO.
- ✓ Diseñar el almacenamiento y procesar el CUBO.
- ✓ Examinar los datos del CUBO.

A - ESCENARIO

FoodMart es una cadena de supermercados con presencia en los 50 estados de EE.UU. El departamento de Marketing (comercio) desea analizar todas las promociones de ventas realizadas durante el año 1997.

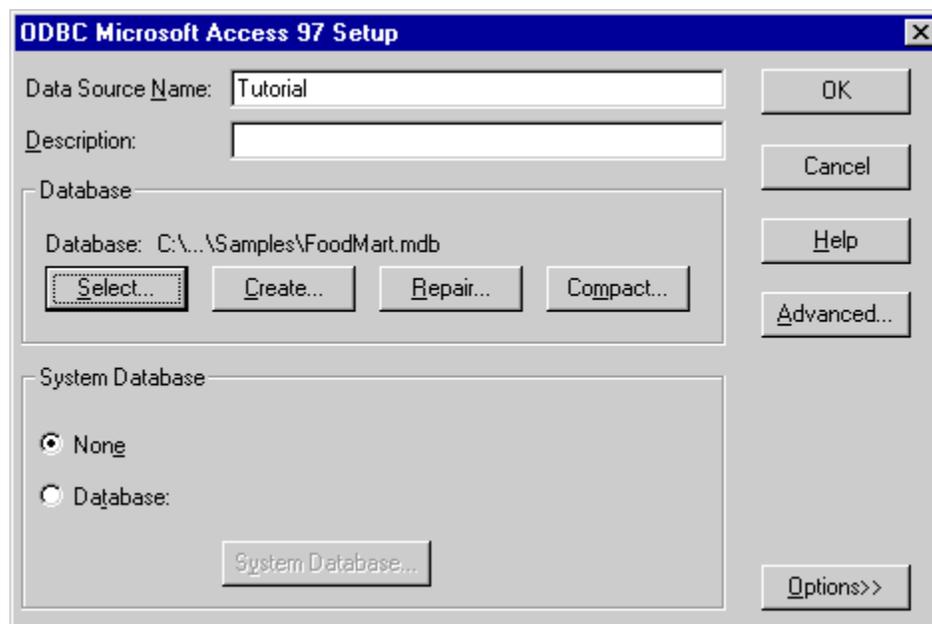
B - GUÍA PRÁCTICA (EJEMPLO)

Como ejemplo construiremos un cubo, configuraremos las conexiones necesarias para tener acceso a los datos, definiremos el cubo y averiguaremos el método más eficaz para analizar los datos.

Antes de comenzar a trabajar con el Administrador de OLAP deberemos configurar las conexiones con el origen de los datos en el Administrador de orígenes de datos ODBC.

- **Para configurar el Nombre de origen de datos (DSN) del sistema**

Aquí llevamos a cabo una serie de pasos en los que se elige entre otras cosas el Controlador del tipo de base de datos que estamos utilizando, en este caso para SQL server establecemos el nombre del origen de datos y seleccionamos la base de datos a utilizar, todo esto a través del Origen de datos ODBC, ubicado en panel de control.



- **Para iniciar el Administrador de OLAP**

El Administrador de OLAP es un complemento que se ejecuta en Microsoft Management Console (MMC).

En el menú Inicio de Windows, seleccione Programas, haga clic en Microsoft SQL Server 7.0, seleccione Servicios OLAP y, finalmente, seleccione Administrador de OLAP.

- **Configurar la base de datos y el origen de datos**

Antes de diseñar el cubo, deberá configurar una estructura de base de datos y conectarla con el origen de datos que configuró anteriormente en el Administrador de orígenes de datos ODBC.

Para configurar la estructura de la base de datos entramos al administrador de OLAP y hacemos la conexión con el servidor luego creamos una nueva base de datos y la conectamos con el origen de datos.



- **Generar el cubo**

Usted es un administrador de bases de datos que trabaja en la compañía FoodMart. FoodMart; con la información almacenada en el almacén de datos de la compañía, construirá una estructura de datos multidimensional (un cubo) para acelerar los tiempos de respuesta cuando los analistas del departamento de comercio consulten la base de datos.

Para abrir el Asistente para cubos. En la vista de árbol del Administrador de OLAP, en la base de datos creada, haga clic con el botón secundario del Mouse en la carpeta Cubos, seleccione Nuevo cubo y, finalmente, escoja Asistente en el menú contextual.

Un cubo consta de medidas (o datos cuantitativos como ventas o costos) y dimensiones (o datos descriptivos del negocio, como regiones geográficas, fechas o información demográfica).

Las medidas se eligen con los datos numéricos que se desean medir. En nuestro caso son las ventas realizadas en el año 1997.

Para darle las dimensiones al cubo se utiliza el asistente de dimensiones y se seleccionan las características que se desean evaluar en el cubo esto se realiza dando clic en el botón nueva dimensión y se elige la tabla que utilizaremos como dimensión si la dimensión es tabla de una sola dimensión o si es tablas de dimensiones múltiples lo seleccionamos según se desee. Además las dimensiones del cubo se deben elegir en orden jerárquico.

Se debe indicar que deseamos compartir la dimensión todo esto lo hacemos con el asistente para dimensiones.

Esto lo debemos de hacer para cada una de las dimensiones que le daremos al cubo.

Una vez que terminemos de seleccionar las dimensiones damos clic en el botón siguiente y escribimos el nombre del cubo y damos clic en finalizar.

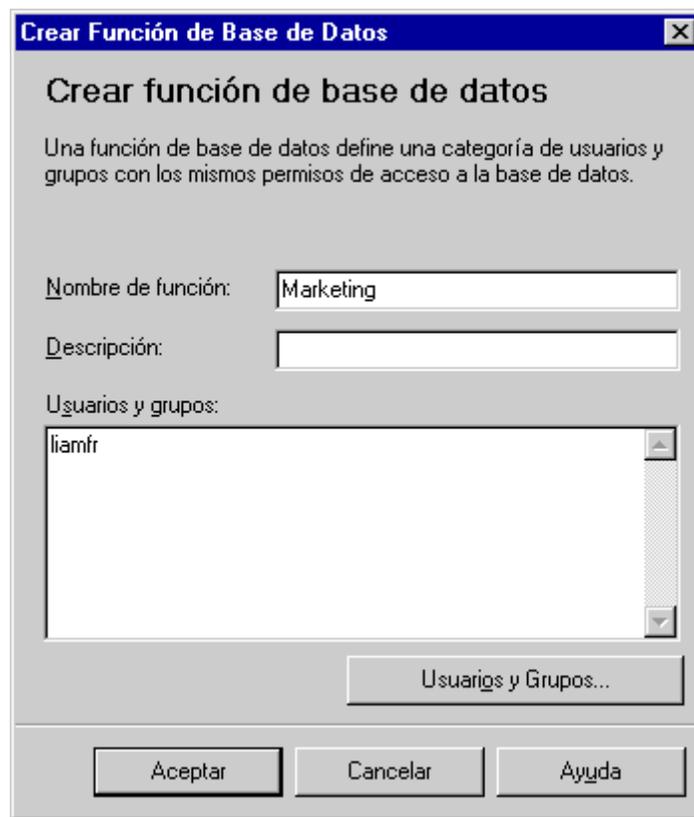


Una vez terminado el cubo, estando en el editor podemos valorar la estructura del cubo en el menú herramientas.

- **Agregar funciones al cubo**

Las funciones del cubo definen qué usuarios o grupos de usuarios podrán tener acceso y consultar los datos contenidos en el cubo. Ahora que el cubo está completamente diseñado, agregará una función al cubo. En este ejemplo creará una función denominada "Comercio".

Para abrir el cuadro de diálogo **Funciones del cubo** En el Editor de cubos, seleccione Administrar funciones en el menú Herramientas. En el cuadro de diálogo Funciones del cubo, haga clic en Nueva función. En el cuadro de diálogo Crear una función de base de datos, escriba "Comercio" en el cuadro Nombre de función. En el cuadro Grupos y usuarios, escriba su nombre de usuario y, a continuación, haga clic en Aceptar.

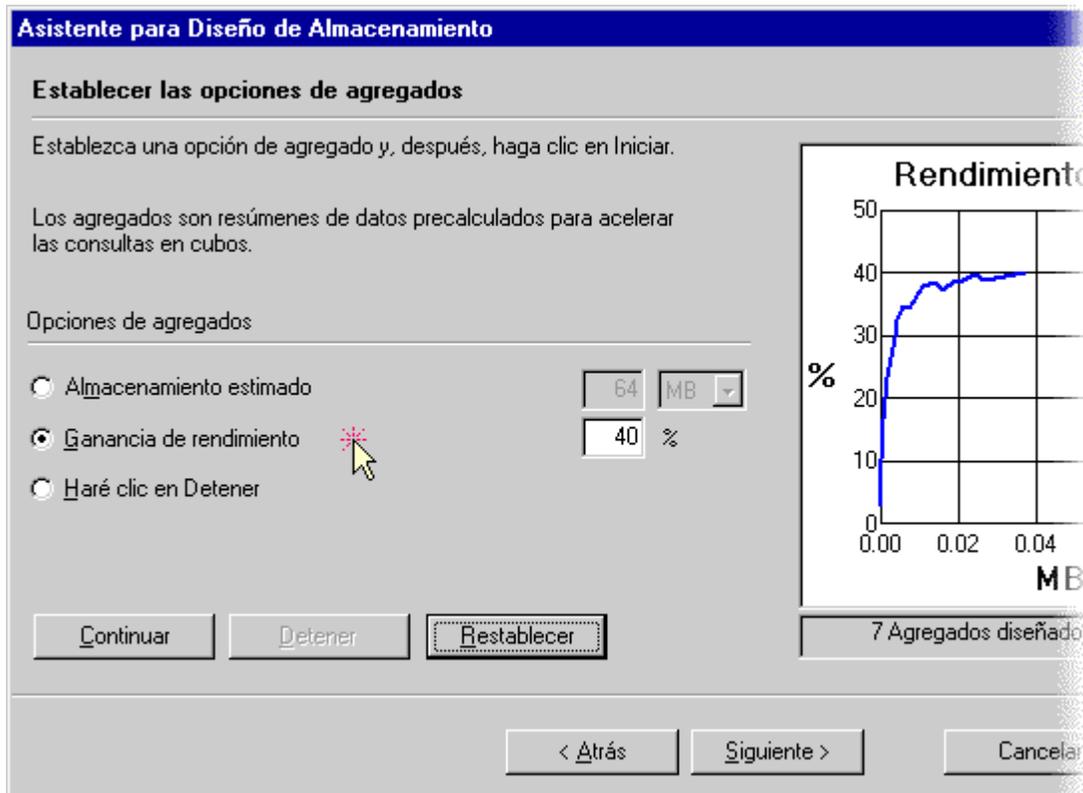


- **Diseñar el almacenamiento y procesar el cubo**

Los Servicios OLAP le permitirán configurar agregados, que mejoran significativamente la eficiencia de las consultas y reducen sus tiempos de respuesta. Utilice el **Asistente para almacenamiento y agregado de datos** para especificar la forma en que los Servicios OLAP almacenarán estos agregados, que optimizarán el rendimiento del proceso de las consultas realizadas al cubo.

Para iniciar el Asistente para almacenamiento y agregado de datos En el Editor de cubos, en el menú Herramientas, seleccione Diseñar almacenamiento. Haga clic en Siguiente. Seleccione MOLAP como el tipo de almacenamiento de datos y, a continuación, haga clic en Siguiente. En Opciones de agregado, seleccione Ganancia de rendimiento. En el cuadro de texto, escriba "40".

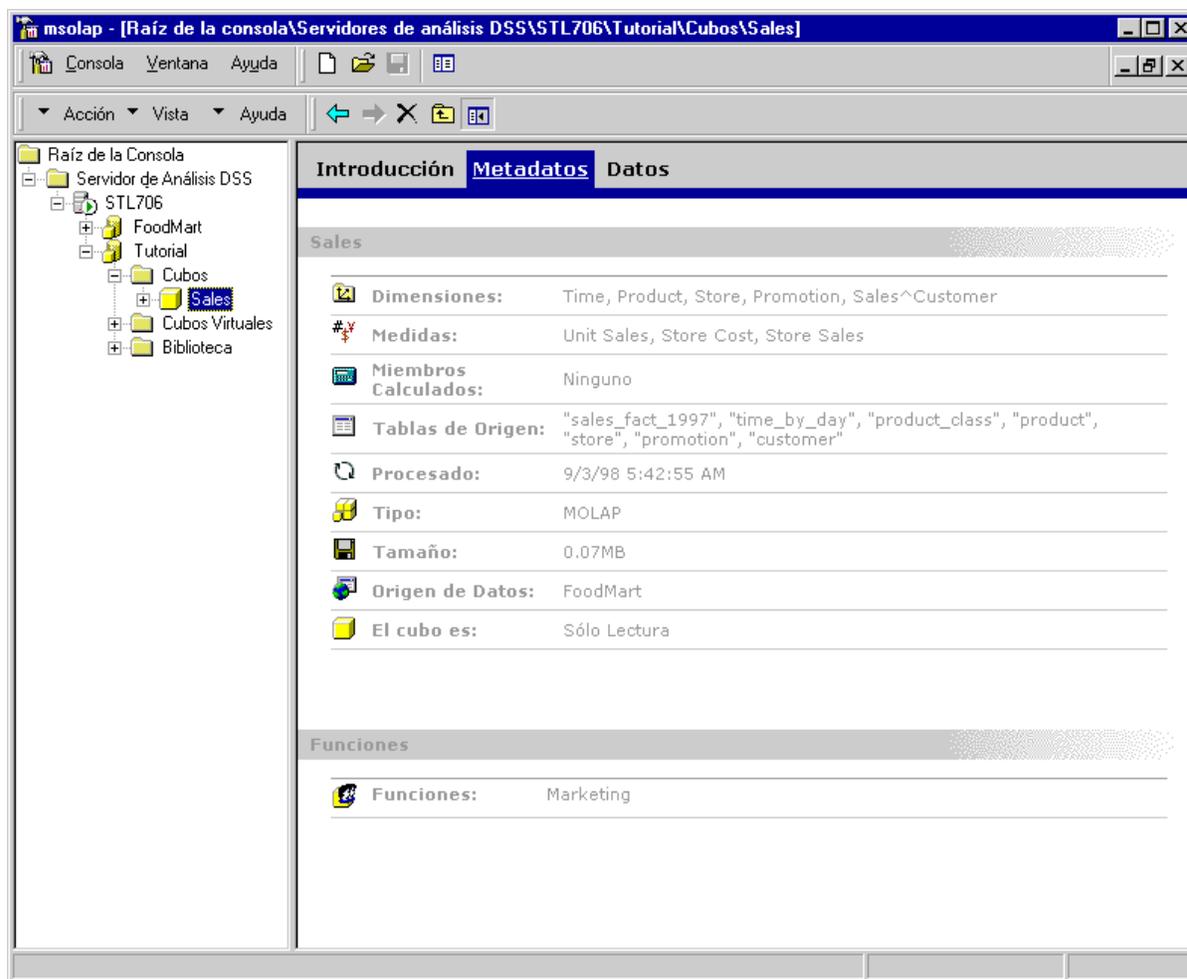
Así, indica a los Servicios OLAP que desea un incremento del 40 por ciento en el rendimiento, independientemente del espacio en disco que se requiera para ello y luego clic en iniciar.



En **¿Qué desea hacer?**, seleccione **Procesar ahora** y, a continuación, haga clic en Finalizar.

- **Presentar los metadatos del cubo**

Los Servicios OLAP le permitirán ver información detallada acerca de la configuración, o los metadatos, del cubo que acabamos de crear. Esta información aparece en el panel de la derecha del Administrador de OLAP.



- **Examinar los datos del cubo**

Ahora está preparado para examinar los datos contenidos en el cubo con el Examinador de cubo, podrá examinar los datos de varias maneras, podrá filtrar la cantidad de datos de dimensiones que son visibles y aumentar el nivel de detalle para ver más (o disminuir el nivel de detalle para ver menos).

Aparecerá el Examinador de cubo, que muestra una cuadrícula formada por una dimensión y las medidas del cubo. Cuatro dimensiones adicionales aparecerán en la parte superior del examinador.



Para sustituir una dimensión de la cuadrícula por otra, arrastre la dimensión desde el cuadro superior y colóquela directamente en la parte superior de la columna con la que desee intercambiarla. Asegúrese de que el puntero adopta la forma de una flecha con dos puntas durante este proceso.

- **Crear una partición para el cubo**

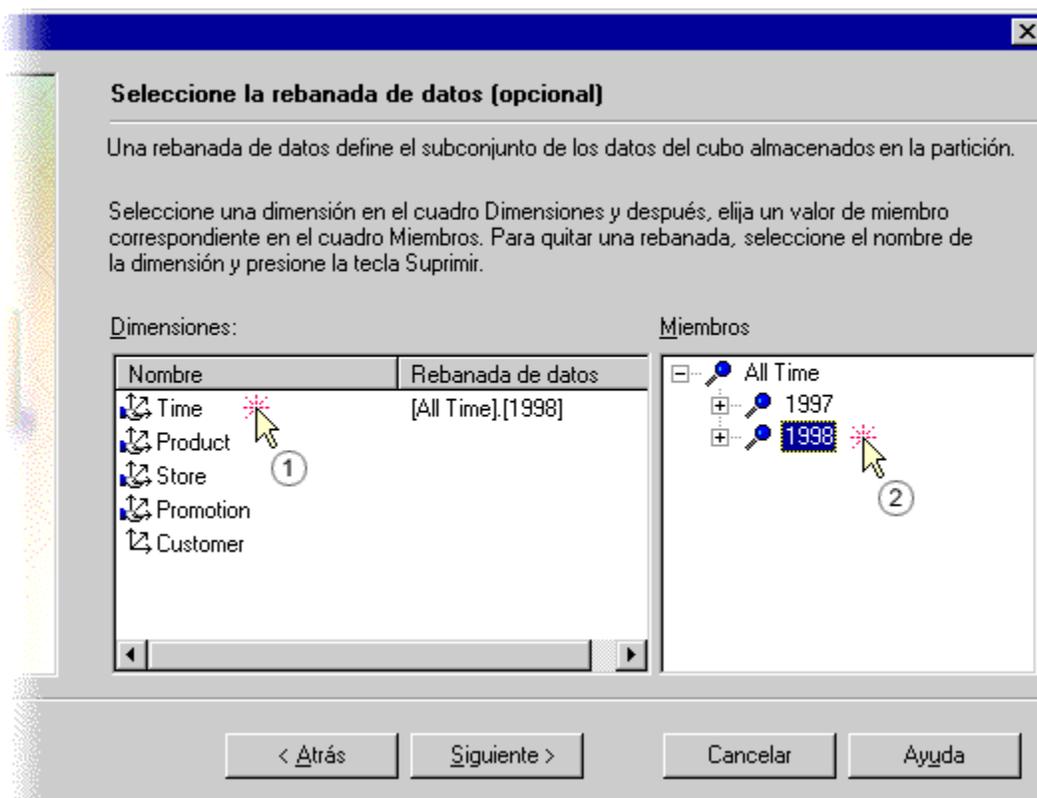
Las particiones definidas por el usuario están disponibles en la edición Enterprise de los Servicios OLAP de Microsoft SQL Server.

Las particiones le ayudan a separar un cubo lógico en varias particiones físicas independientes para mejorar la flexibilidad en el almacenamiento de datos, la ubicación de orígenes de datos y, en definitiva, para mejorar el rendimiento de las consultas.

Para iniciar el Asistente para particiones. En la vista de árbol del Administrador de OLAP, abra la carpeta Cubos en la base de datos. Abra el cubo y Haga clic con el botón secundario del Mouse (ratón) en la carpeta Particiones y, a continuación, seleccione Nueva partición en el menú contextual.

En el paso Especificar el origen de datos y la tabla de hechos, haga clic en Cambiar. En el cuadro de diálogo Seleccionar una tabla de hechos, expanda el origen de datos (base de datos), seleccione la tabla que actuara como tabla de hechos (hechos de ventas de 1998), luego aceptar.

En el paso Seleccionar la rebanada de datos, seleccione la dimensión Time (Fecha) en el cuadro Dimensiones. En la vista de árbol Miembros, expanda el nivel de fecha All (Todas), seleccione 1998 y, a continuación, haga clic en Siguiente.



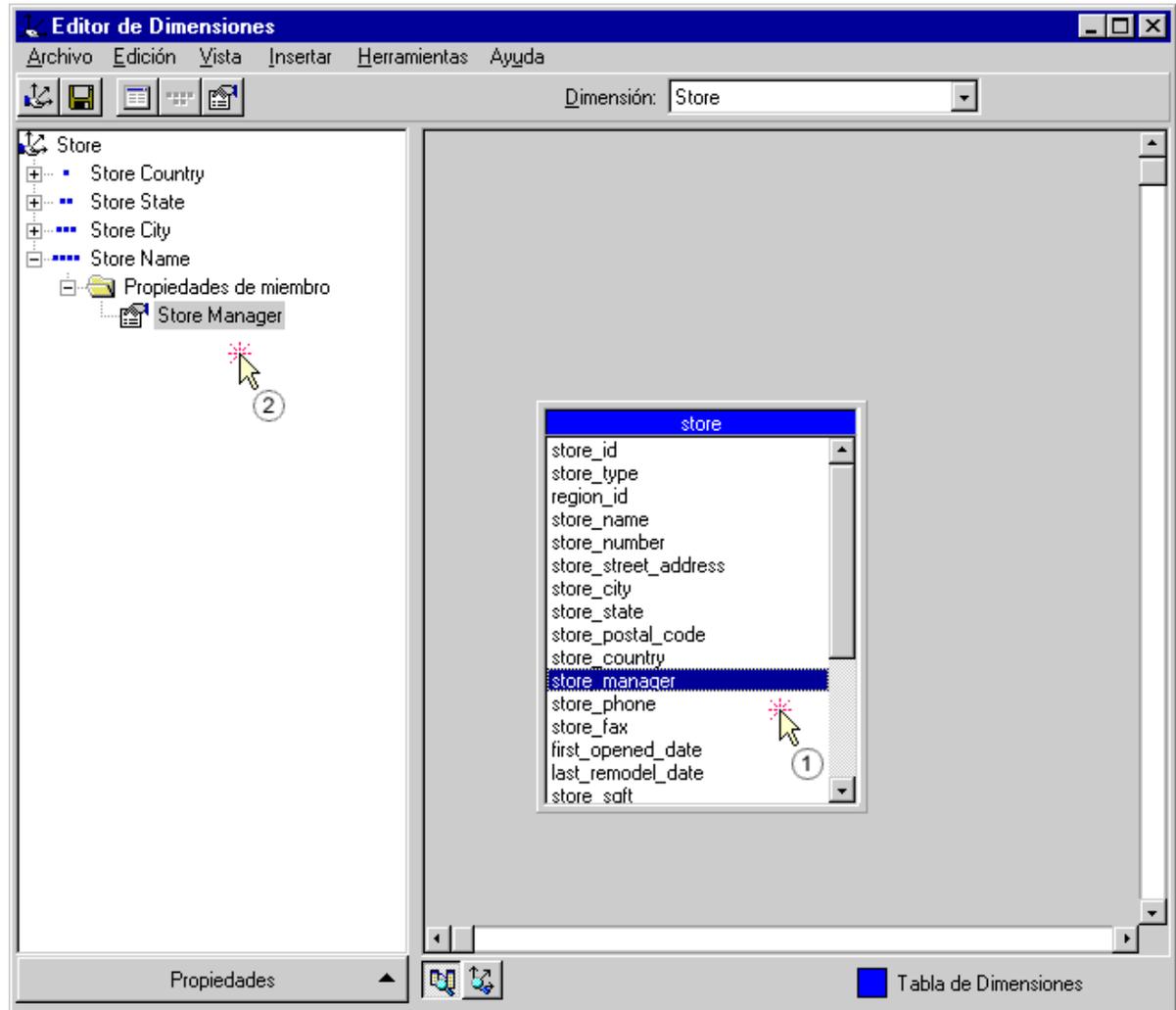
Luego Finalizar el Asistente para particiones, en el cuadro Nombre de partición, escriba el nombre que desee que lleve la partición. En ¿Qué desea hacer?, seleccione Copiar el diseño de los agregados de una partición existente. Active la casilla de verificación Procesar la partición al terminar y luego finalizar.

- **Crear propiedades de miembros y dimensiones virtuales**

Las propiedades de miembros le permitirán definir información adicional acerca de los niveles de una dimensión. Con esta propiedad de miembro, creará posteriormente una

dimensión virtual. La dimensión virtual mostrará información del administrador del almacén cuando usted vea los datos del cubo.

Para crear una propiedad de miembro En la vista de árbol del Administrador de OLAP, expanda la carpeta Biblioteca. Expanda la carpeta **Dimensiones compartidas** luego haga clic con el botón secundario del Mouse en la dimensión que desee modificar y seleccione **Modificar** en el menú contextual. En el editor de dimensiones expanda nombre de la dimensión que esta tratando y podrá ver las propiedades correspondientes. En el panel **Esquema**, arrastre la columna gerente _ almacén de la tabla Almacén a la carpeta **Propiedades de miembro** de **Nombre de almacén**.



Guarde y luego la opción salir.

Para crear una dimensión virtual. En la vista de árbol del Administrador de OLAP, haga clic con el botón secundario del Mouse en la carpeta Dimensiones virtuales y, a continuación, haga clic en **Nueva dimensión virtual** en el menú contextual. En el paso **Seleccione la propiedad de miembro para la dimensión virtual**, expanda la dimensión que desee. Haga clic en la propiedad de miembro nombre de la dimensión. Luego damos clic en siguiente y antes de finalizar escribimos el nombre de la dimensión virtual.

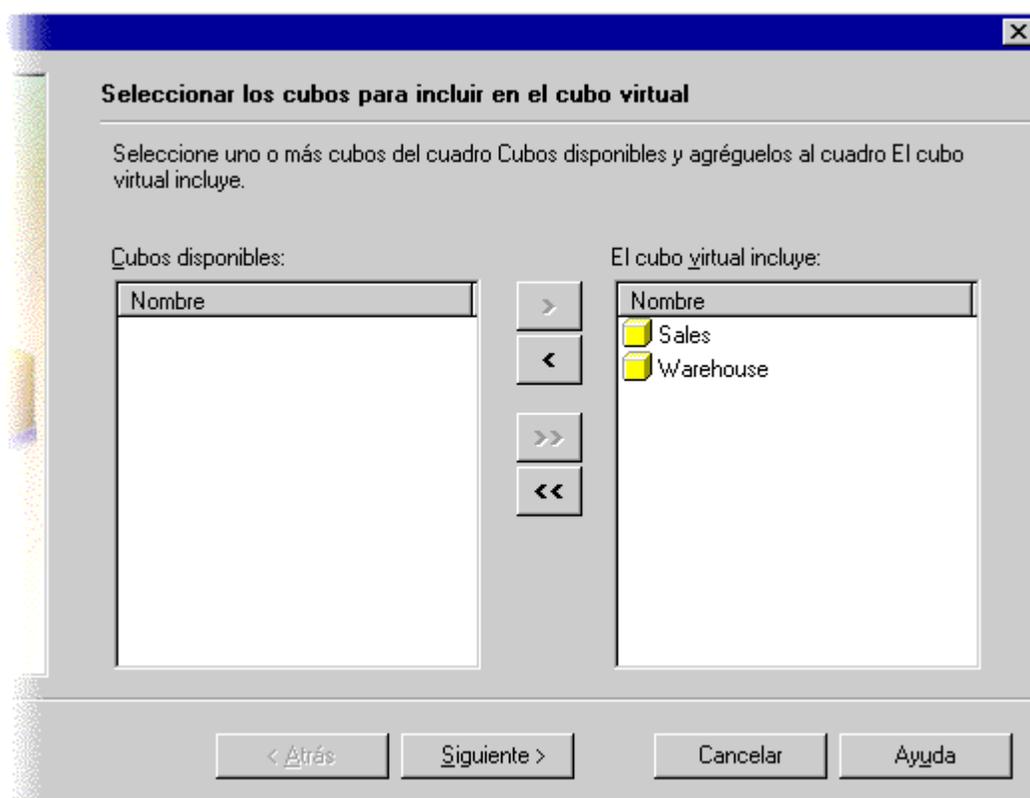
Nota: podrá agregar dimensiones virtuales a un cubo si utiliza el paso Seleccionar dimensiones del Asistente para cubos o el Administrador de dimensiones del Editor de cubos.

- **Crear y procesar un cubo virtual**

Una de las áreas principales donde se pueden reducir los costos de FoodMart son las operaciones de almacén. Para realizar esta operación, creará un cubo virtual.

Un cubo virtual le ayuda a crear una vista más amplia de los datos multidimensionales sin necesidad de utilizar espacio físico de almacenamiento adicional en la unidad de disco.

Para iniciar el Asistente para cubo virtual. En la vista de árbol del Administrador de OLAP, expanda la base de datos FoodMart. Haga clic con el botón secundario del Mouse en la carpeta Cubos virtuales y, a continuación, haga clic en Nuevo cubo virtual en el menú contextual, Para generar el cubo virtual : En el paso Seleccionar los cubos para incluir en el cubo virtual, haga doble clic en el cubo que creo anteriormente en el cuadro de la izquierda Cubos disponibles. Se moverá al cuadro Cubos virtuales incluidos que se encuentra a la derecha.



En el paso Seleccionar las medidas para el cubo virtual, haga doble clic en la medida Ventas de almacén en el cuadro Medidas disponibles de la izquierda. Repita este proceso para las medidas que desee agregar.

En el paso Seleccionar las dimensiones para el cubo virtual, haga doble clic en la dimensión tiempo (Fecha) en el cuadro Dimensiones disponibles de la izquierda. Escriba el nombre del cubo virtual y haga clic en finalizar. A continuación procesar el cubo.

Los analistas de FoodMart podrán ahora consultar el cubo virtual

- **Examinar los datos del cubo virtual**

Ahora podrá examinar los datos contenidos en el cubo virtual que creo.

Para ver los datos en el cubo virtual expanda cubos virtuales y seleccione el cubo que creo y luego explorar.

The screenshot shows a window titled "Examinador de cubo - Cost versus Sales". At the top, there are two dropdown menus: "Time" set to "1997" and "Warehouse" set to "All Warehouses". Below these is a table with three columns: "Store Country", "Store Sales", and "Warehouse Cost". The table has four rows: "All Stores", "Canada", "Mexico", and "USA". The "All Stores" row shows "Store Sales" as 565,238.13 and "Warehouse Cost" as 89,043.25. The "USA" row also shows "Store Sales" as 565,238.13 and "Warehouse Cost" as 89,043.25. The "Canada" and "Mexico" rows are currently empty. At the bottom of the window, there is a text instruction: "Haga doble clic en un miembro para elegir el nivel de detalle." and two buttons: "Cerrar" and "Ayuda".

| | MeasuresLevel | |
|-----------------|---------------|----------------|
| + Store Country | Store Sales | Warehouse Cost |
| All Stores | 565,238.13 | 89,043.25 |
| + Canada | | |
| + Mexico | | |
| + USA | 565,238.13 | 89,043.25 |

V - CONCLUSIONES

- El Datawarehouse es un tema muy complejo y extenso, pero durante su estudio, se ha podido comprender paso a paso sus utilidades y funcionamientos.
- Una de las principales ventajas que se observa en el uso del datawarehouse es su capacidad para ocultar la complejidad de la base de datos al usuario.
- Se comprobó que la agregación previa de los datos, permite que ante una consulta, el tiempo de respuesta sea corto.
- Hemos observado claramente en el transcurso de la realización del presente trabajo, que la construcción del Datawarehouse es un proceso arduo que puede tomar mucho tiempo.
- Como sabemos las herramientas de hoy en día hacen fácil la implementación, pero como siempre, si no existe un buen análisis de requerimientos el proyecto no saldrá a flote.

VI - RECOMENDACIONES

A los posibles usuarios se les recomienda:

- Mantener una buena organización de la información de la empresa.
- Aplicar técnicas de análisis para hacer una mejor explotación del datawarehouse.
- Basar el diseño del datawarehouse en base a una plataforma adecuada. Tomar en cuenta el Sistema Operativo, la Gestión de Base de Datos y las utilidades para analizar y generar reportes.

VII - BIBLIOGRAFIA

1. System Service Corporation. An Introduction to datawarehouse by Vivek R. Gupta.

2. Web:

<http://mailweb.udlap.mx/~is109391/tesis/propuestaformal.html>

www.monografias.com

www.datawarehouse.com

<http://www.dwinforcenter.org>

<http://www.dw-institute.com>