

UNIVERSIDAD DEL BIOBIO
FACULTAD DE CIENCIAS EMPRESARIALES
DEPARTAMENTO DE AUDITORÍA E INFORMÁTICA



**“DESAROLLO DE INFORMES CONTABLES A
TRAVÉS DE TECNOLOGÍA OLAP”**

CECILIA ERNESTINA URREA GUTIÉRREZ

Memoria para optar al título de
Ingeniero de Ejecución en Computación e Informática

CHILLAN 2006

UNIVERSIDAD DEL BIOBIO
FACULTAD DE CIENCIAS EMPRESARIALES
DEPARTAMENTO DE AUDITORÍA E INFORMÁTICA



**“DESAROLLO DE INFORMES CONTABLES A
TRAVÉS DE TECNOLOGÍA OLAP”**

CECILIA ERNESTINA URREA GUTIÉRREZ

PROFESOR GUÍA : SR. HÉCTOR VILLAR RODRÍGUEZ
PROFESOR INFORMANTE : SRA. MARCELA PINTO FERNÁNDEZ
NOTA FINAL EXAMEN TÍTULO _____

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO DE
EJECUCIÓN EN COMPUTACIÓN E INFORMÁTICA**

CHILLAN 2006

Resumen

Gran parte de los productos generados por tecnologías de información no es precisamente información lo que entregan, sino datos en bruto, los que 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 añaden ideas y capacidad analítica.

En esta memoria se detallan las motivaciones y necesidades de contar con un sistema de apoyo a las decisiones y se justifica el uso del modelo utilizado. Además de sugerir mejoras a los problemas encontrados en la empresa. Por otra parte se analizan las opciones existentes para transformar un conjunto de datos en conocimiento en base a los recursos con los que cuenta ésta cuenta.

La solución propuesta se basa en el desarrollo de base de datos OLAP, sobre la cual se desarrollaron dos cubos orientados a solucionar los problemas de cálculo que presenta actualmente la empresa, para lo cual se utilizó SQL Server Analysis Service 2000.

La alta capacidad analítica de Analysis Service permitió detectar una serie de anomalías a través de la emisión de un resumen macro de la información que se había ingresado durante años en los sistemas, por otra parte se redujo los tiempos de procesamiento considerablemente, además de mostrar la información desde nuevas perspectivas, al encontrar patrones que con informes simples no había sido posible detectar.

Las estructuras multidimensionales ofrecen importantes aportes para el análisis de la información pero antes de llevar a cabo una implementación de éste tipo se aconseja hacer un estudio de si es realmente la solución que necesita la organización.

Agradecimientos

A mi mentor, Don Diego de la Sotta S., por la paciencia y tiempo dedicado, por su capacidad motivadora, incondicional apoyo y confianza principalmente en los momentos difíciles.

Introducción

El alto grado de competencia del mercado revela una necesidad clara de estar continuamente mejorando los procesos del negocio, ya que éstos se consideran como el pilar de éxito de una organización. El ser esta capaz de reaccionar rápido a los continuos cambios y salir exitosamente de los problemas y retos hace denominarlas como "Empresas inteligentes".

La utilización de tecnologías de información permite una integración funcional departamental de los datos y las aplicaciones para que las organizaciones realicen una mejor toma de decisiones y planeación a través de la mejora de sus procesos y capacidades competitivas.

Las organizaciones inteligentes para tomar decisiones se basan en la información que le entregan sus sistemas de información para planear el futuro y asignar los recursos eficientemente. Hasta hace algunos años, la información había sido entregada en reportes impresos en papel, los cuales eran generados de manera centralizada por el departamento de tecnología de información. En la actualidad cada vez más empleados calificados necesitan datos en otros formatos, más fáciles de usar, y que les puedan ofrecer un mayor nivel de detalle y flexibilidad.

Hoy en día toda empresa para la toma de decisiones necesita depositar mucha confianza en la información (hechos y cifras) que le entregan los sistemas con los que cuenta, sobre todo si se tiene en cuenta que la competencia crece en todo momento y las decisiones se deben tomar, en muchos casos de forma acelerada. El problema generalmente surge cuando se tiene una montaña de información por analizar y poco tiempo para entregar una respuesta.

Ante estos casos se requiere de herramientas que permitan minimizar los tiempos de análisis de elevados volúmenes de información con mayor velocidad y mejorar la precisión de los informes, objetivos que permitirán mantener la competitividad, de modo de estar atentos y saber reaccionar ante los cambios del mercado.

Esto nos muestra que las empresas invierten en tecnología y soluciones con las cuales se mantienen vigentes en un mercado cambiante, ahora las empresas no dependen tan solo de factores como ubicación, productos, etc., sino también del conocimiento. Tal conocimiento basado en información comprensible, detallada y relevante es crucial para lograr y sostener ventajas competitivas. El poseer conocimientos correctos significa tener respuestas correctas y realizar decisiones estratégicas para la ejecución de la empresa. Pero las tareas de recolectar, procesar, limpiar y transformar la información necesaria para la toma de decisiones no es una tarea sencilla más si se considera que una empresa tiene distintas áreas.

La finalidad de ésta tesis es buscar mejoras en la calidad de la información que se les entrega a las empresas que conforman el grupo de Forestales Regionales S.A. Éstas actualmente cuentan con un sistema que genera un listado de datos sobre los cuales se toman decisiones, pero ¿Qué tan efectivos y certeros son éstos?, ¿Realmente es lo que necesitan los directivos para hacer sus evaluaciones y analizar la situación de la empresa?, ¿Los tiempos que involucran estos análisis podrían ser reducidos para agilizar los procesos?

Estas interrogantes son algunos de los puntos que se pretende atacar en esta memoria, buscando nuevas metodologías con la finalidad de transformar el alto volumen de datos en conocimiento, a través de informes resumidos, ordenados y flexibles, que permitan ver la situación de la empresa desde nuevas perspectivas.

Entre los puntos a tratar en este trabajo se encuentran un estudio sobre las opciones existentes en el mercado orientas a dar solución a los problemas planteados anteriormente, el cómo se abordarán éstos y la metodología utilizada para solucionarlos, además de instruir sobre la herramienta seleccionada y dar observaciones sobre los resultados obtenidos versus los requerimientos planteados por la organización.

INDICE

CAPÍTULO 1 - MARCO TEÓRICO	10
1.1 NEGOCIO INTELIGENTE O BUSSINES INTELLIGENCE.....	10
1.1.1 <i>Introducción a los almacenes de datos</i>	10
1.1.2 <i>Bodega de datos o Data warehouse (DW)</i>	11
1.1.2.1 Definición.....	11
1.1.2.2 ¿Qué es una bodega de datos?	12
1.1.2.3 Características del Data warehouse o bodega de datos	13
1.1.2.3.1 Orientado a temas.....	14
1.1.2.3.2 Integrado	14
1.1.2.3.3 Variable en el tiempo	15
1.1.2.3.4 No volátil.....	15
1.1.2.4 Características generales de una bodega de datos.....	16
1.1.2.5 ¿Cómo se puede conformar la bodega de datos?	16
1.1.2.6 ¿Qué pueden ofrecer las bodegas de datos?.....	17
1.1.2.7 ¿Cuál es el retorno de la inversión?	18
1.1.4 <i>Data Mining</i>	19
1.1.4.1 Objetivos	19
1.1.4.2 El Alcance de Data Mining.....	20
1.1.5 <i>Data Mart</i>	21
1.1.6 <i>OLAP</i>	22
1.1.7 <i>¿Qué es OLAP?</i>	23
1.1.7.1 ¿Porque utilizar OLAP?.....	24
1.1.7.2 Beneficios que trae la utilización de OLAP.....	25
1.1.7.3 Consideraciones que se deben tener para implantaciones OLAP	26
1.1.8 <i>Uso de herramientas OLAP</i>	27
1.1.8.1 Modos de Almacenamiento	27
1.1.9 <i>Arquitectura de cubos</i>	29
1.1.10 <i>Modelado dimensional</i>	30
1.1.10.1 Esquema de estrella	30
1.1.10.2 Esquema copo de nieve	31
1.1.11 <i>Dimensiones</i>	31
1.1.11.1 ¿Qué son las dimensiones?	31
1.1.11.2 Niveles.....	33
1.1.11.3 Jerarquías.....	34
1.1.11.4 Medidas	35
1.1.11.5 Miembros calculados	35

1.1.11.6 Partición.....	36
1.1.11.7 Agregaciones	36
1.1.11.8 Hechos	36
<i>1.1.12 Operaciones analíticas básicas</i>	<i>37</i>
1.1.12.1 Drill down.....	37
1.1.12.2 Drill Up.....	37
1.1.12.3 Drill Through.....	37
1.1.12.4 Slicing and dicing	37
1.1.12.4.1 Tajadas	37
1.1.12.4.2 Slicing o rebanar.....	38
1.1.12.4.3 Dicing o Cortar en cubitos.....	39
1.1.12.5 Rotación.....	40
<i>1.1.13 Lenguaje MDX.....</i>	<i>40</i>
1.1.13.3.1 Consultas MDX	40
1.1.13.3.2 Expresiones de cálculo	41
1.1.13.4 Sintaxis básica de una consulta.....	41
CAPÍTULO 2 – DESCRIPCIÓN DE LA EMPRESA	44
1.1 INTRODUCCIÓN	44
<i>1.1.1 Historia de la empresa</i>	<i>44</i>
2.2.2 Visión de la empresa.....	45
2.2.3 Organigrama de TDIFOR	46
2.2.4 Situación de la empresa.....	46
2.3 CARACTERÍSTICAS DEL SISTEMA CONTABLE	47
2.3.1 El libro diario	48
2.3.2 Libro Mayor.....	48
2.3.3 Libros de Balances	48
2.3.4 Cuentas de pérdidas y ganancias	49
2.3.5 Problemas Detectados en el sistema contable.....	53
2.4 ANÁLISIS DE REQUERIMIENTOS.....	54
2.5 OBJETIVOS DEL PROYECTO	54
2.3 CARACTERÍSTICAS GENERALES DE LA RED	55
2.3.1 Características técnicas de los servidores.....	57
CAPÍTULO 3 - REQUERIMIENTOS.....	60
4.1 INTRODUCCION	60
4.2 REQUERIMIENTOS.....	60
4.2.1 Requerimientos funcionales.....	60
4.2.2 Requerimientos no funcionales.....	61

CAPÍTULO 4 – SOLUCIÓN PROPUESTA	64
4.1 INTRODUCCIÓN	64
4.1.1 Necesidad de información y conocimiento en la empresa	64
4.1.2 Tipo de información que necesitan las empresas	66
4.2 Propuesta de desarrollo	66
4.3 ANÁLISIS DEL PROBLEMA	69
4.5 SOLUCIÓN PROPUESTA	69
4.6 TECNOLOGÍA A UTILIZAR	71
4.7 MODALIDAD DE CONEXIÓN	72
4.8 SUGERENCIAS AL SISTEMA ACTUAL.	73
4.8.1 Normalización de las tablas de comprobante contable.	73
4.8.1.2 Campos de tabla	76
4.8.1.3 Definición de los atributos.....	77
CAPÍTULO 5 – PRUEBAS DEL SISTEMA	79
5.1 INTRODUCCIÓN	79
5.1.1 Pruebas de integridad de la información	79
5.1.2 Pruebas de rendimiento.....	83
CAPÍTULO 6 – CONCLUSIONES	87
6.1 INTRODUCCIÓN	87
6.2 VENTAJAS Y DESVENTAJAS DE LA HERRAMIENTA SQL SERVER ANALYSIS SERVICE (SSAS)	87
6.3 RECOMENDACIONES.....	88
6.4 RESULTADOS OBTENIDOS.....	90
6.5 OBJETIVOS FUTUROS	91
ANEXO 1 – CREACIÓN DEL CUBO OLAP CUENTAS CORRIENTES.....	93
1.1 INTRODUCCIÓN	93
1.1.1 Configuración del origen de datos.	93
1.1.2 Creación de las dimensiones del cubo	96
1.1.2.1 Creación de la dimensión “Cuentas corrientes”.....	97
1.1.2.2 Creación de la dimensión “Fecha”.....	103
1.1.2.3 Creación de la dimensión “Rut”	106
1.1.3 Creación del cubo “Cuentas corrientes”.....	109
1.1.4 Proceso de creación de los miembros calculados.....	114
ANEXO 2 – CREACIÓN DEL CUBO ANÁLISIS CUENTA.	119
1.1 INTRODUCCIÓN	119
1.1.1 Creación de las dimensiones del Cubo.....	119

1.1.1.1	Configuración del origen de datos.....	119
1.1.1.2	Creación de las dimensiones de Análisis.....	122
1.1.1.3	Creación de las dimensiones de Valor.....	129
1.1.1.4	Creación de la dimensión Fecha.....	135
1.1.1.5	Creación de la dimensión Cuenta	139
1.1.2	<i>Creación del cubo “Análisis Cuenta”</i>	<i>145</i>
1.1.2.1	Creación del miembro calculado “Saldo débito”.....	152
1.1.2.2	Creación del miembro calculado “Saldo crédito”	153
ANEXO 3 - GLOSARIO		156
BIBLIOGRAFÍA.....		159

Capítulo 1

Marco teórico

Capítulo 1 - Marco teórico

1.1 Negocio inteligente o Bussines Intelligence

El tener respuestas correctas para tomar decisiones estratégicas requiere de un alto grado de conocimiento de la situación actual de la empresa, pero las tareas de recolectar, procesar, limpiar y transformar la información necesaria para brindar apoyo no son nada sencillas, más aún si se considera que una empresa tiene distintas áreas que a veces se encuentran alejadas de los ejecutivos del negocio.

Lo anterior ha llevado a las empresas a invertir en tecnologías de información y a que transformen las montañas de información que poseen en conocimiento, lo que puede verse representado en información comprensible, detallada, relevante y crucial para lograr una ventaja competitiva frente a la competencia.

En el siguiente capítulo se hablará de las distintas soluciones existentes en el mercado para dar apoyo a la toma de decisiones y en qué consiste cada una de éstas.

1.1.1 Introducción a los almacenes de datos

El objetivo de un almacén de datos es consolidar y organizar estos datos para poder analizarlos y utilizarlos con el fin de apoyar las decisiones empresariales.

Los almacenes de datos normalmente contienen información histórica, generalmente recopilados a partir de varios orígenes distintos como sistemas de procesos de transacciones en línea (OLTP, Online Transaction Processing), sistemas heredados, hojas de cálculo o archivos de texto. Un almacén de datos combina estos datos, los limpia para obtener precisión, coherencia, además los organiza para facilitar la realización de consultas y mejorar su eficacia.

1.1.2 Bodega de datos o Data warehouse (DW)

"Es un conjunto de datos integrados o orientados a una materia que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de la administración y está orientada al manejo de grandes volúmenes de datos provenientes de diversas fuentes o diversos tipos.

Estos datos cubren largos períodos de tiempo lo que trae consigo que se tengan diferentes esquemas de los datos fuentes, la concentración de esta información esta orientada a su análisis para apoyar la toma de decisiones oportunas y fundamentadas, previo a su utilización se debe aplicar procesos de análisis, selección y transferencia de los datos seleccionados." (VALLE,2005) ¹

Los sistemas de bases de datos han representado una gran herramienta en la administración y gerencia, llegando a transformarse en la solución ideal para almacenar la información.

En la actualidad los motores de bases de datos, son herramientas de software muy sofisticadas que han ocultado al usuario su complejidad y potencializado su implantación.

Considerando lo anterior los sistemas de información deben dejar de ser estructuras estáticas de procesamiento y almacenamiento, para convertirse en robustas bases de datos que integren y analicen la información del negocio, para posteriormente transformarse en estructuras globales dinámicas, variantes en el tiempo y con comportamientos propios. Tales características se pueden encontrar si se diseña un nivel superior de bases de datos llamadas Bodegas de datos o Data warehouse (DW).

1.1.2.1 Definición

Los data warehouse se pueden definir como: "Una gran base de datos, que habilita espacial e históricamente la información del negocio, desde una perspectiva holística orientada a las

¹ Valle J., Bodega de datos, (2005).

áreas del negocio, no a las aplicaciones transaccionales diarias, con la finalidad de servir de apoyo a la toma de decisiones". (BOHORQUEZ,1999)²

"Es un proceso, no un producto. Es una técnica para consolidar y administrar datos de variadas fuentes, con el propósito de responder a preguntas de negocios y tomar decisiones, de una forma que no era posible." (FERNÁNDEZ,2000)³

"Manejar grandes volúmenes de datos de una forma que no era posible, o no era costo efectiva." (ALMONACID,2000)⁴

1.1.2.2 ¿Qué es una bodega de datos?

"Una bodega de datos es una colección de datos orientados a temas, integrados, no volátiles y variantes en el tiempo, organizados para soportar necesidades empresariales." (INMON,1992)⁵

"Es 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." (OSTERFELDT,1993)⁶

Otros lo consideran como una base de datos que:

- Es organizada como servidor central de almacenamiento de los datos.
- Es usada para ingeniería de datos (Data Mining) y otras aplicaciones.

² Bohorquez J., Aproximación metodológica de un Spatial Data Warehouse, (1999).

³ Fernández A., Data Warehouse (2000).

⁴ Almonacid F., Data Warehouse (2000).

⁵ Inmon H., La Tecnología Datawarehousing (1992).

⁶ Osterfeldt M., Bodega de Datos (1993).

- Reúne un específico juego de requerimientos del negocio.
- Usa los datos que agrupan un predefinido juego de criterios del negocio.

La siguiente figura muestra una representación gráfica del funcionamiento de un almacén de datos.

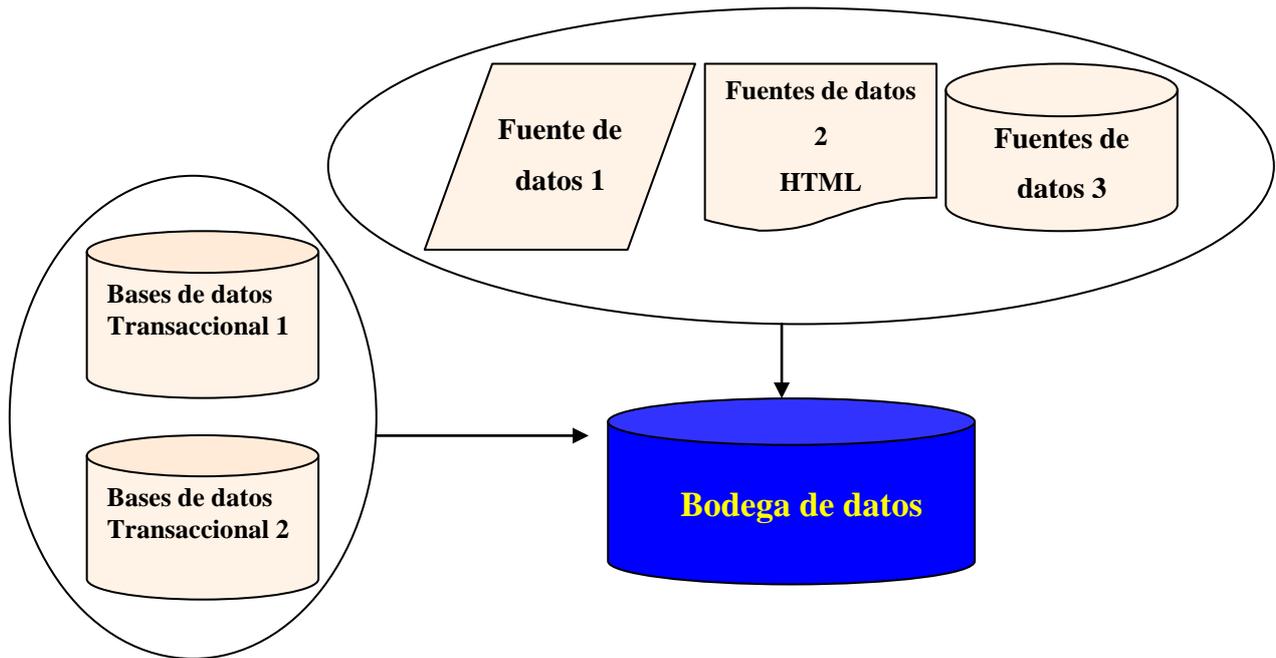


Figura 1.1 Representación gráfica de un almacén de datos.

1.1.2.3 Características del Data warehouse o bodega de datos

Una bodega de datos es el "corazón" de la arquitectura de un sistema de soporte de decisiones (DSS). Es la parte fundamental del funcionamiento de un DSS, dado que es una única fuente integrada de datos que son realmente accesibles permitiendo a los analistas trabajar en un ambiente mucho más fácil.

El sistema de soporte de decisiones provee los mecanismos de acceso a los datos y el análisis de los mismos. El DSS es el "cerebro" de la arquitectura, el motor que traducirá los requerimientos de los usuarios en las correspondientes sentencias de consulta para la bodega de datos y el que interpretará los resultados devueltos por éste para mostrarlos según lo solicitado por el usuario. Realmente es el componente que concentra la inteligencia del sistema, pues diferentes DSS pueden cambiar notablemente la explotación y el comportamiento de una bodega de datos.

Un sistema de soporte de decisiones basado en técnicas de bodegas de datos , transforma los datos que tiene una organización en información, con el objetivo de tomar mejores decisiones de negocio basados en un análisis dimensional.

1.1.2.3.1 Orientado a temas

Los datos son categorizados y almacenados por temas del negocio y no por la aplicación, es decir, la información es organizada, almacenada y analizada por áreas, y no por los análisis que sean el resumen de información de distintos enfoques del negocio, la idea es poder mirar los datos de manera abstracta, agrupados por temas del negocio y a partir de ellos deducir el comportamiento de la empresa para un determinado período.

En una aplicación normal la información es organizada de tal manera de poder extraer de forma directa los datos, en cambio una bodega de datos tiene una base de datos diseñada, estructurada y organizada por áreas, independiente de las aplicaciones.

1.1.2.3.2 Integrado

Los datos son almacenados una sola vez de acuerdo al área temática a la que pertenecen, como en el ámbito de las bodegas de datos no se realiza un diseño de acuerdo a las aplicaciones, si no de acuerdo a las áreas del negocio, los datos son integrados y estructurados, lo que permite garantizar un juego de información estándar, consistente, exacta y confiable para las aplicaciones, procesos y análisis del negocio.

1.1.2.3.3 Variable en el tiempo

Una de las ventajas es que los datos se almacenan con sus respectivas fechas de inserción, lo que garantiza poder desarrollar análisis de éstos en una serie de instancias representadas en un período de tiempo.

El almacenar información de manera histórica es lo que permite desarrollar pronósticos y análisis de patrones en forma estadística.

Lo anteriormente mencionado se puede graficar de la siguiente forma.

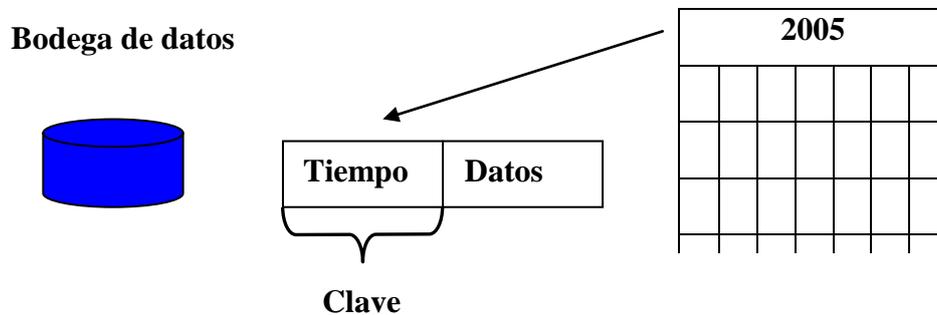


Figura 1.2: Muestra la forma en que almacena los datos en un data warehouse

1.1.2.3.4 No volátil

Generalmente en estas estructuras los datos no son actualizados desde los sistemas, sino que son insertados en nuevos registros, para no sobre escribir los existentes. En otras palabras en este tipo de bases de datos no hay ni actualizaciones de registros, ni borrado, solamente consultas e inserción.

En la siguiente figura se muestra un esquema del funcionamiento de una bodega de datos.

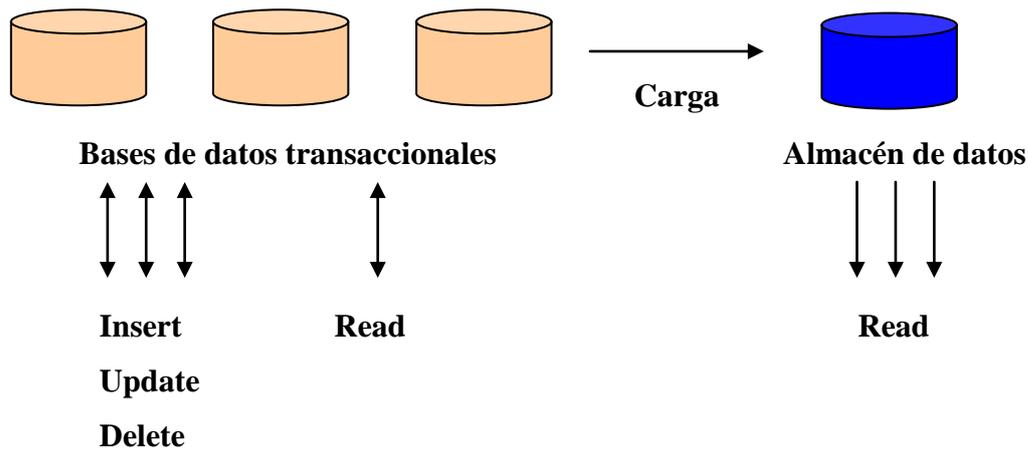


Figura 1.3: Esquema de funcionamiento de un almacén de datos

1.1.2.4 Características generales de una bodega de datos

- **Sistema operativo:** La bodega de datos se encuentra sobre la plataforma del sistema operativo. La seguridad representada en la disponibilidad, confidencialidad y controles de accesos y privilegios sobre las áreas de almacenamiento y procesamiento son en gran medida dependientes de la plataforma.
- **Red:** Es la infraestructura de comunicación que permite que las diferentes componentes intercambien información. La cantidad de datos contenidos en un data warehouse incrementa su importancia.
- **Riesgos:** Entre los posibles riesgos se encuentran los accesos al sistema desde elementos externos sin autorización (aplicaciones, personas, etc.), que la red se convierta en un cuello de botella para la operación del sistema y la inexistencia de elementos que respalden un componente que falle.

1.1.2.5 ¿Cómo se puede conformar la bodega de datos?

Existen tres aspectos muy importantes que intervienen en el establecimiento y el uso de una bodega de datos:

- El diseño,
- El mantenimiento,
- Y el uso de esta.

El diseño requiere de un análisis profundo de la organización o de las áreas funcionales responsables de la utilización de la bodega, las fuentes de datos que alimentarán la bodega.

Una decisión mal tomada en este aspecto puede significar demoras de días en vez de horas o minutos para la obtención de las respuestas, o incluso, que la bodega no esté en capacidad de responder las preguntas claves para la organización

Por último, están las decisiones que se tomarán sobre el uso de la bodega de datos. Cada empresa debe definir en forma muy clara, el ámbito empresarial en el que operará la bodega, las dependencias administrativas de la organización que van a tener acceso a ésta y las consultas que ésta debe responder inicialmente.

1.1.2.6 ¿Qué pueden ofrecer las bodegas de datos?

El objetivo de las bodegas de datos es centralizar una gran variedad de datos e información, para posteriormente interpretarla y darle un valor agregado para beneficio del negocio, todo ello por supuesto, con un fácil acceso y visualización por parte de los usuarios.

Algunos procesos que se realizan en estos escenarios son:

- Transformación de datos. Se obtienen datos e información de diferentes fuentes de almacenamientos y se aplican una serie de reglas definidas que convierten los datos en información útil para la toma de decisiones.

- Repositorios y metadatos. Más importante aún que el flujo de datos, es entender el origen y la descripción de éstos de una forma que sea común para toda la organización, ya que la idea no es tener más datos, sino entender lo que en conjunto representan.
- Procesamiento analítico en línea (OLAP). Provee el medio para obtener visualizar y analizar información con alto rendimiento y flexibilidad. OLAP presenta la información a los usuarios de una forma natural e intuitiva. De esta manera los usuarios pueden ser más efectivos en reconocer el valor de dicha información.
- Visualización. Generalmente los datos pueden ser mejor entendidos si los números son combinados de diferentes formas y presentados visualmente en forma de histogramas y varios tipos de gráficas. La visualización puede ser especialmente útil en identificar rápidamente cuales datos pueden tener un análisis especial.

1.1.2.7 ¿Cuál es el retorno de la inversión?

Entre los beneficios que se esperan para una compañía que configure y mantenga adecuadamente la bodega de datos se encuentran:

- Ahorros en costos, al igual que el aumento en ingresos son algunos de los beneficios tangibles que se pueden presentar en las empresas.
- Adicionalmente se pueden efectuar análisis para identificar y mantener a los clientes que generen la mayor utilidad, mientras se mantiene un mejor conocimiento de quienes son los clientes de la empresa.
- Una compañía no se debe olvidar que el objetivo de cualquier proyecto para una bodega de datos es reducir los costos operativos y generar ingresos. Esto es una inversión y se debe poner a esperar un retorno cuantificable a esta inversión en el tiempo.

- Por último, una bodega de datos bien implementada y mantenida trae grandes beneficios para la institución, cualquier falla en su implementación, garantizará que el dinero invertido se ha despilfarrado.

1.1.4 Data Mining

Data mining es una tecnología de soporte para el usuario final, cuyo objetivo es extraer conocimiento útil y utilizable a partir de la información contenida en las bases de datos de las empresas.

1.1.4.1 Objetivos

La extracción de información oculta y predecible de grandes bases de datos, para ayudar a las compañías a concentrarse en la información más importante de sus almacenes de datos (Data Warehouse).

Las herramientas de Data Mining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones conducidas por un conocimiento acabado de la información. Además pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas y a los cuales los usuarios de esta información casi no están dispuestos a aceptar. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas.

Las técnicas de Data Mining pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea. Una vez que las herramientas de Data Mining fueron implementadas en computadoras cliente servidor, pueden analizar bases de datos masivas para brindar respuesta a diversas pregunta y presentar los resultados en formas de tablas, con gráficos, reportes, texto, hipertexto, etc.

1.1.4.2 El Alcance de Data Mining

El nombre de Data Mining deriva de las similitudes entre buscar valiosa información de negocios en grandes bases de datos por ej.: encontrar información de la venta de un producto entre grandes montos de Gigabytes almacenados y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar exactamente donde residen los valores. Dadas las bases de datos de suficiente tamaño y calidad, la tecnología de Data Mining puede generar nuevas oportunidades de negocios al proveer estas capacidades:

- **Predicción automatizada de tendencias y comportamientos:** Data Mining automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos. Un típico ejemplo de problema predecible es el marketing apuntado a objetivos (targeted marketing). Otros problemas predecibles incluyen pronósticos de problemas financieros futuros y otras formas de incumplimiento, e identificar segmentos de población que probablemente respondan similarmente a eventos dados.
- **Descubrimiento automatizado de modelos previamente desconocidos:** Las herramientas de Data Mining barren las bases de datos e identifican modelos previamente escondidos en un sólo paso. Otros problemas de descubrimiento de modelos incluye detectar transacciones fraudulentas de tarjetas de créditos e identificar datos anormales que pueden representar errores de tipeado en la carga de datos.

Cuando las herramientas de Data Mining son implementadas en sistemas de procesamiento paralelo de alto alcance, pueden analizar bases de datos masivas en minutos. Procesamiento más rápido significa que los usuarios pueden automáticamente experimentar con más modelos para entender datos complejos. Alta velocidad hace que sea práctico para los usuarios analizar inmensas cantidades de datos. Grandes bases de datos, a su vez, producen mejores predicciones.

Las técnicas más comúnmente usadas en Data Mining son:

- Redes neuronales artificiales: modelos predecibles no-lineales, que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica.
- Árboles de decisión: estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen árboles de clasificación y regresión (CART: Classification And Regression Tree) y detección de interacción automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection).
- Algoritmos genéticos: técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.
- Método del vecino más cercano: una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases del o de los k registros más similares a él en un conjunto de datos históricos.
- Regla de inducción: la extracción de reglas if-then de datos basados en significado estadístico.

Muchas de estas tecnologías han estado en uso por más de una década en herramientas de análisis especializadas que trabajan con volúmenes de datos relativamente pequeños. Estas capacidades están ahora evolucionando para integrarse directamente con herramientas OLAP y de Data Warehousing.

1.1.5 Data Mart

Uno de los problemas que presentan las bodegas de datos , es que al crecer el rendimiento de sus consultas decae y el modelo centralizado deja de ser óptimo.

Para dar solución a lo anterior se pueden crear almacenes de datos especializados para cada área, los que reciben los datos de un almacén centralizado, conocidos con el nombre de Data

Marts. También los podemos definir como un subconjunto de una bodega de datos para un propósito específico, entre los cuales destacan el disminuir la demanda de la bodega y reducir el tráfico de la red.

La siguiente figura contiene una representación gráfica de la estructura de un almacén de datos conformados por varios Data Mars.

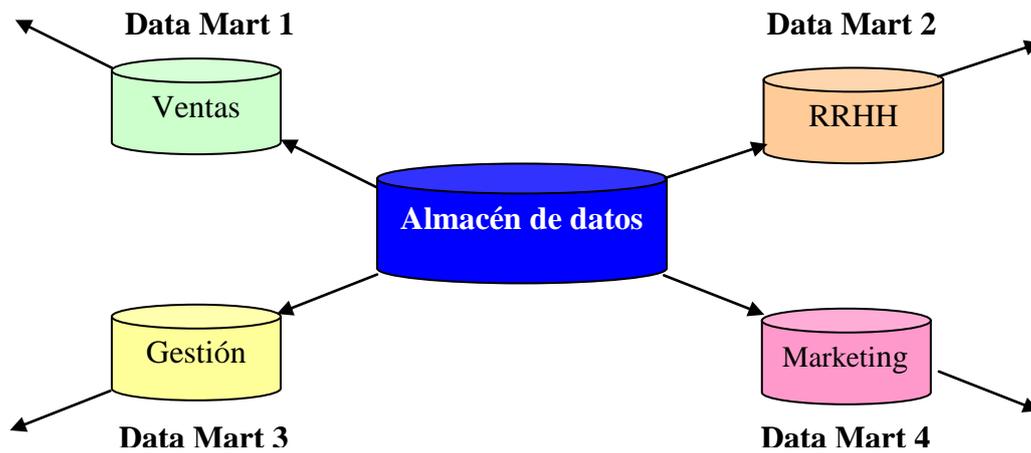


Figura 1.4: Almacén de datos conformado por varios data marts.

Existen dos tipos de Data Marts:

- Dependientes : Donde los datos son extraídos de un Data warehouse.
- Independientes: Los datos son extraídos de los sistemas operacionales y no requieren de la existencia de una bodega de datos.

1.1.6 OLAP

Es una componente de las bodegas de datos orientada al procesamiento analítico en línea. Este cuenta con un motor de consultas especializado, orientado al análisis en línea, administración y ejecución de consultas complejas que permitan deducir información del comportamiento del negocio.

En el siguiente punto se profundizará en lo que es OLAP, sus características y las múltiples alternativas que entrega para consultar los datos desde distintas perspectivas.

"OLAP proporciona una presentación multidimensional de los datos de un almacén mediante la creación de cubos que organizan y resumen los datos para mejorar la eficiencia de las consultas analíticas. El diseño de la estructura del almacén de datos puede afectar a la facilidad con la que se podrán diseñar y construir estos cubos". (MICROSOFT,2000) ⁷

1.1.7 ¿Qué es OLAP?

"Es un motor de consultas especializadas de las bodegas de datos, las herramientas OLAP son una tecnología de software para análisis en línea, administración y ejecución de consultas para inferir información del comportamiento del negocio." (BOHORQUEZ,1999) ⁸

Un sistema OLAP se puede entender como la generalización de un generador de informes. Las aplicaciones informáticas clásicas de consulta, orientadas a la toma de decisiones, deben ser programadas y orientadas a atender las necesidades del usuario, se crea una u otra interfaz. Sin embargo, muchos desarrolladores se dieron cuenta de que estas aplicaciones eran susceptibles de ser generalizadas y servir para casi cualquier necesidad. Los sistemas OLAP evitan la necesidad de desarrollar interfaces de consulta, y ofrecen un entorno único válido para el análisis de cualquier información histórica, orientado a la toma de decisiones. A cambio, es necesario definir dimensiones, jerarquías y variables, organizando de esta forma los datos.

Para los desarrolladores de aplicaciones acostumbrados a trabajar con bases de datos relacionales, el diseño de una base de datos multidimensional puede ser complejo o al menos, extraño. Pero en general, el diseño de dimensiones y variables es mucho más sencillo e

⁷ Microsoft., Toolkit para migración de datos y metadata estructuras multidimensionales, (2000).

⁸ Bohorquez. J., Aproximación metodológica de un Spatial Data Warehouse (1999).

intuitivo que un diseño relacional. Esto es debido a que las dimensiones y variables son reflejo directo de los informes utilizados por la organización.

Entre las fortalezas de esta herramienta se encuentra brindar información histórica del comportamiento de la empresa, pero desde un punto de vista multidimensional, es decir combinando múltiples variables simultáneamente, además de realizar análisis de información desde diversas perspectivas que aparentemente no se relacionan.

1.1.7.1 ¿Porque utilizar OLAP?

El competitivo mundo de los negocios exige a las empresas contar con análisis cada vez más sofisticados y oportunos para dar respuesta a sus constantes consultas.

Los consumidores de información necesitan estar en capacidad de interactuar con los resultados obtenidos, para poder tomar sus propias decisiones, basándose en el nuevo conocimiento que tienen a su disposición. La tecnología OLAP, suple la necesidad de reportes dinámicos, facilitando la toma de decisiones de manera inteligente, a lo largo y ancho de la organización.

Una de las fortalezas de OLAP, es que brinda rápidas respuestas a múltiples consultas, además de mostrar el comportamiento del negocio desde diversas perspectivas.

Comúnmente, los sistemas OLAP ofrecen una estructura de datos en forma de cubo que virtualmente contiene miles de resúmenes de reportes almacenados. Este cubo es multidimensional y ofrece a los usuarios finales diferentes alternativas y perspectivas de ver los datos. Por ejemplo, un cubo de reporte que contiene datos de ventas, puede producir reportes tan diversos como: ventas totales por trimestre, porcentaje de ventas por persona, y ventas promedio por región geográfica, además de cálculos comunes como sumas, conteos, promedios y porcentajes, y otros cálculos que también pueden programarse en el cubo.

Otra característica importante de mencionar es que los datos pueden ser vistos de acuerdo a las necesidades de los usuarios finales a través de opciones como: Rebar, Rotar y Taladrar el cubo, las que serán analizadas más profundamente más adelante.

Las aplicaciones OLAP generalmente son herramientas muy veloces que hacen de los reportes y de los análisis básicos un proceso más rápido que cualquier otro método tradicional de reporte.

OLAP ofrece medios más eficientes para tener acceso a la información, y así entender qué pasó y a partir de esto tomar mejores decisiones.

1.1.7.2 Beneficios que trae la utilización de OLAP.

Entre los beneficios de los reportes OLAP, sobre otras herramientas que permitan generar información para la toma de decisiones, podemos mencionar los siguientes puntos:

- **Mayor flexibilidad:** Generalmente los departamentos de tecnologías de información son los responsables de controlar el acceso a los datos, determinar cuales serán los datos que se desplegarán en los informes, siendo generalmente estas decisiones no las más correctas. Con los reportes OLAP en cambio, existe una gran variedad de configuraciones que permiten dar a los usuarios un mayor control en el manejo de la información, permitiéndoles decidir cual es la información que realmente necesitan.
- **Apoyo a la toma de decisiones:** Los reportes OLAP facilitan la toma de decisiones ya que la generación y modificación de reportes se hace de manera rápida. La creación y modificación de cubos puede ser realizada de manera más rápida y ágil que si se estuviera utilizando otra metodología para generar los informes.
- **Decisiones inteligentes:** Los reportes OLAP ofrecen mayor poder analítico ya que los reportes contienen cálculos que van más allá de los simples conteos, promedios y sumas. Hay un camino de migración claro y sencillo hacia herramientas más avanzadas de modelamiento y Data Mining.
- **Alto retorno en la inversión:** Los reportes OLAP son menos costosos que la opción de implementar otros sistemas tradicionales. Los ahorros se deben principalmente a su rápida implementación. También se hacen ahorros significativos en lo relacionado con la administración del sistema, no sólo en la implementación inicial del sistema, sino en el mantenimiento posterior de este. Esto debido a su sencilla arquitectura. Además, al ser tan

rápida y simple su instalación permite empezar a realizar los primeros análisis de manera casi inmediata.

- Los reportes OLAP se pagan por sí solos: Los reportes OLAP traen múltiples análisis para entregar a los usuarios más respuestas a partir de sus datos, adicionalmente una distribución rápida y fácil de los reportes y los resultados. Cada una de las dimensiones del cubo permite tener una visión simplificada del problema que se quiere solucionar, permitiendo descubrir en el camino información a través de las opciones que éste brinda, por ejemplo rotación, rebanas y taladraje del cubo, temas que serán analizados con mayor profundidad más adelante.

1.1.7.3 Consideraciones que se deben tener para implantaciones OLAP

Para desarrollar una estructura multidimensional se parte de la base que los datos proporcionados por el almacén son exactos, estables y tienen integridad referencial, además se deben tener las siguientes consideraciones:

- Preferir el esquema de estrella, con la finalidad de minimizar el número de tablas en las que se basarán las dimensiones.
- Diseñar tablas de dimensiones para los usuarios. La idea de esto es incluir información significativa sobre los datos que los usuarios deseen explorar.
- Aplicar normas de sentido común en el diseño de tablas de dimensiones, en otras palabras evitar repetir datos en las tablas de dimensiones.
- Utilizar una estructura común de tabla de hechos para datos similares.
- Los datos que se vayan a utilizar en el mismo cubo se pueden almacenar en varias tablas de hechos, siempre que éstas tengan la misma estructura.
- Evitar crear tablas auxiliares de datos resumidos.

- Crear índices en los campos de claves para mejorar el tiempo de las consultas.
- Crear índices en las columnas claves de las dimensiones y para cada tabla de hechos, esto permite al cargar las estructuras de datos multidimensionales una mejora de forma significativa en el rendimiento del proceso del cubo.
- Asegurar la integridad referencial, con esto se quiere cerciorar que los hechos se encuentren representados en todas las tablas de dimensiones. Los hechos almacenados en una tabla de hechos que no tengan una clave correspondiente en una tabla de dimensiones pueden originar errores o hacer que las filas de la tabla de hechos se pasen por alto si las tablas de hechos y de dimensiones se utilizan en el mismo cubo.
- Diseñar una estrategia de actualización de datos.
- Al momento de agregar o modificar los datos de un almacén, se deberán actualizar los cubos que se hayan creado a partir de datos anteriores para que los usuarios puedan disponer de estos nuevos datos. Incorporar datos adicionales a un cubo requiere menos tiempo que reconstruir el cubo cuando se modifican los datos existentes.

1.1.8 Uso de herramientas OLAP

1.1.8.1 Modos de Almacenamiento

Las opciones de almacenamiento físico afectan el rendimiento de los cubos que se implemente, por ello es necesario tener muy claro el funcionamiento de cada una de estas metodologías de modo de poder determinar cual es el que se apega más a las necesidades de la organización.

Para los cubos, se ofrecen tres formas de almacenar su información, las cuales afectan el rendimiento y los requerimientos de almacenaje de los cubos. Existen tres tipos de almacenamiento:

1. MOLAP (OLAP multidimensional): " Los datos fuente del cubo son almacenados junto con sus agregaciones en una estructura multidimensional de alto rendimiento. El almacenaje de

MOLAP, provee excelente rendimiento y compresión de datos. Tiene el mejor tiempo de respuesta, dependiendo solo en el porcentaje y diseño de las agregaciones del cubo. En general este método, es muy apropiado para cubos con bajo porcentaje de modificación de los datos y de uso frecuente por su rápida respuesta". (PRADOLAND CONSULTAORES,2000)⁹

2. ROLAP (OLAP relacional): " Toda la información del cubo, sus datos, agregaciones, sumas etc., son almacenadas en una base de datos relacional (ROLAP), cabe señalar que ésta no almacena una copia de la base de datos, sino que accesa las tablas originales cuando necesita responder a preguntas, es generalmente, mucho más lenta que las otras dos estrategias de almacenaje. Generalmente ROLAP se usa, para largos conjuntos de datos que no son frecuentemente buscados, tales como datos históricos de los años más recientes".(PRADOLAND CONSULTAORES,2000)¹⁰

3. HOLAP (OLAP Híbrido): "Combina atributos de MOLAP y ROLAP, la agregación de datos es almacenada en una estructura multidimensional usada por MOLAP, y la base de datos fuente, en una base de datos relacional. Para procedimientos de búsqueda que accesan datos. sumarizados, HOLAP es equivalente a MOLAP, por el contrario si estos procesos accesarán datos fuentes como los drill down, estos deben de buscar los datos en la base de datos relacional y esto no es tan rápido comparado a si los datos estuvieran almacenados en una estructura MOLAP.

Los cubos almacenados como HOLAP, son más pequeños que los MOLAP y responden más rápido que los ROLAP." (PRADOLAND CONSULTAORES,2000)¹¹

Existen diversas opiniones sobre cual de las tecnologías MOLAP o ROLAP es mejor. Por lo general, las implementaciones de MOLAP presentan mejor rendimiento que la tecnología

⁹ PradoLand Consultores., Cubos Conceptos (2000).

¹⁰ PradoLand Consultores., Cubos Conceptos (2000).

¹¹ PradoLand Consultores., Cubos Conceptos (2000).

relacional; sin embargo, también tienen problemas de escalabilidad. Por otra parte, las implementaciones de ROLAP son más escalables y, a menudo, son más atractivas para el cliente debido a que aprovechan las inversiones efectuadas en tecnología de bases de datos relacionales ya existentes.

1.1.9 Arquitectura de cubos

Un cubo es un conjunto de datos, dimensiones, medidas y particiones organizadas multidimensionalmente. El conjunto de tablas del que derivan las dimensiones y medidas de un cubo se denomina esquema del cubo. Las medidas del cubo derivan de las columnas de la tabla de hechos. Las dimensiones del cubo derivan de columnas de las tablas de dimensiones.

En la siguiente figura se muestra una representación gráfica de un cubo, el cual está conformado por las dimensiones de Almacén, Producto y Tiempo:

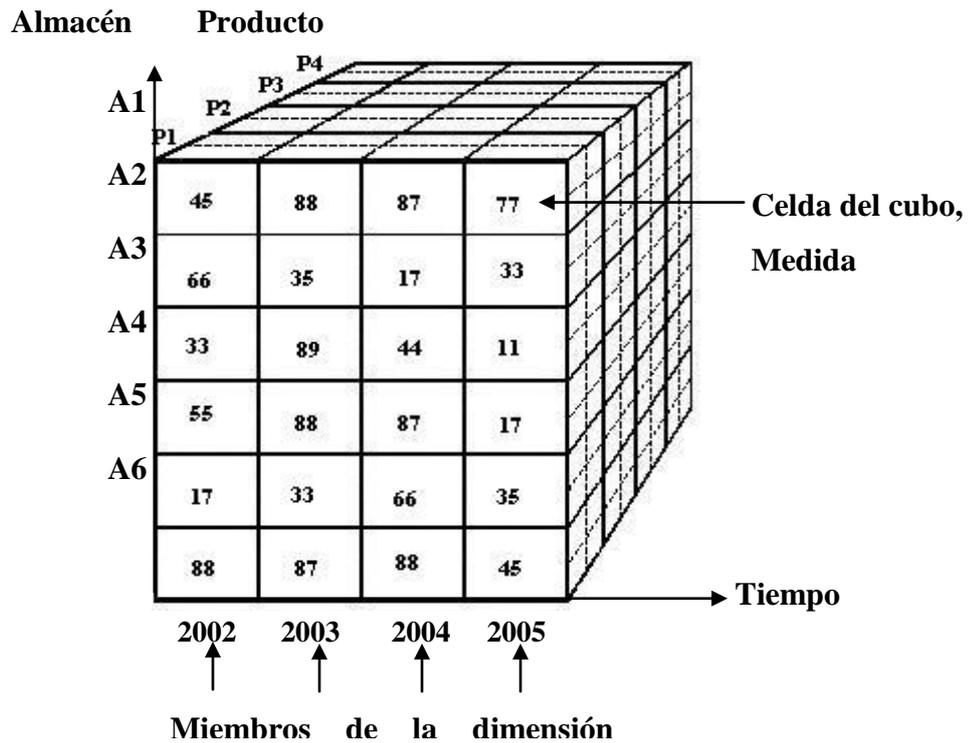


Figura 1.5 Esquema de un cubo de tres dimensiones.

1.1.10 Modelado dimensional

El diseño del esquema de la base de datos del almacén de datos debe incorporar los principios de modelado dimensional de forma que las tablas de dimensiones y las tablas de hechos representen los datos del negocio. Generalmente, el modelado dimensional proporciona esquemas de base de datos de estrella o de copo de nieve. Estos esquemas simplifican el diseño del cubo y reducen el número de combinaciones de varias tablas al consultar la base de datos para procesar las dimensiones y los cubos. Normalmente, las tablas de hechos incluyen la gran mayoría de los datos del almacén de datos.

1.1.10.1 Esquema de estrella

Este esquema facilita el acceso y análisis de los datos al centralizar los hechos en una única tabla de dimensiones. La profundidad de la dimensión depende del número de niveles. Cada nivel se deriva de una columna.

Cuando se agrega la dimensión a un cubo, la tabla de dimensiones se combina con la tabla de hechos. Si cada una de las dimensiones del cubo se basa en una única tabla, el cubo tendrá un esquema de estrella.

En la siguiente figura se muestra una representación de cómo sería un esquema de estrella.

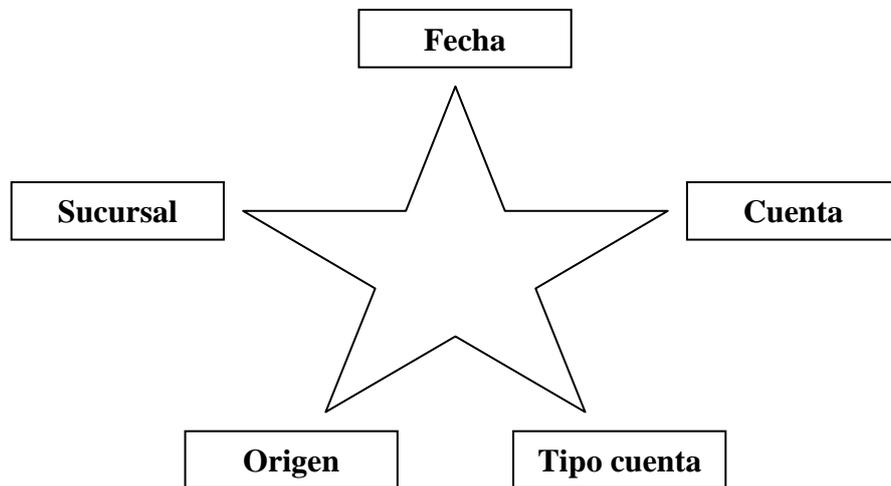


Figura 1.5 Muestra una representación gráfica de un esquema de estrella

1.1.10.2 Esquema copo de nieve

Crea una dimensión regular basada en varias tablas de dimensiones combinadas. La profundidad de la dimensión depende del número de niveles que se deseen crear. Cada nivel se deriva de una columna de una tabla.

Cuando se agrega la dimensión a un cubo, sólo se combina una de sus tablas de dimensiones con la tabla de hechos. Por lo tanto, el cubo adopta un esquema de copo de nieve.

1.1.11 Dimensiones

1.1.11.1 ¿Qué son las dimensiones?

Las Dimensiones son categorías descriptivas por las cuales los datos numéricos (Las Mediciones) en un cubo, son separados para su análisis. Por ejemplo, si una medición de un

cubo son las remuneraciones de los empleados, las dimensiones son tiempo, RUT del empleado y el departamento al que pertenece.

Una dimensión puede ser creada para usarse en un cubo individual o en múltiples cubos. Una dimensión creada para un cubo individual, es llamada dimensión privada. Por el contrario si ésta puede ser usada por múltiples cubos, se le llama dimensión compartida. Estas podrán ser usadas dentro de todo cubo en la base de datos, así se optimiza el tiempo y se evita duplicar dimensiones.

Las dimensiones compartidas, también habilitan la estandarización de las métricas de negocios entre cubos. Por ejemplo, al estandarizar las dimensiones compartidas para el tiempo y departamento, se asegura que los datos analizados, desde diferentes cubos, estén organizados similarmente.

Existen otros tipos de dimensiones, las conocidas como "Dimensión Virtual". Las dimensiones virtuales, pueden ser creadas desde propiedades de miembros, asociadas con dimensiones compartidas. No requieren un almacenamiento físico en el cubo, funcionan de manera similar a las dimensiones reales.

A continuación se muestra una representación espacial de una variable multidimensional con una, dos y tres dimensiones. En esta figura los cubos pequeños representan valores de dimensión y las esferas son datos.

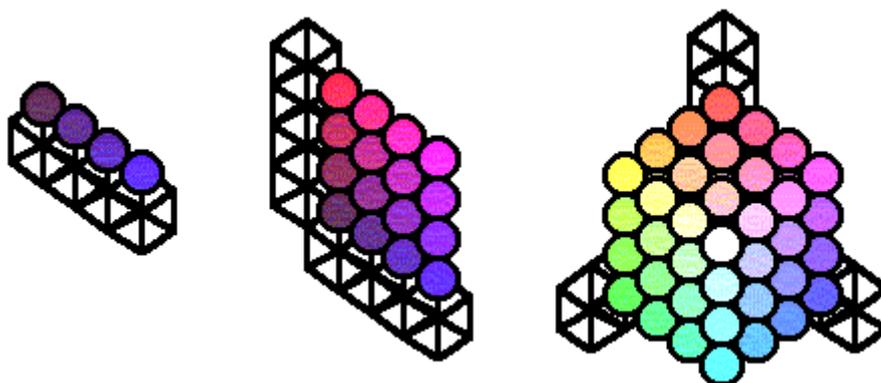


Figura. 1.6 Variables con una, dos y tres dimensiones

Una variable unidimensional podría ser el cambio de peso chileno a dólar, la que sólo varía en la dimensión <Tiempo>. Los cubos pequeños serían, por ejemplo, los meses del año y las esferas serían los valores numéricos correspondientes al cambio monetario en cada momento. Un ejemplo de variable de dos dimensiones es el número de habitantes, que se mueve por las dimensiones <Geografía> y <Tiempo>. Finalmente, los ingresos de una organización podrían almacenarse mediante una variable de tres dimensiones: <Producto>, <Geografía> y <Tiempo>.

Normalmente los elementos de una dimensión forman una jerarquía, creando niveles dentro de la dimensión donde existen nodos padres y nodos hijos.

1.1.11.2 Niveles

Un nivel es un elemento de una jerarquía de dimensiones. Los niveles describen la jerarquía desde el nivel de datos más alto (más resumido) al más bajo (más detallado).

Los niveles existen únicamente en las dimensiones. Se basan en columnas de una tabla de dimensiones o en propiedades del miembro de una dimensión. Además los niveles se definen en una dimensión para especificar el contenido y la estructura de la jerarquía de las dimensiones, es decir, las definiciones de nivel determinan los miembros que se incluyen en la jerarquía y sus posiciones en la jerarquía.

Después de crear las dimensiones compartidas y sus niveles, se suelen crear los objetos. Sin embargo, se crean en el proceso de creación de los cubos que los contienen. Por ejemplo, una dimensión Tiempo contiene los niveles año, trimestres, mes y día. La relación entre los niveles y los miembros de la dimensión tiempo se puede representar de la siguiente forma:

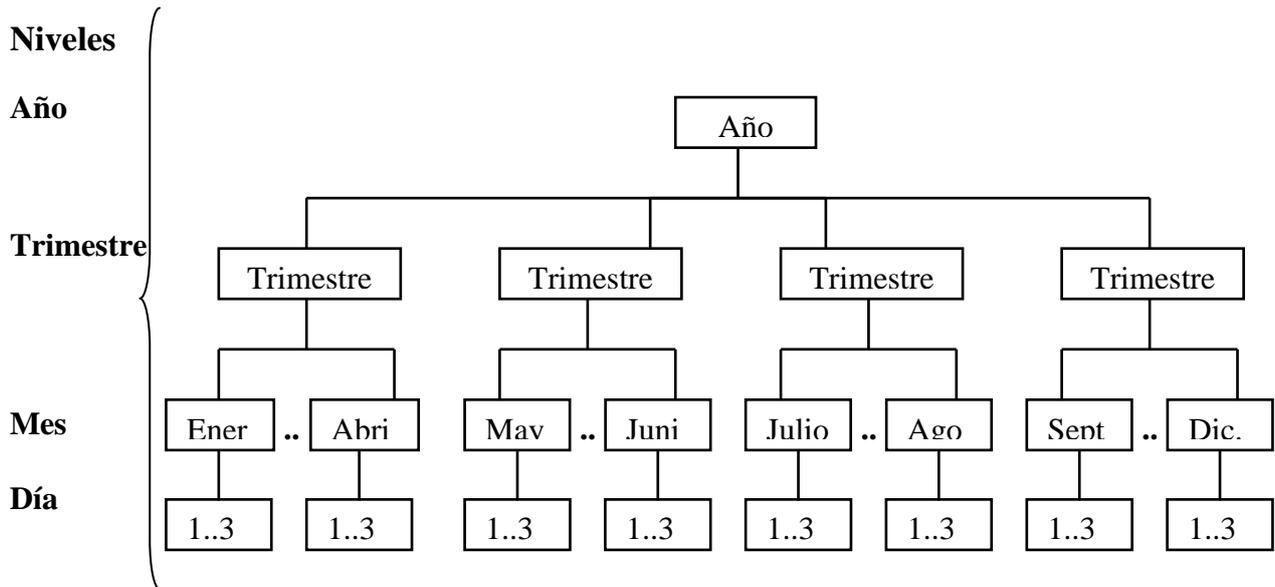


Figura 1.7. La siguiente figura representa gráficamente las jerarquías y niveles de una dimensión OLAP.

1.1.11.3 Jerarquías

Las dimensiones son un atributo estructural de los cubos. Están organizadas en jerarquías de categorías y (niveles) que describen los datos de las tablas de hechos. Estas categorías y niveles describen conjuntos similares de miembros en los que el usuario desea basar un análisis.

Las dimensiones también pueden estar basadas en modelos de minería de datos OLAP. Pueden utilizarse para almacenar los resultados de un análisis de modelo de minería de datos y pueden examinarse dentro del contexto de un cubo virtual.

Todas las dimensiones se basan directa o indirectamente en tablas. Cuando se crea una dimensión a partir de una tabla, se seleccionan las columnas que la definen. El orden de selección de las columnas es importante, ya que afecta a la colocación de los miembros dentro de la jerarquía de la dimensión. No obstante, el orden de los miembros de una dimensión creada mediante un análisis de modelo de minería de datos está determinado por el análisis. Esto se debe a que la dimensión representa el contenido del modelo.

Las dimensiones son jerárquicas y, en la mayoría de los casos, sus miembros están organizados en una configuración de tipo piramidal. Los valores de columnas con el mismo nivel en la jerarquía de dimensiones da lugar a la disposición horizontal y los valores de columnas con distintos niveles dan lugar a la disposición vertical.

La ventaja de manejar jerarquías radica en poder analizar de lo más general a lo más particular, lo que se conoce como Drill (Drill Down y Drill Up), en otras palabras ir taladrando en el detalle de los cubos o ir de lo más general a lo más particular.

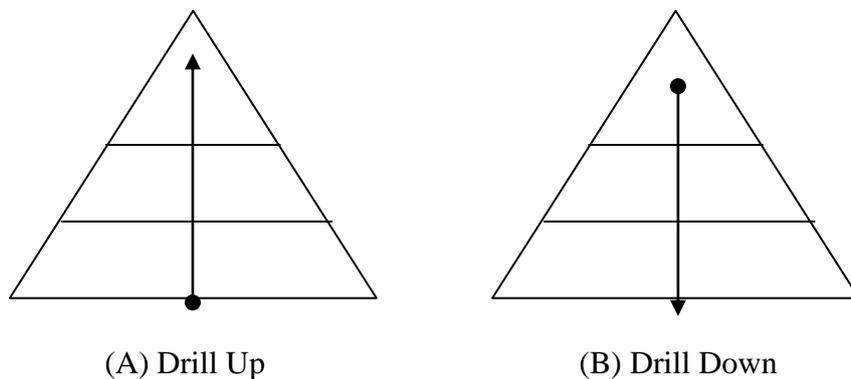


Figura 1.8 Las siguientes figuras grafican la jerarquía de las consultas.

1.1.11.4 Medidas

Las medidas identifican valores numéricos de interés que son sumados para su posterior análisis. Algunas medidas comunes son Ventas en unidades, ventas en pesos, costo de ventas, gastos, presupuestos, etc. Estas son usadas por el procedimiento de agregación de los servicios de OLAP y almacenadas para su rápida respuesta a las peticiones de los usuarios.

1.1.11.5 Miembros calculados

Se puede crear una medida calculada y calcular miembros de dimensiones, combinando expresiones multidimensionales (MDX), fórmulas matemáticas y funciones definidas por el usuario. Esta facilidad, habilita a usted a definir nuevas medidas y miembros de dimensión, basados sobre una sintaxis de fórmulas sencillas.

1.1.11.6 Partición

"Son contenedores de almacenamiento multidimensional que mantienen los datos del cubo. Cada cubo contiene al menos una partición y un cubo de datos, puede ser combinado desde múltiples particiones, cada partición podrá tomar datos de diferentes fuentes de datos y puede ser almacenada en localizaciones separadas. Una partición de datos, puede ser actualizada independientemente de otras particiones en un cubo. Las particiones son invisibles a los usuarios, en el cubo se presentan como un simple objeto." (PROLAND CONSULTORES,2000) ¹²

1.1.11.7 Agregaciones

"Las agregaciones son resúmenes precalculados de los datos del cubo que ayudan a proporcionar respuestas rápidas a las consultas." (NADER,1998) ¹³

El objetivo es diseñar el número óptimo de agregaciones. Este número no sólo proporcionará tiempos de respuesta satisfactorios, sino que también evitará que el tamaño de la partición sea excesivo. Cuanto mayor sea el número de agregaciones, menor será el tiempo de respuesta, pero también se necesitarán más espacio de almacenamiento. Además, a medida que el asistente diseña más agregaciones, las agregaciones más antiguas producen mayores mejoras de rendimiento que las agregaciones más recientes.

1.1.11.8 Hechos

Los hechos comprenden el volumen de la bodega de datos y pueden estar compuestos por millones de registros dependiendo de los niveles de detalle de los datos (granularidad) y los intervalos de tiempo de estos. Cabe señalar que los hechos son accedidos por las dimensiones, un hecho es un dato instantáneo en el tiempo bajo condiciones definidas por las dimensiones. Un ejemplo sería los saldos de una cuenta corriente, donde las dimensiones serían la fecha hasta la cual se quiere ver el saldo y la cuenta para la cual se quiere consultar.

¹² PradoLand Consultores., Cubos Conceptos (2000).

¹³ Nader J., Sistema de Apoyo Gerencial Universitario (1998).

1.1.12 Operaciones analíticas básicas

Los datos pueden ser vistos desde distintas perspectivas o dimensiones, pudiendo probar distintas combinaciones, permitiendo obtener diferentes análisis. En este modelo los datos son vistos como cubos, los que consisten en categorías descriptivas (dimensiones) y valores cuantitativos (medidas).

1.1.12.1 Drill down

Aumenta los detalles de los miembros de un conjunto a un nivel por debajo del nivel más bajo representado en el conjunto, al permitir a partir de información visualizada en el pantalla, "sumergirse o bucear" en esa información con el fin de conocer el detalle de los datos que han servido inicialmente para crearla. Por ejemplo, a partir de la cifra de negocios de una región se visualiza la lista de las cifras de negocios de las sucursales que componen esa región.

1.1.12.2 Drill Up

Disminuye el nivel de detalles de los miembros por sobre un nivel al actualmente representado.

1.1.12.3 Drill Through

El consiste en visualizar bajo varios ángulos distintos una misma información. Por ejemplo, a partir de la visualización de la cifra de negocios de productos financieros en el último mes de Marzo de una sucursal, se desea presentar la misma información para otra sucursal.

1.1.12.4 Slicing and dicing

1.1.12.4.1 Tajadas

Una tajada corresponde a la visualización de los datos desde una dimensión. En la siguiente figura se muestra una representación gráfica de las tajadas de un cubo, las que se encuentran diferenciadas por colores.

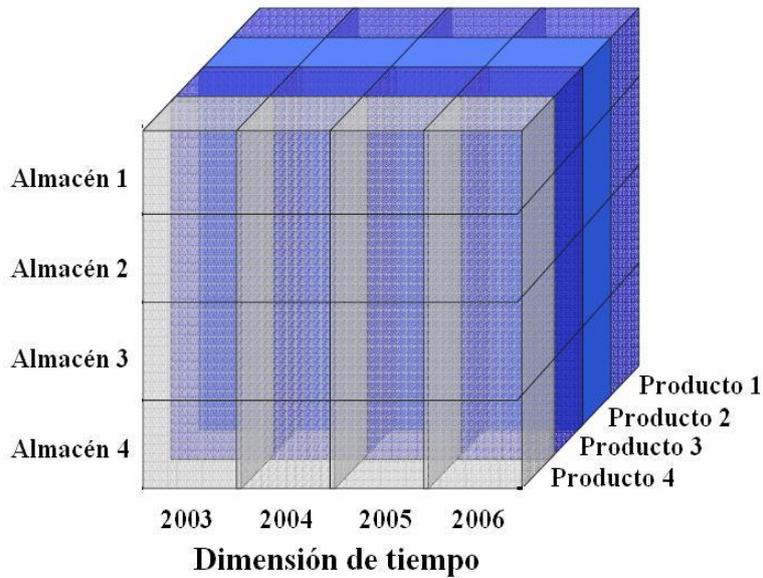


Figura 1.8 Representación gráfica de las tajadas de un cubo

1.1.12.4.2 Slicing o rebanar

Consiste en cambiar una dimensión por otra para poder ver el cubo desde otra perspectiva, en otras palabras ver los resúmenes de información analizados desde otra cara. En la siguiente figura cada uno de los colores representa una rebanada.

Es importante destacar que de un cubo podemos obtener múltiples consultas, todas ellas visualizadas de distintas perspectivas. Un ejemplo práctico de lo anteriormente mencionado se muestra en la siguiente figura, donde se aplica el concepto de rebanas.

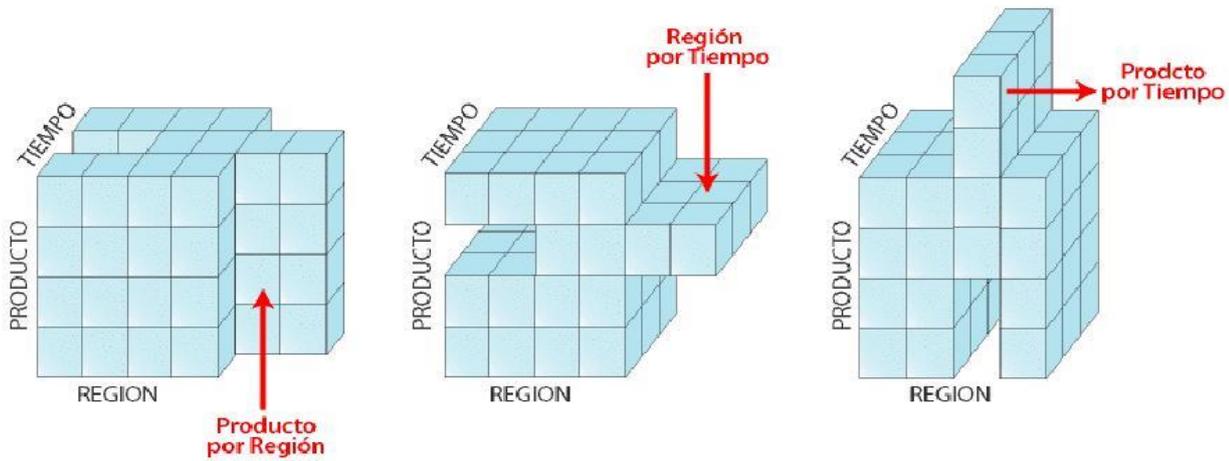


Figura 1.9 La siguiente figura muestra las distintas perspectivas de las que puede ser analizado un cubo, también llamadas rebanas.

1.1.12.4.3 Dicing o Cortar en cubitos

Es la capacidad de visualizar los datos desde distintas perspectivas.

En la siguiente figura se muestra una representación gráfica de la operación anteriormente explicada.

Ventas

	Productos	Store 1	Store 2
Q1	Electronics	\$5,20	\$5,60
	Toys	\$1,90	\$1,40
	Clothing	\$2,30	\$2,60
	Cosmetics	\$1,10	\$1,10
Q2	Electronics	\$8,90	\$7,20
	Toys	\$0,75	\$0,40
	Clothing	\$4,60	\$4,60
	Cosmetics	\$1,50	\$0,50

➔

Ventas

	Productos	Store 1
Q1	Electronics	\$5,20
	Toys	\$1,90
Q2	Electronics	\$8,90
	Toys	\$0,75

Figura 1.8. En la siguiente figura se muestra un ejemplo práctico de cortar en cubitos.

1.1.12.5 Rotación

Se utiliza para indicar el orden de visualización de los datos, permitiendo ver los resúmenes de información desde distintas perspectivas. En la siguiente figura se da una representación gráfica de cómo se vería un cubo en caso de ser rotado.

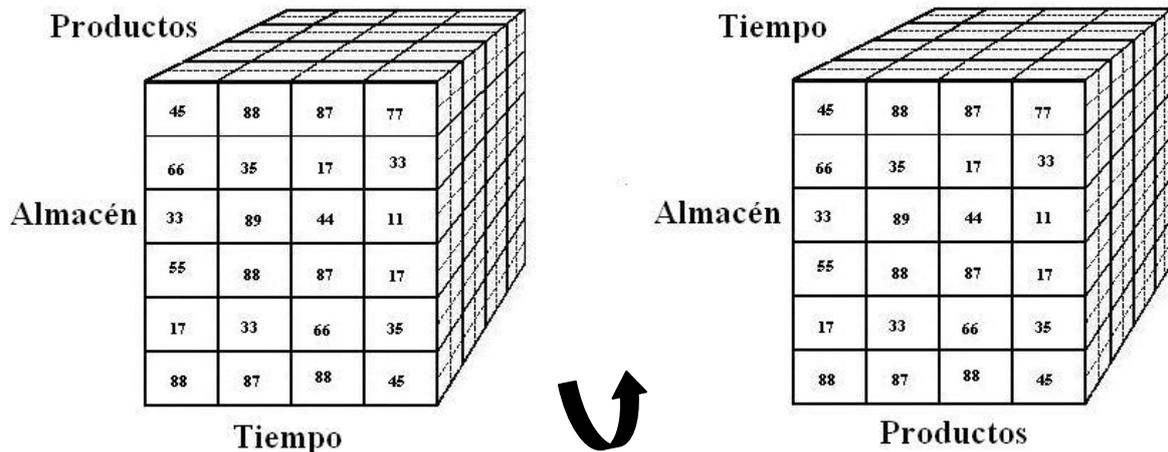


Figura 1.10: Muestra como se vería un cubo de tres dimensiones (Productos, Almacén, Tiempo), luego de ser rotado (Tiempo, Productos, Almacén).

1.1.13 Lenguaje MDX

Es un lenguaje multidimensional de expresión, el equivalente multidimensional del SQL. Es utilizado para obtener la información almacenada en estructuras multidimensionales a través de consultas y cálculos de datos con expresiones.

1.1.13.1 Consultas MDX

Es un lenguaje para consultas multidimensionales, así como SQL es un lenguaje para consultas estructuradas correlativas.

1.1.13.3.2 Expresiones de cálculo

Son expresiones para realizar cálculos multidimensionales, así como las hojas de cálculo tienen sus fórmulas. Generalmente son utilizadas en miembros calculados, miembros predefinidos en las dimensiones y propiedades del cubo.

1.1.13.4 Sintaxis básica de una consulta

[**With** FormulaEspecial 1[...FormulaEspecial N]

SELECT [AxisEspecial]

FROM [Nombre del cubo]

[**WHERE** [Slicer]]

[Propiedades de las celdas]

En la sección FormulasEspecial se especifican cálculos temporales de miembros, los cuales pueden ser utilizados para listar datos en los ejes de la consulta o en la cláusula where de ésta. A continuación se especifica un ejemplo:

Member ParentMember.MemberName As

'ValueExpression'[, SOLVE ORDER = M][, CellProperty=ValueExpression ...]

El AxisEspecial especifica los bordes de la consulta. Un borde es un límite asignado a un eje. Los ejes de las columnas pueden ser presentados en cualquier orden, generalmente estos presentan los miembros calculados del cubo (columnas), los cuales son abiertos por los datos por los cuales se desea analizar la información (filas).

Adicionalmente existen una serie de propiedades que pueden ser utilizadas para filtrar la información comprendida en las filas o columnas.

En la cláusula Where se especifica cuales son las tajadas del cubo que se desean listar, generalmente corresponden a niveles de una dimensión que permite ver la información filtrada sobre un sección específica.

A continuación se especifica un ejemplo de consulta MDX en base al cubo presentado en el punto 1.1.10.5. Este se encuentra formado por tres dimensiones que son Productos, Tiempo y Almacén, supondremos que la dimensión de Productos esta compuesta por tres nivel que son Juguetería, Zapatería y Vestuario, por otra parte la dimensión de tiempo contempla los niveles especificados en la figura 1.6, abarcando desde el año 2000 al 2006.

La dimensión de almacén esta dividida en tres niveles que son Almacén 1, Almacén 2 y Almacén 3.

Se quiere obtener el total de las ventas para los distintos períodos del Almacén 1, para todos los productos existentes en el año 2001.

Select

{[Measures].Ventas} On Columns,

{[Productos].Children} On Rows

From

Market

Where

([Almacén].[Almacén 1], [Fecha].[2005])

Parte 2

Descripción de la empresa

Capítulo 2 – Descripción de la empresa

1.1 Introducción

El presente capítulo tiene por objeto describir a la empresa para la cual se desarrollará el proyecto, entre los puntos a tratar destacan los objetivos de la organización y la descripción de su funcionamiento.

A continuación se presenta una breve reseña que permitirá explicar como nace la organización, para posteriormente describir el ambiente en el que se desarrolla, sus objetivos y metas, además presentar un estudio de la situación actual y de los problemas detectados en las metodologías existentes.

1.1.1 Historia de la empresa

Hasta antes del año 2000 el grupo Forestales Regionales S.A., al que pertenecen empresas como Forestal Diguillín, Forestal Probosque, Procer, Maderas Diguillín y Andeanpine, contaba con departamentos de informática en cada una de las sociedades que lo conformaban. La metodología de trabajo de los distintos grupos informáticos estaba orientada a las necesidades específicas de la empresa a la que pertenecían, en muchas ocasiones estas necesidades eran compartidas, pese a ello se construían diversas aplicaciones para solucionar un mismo problema. Esta forma de trabajo tan poco conectada provocaba la existencia de implementaciones duplicadas, además de elevados costos de mantención y desarrollo de los proyectos.

Lo anterior lleva a los directivos a considerar la idea de fusionar los distintos departamentos naciendo lo que actualmente se conoce como TDIFOR (Tecnologías de información forestal S.A.), esta nueva entidad tendría la finalidad de desarrollar software para las empresas, con la finalidad de obtener un producto único y estandarizado que pueda ser de uso masivo. Además

de poder prestar asesorías a los usuarios frente a los distintos problemas a los que pudieran verse enfrentados.

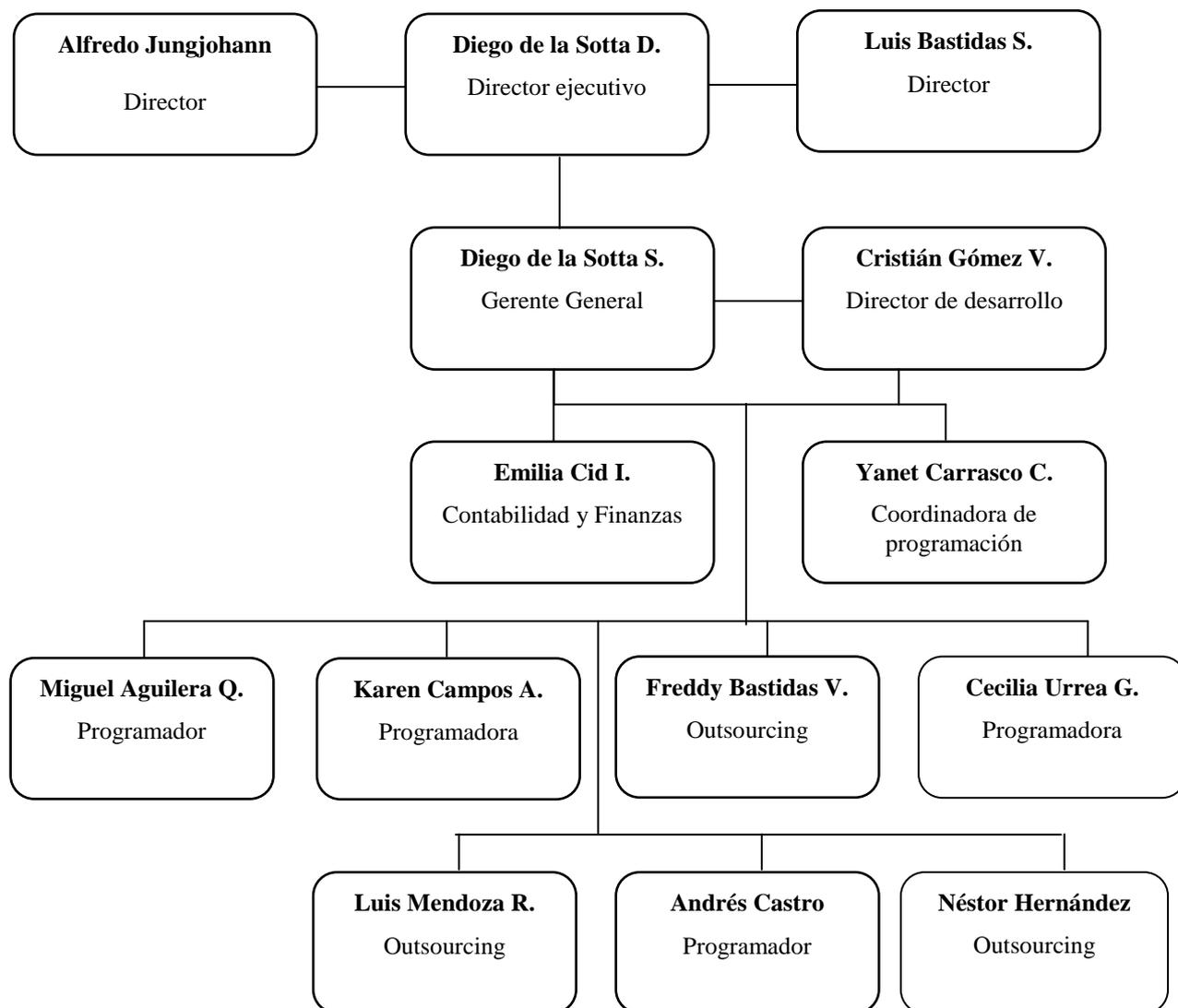
El primer desarrollo de gran envergadura actualmente se conoce como SGF (Sistema de gestión forestal) siendo la finalidad de éste el prestar apoyo a las distintas áreas de las empresas.

Hasta la fecha la TDIFOR ha obtenido grandes avances en materia de desarrollos, contando con productos que han sido vendidos a empresas externas al grupo con gran éxito.

2.2.2 Visión de la empresa

Posesionar y mantener a sus clientes como líderes en el ámbito de tecnologías de información forestal y servicios técnicos, además de poder consolidar sus productos entre la VII y la VIII región.

2.2.3 Organigrama de TDIFOR



2.2.4 Situación de la empresa.

En la actualidad la empresa se encuentra desarrollando un software llamado GCC (Gestión contable computacional), este abarca todas las áreas de las distintas empresas como por ejemplo Recursos Humanos, Gestión y Contabilidad. La finalidad principal de este desarrollo es migrar del antiguo sistema llamado SGF, incorporando las ideas implementadas en éste pero agregando mejoras. La GCC cuenta con un completo sistema de contabilidad, denominada ACC (Administración contable computacional). Esta cuenta con un sin número

de ventanas que permiten a los usuarios llevar una contabilidad ordenada de forma muy sencilla.

El desarrollo de éste módulo ha llevado alrededor de 3 años, pese a ello tiene algunas falencias provocadas por las malas decisiones tomadas al inicio del proyecto. Es precisamente en esta área en donde se enfocará esta tesis con la finalidad de mejorar el actual sistema de toma de decisiones.

2.3 Características del sistema contable

El sistema contable de la empresa permite llevar un completo control de las transacciones que se realizan en las distintas empresas.

Entre las características de éste sistema se encuentran la emisión de distintos informes que permiten a los usuarios ver una serie de resúmenes como por ejemplo: los saldos de los distintos análisis y cuentas que posee la empresa y los movimientos que se han realizado en ellas hasta una determinada fecha, además de permitir llevar un completo control de los pagos, las ventas y compras que se hayan realizado a una determinada entidad o empresa, en resumen contempla todos los pasos básicos para establecer un ciclo contable completo. Entre los informes que se pueden visualizar en el sistema se encuentran los libros de contabilidad, estos son:

- Libro diario
- Libro mayor
- Libro de balances

Además existen informes que permiten ver las cuentas de pérdida y ganancias, obtención de saldos de los análisis para distintos períodos, destacando la ventana de comprobantes contables, encargada de ingresar, eliminar y modificar los movimientos contables que se realizan en la empresa.

2.3.1 El libro diario

En el libro diario se registran todos los hechos y operaciones contables. Este registro se realiza mediante los asientos contables, los cuales se ordenan por fecha de creación.

Al conjunto de asientos que quedan registrados durante un período de tiempo se le llama Libro Diario. El libro diario simplemente informa del movimiento en sí, o sea, qué cuentas han intervenido, la fecha, las cantidades y el concepto.

2.3.2 Libro Mayor

En el libro diario se muestra una ficha individual que registra las cantidades de una cuenta en el debe y en el haber, así como el saldo de la misma. El libro mayor precisamente lo que hace es eso; anotar las cantidades que intervienen en los asientos en su correspondiente cuenta del libro mayor, representada por una T.

En las siguientes figuras se muestran como se registran los asientos en un libro mayor.

<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%; text-align: center;">Debe</td> <td style="width: 33%; text-align: center;">Cientes</td> <td style="width: 33%; text-align: center;">Haber</td> </tr> <tr> <td style="border-right: 1px solid black; height: 40px;"></td> <td style="text-align: center; vertical-align: middle;">5000</td> <td style="border-right: 1px solid black;"></td> </tr> <tr> <td style="border-right: 1px solid black; border-top: 1px solid black;"></td> <td style="border-top: 1px solid black;"></td> <td style="border-top: 1px solid black;"></td> </tr> </table>	Debe	Cientes	Haber		5000					<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%; text-align: center;">Debe</td> <td style="width: 33%; text-align: center;">Caja</td> <td style="width: 33%; text-align: center;">Haber</td> </tr> <tr> <td style="text-align: center; vertical-align: middle;">5000</td> <td style="border-right: 1px solid black;"></td> <td style="border-right: 1px solid black;"></td> </tr> <tr> <td style="border-top: 1px solid black;"></td> <td style="border-top: 1px solid black;"></td> <td style="border-top: 1px solid black;"></td> </tr> </table>	Debe	Caja	Haber	5000					
Debe	Cientes	Haber																	
	5000																		
Debe	Caja	Haber																	
5000																			
(a)	(b)																		

Figura 2.1 En estas dos cuentas (libro mayor) del asiento anterior, se tiene un saldo deudor en caja (positivo) y un saldo acreedor en la cuenta de clientes.

2.3.3 Libros de Balances

Reflejan la situación del patrimonio de la empresa en una fecha determinada. Los Balances se crean cuando se han pasado las cantidades de las cuentas de los asientos a su libro mayor. Existen distintos tipos de balances entre los que podemos mencionar:

- Balance de apertura: es el balance de apertura de una nueva empresa. Casi todo el dinero que se posee son recursos propios de la empresa.

- Balance de sumas y saldos: aparecen las sumas y los saldos de cada cuenta. Se puede consultar en cualquier momento, y se recogen tanto cuentas de gestión (ingresos y gastos) como cuentas patrimoniales (activo y pasivo). Este balance no refleja los resultados de la empresa (ganancias o pérdidas).
- Balance de situación: aparecen las cuentas patrimoniales (activo y pasivo) y el resultado del ejercicio.
- Balance de liquidación: es el que presenta el estado de la empresa cuando queremos venderla

Cabe mencionar que de los balances antes nombrados se encuentra implementado en el sistema el balance de sumas y saldos.

2.3.4 Cuentas de pérdidas y ganancias

La cuenta de pérdidas y ganancias se encarga de mostrar el beneficio o pérdida de la empresa. Aunque esta cuenta se puede consultar en cualquier momento del ejercicio, normalmente se utiliza al final del mismo. La cuenta de pérdidas y ganancias se calcula a partir de las siguientes partidas:

1. Resultados de Explotación
2. Resultados Financieros
3. Resultados Extraordinarios

1. Resultados de Explotación: recoge todos los movimientos de ingresos y gastos naturales, y que son necesarios para la actividad normal de la empresa, como pueden ser: compras, ventas, tributos, salarios, transportes, etc.

Las cantidades utilizadas en compras y gastos se colocan en el debe de la cuenta de pérdidas y ganancias.

Las cantidades utilizadas en ventas e ingresos se sitúan en el Haber de dicha cuenta.

2. Resultados Financieros: recoge los beneficios y pérdidas originados por los dividendos, intereses y otros, de la empresa.

3. Resultados Extraordinarios: se obtienen por métodos diferentes a la actividad normal de la empresa, como por ejemplo, beneficios o pérdidas por la venta de un local.

El esquema de la cuenta de pérdidas y ganancias sería el siguiente:

Debe	Haber
Gastos de Explotación	Gastos Extraordinarios
Gastos Financieros	Ingresos de Explotación
Ingresos Extraordinarios	Ingresos Financieros
Beneficio del ejercicio	Pérdida del ejercicio

Figura 2.2: Esquema de cuentas de perdida y ganancia.

El sistema de la empresa además cuenta con una serie de ventanas orientas al ingreso de información tipo maestros, entre las que se encuentran ventanas de cuentas corrientes de la empresa, ventanas de análisis contables y de conceptos, cabe destacar que esta información es ingresada por el contador de cada una de las empresas.

Un ejemplo muy recurrente de análisis contables es el análisis del "Item gastos", el cual puede tener asociado distintos conceptos como por ejemplo: "Arriendos", "Capacitación", "Consumos básicos", etc. Es importante destacar que el sistema permite sólo un máximo de 20 análisis los que son distintos para cada una de las empresas ya que dependen del contador.

Una de las ventanas más importantes y utilizadas en el sistema es el comprobante contable. La información de esta ventana es grabada en dos tablas una denominada `tcomprobantes_contables` y `dlin_comprobantes_contables`, en la primera se almacena información general del comprobante como la fecha de emisión, el tipo de asiento y el

número del comprobante contable que se está emitiendo, lo que se considera la cabecera, por otra parte está el detalle donde se ingresan datos como el código de la cuenta, la imputación (debe o haber), la glosa, los montos y los análisis de la cuenta en caso de que los tenga, además de los conceptos asociados al análisis y un posible valor que puede corresponder por ejemplo a un número de una factura.

Para poder filtrar los datos que necesitamos para generar los distintos informes nos basaremos en la información asociada al análisis de ID o Rut, a través del cual se podrá llegar a deducir información como por ejemplo: cuanto le adeuda la empresa a un proveedor.

En la siguiente figura se muestra el modelo de datos con las tablas del área contable que se utilizarán para la realización de este proyecto.

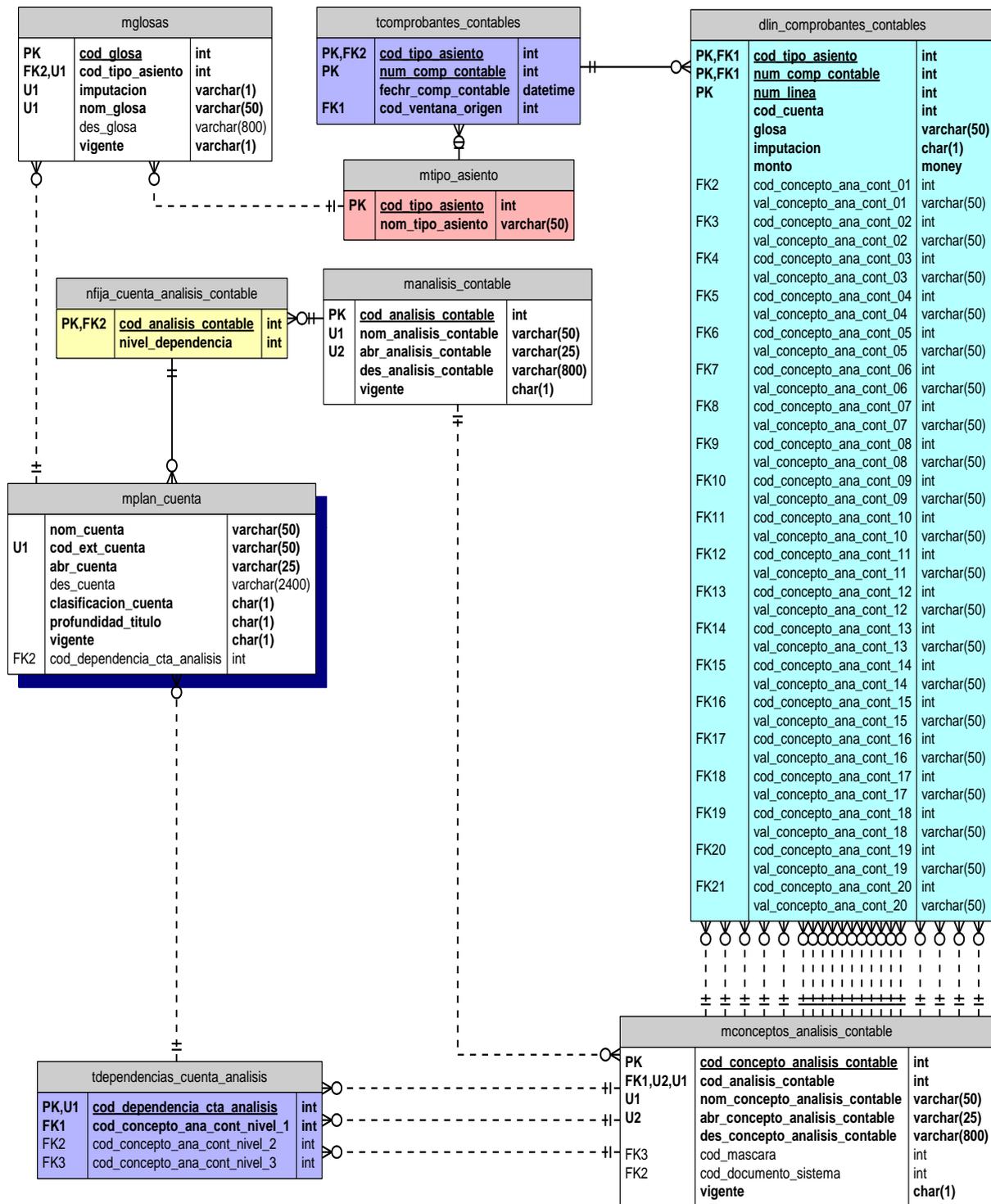


Figura 2.3: Modelo de datos del área contable que incluye las tablas más importantes de las que se sacará información para la generación de los cubos.

Además existen otras 5 tablas que contienen información de los distintos integrantes de la empresa, en ésta clasificación se encuentran los clientes, proveedores, personal y contactos, éstos últimos son personas que pertenecen al personal de un cliente, un proveedor o de la empresa misma. La relación de todas estas tablas con la parte contable es a través del Id o RUT, el que puede estar ingresado en uno de los análisis contables de la tabla `dlin_comprobantes_contables`.

Debido a que en las empresas no fue ingresado el análisis de Id o Rut en la misma posición no se puede mostrar en el modelo de datos la relación existente entre la tabla de comprobantes contables y los integrantes, ya que ésta puede corresponder a cualquiera de los veinte análisis del comprobante contable.

2.3.5 Problemas Detectados en el sistema contable

- En la tabla de `dlin_comprobantes_contables` (mencionada anteriormente) no se encuentra normalizada, esta cuenta en la actualidad con un total de 47 columnas. Cabe señalar que en ella se basa la contabilidad de la empresa.
- Para poder generar informes como por ejemplo informes de saldos de los análisis se crearon tablas auxiliares denominadas `isaldos_analisis_mensuales` e `isaldos_analisis_anuales`, estas contienen los saldos de los análisis de la empresa y las tablas `isaldos_anuales` e `isaldos_mensuales`, que contiene resúmenes de los saldos de las cuentas. Estas tablas son actualizadas automáticamente a través de triggers, lo que provoca la duplicación de la información y un tiempo de espera mayor al momento de guardar, modificar o eliminar un comprobante contable.
- Cada cierto tiempo es necesario hacer un reprocesamiento de la información debido a un descuadre en los saldos, por causas que aún se desconocen. Éste reprocesamiento tarda

alrededor de 30 hrs. en una de las empresas cuya tabla de `dlin_comprobantes_contables` tiene más de 370000 registros.

- Más del 50% de la tabla `dlin_comprobantes_contables` se encuentra llena con nulos, lo que hace trabajar mayormente al servidor al momento de interpretar éste tipo de dato.
- Las lecturas y escrituras se realizan directamente a la base de datos transaccional, provocando posibles deadlock.

2.4 Análisis de Requerimientos.

Debido al mal diseño de las tablas del modelo de datos y a la mala implementación de algunas de las ventanas de ingreso de datos, se hace necesario buscar opciones que permitan mantener de forma sencilla y más eficiente los complejos sistemas de generación de informes de contables.

La meta es cambiar la forma en que se procesa la información buscando mecanismos que ayuden a simplificar el proceso de resumen de información de manera de poder eliminar los complejos sistemas que actualmente posee la empresa, y de paso obtener información certera y oportuna a un menor costo.

2.5 Objetivos del proyecto

La idea es dejar de contar con las tablas auxiliares de saldos y los triggers que las mantienen, simplificando la mantención de la información.

Para lograr lo anterior es necesario realizar un óptimo diseño de las estructuras multidimensionales que se desean implementar para simular las tablas de saldos.

- Diseñar estructuras multidimensionales que permitan optimizar los tiempos de las consultas de las distintas ventanas del área contable.

- Obtener óptimos tiempos de respuesta en las consultas a las estructuras.
- Disminuir los tiempos en las transacciones que involucren la utilización del comprobante contable, al eliminar la actualización de las tablas auxiliares a través de triggers.
- Simplificar la mantención de ventanas que obtengan información del comprobante contable.
- Evitar el reprocesamiento de la información al contar con un mecanismo que entregue información certera de la actual situación de la empresa.
- Simplificar los mecanismos de resúmenes de información de las distintas cuentas de la empresa, así como de los análisis contables que se ven involucrados en las distintas transacciones y los Id o Rut que poseen movimientos contables, disponiendo de información resumida en el momento en que se requiera.

2.3 Características generales de la red

El grupo de Forestales Regionales S.A. cuenta con tres servidores de bases de datos y tres servidores Web para alimentar los sistemas de las distintas empresas. Debido a que en la implementación de esta memoria no se realizaron cambios en los servidores Web, ni interfirieron con el desarrollo de ésta, sólo se dará a conocer las características generales de los servidores de base de datos y la arquitectura lógica que conforman los equipos en la red. Esta definición abarcará solamente, por motivos de seguridad, como se encuentran conectados los servidores con el equipo cliente.

TDIFOR ha dispuesto los servidores en distintos puntos de la ciudad de Los Ángeles. Estos servidores actualmente se reparten la administración de 23 bases de datos, proporcionalmente a la cantidad de usuarios que posee cada una de las empresas.

El siguiente diagrama muestra cómo se conecta una aplicación cliente a la red.



Figura 2.4 Configuración de los equipos de la red.

2.3.1 Características técnicas de los servidores

A continuación se muestra un detalle con las características de los tres equipos.

Por requerimiento de la empresa los nombres de los servidores serán modificados.

<i>Nombre</i>	<i>Características</i>
<i>Servidor A</i>	<i>HP Proliant ML350 G4</i> <i>CPU Intel Xeon 3,20 Ghz</i> <i>RAM 2 GB</i> <i>Microsoft Windows 2003 SP1</i> <i>Disco Duro: 140 GB</i>
<i>Servidor B</i>	<i>Comapq Proliant ML 350 G3</i> <i>CPU Intel Xeon 2.2 Ghz</i> <i>RAM 2 GB</i> <i>Microsoft Windows NT</i> <i>Disco duro 70 GB</i>
<i>Servidor C</i>	<i>HP Proliant ML350 G4</i> <i>CPU Intel Xeon 3,20 Ghz</i>

RAM 3 GB

Microsoft Windows 2003 SP1

Disco Duro: 140 GB

Parte 3

Requerimientos

Capítulo 3 - Requerimientos

4.1 Introducción

En este capítulo se detallan los requerimientos presentados por la empresa las cuales reflejarán las limitantes sobre las que se deberá implementar el nuevo sistema.

4.2 Requerimientos

4.2.1 Requerimientos funcionales

Motor de Base de Datos relacional. Integrarse nativamente con el motor de base de datos, sin requerir ninguna función adicional de administración. Además debe incluir las herramientas de extracción, transformación y procesamiento de datos.

Conectividad. Debe soportar los estándares de conectividad de las aplicaciones existentes de forma segura, además de permitir realizar consultas MDX de forma dinámica, las cuales dependerán de los parámetros que especifique el usuario en la ventana.

Procesamiento de la información. Permitir cálculos por múltiples dimensiones y atributos.

- Recuperar conjuntos de resultados personalizados con cálculos complejos.
- Permitir cambiar la estructura de las dimensiones sin hacer un reproceso de los cubos.

- Permitir dimensiones cambiantes.
- Permitir hacer afinamientos de los cubos de acuerdo a criterios tales como: tipos de almacenamiento.
- Realizar actualizaciones automáticas de los cubos, sin necesidad de procesar su estructura.

4.2.2 Requerimientos no funcionales

- **Evolución de la plataforma:** La nueva tecnología debe ser compatible con la plataforma tecnológica con la que cuenta actualmente la empresa y sus futuras versiones.

Robustez tecnológica. La solución propuesta deberá ser estable y capaz de soportar la carga del sistema de información sin mayores sobresaltos operativos.

Concurrencia. El sistema deberá soportar a múltiples usuarios simultáneamente.

Escalable. La nueva solución deberá garantizar un buen rendimiento frente al crecimiento de la información.

Facilidad de implementación. Se deberán desarrollar soluciones o aplicaciones rápida y fácilmente implementables y administrables.

Interfaz Estándar. La implementación deberá tener una interfaz gráfica estándar, de acuerdo a todos los sistemas de la empresa, permitiendo a los usuarios una interacción fluida que reducirá los tiempos de entrenamiento, soporte y pruebas.

Acceso remoto. La solución propuesta debe permitir a los usuarios tener acceso a los datos remotamente.

Mantenible: que cuente con una documentación técnica completa de tal manera que el personal de tecnologías de información pueda entender su funcionamiento y hacer los ajustes necesarios.

Sistemas de seguridad. El sistema deberá controlar el acceso de los usuarios y garantizar la seguridad y confiabilidad de los datos.

Perfiles de usuario. Permitir la definición de la seguridad del sistema en base perfiles de usuario existentes actualmente en las bases de datos de la empresa.

Identificación del Usuario: Se debe establecer la identidad del usuario a lo largo del todo el sistema.

Parte 4

Solución propuesta

Capítulo 4 – Solución propuesta

4.1 Introducción

Gran parte de los productos generados por tecnologías de información no es precisamente información lo que entregan, sino datos en bruto, los que 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 añaden ideas y capacidad analítica.

En este capítulo se detallan las motivaciones y necesidades de contar con un sistema de apoyo a las decisiones y se justifica el uso del modelo utilizado. Además de sugerir mejoras a los problemas encontrados en la empresa.

4.1.1 Necesidad de información y conocimiento en la empresa

Desde que las organizaciones comenzaron a guardar los datos de sus operaciones en medios de almacenamiento físico, con el fin de permitirles una mayor administración y control de la información, ha existido una necesidad de utilizarla para atender las necesidades propias del negocio.

Actualmente, se considera a la información como el principal conocimiento que sostiene el negocio. Existen empresas que, de modo predominante, ofrecen servicios y giran su negocio principal sobre el manejo de la información (bancos, aseguradoras, Internet, etc.), en ellas es fácil identificar la importancia de la información, si no existiera ésta dejarían de existir. Sin embargo, hay otras en las que su giro principal es alrededor de la producción, en ellas la información debe identificarse para analizar y perfeccionar su producción (líneas de producción, distribución de materias, suministro, inventarios, publicidad, preferencias del

cliente, etc.). De hecho, en cualquier empresa se está tratando de transformar esa información en conocimiento, con la finalidad de mejorar los procesos dando ventajas competitivas a la empresa frente a la competencia.

Lo anterior llevó a la necesidad de dar respuesta rápida a los constantes requerimientos de información para la toma de decisiones, con la finalidad de ayudar a mejorar de alguna manera los procesos internos de negocio.

En la actualidad las empresas están apostando mucho por la tecnología y los individuos para que juntos tengan un conocimiento suficiente que acerque la visión interna de ambos a la realidad del entorno, en la medida que esa brecha disminuye, las decisiones tomadas se acercan más a la realidad exterior, generando decisiones más correctas y en menos tiempo.

“La importancia de una buena información puede ser vista como la diferencia en valor entre una decisión correcta y una decisión equivocada, en donde la decisión está basada en esa información. Mientras más grande sea esa diferencia entre decisión correcta y errónea, mayor será la importancia de contar con una buena información”. (BITAM,2005) ¹⁴

La información que fluye en una empresa está destinada a responder a diversos tipos de preguntas de sus usuarios, de ahí la necesidad que existan sistemas de información para requerimientos muy específicos que permitan la recolección y el manejo de datos.

Los sistemas de procesamiento de datos (OLTP) hacen uso de medios de almacenamiento y técnicas para poblarlos. La mayoría de las empresas, por la cantidad de información que manejan, se basan en los OLTP para guardar muchos datos y tener tiempos de respuesta cortos a los cientos de transacciones realizadas cotidianamente. La tecnología ha tenido que adaptar los medios con los que cuenta para que sean eficientes en el ámbito en que se requieren, tanto para el diseño de estructuras de datos que organizan la información como se desea, como para las herramientas o software que permite solucionar en tiempo y forma lo que el usuario demanda. Es importante resaltar que todos los sistemas de información tienen un fin muy particular, y se complementan para sostener, de la manera más eficiente, un

¹⁴ BITAM, Business Intelligence, (2005).

negocio, cabe señalar que no todos ellos pueden solucionar las distintas demandas que presentan los usuarios.

4.1.2 Tipo de información que necesitan las empresas

Hoy en día las empresas se han dado cuenta de la importancia de la calidad y certeza de la información para la toma de decisiones. Si analizamos el flujo de la información, éste recorre todos los niveles de la organización con diferentes fines (comunicación, control, administración, evaluación, etc.), es por ello que la responsabilidad al tomar decisiones recae sobre toda la organización ya que dentro de ésta existen más personas que toman decisiones y, a pesar de que éstas no tienen un impacto global, deben ser también correctas y oportunas, pues ciertos grupos dependen de las mismas.

Todos las personas dentro de la empresa que toman decisiones deben tener suficiente información para apoyarse en su trabajo cotidiano, el lugar que ocupen en la pirámide organizacional se vuelve secundario cuando el enfoque es hacia el manejo de procesos y todos los puestos tienen cierta relación y dependencia entre si.

4.2 Propuesta de desarrollo

La mayoría de las organizaciones cuentan con datos de los sistemas de ingreso de transacciones, vitales para registrar las operaciones que sostienen a una empresa. A pesar de la riqueza de estos datos, no se puede recurrir a ellos con facilidad cuando necesitamos encontrar respuestas sobre el estado de la organización, como por ejemplo:

- ¿Cuántos proveedores tienen una deuda con la empresa, o cuánto les adeuda la empresa a ellos a una determinada fecha?
- ¿Cuáles son los saldos de las cuentas contables que posee la empresa?
- ¿Cuáles son los saldos de los distintos análisis contables de una cuenta?
- ¿Cuáles son los productos más vendidos durante el 2005?

- ¿Qué proveedor vendió la mayor cantidad de metro ruma durante el primer trimestre del 2006?

Si bien con los sistemas tradicionales se pueden preparar reportes para encontrar las respuestas a algunas de estas preguntas de manera sencilla, se necesita mucho tiempo y recursos del departamento de sistemas para poder responderlas.

Además interfiere en el procesamiento de los sistemas transaccionales, aumentando los tiempos de respuestas de los mismos. Con sistemas inteligentes, se puede formular y responder las preguntas claves sobre el funcionamiento de la empresa, permitiendo encontrar factores que maximicen los beneficios.

Para esto es necesario implementar un sistema de apoyo a la toma de decisiones. El sistema debe estar orientado a brindar información interrelacionada para quienes tienen responsabilidades en el ámbito estratégico y táctico de la organización.

Un Sistema de apoyo a la toma de decisiones debe generar una variedad de productos de información, con el fin de satisfacer las necesidades cambiantes de las personas que toman decisiones a través de toda la empresa. Los encargados de tomar decisiones a nivel estratégico requieren informes, pronósticos e información más resumida. Por otra parte, es posible que las personas que toman decisiones a nivel táctico requieran informes especificados donde se haga énfasis en comparaciones detalladas de datos históricos y actuales que sustenten sus responsabilidades más estructuradas en las operaciones diarias.

Existen varios tipos de sistemas de información, los cuáles se deben integrar o fusionar en un sistema de apoyo gerencial y estos son:

- Sistemas de Información Gerencial (MIS)
- Sistemas de Apoyo a las Decisiones (DSS)

- Sistemas de Información Ejecutiva (EIS).

A continuación se presenta una tabla 3.1 muestra un resumen de las principales características de cada tipo de sistema.

Características	Tipo de sistemas		
	MIS	DSS	EIS
Apoyo a las decisiones que suministra.	Suministra información sobre el desempeño táctico de la información.	Suministra información y técnicas de apoyo a las decisiones para analizar problemas u oportunidades específicas.	Suministran información sobre el desempeño estratégico (factores claves) de la organización.
Forma y frecuencia de la información.	Respuesta e informes periódicos.	Consultas y respuestas Interactivas.	Consultas y respuestas Interactivas.
Formato de la información.	Formato fijo especificado con anterioridad	Consultas y respuestas interactivas	Formato ad-hoc, flexible y adaptable. Formato muy resumido y preferentemente gráfico.
Metodología de procesamiento de la información.	Información generada mediante extracción y manipulación de datos empresariales	Información generada mediante modelación analítica de los datos empresariales	Información generada mediante extracción y manipulación de datos empresariales como así También análisis de tendencias.

Tabla 3.1 Resumen de las principales características de cada tipo de sistema.

4.3 Análisis del problema

Una vez planteados los problemas y necesidades de la empresa se realizó un análisis que determinó dos grandes objetivos:

1. Crear un sistema que permita obtener el estado las cuentas corrientes de los distintos proveedores o clientes que posee la empresa en forma rápida y precisa, permitiendo ver el detalle de los movimientos de las cuentas a una fecha determinada con sus respectivos saldos, en otras palabras, simular una cartola de banco, la que debe estar actualizándose constantemente en base a las modificaciones que se realicen en el origen de datos, de modo de representar el estado de la cuenta de los distintos integrantes.
2. Sustituir las actuales tablas de saldos, mencionadas en el capítulo 2, por otra modalidad más certera que entregue los saldos de las cuentas y análisis contables, con sus respectivos movimientos a un período determinado, mostrando el detalle de éstas en profundidad para una determinada cuenta. Asimismo se quiere poder aumentar el detalle de búsquedas de datos a través de filtros de búsquedas de información sobre los conceptos y datos asociados a los análisis de una cuenta.

4.5 Solución propuesta

Para solucionar el punto 1 del ítem anterior se creará un cubo de tres dimensiones, el que se llamará "Cuentas corrientes", éste estará constituido por tres dimensiones que permitirán obtener información, por Rut y cuenta contable, de los abonos, cargos y saldos para un período determinado.

Para poder obtener los análisis anteriormente mencionados se creará la dimensión de "Ruts", que estará conformada por todos aquellos Rut que tenga movimientos contables asociados. Además, estará la dimensión de "Fecha", que contendrá todas las fechas en que se realizaron movimientos contables, y finalmente la de "Cuentas corrientes" que sólo contendrá las cuentas que tienen un Id o Rut asociado y que representan ingresos o costos para la empresa.

La información sobre la cual el cubo realizará sus cálculos corresponde a los montos ingresados en cada línea de un comprobante. Para calcular los saldos se ha definido el siguiente criterio:

- Los contadores de cada empresa deben especificar si la cuenta tendrá imputaciones al debe o al haber, lo que indicará si el monto corresponde a un cargo o a un abono para la cuenta. Generalmente las imputaciones al debe corresponden a cargos y las imputaciones al haber a abonos, pero existen ciertos casos extraordinarios donde esto no se cumple, por ello en base a la ventana de parámetros cuentas corrientes se definirá la clasificación de la imputación, datos que se tomarán como base para realizar los cálculos de saldos.
- Las medidas que contendrá el cubo serán tres: Los abonos, los cargos y el saldo que corresponde a la diferencia entre los abonos menos los cargos.

Para solucionar el problema de las tablas de saldos se creará otro cubo llamado “Análisis cuenta”, el que estará compuesto principalmente por todas las columnas de análisis y valores que componen la tabla de detalle del comprobante contable mostrada en el capítulo 2. Además existirá una dimensión de cuentas la cual contendrá todas las cuentas contables que hayan sido ingresadas al sistema y la dimensión de fechas que será compartida por ambos cubos ya que contendrá los mismos datos (Para ver la creación de este cubo dirigirse al anexo 2).

Este cubo estará compuesto por dos medidas: Montos debe y haber, que corresponde a los movimientos de un período de una cuenta, y por dos miembros calculados que corresponderán a los saldos débitos y créditos acumulados a una fecha.

Para el cálculo de éstos últimos se ha definido lo siguiente: Si la cuenta por la cual se desean ver los saldos es de activo o pasivo entonces se sumarán los movimientos desde el inicio de la contabilidad, en cambio si son de resultado, pérdida o ganancia, sólo se sumarán los movimientos del año. La diferencia entre el total de montos debe y haber para un período

indicará si el saldo es Deudor o Acreedor, en otras palabras, si consideramos a A como la suma de los movimientos al debe y a B como la suma de los movimientos al haber, entonces si $A - B > 0$ se considerará saldo deudor y saldo acreedor en caso contrario.

4.6 Tecnología a utilizar

Podemos definir varios tipos de estructuras OLAP, estas variarán de acuerdo a las técnicas que se utilicen a la hora de obtener los datos, la forma en la que se estructurarán para manejar la información, etc.

Entre los tipos de almacenamiento que proporciona OLAP se encuentran:

- Herramientas ROLAP (Relational On-line Analytical Process). "Son herramientas OLAP que crean vistas multidimensionales extrayendo los datos de bases de datos SQL ordinarias, es decir, relacionales. Estas herramientas simulan los datos multidimensionales usando sofisticadas técnicas de indexación, cachés, metadata, etc." (BORJA,2002) ¹⁵
- Herramientas MOLAP (Multidimensional On-line Analytical Process). "Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos. En otras palabras copian los datos del modelo relacional y construyen una nueva estructura pero multidimensional." (BORJA,2002) ²
- Herramientas HOLAP (Hybrid On-line Analytical Process). "Nos permiten un análisis híbrido de la información, es decir que une lo mejor de los dos tipos anteriores. El análisis HOLAP nos ayudará a reducir costes de hardware ya que necesitaremos menos disco que en las bases de datos relacionales. Además, la respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto

¹⁵ Borja R., La empresa multidimensional: OLAP (2002).

negativo, los datos multidimensionales deben ser cargados antes de ser consultados y refrescados cuando se actualizan los datos de la organización." (BORJA,2002) ¹⁶

Para tomar la decisión sobre que tipo de herramienta OLAP utilizar se partió de la base que constantemente se están realizando modificaciones en los datos de la empresa. Al seleccionar una herramienta MOLAP para implementar el cubo se tendría que estar procesando constantemente para que éste actualizara sus datos, ello implicaría un tiempo de espera para el usuario de alrededor de 1 minuto para el cubo de "Cuentas corrientes", y de 5 minutos para el cubo de "Análisis cuenta". Lo anterior afecta a cada cambio que se realice en los datos en los que se basa el OLAP para sacar información, resultando el sistema mucho más lento ante modificaciones que se realicen en el comprobante contable. Dado lo anterior, es aconsejable utilizar una herramienta que permita una actualización automática de los datos, evitando posibles errores ante fallas en las técnicas de actualización del cubo.

ROLAP permite dar una solución a todas las problemáticas anteriormente planteadas, ya que permite mantener los datos actualizados sin necesidad de crear procesos que se encarguen de ello, por ésta razón la solución se implementará mediante esta estrategia.

4.7 Modalidad de conexión

OLAP proporciona sus usuarios la posibilidad de conectarse directamente a los datos que contiene a través de distintas herramientas, como por ejemplo desde una planilla Excel. Para ello sólo basta cumplir con dos requerimientos:

1. Pertenecer a la red donde se encuentra el servidor de base de datos de la empresa.
2. Haber sido creado como miembro de administradores OLAP en el servidor de bases de datos.

¹⁶ Borja R., La empresa multidimensional: OLAP (2002).

Lo anterior trae ciertas complicaciones y limitantes que no permitirían sacar todo el provecho a las aplicaciones que puedan funcionar en base la información que entreguen los cubos, entre ellas podemos encontrar las siguientes:

1. No es posible acceder a los datos desde fuera de la red.
2. Los usuarios deben desarrollar conocimientos adicionales sobre consultas a cubos OLAP (Desarrollar consultas MDX), las cuales son muy complejas, considerando además que el poco conocimiento de estas puede provocar que las respuestas entregadas por la base de datos OLAP sean muy lentas.
3. La alta cantidad de información que pueden almacenar las estructuras multidimensionales, complicará la obtención de información por parte de los usuarios, pudiendo provocar su rechazo.

Dado lo anterior se han definido los siguientes criterios:

1. Sólo podrán conectarse al cubo aquellas personas que poseen cuentas de usuario en el servidor de base de datos de la empresa.
2. Las consultas al cubo deben realizarse a través de procedimientos almacenados, puesto que en esta modalidad no se producen errores de conexión como ocurre con otras aplicaciones, además son más eficientes y seguras ya que no es necesario abrir los puertos para admitir conexiones de clientes.
3. Los usuarios podrán conectarse al sistema desde cualquier punto, independientemente si se encuentran dentro o fuera de la red de la empresa.

4.8 Sugerencias al sistema actual.

4.8.1 Normalización de las tablas de comprobante contable.

Entre los problemas detectados se encuentra la desnormalización de las tablas de comprobante contable, para solucionar esto se propuso a la empresa considerar la opción de migrar a una

tabla que se encuentre en tercera forma normal, con la finalidad de hacer más rápidos los accesos a los datos y hacer más escalable el sistema.

La estructura propuesta cuenta de tres tablas, la primera llamada tcomprobantes_contables, mantiene su cantidad de atributos que posee actualmente, pero se sugiere modificar la clave compuesta existente, por una clave única que corresponde al número del comprobante contable.

Respecto a la tabla dlin_comprobantes_contables se eliminan las columnas de conceptos y valores, traspasándolas a otra tabla llamada dlin_analisis_comp_contables.

Entre los atributos de la tabla se encuentran: el número de comprobante contable al que pertenecen, la línea de comprobante, el análisis contable asociado a la cuenta, el concepto asociado al análisis y el valor asociado al concepto contable, si es que tiene uno asociado.

La representación del modelo sugerido es la siguiente:

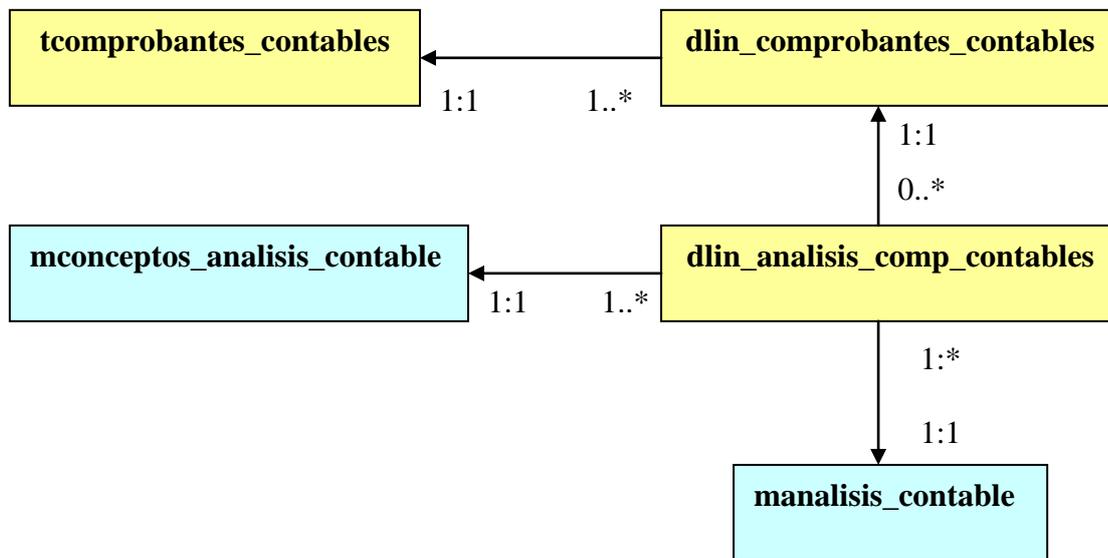


Figura 4.1, Vista de las tablas del comprobante contable normalizadas.

4.8.1.2 Campos de tabla

Entidad	Atributo	Tipo de dato	Clave	Clave foránea
tcomprobantes_contables	cod_comprobante_contables	Int	PK	
	cod_tipo_asiento	Int		FK
	fechr_comp_contable	Datetime		
	cod_ventana_origen	Int		FK
dlin_comprobantes_contables	cod_comprobante_contables	Int	PK PK	FK
	num_linea	Int		
	cod_cuenta	Int		FK
	glosa	Int		
	imputacion	Char		
	monto	Money		
dlin_analisis_comp_contables	<u>cod_comprobante contables</u>	Int	PK	FK
	<u>num linea</u>	Int	PK	
	<u>cod analisis contable</u>	Int	PK	FK
	cod_concepto_analisis_contable	Int		FK
	val_concepto_analisis_contable	Char		

Tabla 4.1, Muestra los atributos de las tablas con sus tipos de datos, claves primarias (PK) y claves foráneas (FK), los atributos que se encuentran en negrita corresponden a campos requeridos.

4.8.1.3 Definición de los atributos

- `cod_comprobante_contables`: Código del comprobante contable ingresado.
- `cod_tipo_asiento`: Indica si el comprobante es un traspaso, un ingreso o egreso. (Hace referencia a la tabla `mtipo_asiento`)
- `fechr_comp_contable`: Fecha en que se emitió el comprobante.
- `cod_ventana_origen`: Ventana en la cual se emitió el comprobante. (Referencia a la tabla `mventanas`)
- `num_linea`: Número de línea de un asiento en un comprobante contable.
- `cod_cuenta`: Cuenta contable que se verá involucrada en la generación del asiento, para una línea de comprobante contable. (Referencia a la tabla `mplan_cuentas`)
- `glosa` : Breve descripción de la operación comercial que se está realizando.
- `imputacion` : Indica si el asiento se realizó al debe o al haber.
- `monto` : Monto asignado a la cuenta.
- `cod_analisis_contable`: Código del análisis asociado a la cuenta contable. (Referencia a la ventana `manalisis_contable`).
- `cod_concepto_analisis_contable`: Código del concepto contable asociado a un análisis de una cuenta contable. (Referencia a la ventana `mconceptos_analisis_contable`).
- `val_concepto_analisis_contable`: Dato asociado al concepto contable.

Se debe aclarar que esta opción ha sido descartada a corto plazo por la empresa, debido a los altos costos que tendría la modificación de las estructuras de las tablas utilizadas en el comprobante contable, ya que más del 60% de las ventanas del sistema las utilizan.

Parte 5

Pruebas

Capítulo 5 – Pruebas del sistema

5.1 Introducción

El siguiente capítulo tiene por finalidad mostrar el rendimiento del servidor ante las consultas realizadas a la base de datos OLAP. La idea es verificar que las respuestas sean entregadas en un tiempo prudente, además de evaluar la calidad de la información que se entrega.

5.1.1 Pruebas de integridad de la información

Las siguientes pruebas fueron hechas sobre datos ficticios dispuestos en base de datos de pruebas existentes en los distintos servidores. La idea es demostrar que no hay retardos en la actualización de la base de datos al realizar inserciones, modificaciones y eliminaciones.

Los siguientes resultados fueron verificados sobre el cubo de cuentas corrientes y comparados con los entregados por las consultas a las base de datos implementadas anteriormente en TDIFOR.

Pruebas realizadas al Servidor A, cubo cuentas corrientes.

N°	Nombre	Entrada	Saldo Esperado	Saldo Real
1	Inserción de un comprobante contable para el proveedor Andrés Caro, por \$500.000 el día 12 de diciembre de 2005.	500.000 pesos	853.000 pesos	853.000 pesos
2	Modificación de un comprobante contable para el cliente Abraham Urrutia, el día 12 de diciembre del 2005, intervienen los análisis de Id o rut, factura proveedor.	4.000.000 pesos	0 pesos	0 pesos
3	Eliminación de un comprobante contable ingresado el día 11 de noviembre del 2005, cliente involucrado Forestal el Bosque.	1.500.000 pesos	-1.500.000 pesos	-1.500.000 pesos

Pruebas realizadas al Servidor A, cubo análisis cuenta.

N°	Nombre	Entrada	Saldo Esperado	Saldo Real
1	Inserción de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por 500.000 pesos	Saldo crédito \$ 12.853.000	Saldo crédito \$12.853.000
2	Modificación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por 100.000 pesos	Saldo crédito \$ 12.153.000	Saldo crédito \$ 12.153.000
3	Eliminación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Saldo crédito inicial \$ 12.153.000	Saldo crédito \$ 0	Saldo crédito \$ 0

Pruebas realizadas al Servidor A, cubo análisis cuenta.

N°	Nombre	Entrada	Saldo Esperado	Saldo Real
1	Inserción de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por 500.000 pesos	Saldo crédito \$ 12.853.000	Saldo crédito \$12.853.000
2	Modificación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por 100.000 pesos	Saldo crédito \$ 12.153.000	Saldo crédito \$ 12.153.000
3	Eliminación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Saldo crédito inicial \$ 800.000	Saldo crédito \$11.353.000	Saldo crédito \$ 11.353.000

Pruebas realizadas al Servidor B, cubo análisis cuenta.

N°	Nombre	Entrada	Saldo Esperado	Saldo Real
1	Inserción de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por 750.000 pesos	Saldo crédito \$ 15.360.000	Saldo crédito \$15.360.000
2	Modificación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por 300.000 pesos	Saldo crédito \$ 12.053.000	Saldo crédito \$ 12.053.000
3	Eliminación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Saldo crédito inicial \$ 11.758.000	Saldo crédito \$ 273.000	Saldo crédito \$ 273.000

Pruebas realizadas al Servidor C, cubo análisis cuenta.

N°	Nombre	Entrada	Saldo Esperado	Saldo Real
1	Inserción de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por \$10.000.000	Saldo crédito \$ 19.005.000	Saldo crédito \$19.005.000
2	Modificación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Imputación al haber por 1.000.000 pesos	Saldo crédito \$ 10.005.000	Saldo crédito \$ 10.005.000
3	Eliminación de un comprobante contable para la cuenta Facturas por pagar, intervienen los análisis Id o rut y documento sistema.	Saldo crédito inicial \$ 1.000.000	Saldo crédito \$ 9.005.000	Saldo crédito \$ 9.005.000

Pruebas realizadas al Servidor B, cubo cuentas corrientes.

N°	Nombre	Entrada	Saldo Esperado	Saldo Real
1	Inserción de un comprobante contable para el proveedor Maderas Pino, por \$453.333 el día 6 de enero del 2005.	\$ 453.333	\$ 956.333	\$ 956.333
2	Modificación de un comprobante contable para el cliente Forestal Austral, el día 9 de diciembre del 2005, intervienen los análisis de Id o rut, factura proveedor.	\$5.111.111	\$ 111.111	\$ 111.111
3	Eliminación de un comprobante contable ingresado el día 15 de enero del 2005, cliente involucrado Maderas Pino.	\$ 100.000	\$ -11.111	\$ -11.111

Pruebas realizadas al Servidor C, cubo cuentas corrientes.

N°	Nombre	Entrada	Saldo Esperado	Saldo Real
1	Inserción de un comprobante contable para el proveedor Cristian Ahumada, por \$453.333 el día 11 de marzo del 2005.	\$ 1.953.333	\$ 333	\$ 333
2	Modificación de un comprobante contable para el cliente Forestal Austral, el día 15 de noviembre del 2005, intervienen los análisis de Id o rut, factura venta exenta.	\$ 500.000	\$ 0	\$ 0
3	Eliminación de un comprobante contable ingresado el día 15 de enero del 2005, cliente involucrado Maderas Pino.	\$ 7.500.000	\$ -7.500.000	\$ -7.500.000

5.1.2 Pruebas de rendimiento.

Análisis de rendimiento al Servidor A, cubo análisis cuenta.

N°	Datos de la prueba	Tiempo esperado	Tiempo Real
1	Generación de informe que entrega los saldos de la cuenta Facturas por pagar, abiertas por los análisis Id o Rut, documento sistema, al 12 de diciembre del 2005.	Menos de 1 min.	43. seg.
2	Generación de saldos de los análisis de la cuenta Costo de ventas, abierta por Centro de costos, Negocios, Item Gastos, al 31 de enero 2006.	Menos de 1 min.	12 Seg.
3	Obtención de saldos de los análisis banco en transito, abierto por los análisis conceptos banca, categoría flujo efectivo, flujo efectivo, al 1 de marzo del 2006.	Menos de 1 min.	15 Seg.

Análisis de rendimiento al Servidor B

N°	Datos de la prueba	Tiempo esperado	Tiempo Real
1	Generación de informe que entrega los saldos de la cuenta Facturas por pagar, abiertas por los análisis Id o Rut, documento sistema, al 12 de diciembre del 2005.	Menos de 1 min.	45. seg.
2	Generación de saldos de los análisis de la cuenta Costo de ventas, abierta por Centro de costos, Negocios, Item Gastos, al 31 de enero 2006.	Menos de 1 min.	11 Seg.
3	Obtención de saldos de los análisis banco en tránsito, abierto por los análisis conceptos banca, categoría flujo efectivo, flujo efectivo, al 1 de marzo del 2006.	Menos de 1 min.	13 Seg.

Análisis de rendimiento al Servidor C

Nº	Datos de la prueba	Tiempo esperado	Tiempo Real
1	Generación de informe que entrega los saldos de la cuenta Facturas por pagar, abiertas por los análisis Id o Rut, documento sistema, al 12 de diciembre del 2005.	Menos de 1 min.	45. seg.
2	Generación de saldos de los análisis de la cuenta Costo de ventas, abierta por Centro de costos, Negocios, Item Gastos, al 31 de enero 2006.	Menos de 1 min.	10 Seg.
3	Obtención de saldos de los análisis banco en transito, abierto por los análisis conceptos banca, categoría flujo efectivo, flujo efectivo, al 1 de marzo del 2006.	Menos de 1 min.	12 Seg.

Parte 6

Conclusiones

Capítulo 6 – Conclusiones

6.1 Introducción

En esta tesis se analizaron las opciones existentes para transformar un conjunto de datos en conocimiento, entregando un aporte real a la toma de decisiones en base a los recursos con los que cuenta la empresa. Para realizar esta tarea se especificaron una serie de requerimientos y restricciones, entre las que se encontraban problemas originados por las malas decisiones tomadas por el personal de TI al inicio del desarrollo de su proyecto (GCC).

En este capítulo se dan a conocer las conclusiones luego de implementar la solución de descrita. Éstas se han organizado

6.2 Ventajas y desventajas de la herramienta SQL Server Analysis Service (SSAS)

SQL Server Analysis Service 2000, es una herramienta muy potente en cuanto a la entrega de resúmenes de datos, ya que en un tiempo menor, se pueden crear resúmenes de altos volúmenes de información. Es importante destacar que para sacar el máximo de provecho a esta herramientas, en conjunto con asegurar la calidad de la información que ésta genera, es necesario contar con tablas cuyas claves principales estén correctamente definidas y los datos ingresados sean correctos.

Entre las desventajas encontradas en ésta herramienta se encuentran ciertas complicaciones que se presentan al tratar de crear dimensiones teniendo una tabla de origen cuya clave primaria sea compuesta, ya que la propiedad que éste tiene para identificar la clave primaria de ésta estructura, sólo admite la selección de un campo. Generalmente se recomienda antes de crear un proyecto del tipo Business intelligence, contar con tablas que estén formadas por

una clave única, esto principalmente porque las búsquedas de datos son más eficientes que al contar con una clave compuesta.

Por otra parte las estructuras ROLAP traen un mayor número de restricciones respecto a las MOLAP, uno de los más importantes y por lo cual se tuvo que dejar todas las dimensiones con un solo nivel, es la falta de la propiedad “Group members”, la cual permite agrupar los miembros de una dimensión, siempre que el número de éstos sea mayor a 64.000. Lo cual mejora los tiempos de búsqueda y extracción de datos.

Es importante destacar que la alta capacidad analítica de Analysis Service, permitió a la empresa detectar y corregir ciertas inconsistencias en la información, provocadas por la falta de validaciones en el sistema que desarrolla actualmente, las cuales habían sido agregadas en versiones posteriores sin investigar previamente si los datos ingresados hasta esa fecha se encontraban correctos, produciendo descuadres en determinados análisis de una cuenta.

Por último, se debe hacer énfasis en que para sacar el máximo de potencial a éste tipo de herramientas es imprescindible contar con el apoyo del personal especializado en los procesos de la empresa sobre los cuales trabajará el cubo, esto principalmente por que el enfoque que se les puede dar a los resúmenes puede ser mucho más especializado, además de permitir hacer consultas más eficientes al agregar filtros que permitan restringirlas a lo que el usuario realmente desea saber.

6.3 Recomendaciones

El proceso de transformar datos en conocimiento es muy complejo, sobre todo si no se cuenta para su desarrollo con personal especializado que tenga claro que es lo que se necesita y se quiere lograr. Una vez listos estos puntos, se debe estudiar completamente la situación actual de la empresa, ya que esto nos entregará las respuestas respecto a si es necesario o no crear un

cubo para mostrar resúmenes de datos, o sólo basta con mejorar las técnicas de extracción de datos de la base de datos transaccional.

A continuación se presentan algunas consideraciones que se deben tener en cuenta para llevar a cabo un proyecto OLAP.

Antes de crear una estructura multidimensional es importante hacer un análisis de las tablas del modelo relacional sobre las cuales trabajará el cubo. Especialmente sobre la existencia de claves primarias y foráneas que permitan unicidad y consistencia en la información que se va a resumir.

Además se debe considerar que si las dimensiones tienden a tener un alto nivel de miembros (mayor a 100.000) se debe volver a analizar la idea de implementar un cubo OLAP, ya que en lugar un análisis en estructuras multidimensionales, tal ves solo requiere de una consulta SQL a una base de datos transaccional.

Es importante mencionar que algunos DBA sugieren desnormalizar las tablas sobre las cuales trabajará un cubo, la idea es crear una imagen que se adapte a la estructura que tendrá éste, con la finalidad de hacer más eficientes las consultas, al tener la información de una dimensión más resumida en tablas independientes.

Otra observación importante es la búsqueda de índices correctos en las distintas tablas, ya que en diversas pruebas se comprobó las notables diferencias de rendimiento de las consultas MDX al contar con tablas indexadas.

Pero lo principal es que para lograr mejores resultados se debe contar con el apoyo de personal especializado en el área sobre la cual trabajará el cubo, para comprobar que la información que refleja éste es correcta, así como también para analizar los datos que se entrega y a partir de ellos inferir nuevas ideas respecto a que tipo de conocimiento sería útil para la empresa.

6.4 Resultados obtenidos.

Los logros alcanzados con los informes generados han cumplido satisfactoriamente con las expectativas, ya que se han logrado realizar análisis que hasta un tiempo eran imposibles. La flexibilidad de la información generada ha llevado a descubrir nuevas formas de ver los datos, obteniendo grandes aportes para la toma de decisiones, sin considerar que los tiempos de ahorro en la generación de resúmenes se han disminuido en algunas de las empresas, hasta en dos días.

Otro punto importante de destacar es la personalización de los informes, ya que se generan dinámicamente en base a los requerimientos de los usuarios, esto principalmente porque es difícil manejar este tipo de estructuras, sobre todo si se considera la poca y casi nula experiencia de los usuarios en éste tema, lo que los hace mucho más amigables y simples de entender siendo éste precisamente el objetivo principal del proyecto.

Con respecto al tema de seguridad se cumplió con lo estándares establecidos por la empresa en el desarrollo existente, ya que se creó un método de conexión que no requería abrir nuevos puertos de comunicación, además de funcionar con el sistema de reconocimiento de usuarios y permisos implementado por ésta.

Cabe mencionar que los aportes que se realizaron en éste proyecto no sólo se redujeron a la emisión de informes, sino que también se realizaron mejoras en algunos de los procesos actuales de la empresa, con la finalidad de evitar posibles errores en el manejo de los datos y disminuir la cantidad de trabajo del servidor de base de datos, por otra parte se hicieron observaciones sobre la estructura de la tabla que contiene la información contable, dejando en claro las desventajas de contar con una tabla desnormalizada y la baja escalabilidad que tiene el proyecto al contar con esta estructura.

6.5 Objetivos futuros

Los aportes realizados con este trabajo, ha llevado a crear nuevos proyectos en los cuales se puedan aprovechar las ventajas de las herramientas OLAP, la idea es crear cubos que procesen la información de los distintos módulos con los que cuentan las empresas, en otras palabras contar con resúmenes especializados, de modo de hacer más eficientes los procesos de captura de datos.

Actualmente se comenzó el desarrollo de un cubo que controle la compra y venta de Metro Ruma en los distintos centros de compras distribuidos a lo largo del país, posteriormente se pretende implementar un cubo que controle las exportaciones y el stock disponible de paquetes de madera.

Para terminar, la exactitud de la información que entregan los cubos ha llevado a masificar su uso dentro la empresa, además que el darle enfoques desconocidos a los datos ha traído nuevas ideas para el negocio, permitiendo en algunos casos sacar deducciones que hasta hace poco tiempo no eran posibles a simple vista o requerían demasiado tiempo.

Anexo 1

Creación del cubo OLAP Cuentas corrientes.

Anexo 1 – Creación del cubo OLAP Cuentas corrientes

1.1 Introducción

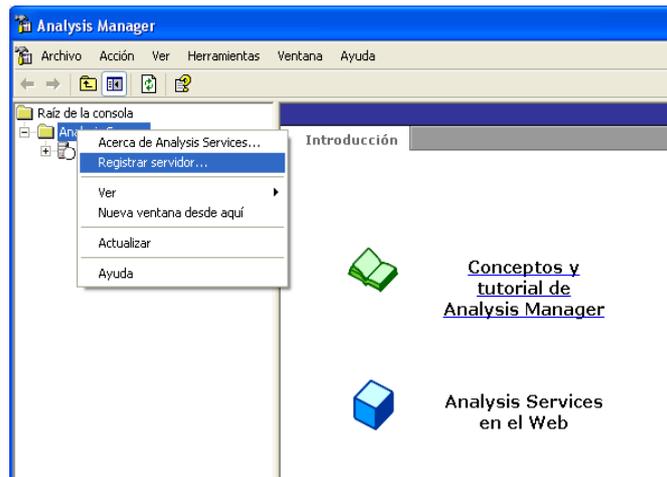
Este cubo tiene la finalidad emitir resúmenes de los saldos de las cuentas corrientes de los distintos integrantes de la empresa a una fecha determina. La idea es simular las cartolas de los Bancos permitiendo a la empresa tener claro cuanto adeuda a un proveedor o cuánto éste le adeuda a la empresa.

1.1.1 Configuración del origen de datos.

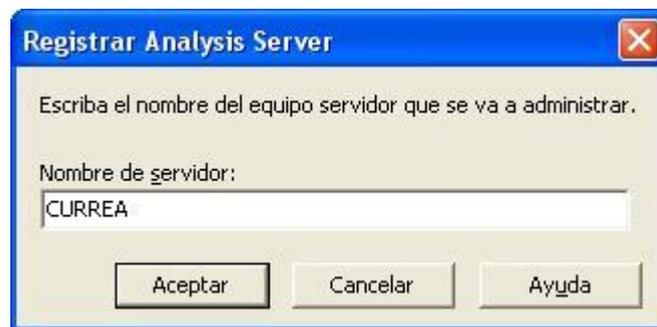
Para crear el cubo debe dirigirse a la barra de inicio, Programas, Microsoft SQL Server, Analysis Services, Análisis Manager.



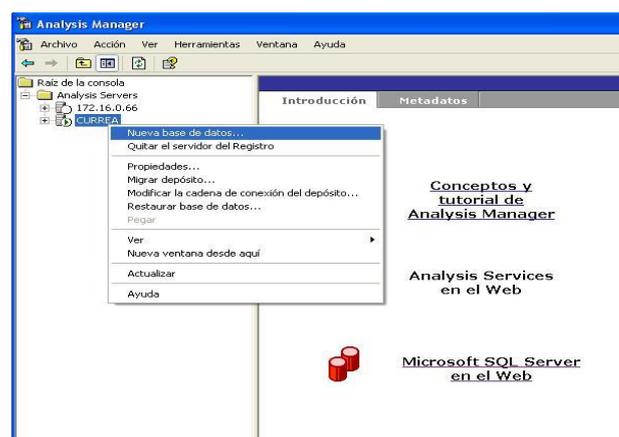
Si no tiene agregado ningún servidor registrado, con el botón derecho del Mouse seleccionar “Registrar Servidor”.



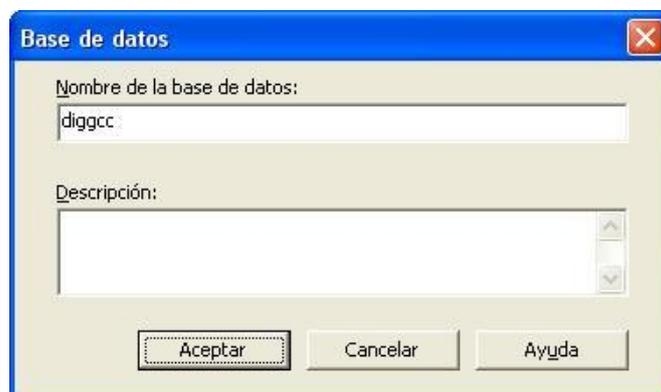
Luego aparecerá el siguiente mensaje, en el cual debe ingresar el nombre del servidor del cual desea obtener la información. Por ejemplo: CURREA



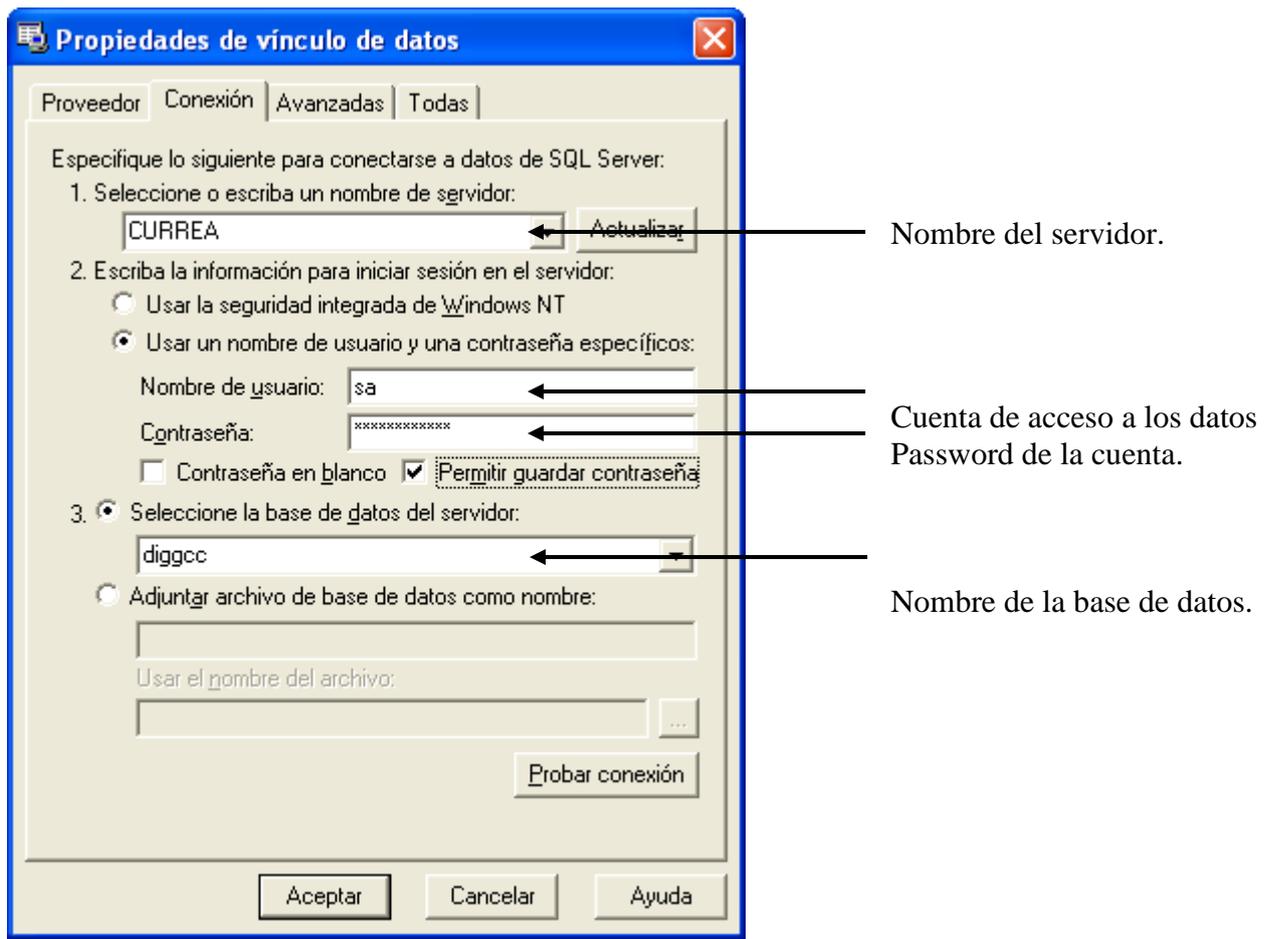
Si no aparece ninguna base de datos asociada al servidor, con el botón derecho del Mouse sobre el servidor seleccione “Nueva base de datos”



Posteriormente aparecerá el siguiente mensaje donde debe indicar el nombre de la base de datos a la cual quiere hacer referencia. Por ejemplo: diggcc, tifgcc, etc.



Una vez creada la base de datos se debe configurar el origen de datos, para ello es necesario verificar que los siguientes campos estén correctamente ingresados.



Posteriormente presionar “Aceptar”.

1.1.2 Creación de las dimensiones del cubo

Una vez terminada la etapa de configuración de la base de datos, se procede a crear el cubo. Lo primero que se debe crear son las dimensiones, ya que éstas contienen la información por la cual se quieren analizar los saldos resumidos.



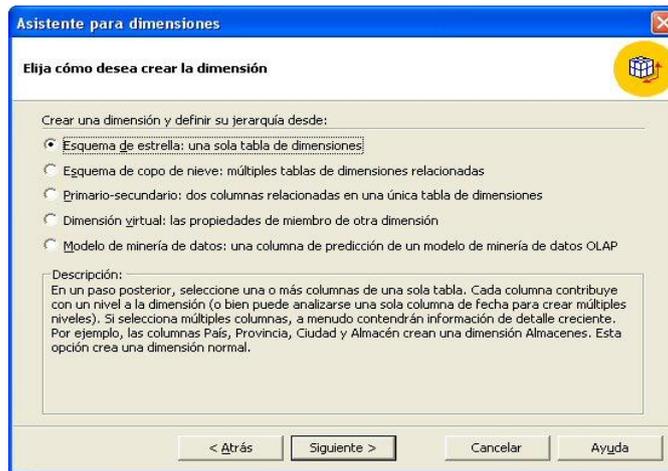
El siguiente paso es crear las dimensiones del cubo. Estas son categorías descriptivas que permitirán clasificar la información por distintos criterios, para su posterior análisis.

1.1.2.1 Creación de la dimensión “Cuentas corrientes”.

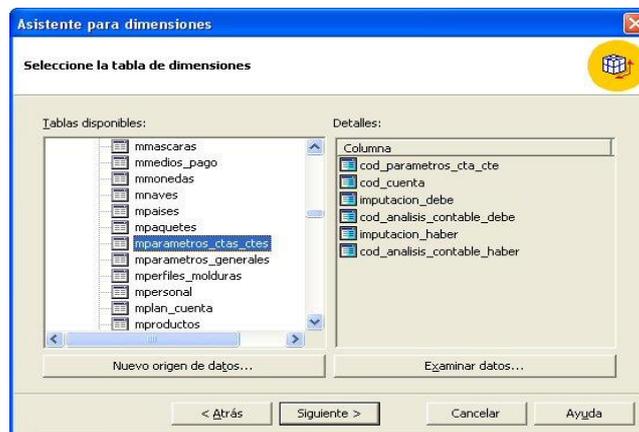
La dimensión de Cuentas corrientes estará compuesta por todos aquellos registros que hayan sido ingresados en la ventana de parámetros cuentas corrientes, los que corresponden a las cuentas que tienen asociado un ID o Rut y además generan costo o ingresos para la empresa.



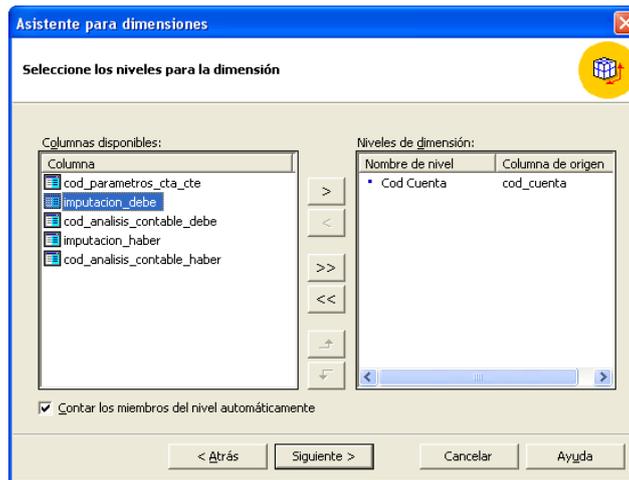
El esquema a utilizar será tipo estrella, ya que la información se obtendrá de una sola tabla.



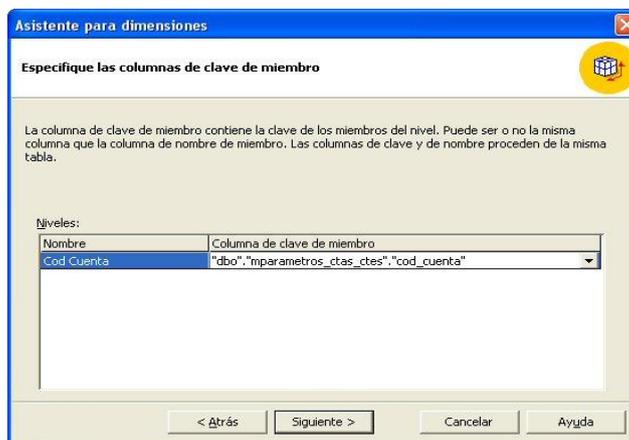
La tabla de la cual se obtendrá la información para la dimensión será mparametros_ctas_ctes, ya que contiene sólo las cuentas que tienen un análisis de Id o rut asociado.



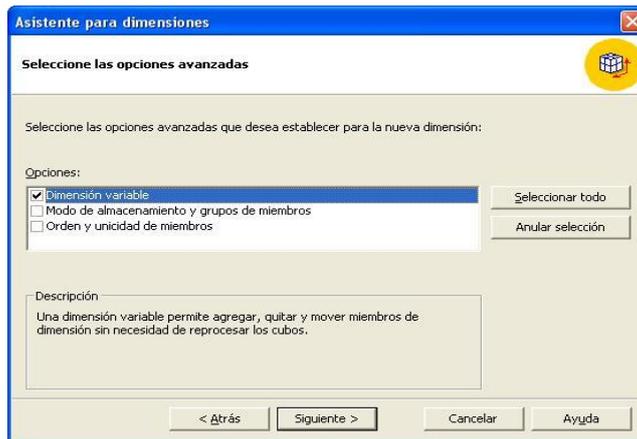
El siguiente paso es seleccionar los niveles de la dimensión, para este caso la dimensión tendrá un solo nivel compuesto por los códigos de las cuentas (cod_cuenta).



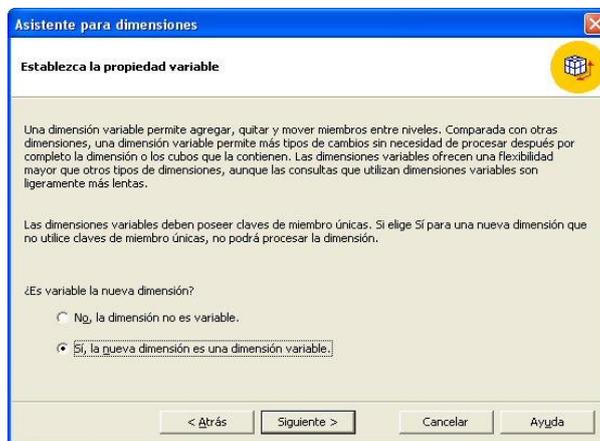
La columna clave del miembro será el mismo código de la cuenta, puesto que permitirá relacionar la dimensión con la tabla de hechos del cubo.



La característica que tendrá esta dimensión es la de ser variable, ya que esto permite la realización de cambios sin la posterior necesidad de procesar por completo la dimensión o los cubos que la contienen.



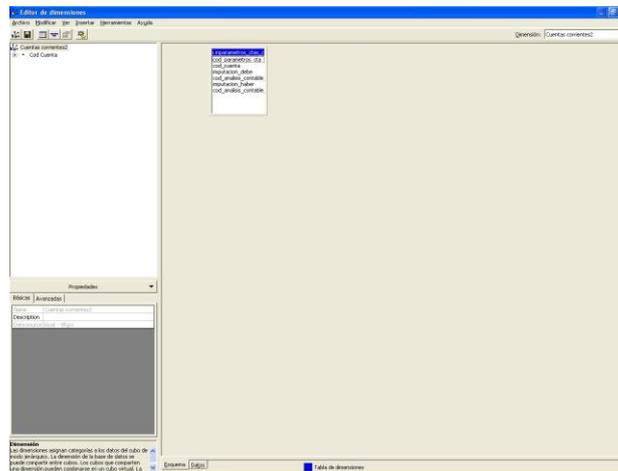
La siguiente opción permite establecer la dimensión como variable, siempre que ésta contenga claves de miembro únicas.



El siguiente paso es indicar el nombre de la dimensión, para este caso "Cuentas corrientes".



Luego se abrirá un menú que permitirá configurar la dimensión de acuerdo a nuestros requerimientos.



Para poder determinar si las imputaciones de la cuenta están clasificadas como cargos o abonos es necesario configurar las propiedades de los miembros de la dimensión.



Una vez configurada la dimensión debe ser guardada.



Posteriormente se procede a procesar el cubo.



Aunque la opción actualización incremental aparece bloqueada, es necesario hacer notar que siempre que se haga un procesamiento del cubo se debe elegir la opción “Volver a generar la estructura de la dimensión”, puesto que la primera duplica la información provocando posibles errores en los cálculos del cubo.

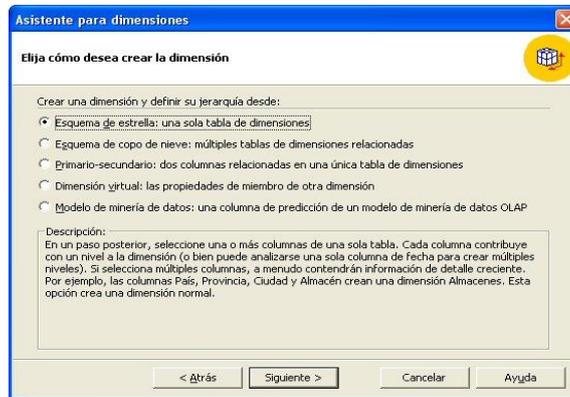


Para que el reprocesamiento sea satisfactorio debe aparecer la siguiente ventana con el mensaje de “Proceso completado correctamente”.

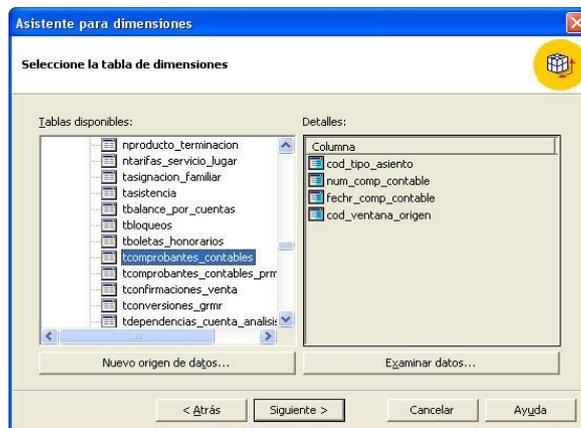


1.1.2.2 Creación de la dimensión “Fecha”

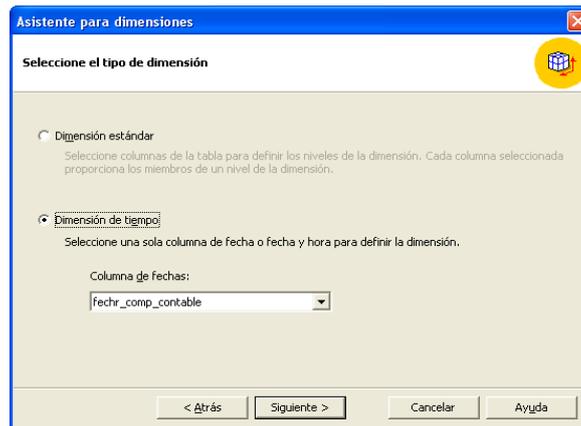
Esta dimensión contendrá todas las fechas en las que se registraron movimientos contables. La información será obtenida de una sola tabla por lo que el esquema a utilizar será tipo estrella.



La tabla de cual sacará información la dimensión de fechas será tcomprobantes_contables.



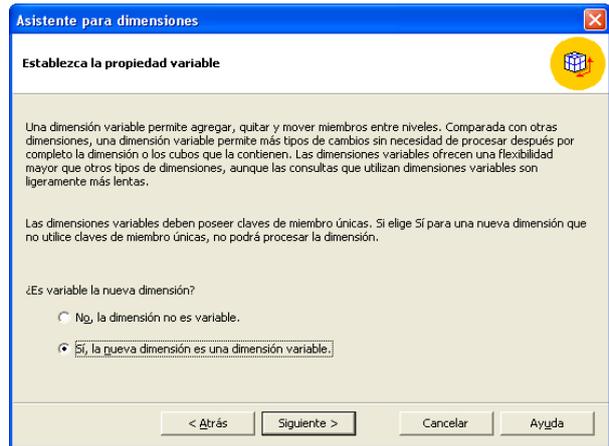
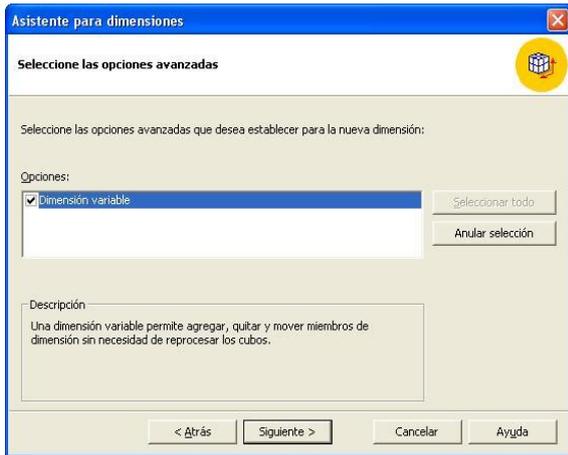
Para crear la fecha se debe seleccionar “Dimensión de tiempo”, esta opción creará automáticamente los distintos niveles de información que contiene esta dimensión, la columna de la cual se sacará la información es fechr_comp_contable, para este caso la única columna de fecha de la tabla.



Luego se debe seleccionar los niveles en los que se desea ver la información, por ejemplo: Año – Mes – Día; Año – Mes, etc. Para este caso se debe seleccionar Año- Trimestre – Mes – Día.



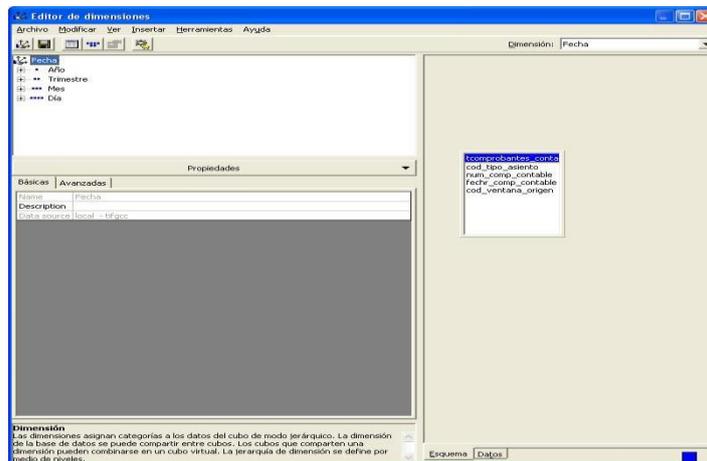
Ya que habitualmente se están ingresando comprobantes contables, es necesario crear la dimensión como variable, pues como se indicó anteriormente esto permitirá que se actualice automáticamente.



En la siguiente pantalla debe ingresar el nombre de la dimensión para este caso será “Fecha”.



Luego se deben configurar las propiedades de la dimensión.



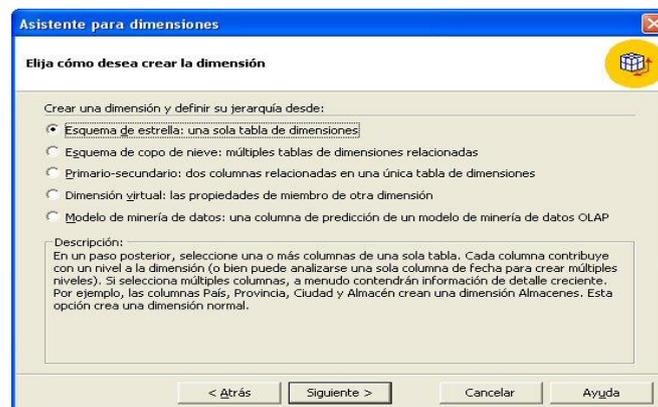
Para este tipo de dimensión los miembros no son únicos por lo que la propiedad Member keys unique debe configurarse en falso, a diferencia de la dimensión de Cuentas corrientes.



Luego se debe guardar la nueva dimensión y se procesa (Repetir pasos de la página 88).

1.1.2.3 Creación de la dimensión “Rut”

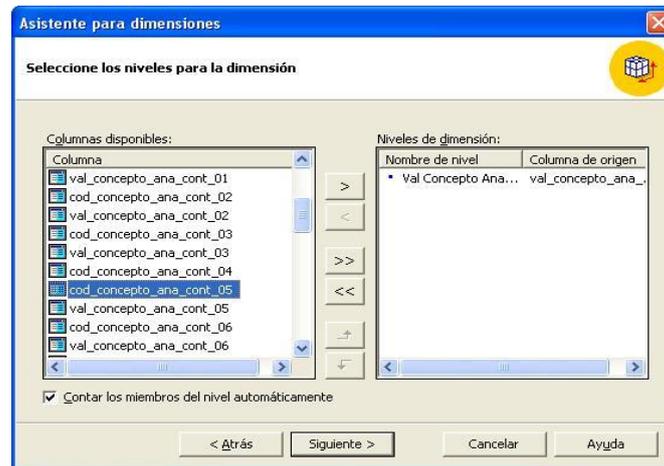
Esta dimensión estará conformada por todos los Id o rut que tengan movimientos en las tablas de comprobantes contables. Al igual que las dimensiones anteriormente creadas su estructura será tipo estrella ya que saca la información de una sola tabla de hechos.



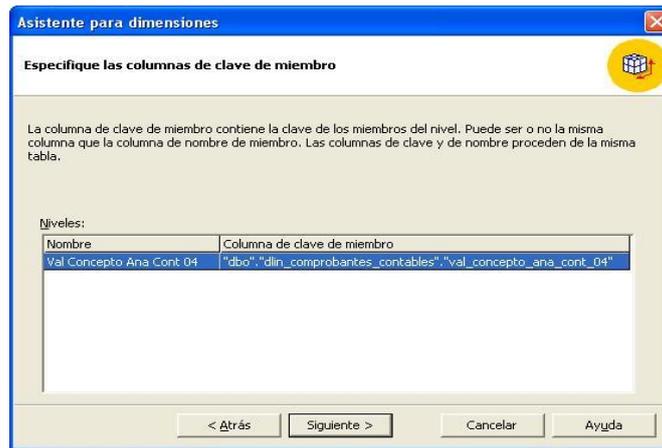
La tabla de origen de los datos será la dlin_comprobantes_contables.



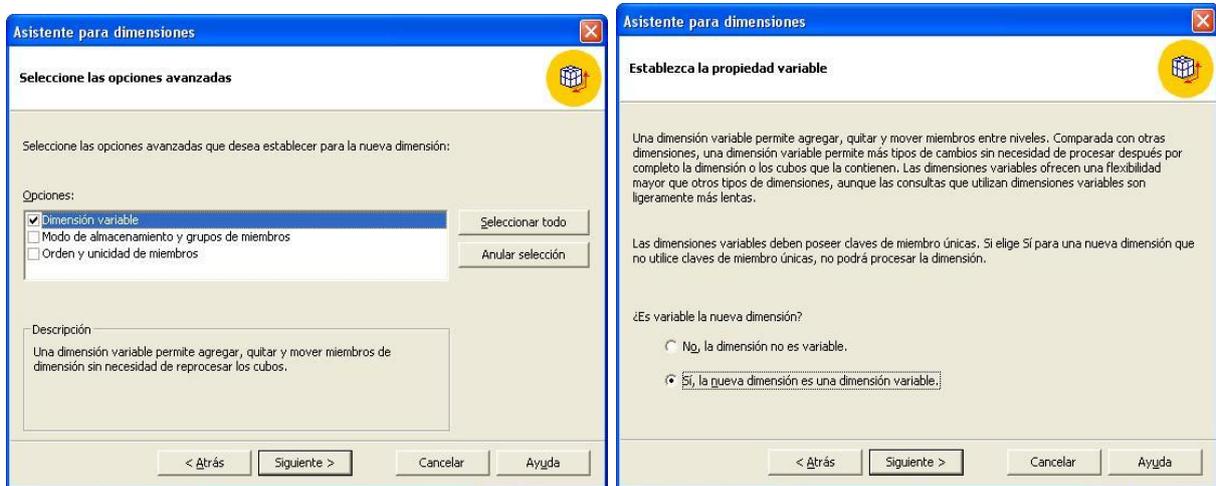
La siguiente etapa es seleccionar el nivel de la dimensión, pero en este caso se debe hacer un análisis mayor sobre cuál es la columna en la que se basará la información de nuestra dimensión ya que esta no es igual para todas las bases de datos, esto debido a que el análisis de Id o rut no ha sido creado en todas las empresas en la misma columna, por ejemplo si creamos el cubo cuentas corrientes para Forestal Diguillín el valor del concepto que deseamos almacenar se encontrará en la columna val_concepto_ana_cont_04, en cambio si estamos en Andenpine el valor del concepto a almacenar se encuentra en la columna val_concepto_ana_cont_10. Por ello antes de crear esta dimensión se debe consultar los distintos análisis que conforman la tabla para identificar cuál es el que contiene la máscara de Id o rut, una vez realizada esta operación se procede a crear la dimensión.



En este caso la clave del miembro es el mismo Id o rut por lo que no hay modificaciones en esta pantalla.



En la siguiente pantalla se debe seleccionar la propiedad de dimensión variable.



En la siguiente pantalla debe ingresar el nombre de la dimensión para este caso "Rut".



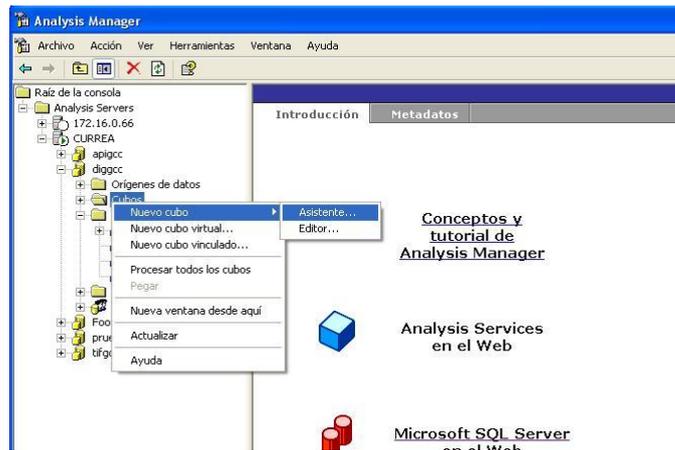
La siguiente etapa es configurar las propiedades de la dimensión.

Básicas	Avanzadas
All Level	Yes
All Caption	Todas Rut
Type	Standard
Default Member	
Depends On Dimension	(None)
Changing	True
Write-enabled	False
Member Keys Unique	True
Member Names Unique	False
Allow Duplicate Names	False
Source Table Filter	
Storage Mode	ROLAP
Enable Real-Time Updates	True
Virtual	False
All Member Formula	

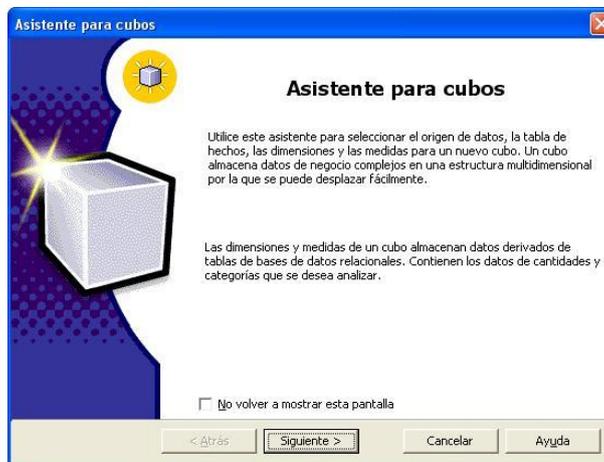
Posteriormente se debe guardar la dimensión y procesar la estructura (Repetir pasos de la página 88).

1.1.3 Creación del cubo “Cuentas corrientes”.

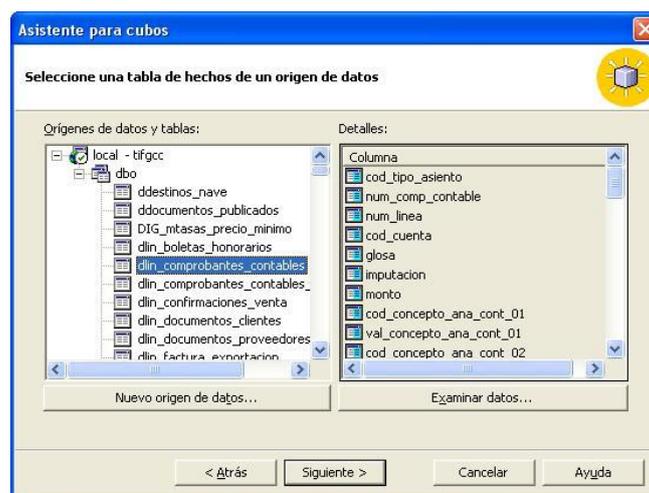
Con el botón derecho sobre la opción cubos seleccione “Nuevo cubo”, “Asistente”.



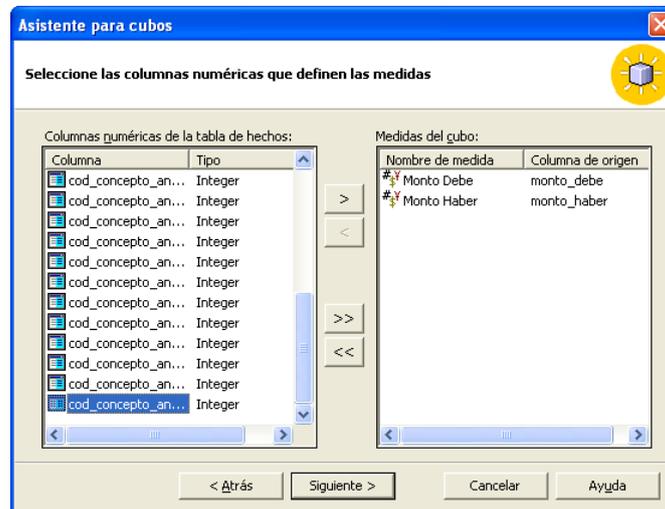
Luego aparecerá la siguiente pantalla para comenzar con la configuración del cubo.



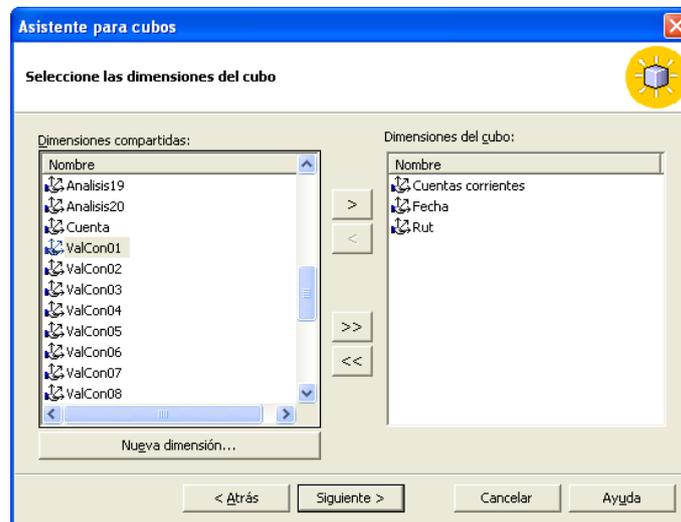
Luego debe seleccionar la tabla de hechos, para este caso dlin_comprobantes_contables.



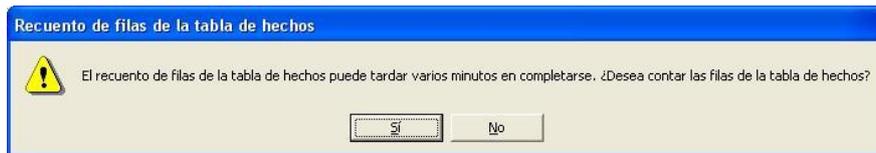
En el siguiente paso se debe seleccionar las columnas numéricas de las cuales se quiere sacar información. Para nuestro caso la columna “monto_debe” y “monto_haber”, a partir del cual se calcularán los saldos asociados a los distintos Id o rut existentes en la dimensión anteriormente creada.



Luego se seleccionan todas las dimensiones que se utilizarán para calcular las distintas agregaciones, para nuestro caso se deben incluir las dimensiones “Fecha”, “Cuentas corrientes” y “Rut”.



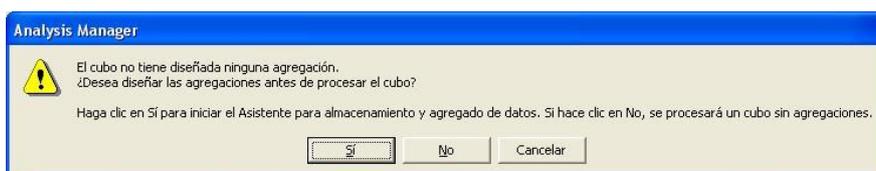
Luego aparecerá el siguiente mensaje en el que se debe seleccionar “Sí”.



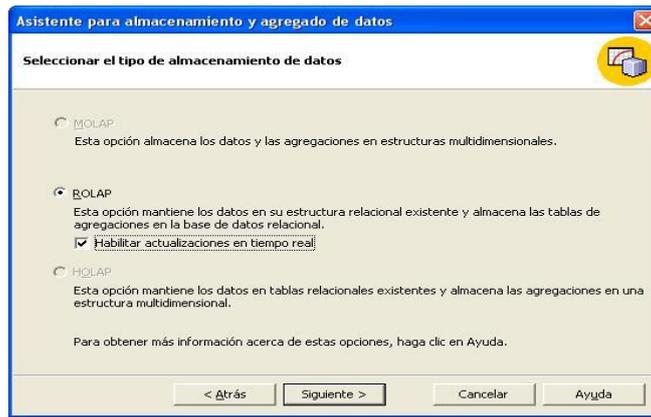
Una vez contadas las filas de las tablas de hechos se debe dar nombre al cubo “Cuentas corrientes”.



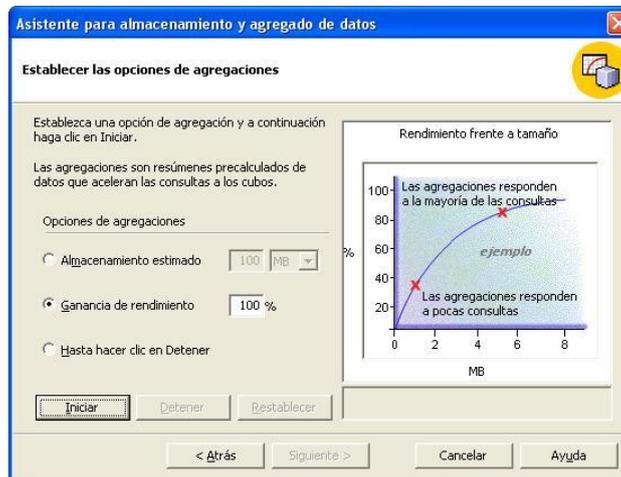
La siguiente etapa es configurar las propiedades del cubo y crear los miembros que a partir de la información de la dimensión de medidas permita calcular los saldos, pero previo a ello se debe procesar el cubo seleccionando la opción del menú “Herramientas”, “Procesar cubo”.



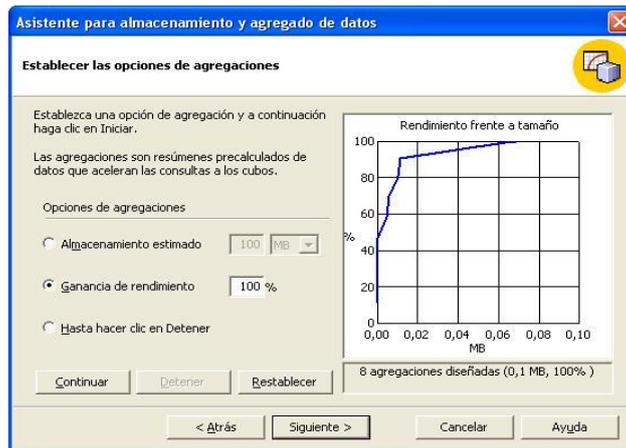
El tipo de almacenamiento por defecto es ROLAP, el cual permitirá mantener la información actualizada ante cualquier cambio que se pueda realizar desde el sistema.



Con respecto a las opciones de agregaciones se le da mayor importancia al rendimiento, desechando la opción de almacenamiento estimado en tamaño. Dado lo anterior se debe ingresar un rendimiento de 100% y presionar aceptar.



Una vez calculadas a las agregaciones aparecerá un gráfico que indicará el rendimiento del cubo frente al tamaño.

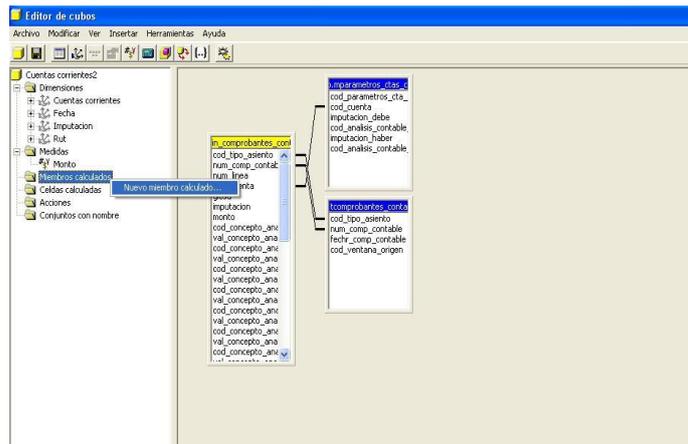


Una vez terminada la etapa de procesamiento se debe seleccionar “Procesar ahora”.



1.1.4 Proceso de creación de los miembros calculados

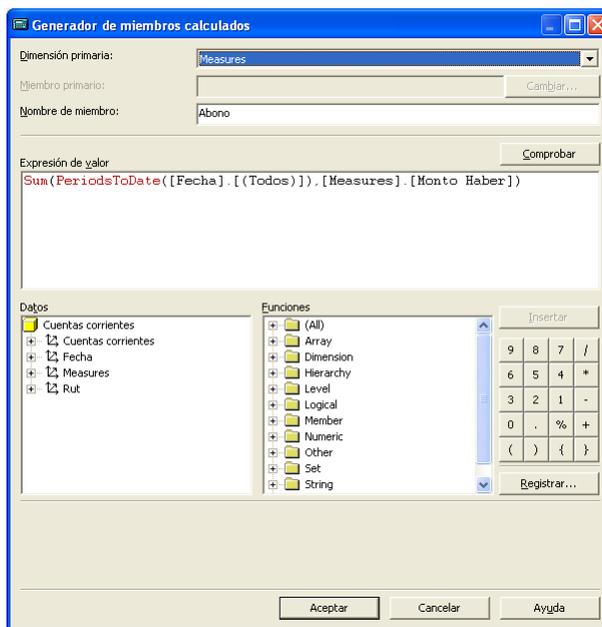
Para crear estos nuevos miembros se debe seleccionar con el botón derecho sobre “Miembros calculados”, “Nuevo miembro calculado”.



Posteriormente aparecerá un editor de formulas MDX.

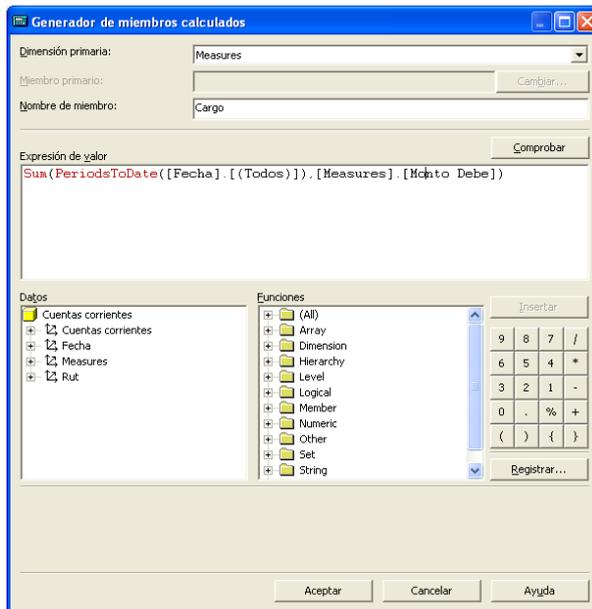
El primer miembro calculado a crear es el “Abono”. La fórmula que se aplicará será la siguiente:

`Sum(PeriodsToDate([Fecha].[(Todos)],[Measures].[Monto Haber])`



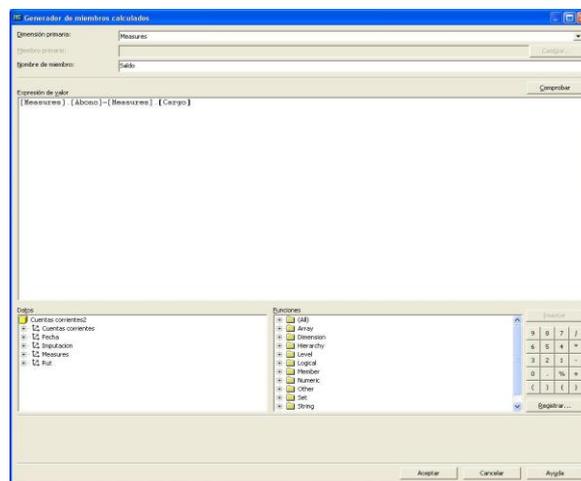
El siguiente miembro calculado a crear es el “Cargo”. La fórmula que se aplicará a éste será la siguiente:

`Sum(PeriodsToDate([Fecha].[(Todos)],[Measures].[Monto Debe])`

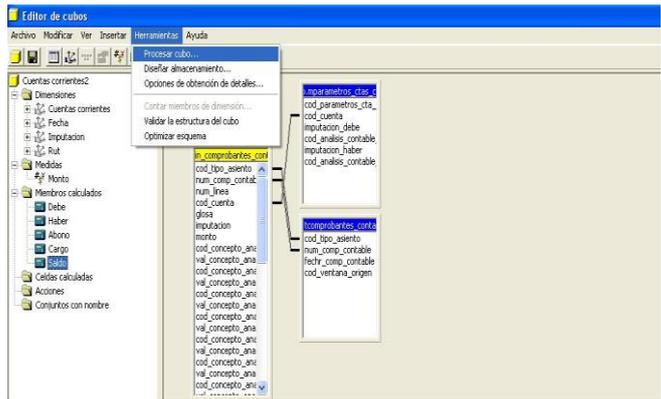


Una vez obtenidos los cargos y abonos podemos calcular el saldo para los distintos Rut que conforman la dimensión de este mismo nombre. La fórmula MDX que calcula estos valores es muy simple, ya que es la diferencia entre los Abonos y Cargos, y se puede escribir de la siguiente forma:

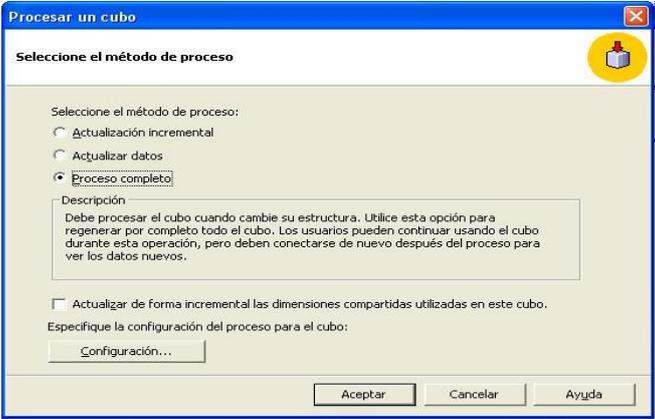
$[Measures].[Abono] - [Measures].[Cargo]$



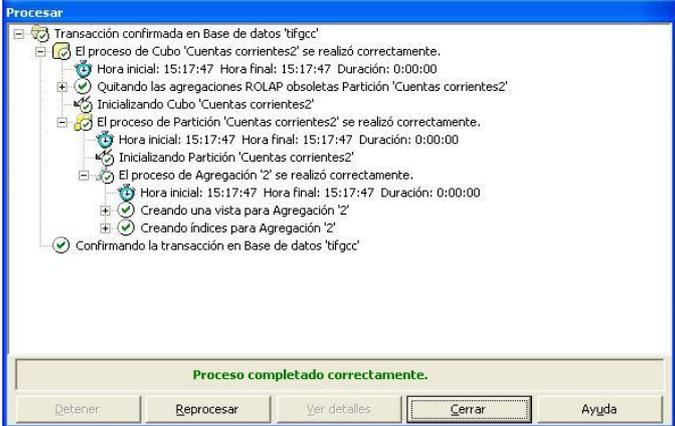
Una vez creadas todos los miembros calculados se debe guardar y procesar el cubo. Para ello se debe ir al menú en “Herramientas”, elegir “Procesar cubo”



Posteriormente seleccionar “Proceso completo”, esto permitirá procesar completamente el cubo.



Para verificar que el cubo ha sido creado satisfactoriamente debe aparecer la siguiente ventana indicando “Proceso completado correctamente”, además de haber realizado todas las transacciones que se listan en el siguiente menú satisfactoriamente.



Anexo 2

Creación del cubo Análisis cuenta

Anexo 2 – Creación del cubo Análisis cuenta.

1.1 Introducción

El siguiente cubo tiene por finalidad permitir ver los montos totales de los análisis de una determinada cuenta a una fecha específica, además de permitir analizar las cuentas por más de un análisis de forma eficiente.

Un ejemplo práctico sería poder analizar la cuenta facturas por pagar por los análisis documento de sistema e Id o Rut para un período determinado. Esto entregará un resumen con los números de los documentos que hasta la fecha especificada no han sido pagados, además de especificar los datos del integrante al cual se encuentra asociado.

1.1.1 Creación de las dimensiones del Cubo

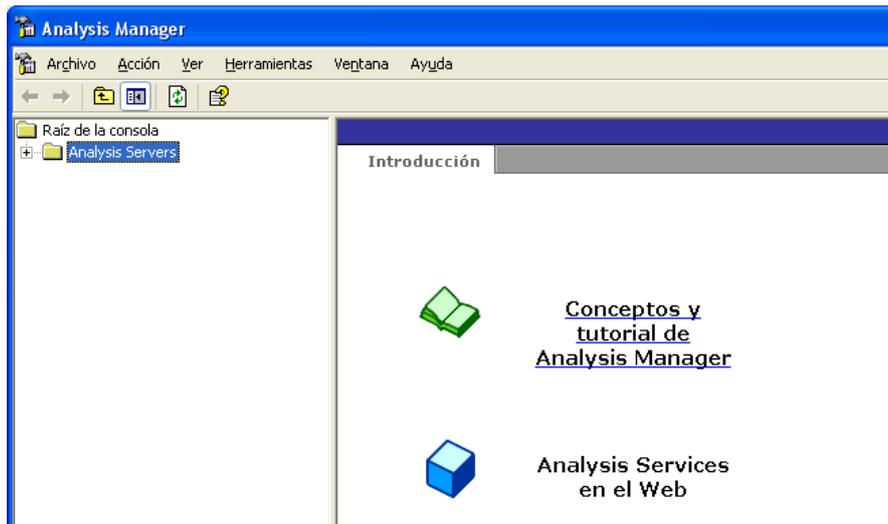
Este cubo tendrá un total de 42 dimensiones, 20 de ellas corresponden a los análisis que conforman la tabla de detalle del comprobante contable. Además existen otras veinte que corresponden a los datos asociados a los conceptos, un ejemplo práctico de esto es el número de una factura, donde el concepto corresponde al tipo de factura que se emitió, por ejemplo factura proveedor.

Otra dimensión es la de Fecha, ésta permite calcular la suma de los montos imputados al debe y al haber para distintos períodos, además de mostrar el saldo hasta una fecha determinada.

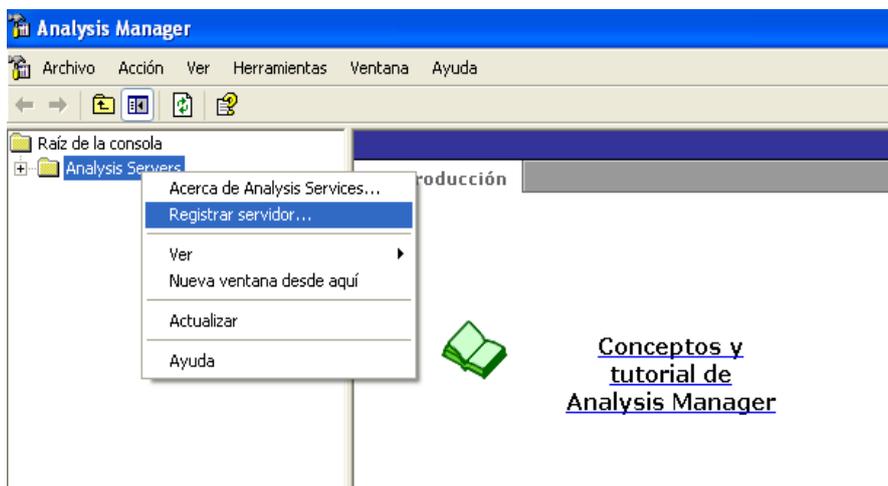
Por último se encuentra la dimensión Cuenta, la cual contiene todas las cuentas de la empresa con sus respectivas características, las que serán explicadas más adelante.

1.1.1.1 Configuración del origen de datos

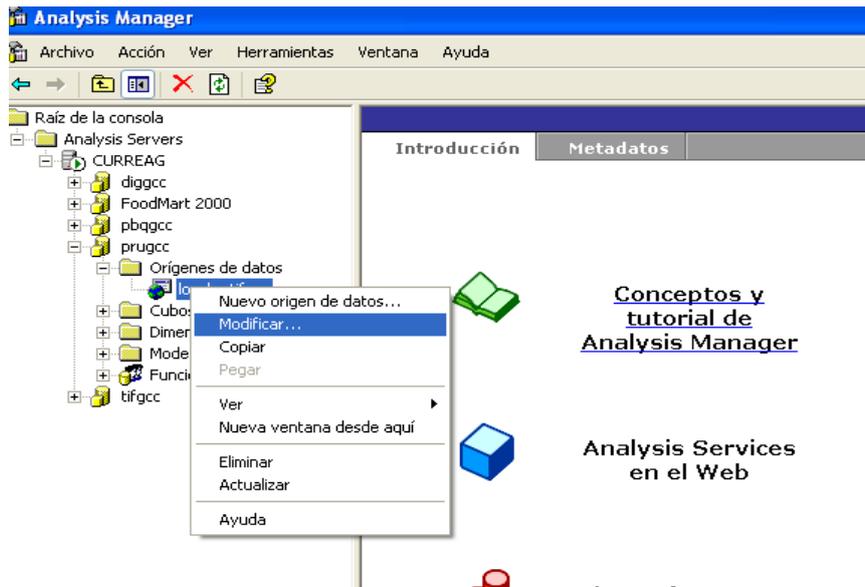
En la barra de inicio seleccionar “Todos los Programas” - “Microsoft SQL Server” –“Analysis Service” - “Analysis Manager”



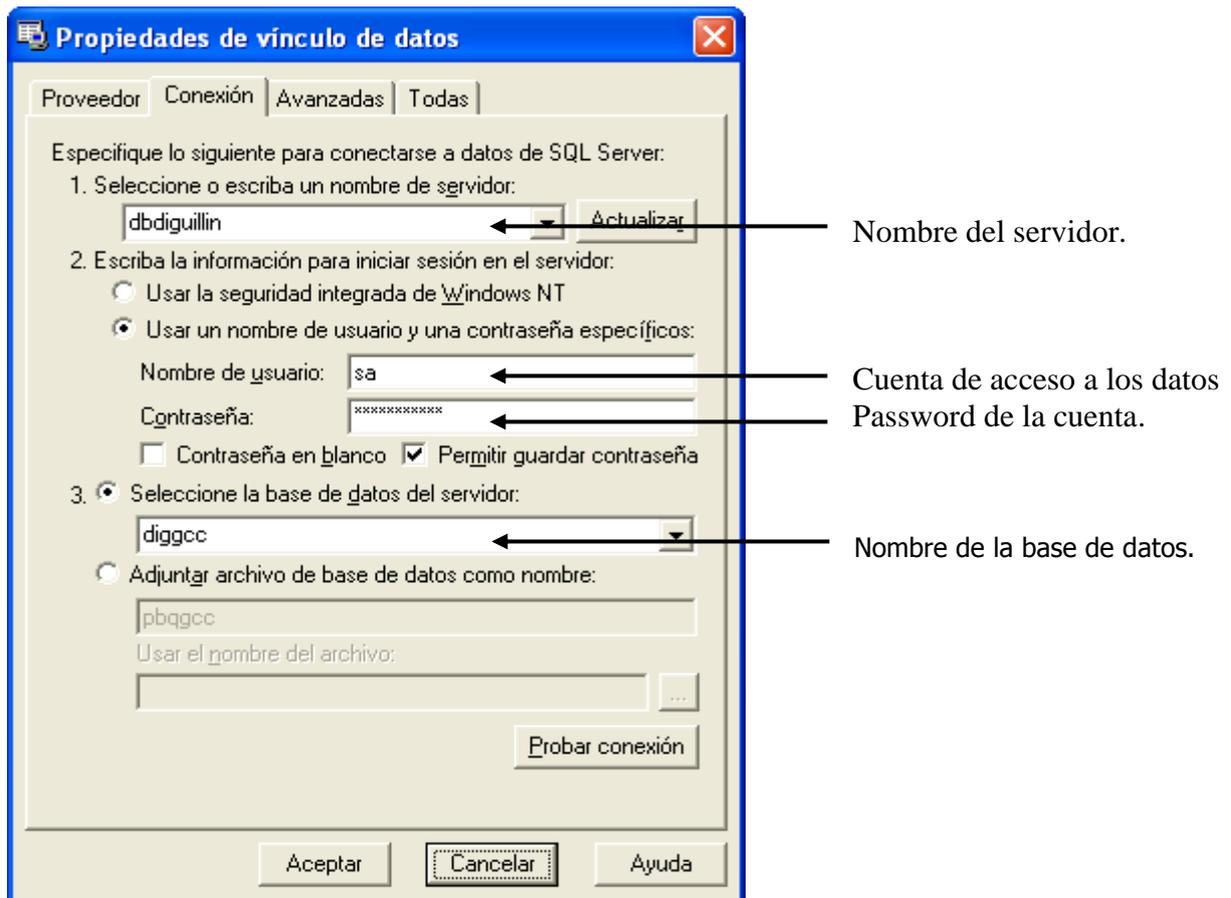
Si no aparece ningún servidor se debe seleccionar con el botón derecho del Mouse “**Análisis servers**” y “**Registrar servidor**”. Este procedimiento debe ser realizado en el mismo computador donde se encuentra la base de datos a la que hace referencia el Cubo OLAP.



Luego sobre la base de datos seleccionar “**Orígenes de datos**” y sobre el origen datos seleccionar “**Modificar**”. .



Debido a que al restaurar la base de datos no se actualiza automáticamente el origen de datos es necesario verificar que los siguientes campos estén correctamente ingresados.



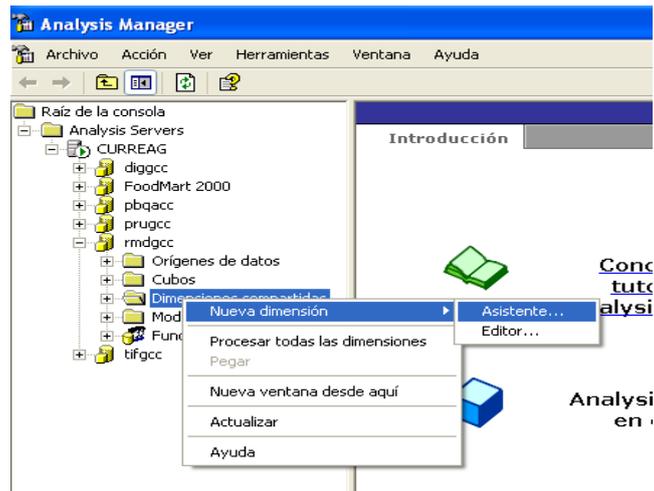
Posteriormente presionar **“Aceptar”**.

1.1.1.2 Creación de las dimensiones de Análisis

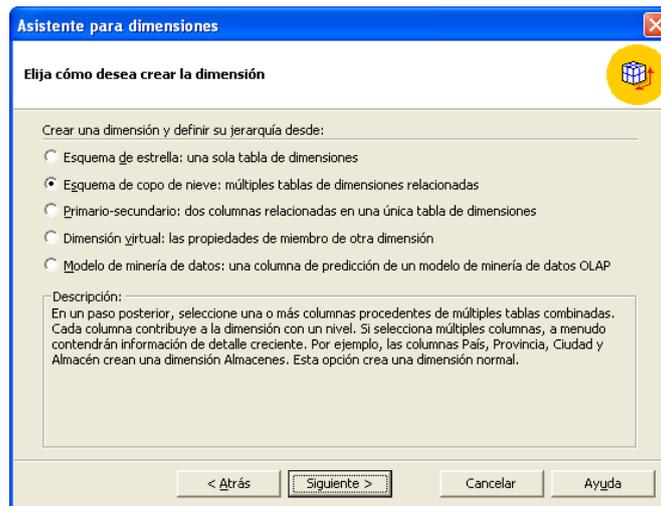
Debido a que se quiere ver la información de las cuentas por más de un análisis es necesario crear dimensiones independientes que permitan mayor flexibilidad a las consultas MDX.

Como el sistema permite un máximo de 20 análisis contables, se crearán este mismo número de dimensiones, la cuales basarán su información en los datos que contenga la tabla de detalle del comprobante contable.

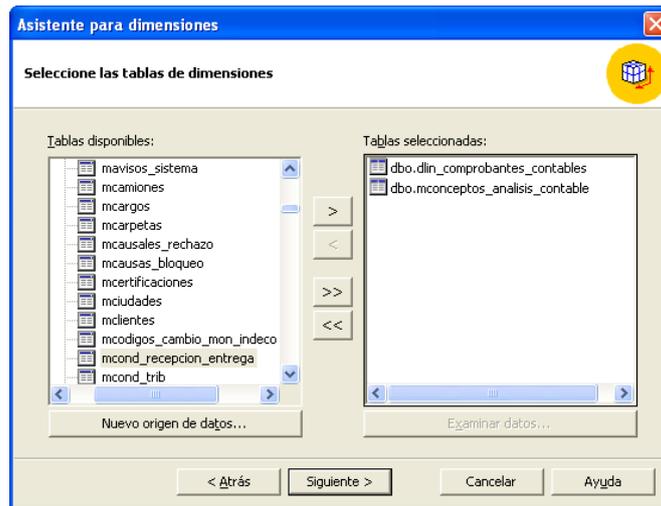
Para crear una dimensión se debe presionar el botón derecho sobre el nivel **“Dimensiones compartidas”**, **“Nueva dimensión”**, **“Asistente”**.



Esta ventana nos permite indicar si la dimensión sacará la información de una o más tablas. Para el caso de los análisis se ocuparán dos, estas son la `dlin_comprobantes_contables` y la `mconceptos_analisis_contable`, donde se listarán sólo aquellos conceptos que tengan algún movimiento contable asociado.



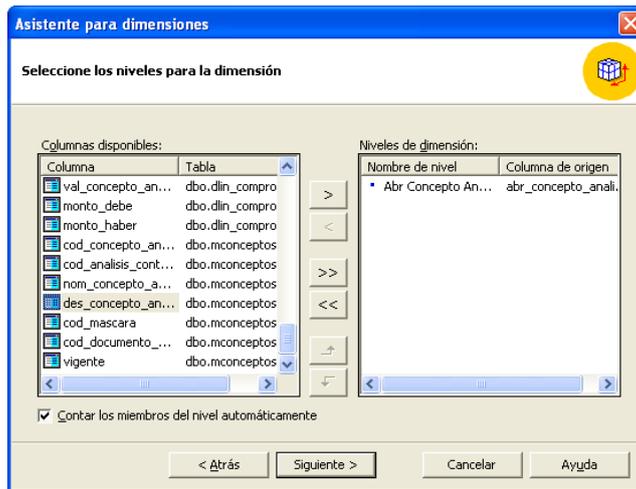
Seleccionar las tablas anteriormente mencionadas, y presionar siguiente.



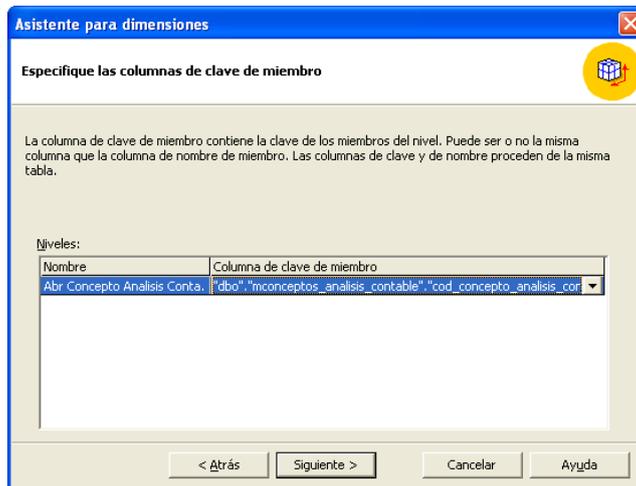
Después de especificar las propiedades del cubo se debe trazar la relación entre el código del concepto de la tabla `mconceptos_analisis_contable` y el `cod_concepto_ana_cont_01` de la tabla de detalle del comprobante, este variará dependiendo de la dimensión que se esté creando, por ejemplo si se crea el “Análisis11” entonces se debe establecer la relación entre `mconceptos_analisis_contable` y el `cod_concepto_ana_cont_11`, posteriormente presionar siguiente.



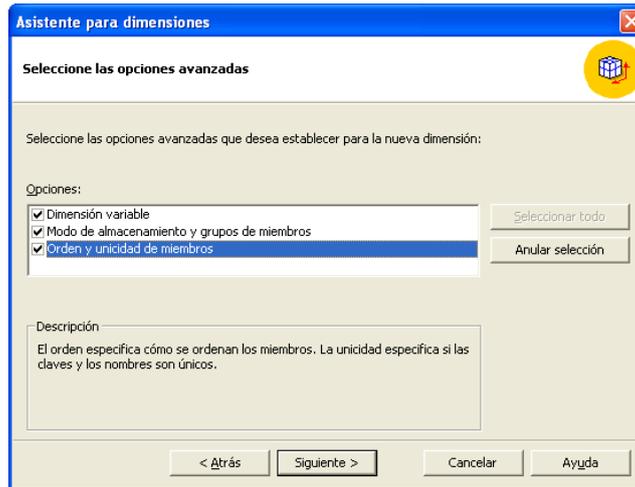
Esta dimensión contará de un sólo nivel, la abreviación del concepto.



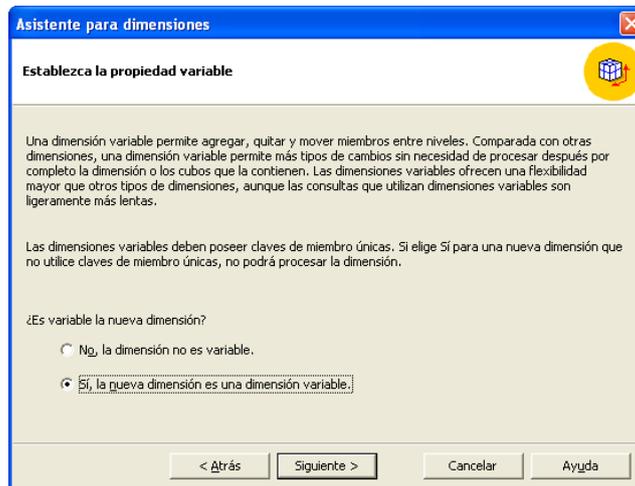
En la siguiente ventana se debe especificar las columnas que son clave de los miembros, en este caso el código del concepto.



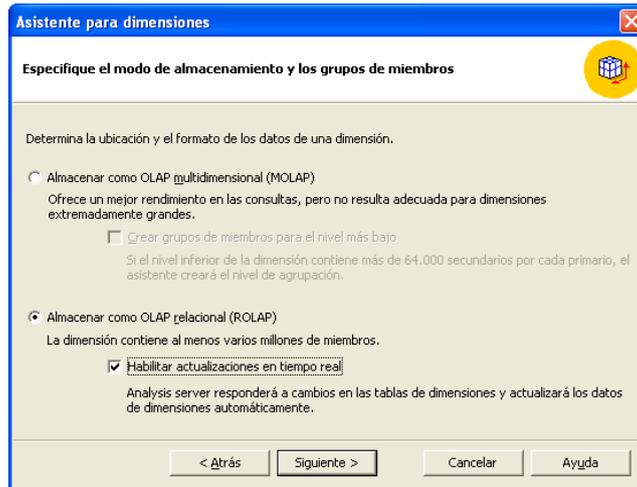
La siguiente ventana tiene por finalidad definir las propiedades de la dimensión.



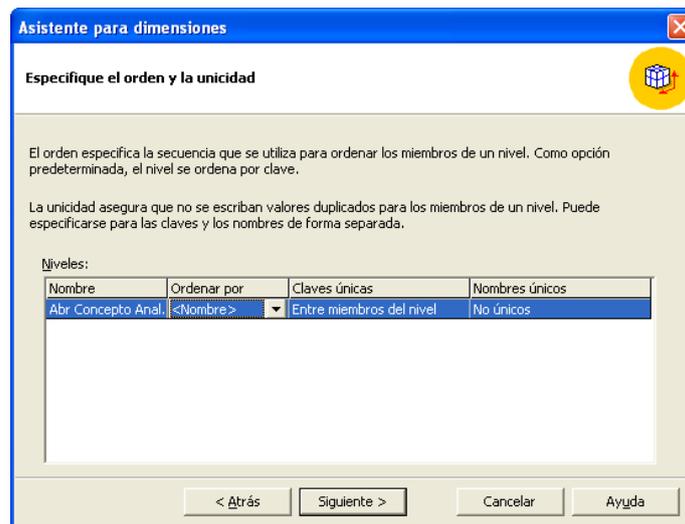
Se debe especificar que la dimensión será variable, o sea que se actualizará automáticamente cada vez que se realicen modificaciones en la base de datos.



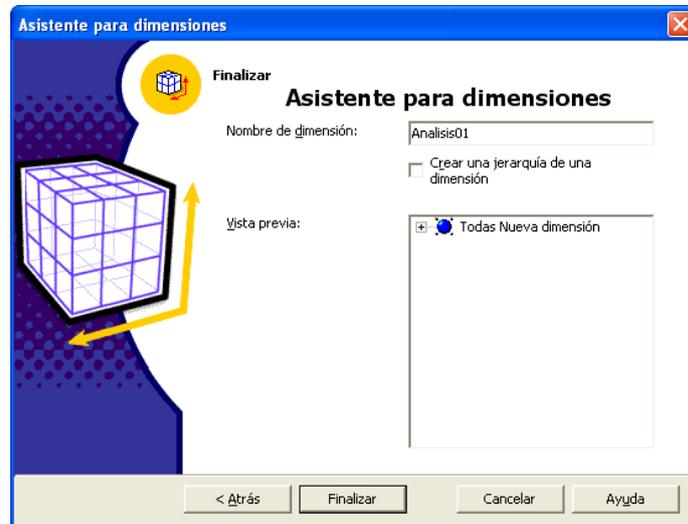
En la siguiente pantalla se selecciona el tipo de almacenamiento de los datos del cubo para nuestro caso ROLAP. Además se debe habilitar las actualizaciones automáticas.



En la especificación del orden se debe indicar el modo en que se listarán los datos, este puede ser ordenado por nombre, columna o clave. Para nuestras dimensiones se ordenará por nombre. También se debe especificar si las claves son únicas para todos los miembros del nivel, ya que no existen dos o más conceptos con la misma clave en la tabla de conceptos. Con respecto a los nombres únicos se debe especificar que estos pueden estar repetidos dentro de la dimensión pero poseer distintas claves.



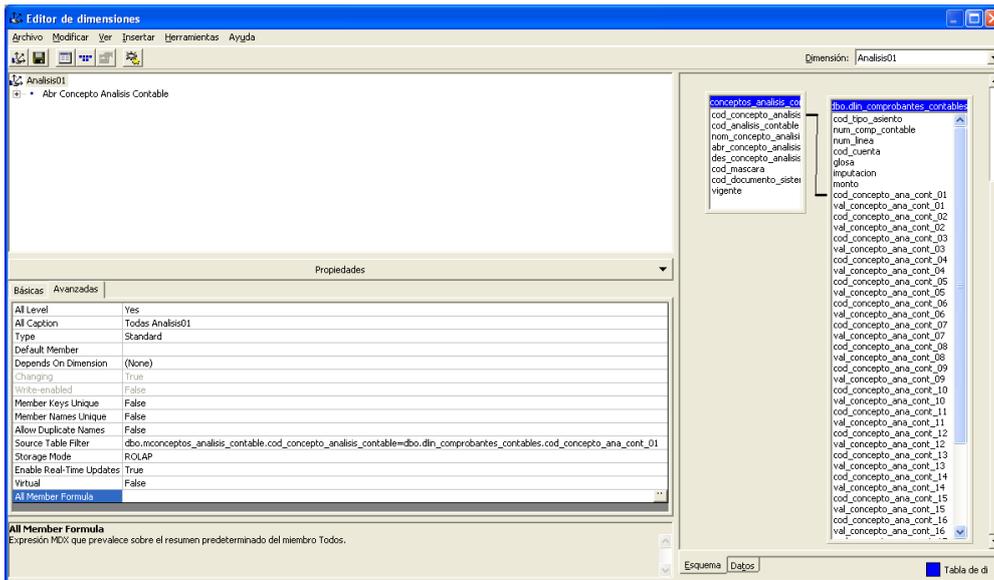
El nombre de la dimensión será “Análisis01”, el número que acompaña la palabra análisis corresponde a la columna seleccionada en la tabla de detalle del comprobante contable, para esta primera dimensión será “Análisis01”



La siguiente ventana muestra una vista de las relaciones de las tablas que componen la dimensión y sus propiedades.

En las etapas anteriores se configuraron todas las propiedades necesarias para la actualización automática del cubo, pero es necesario además establecer las relaciones de las tablas en la propiedad “**Source table filter**”, ya que se quieren listar solo aquellos conceptos que se encuentren ingresados en el comprobante contable, la condición debe ser la siguiente “`dbo.mconceptos_analisis_contable.cod_concepto_analisis_contable = dbo.dlin_comprobantes_contables.cod_concepto_ana_cont_01`”, es decir, el número del concepto debe ser igual al número de la dimensión de análisis que se esté creando.

Posteriormente guardar y procesar.

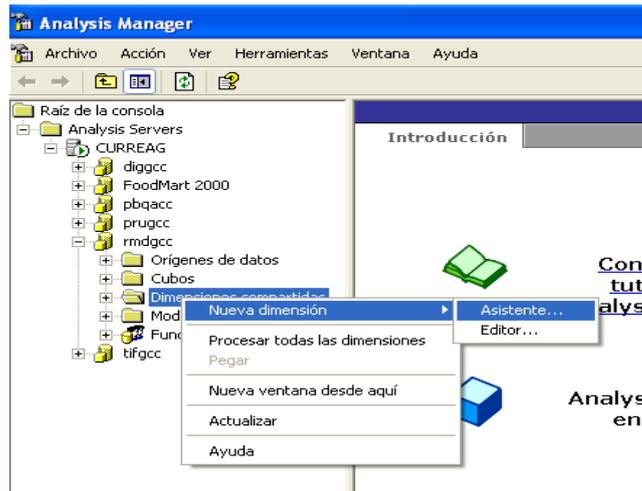


Observación: Para la creación de las dimensiones de análisis se deben seguir los mismos pasos que los utilizados en la dimensión del Analisis01, la única diferencia está en la relación de las tablas, ya que si estamos creando el Análisis15 la tabla de conceptos debe ser unida al cod_concepto_ana_cont_15 de la tabla de dlin_comprobantes_contables.

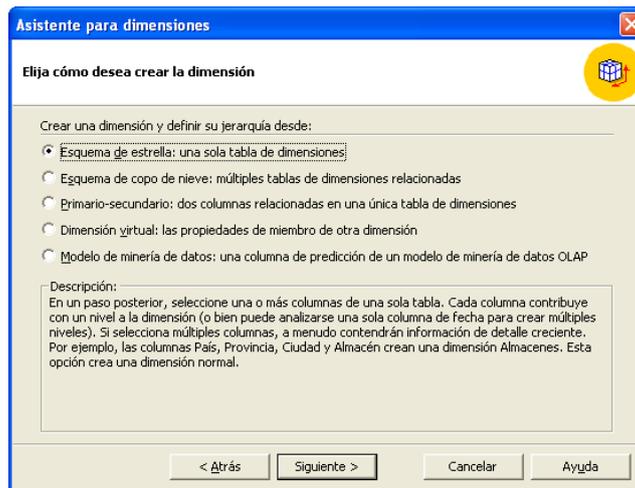
1.1.1.3 Creación de las dimensiones de Valor

La dimensión de valor contiene los datos que se asociaron a un concepto determinado en la ventana de comprobantes contables.

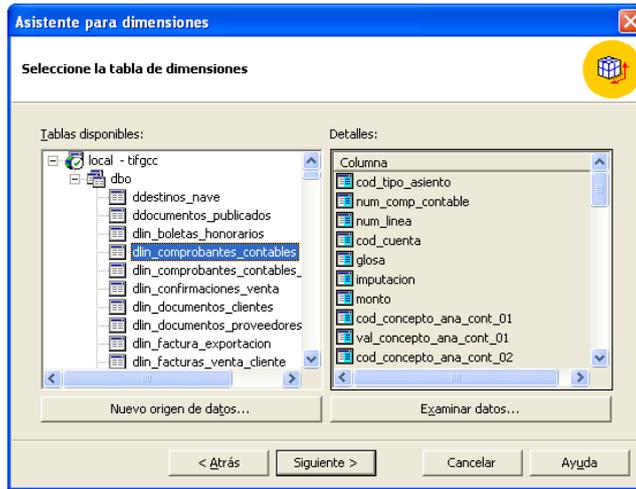
Para crear una dimensión se debe presionar el botón derecho sobre el nivel “Dimensiones compartidas”, “Nueva dimensión”, “Asistente”.



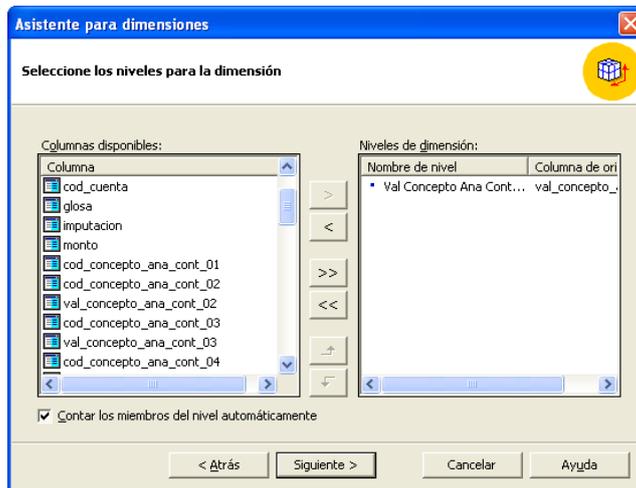
Esta dimensión obtendrá los datos sólo de la tabla dlin_comrpobantes_contables, por lo que se debe seleccionar el esquema tipo estrella.



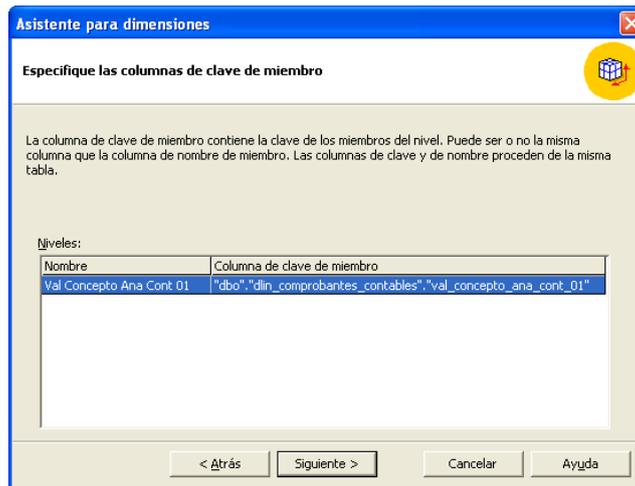
Seleccionar la tabla y presionar siguiente.



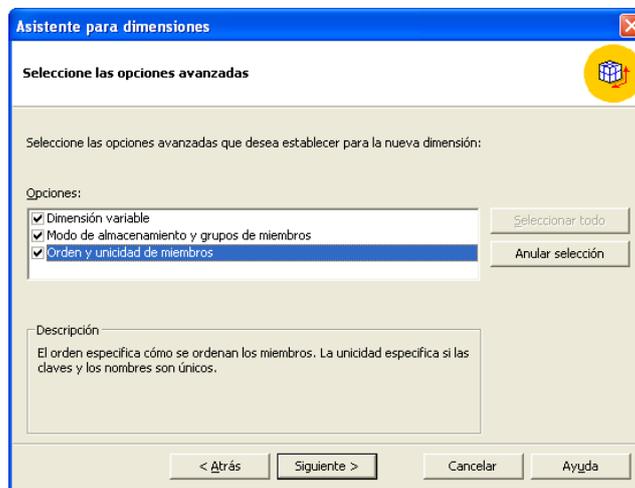
La dimensión contará de un solo nivel, éste corresponde al val_concepto_ana_con_01 si se está creando la primera dimensión de valor.



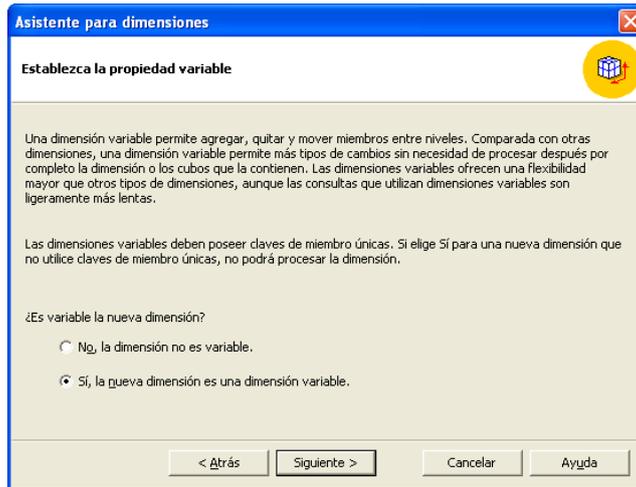
En la siguiente pantalla se debe especificar la clave del miembro, para el caso de los datos de un concepto, el nombre y la clave son el mismo.



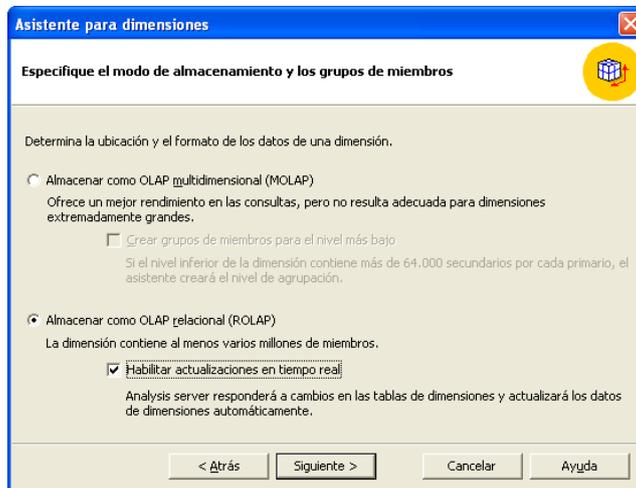
Las siguientes pantallas tienen por finalidad configurar las propiedades del cubo.



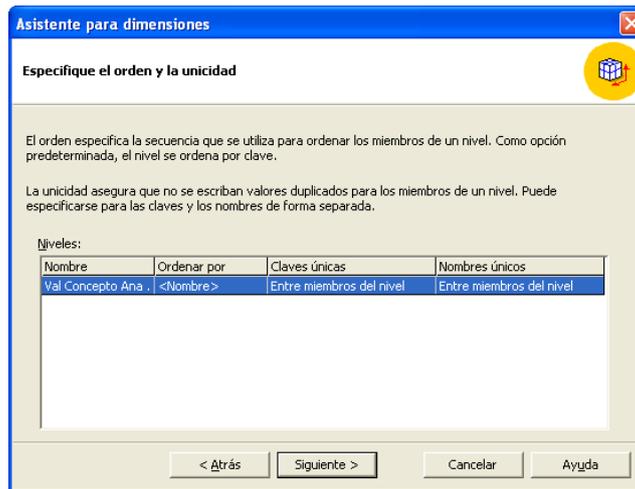
Se debe especificar que la dimensión es variable, esto permite que cada vez que se ingrese un nuevo dato la dimensión se encuentre actualizada.



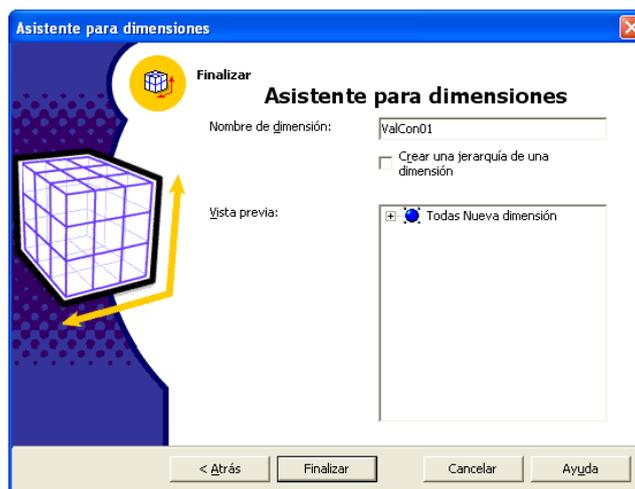
La siguiente pantalla indica que el método de almacenamiento será ROLAP.



Como la dimensión tendrá un solo nivel y además las claves y los nombre corresponden al mismo valor, en la siguiente pantalla se debe especificar que las claves y los nombres serán únicos entre los miembros del nivel.

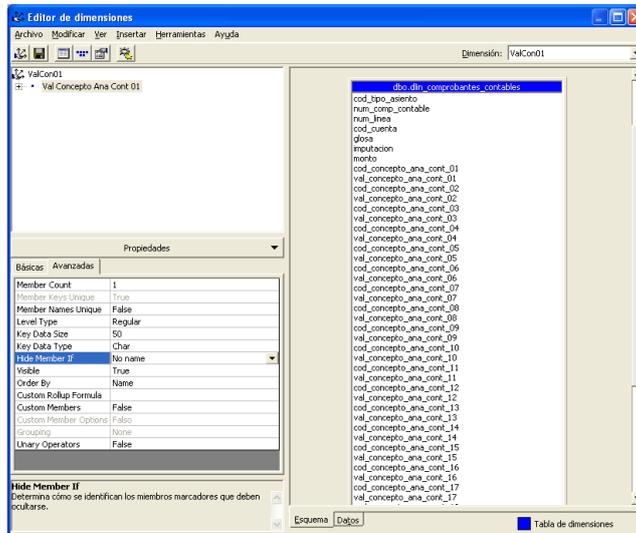


El nombre de la dimensión será ValCon acompañado del número del valor que se haya seleccionado para generar la dimensión, para este caso ValCon01, ya que creamos la dimensión para el val_concepto_ana_con_01.



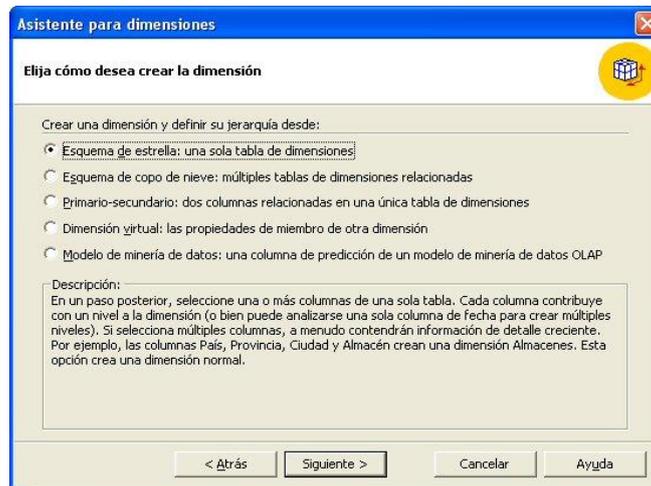
Las propiedades de la dimensión quedaron configuradas en las etapas anteriores, por lo que en la siguiente pantalla sólo se deben cambiar las propiedades del nivel. Para que aparezcan las propiedades del nivel se debe seleccionar Val Concepto Ana Cont 01 (para este caso), donde la propiedad **“Hide Member If”** debe ser igual a **“No name”**.

Una vez terminada la configuración se debe guardar y procesar la dimensión.

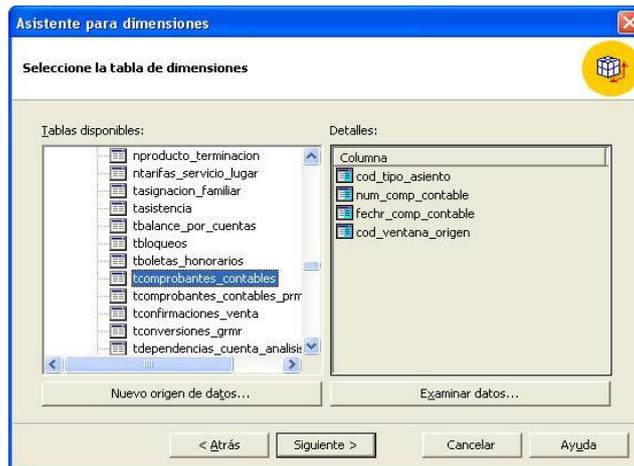


1.1.1.4 Creación de la dimensión Fecha

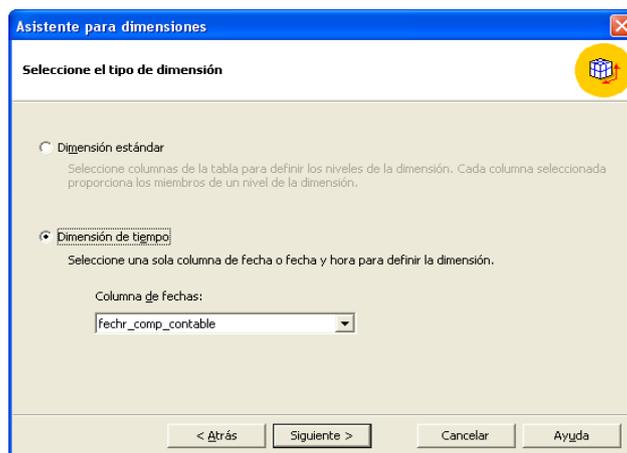
Esta dimensión contendrá todas las fechas en las que se registraron movimientos contables. La información será obtenida de una sola tabla por lo que el esquema a utilizar será tipo estrella.



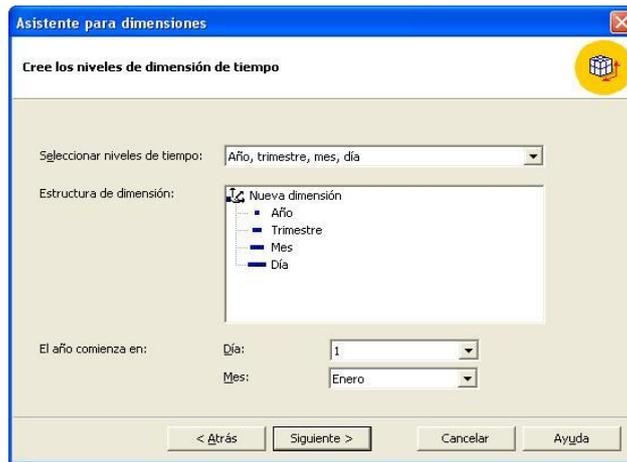
La tabla de la cual se sacará información de la dimensión de fechas será tcomprobantes_contables



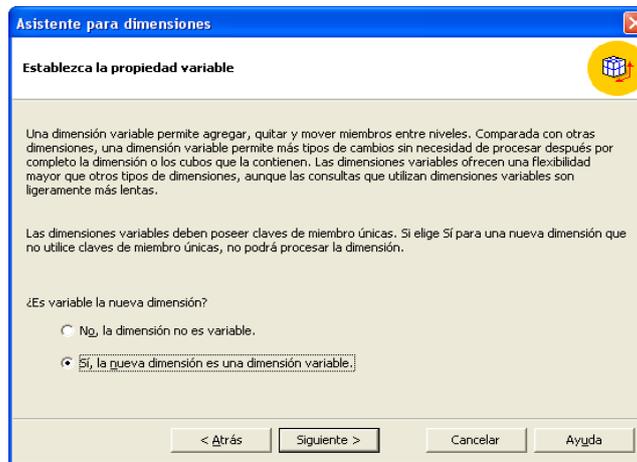
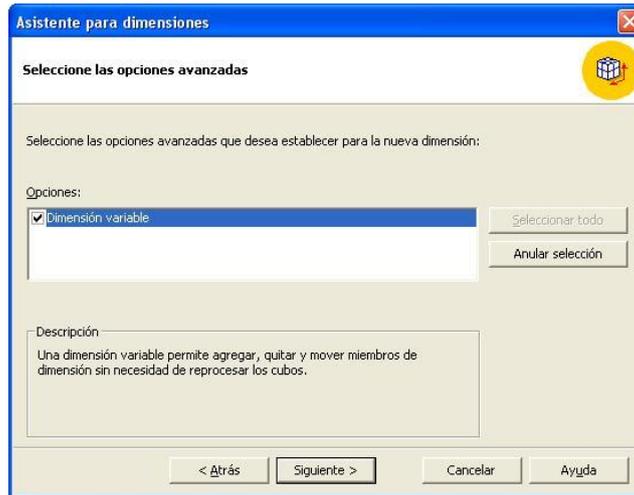
Para crear la fecha se debe seleccionar “Dimensión de tiempo”, esta opción creará automáticamente los distintos niveles de información que contiene esta dimensión, la columna de la cual se sacará la información es fechr_comp_contable, para este caso la única columna de fecha de la tabla.



Luego se deben seleccionar los niveles en los que se desea ver la información, por ejemplo: Año – Mes – Día; Año – Mes, etc. Para este caso se debe seleccionar Año- Trimestre – Mes – Día.



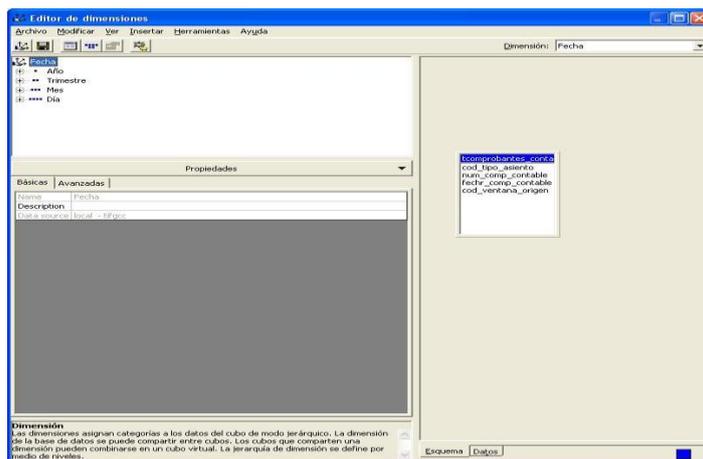
Como habitualmente se están ingresando comprobantes contables, es necesario crear la dimensión como variable, ya que como se indicó anteriormente esto permitirá que se actualice automáticamente.



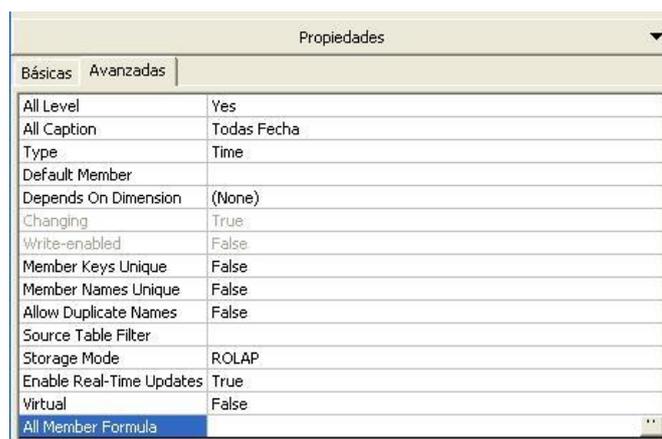
En la siguiente pantalla debe ingresar el nombre de la dimensión para este caso será “Fecha”.



Luego se deben configurar las propiedades de la dimensión.



Para este tipo de dimensión los miembros no son únicos por lo que la propiedad Member keys unique debe configurarse en falso, a diferencia de la dimensión de Cuentas corrientes.

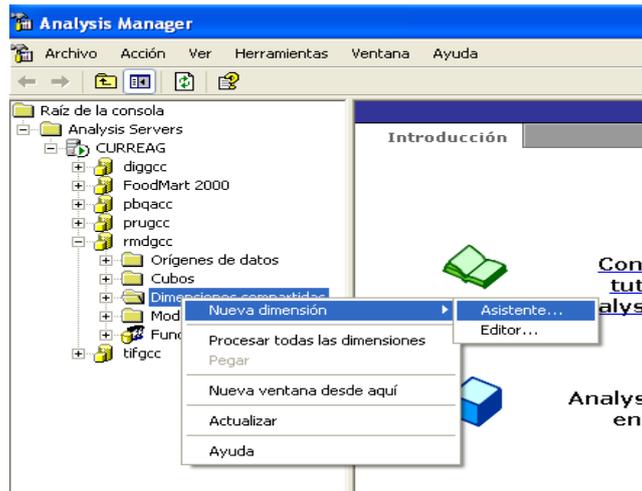


Observación: Si ha sido creado el cubo de cuentas corrientes anteriormente no es necesario crear esta dimensión ya que ésta se encuentra compartida.

1.1.1.5 Creación de la dimensión Cuenta

La dimensión Cuenta está conformada por todas las cuentas de la tabla mplan_cuenta y sus respectivas propiedades.

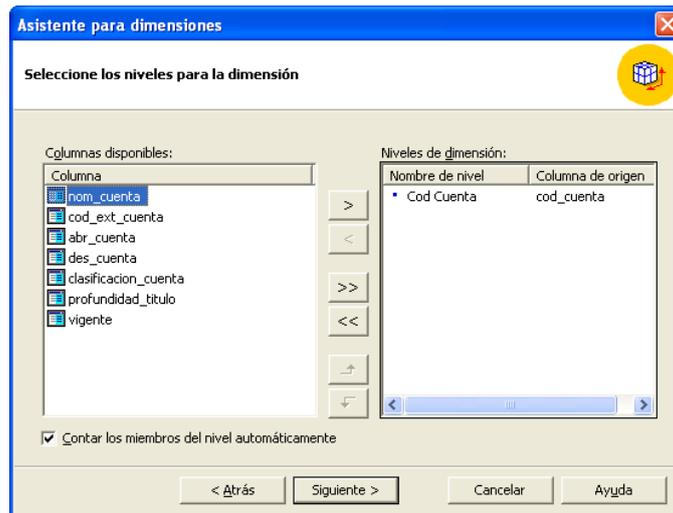
Para crear la dimensión se debe presionar el botón derecho sobre el nivel “Dimensiones compartidas”, “Nueva dimensión”, “Asistente”.



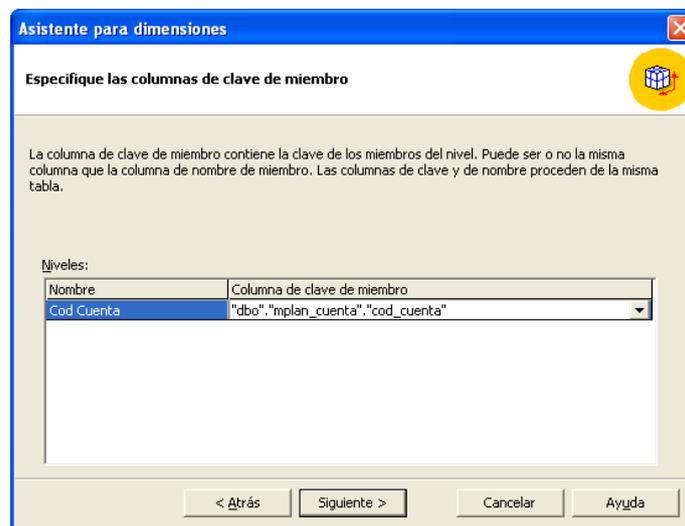
Seleccionar la tabla mplan_cuenta y presionar siguiente.



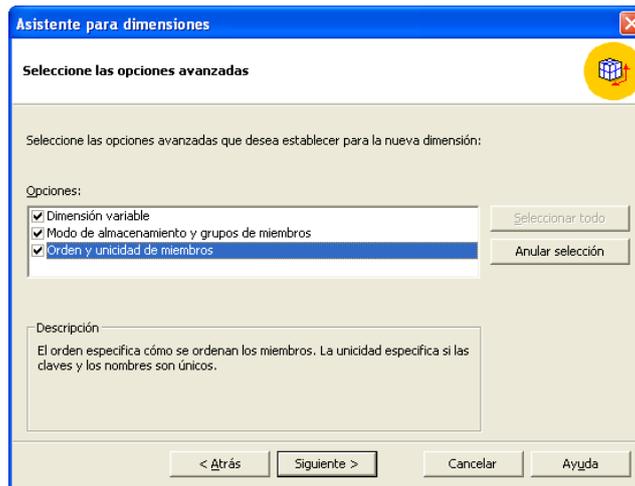
La dimensión tendrá un solo nivel, este corresponderá al código de la cuenta. Seleccionar de la lista de datos cod_cuenta y presionar siguiente.



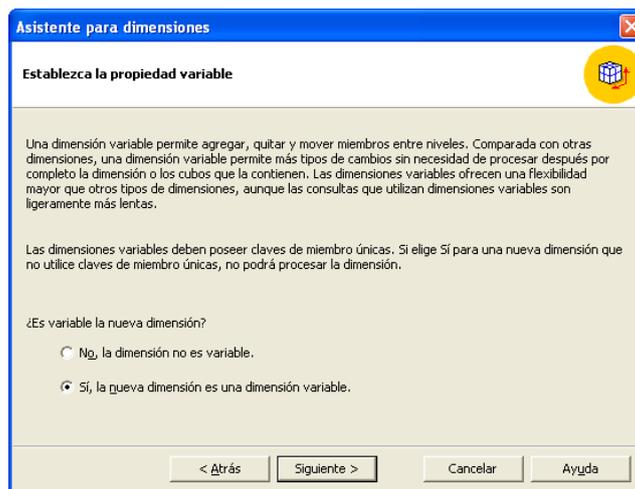
La columna clave del miembro es el mismo código de la cuenta por lo que esta ventana no cambia.



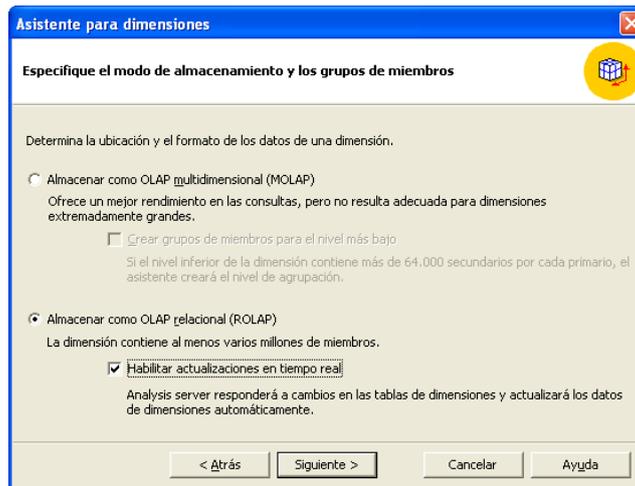
Las siguientes pantallas tienen por finalidad configurar las propiedades del cubo.



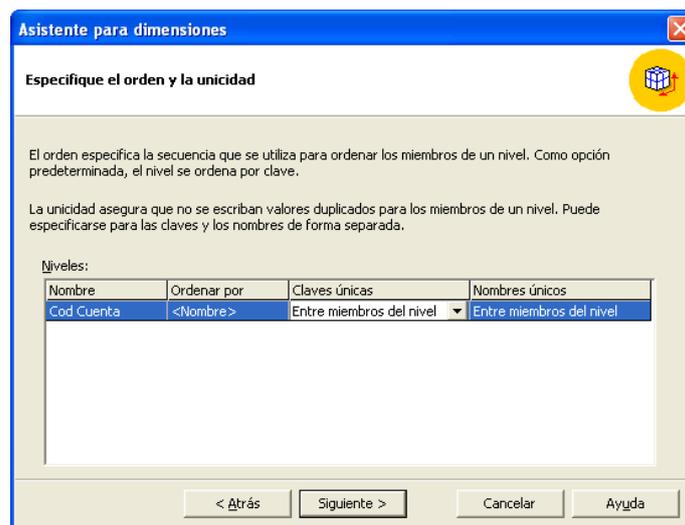
Se debe especificar que la dimensión es variable, esto permite que cada vez que se ingrese una nueva cuenta en la ventana Plan de cuenta los datos de la dimensión se actualizarán automáticamente.



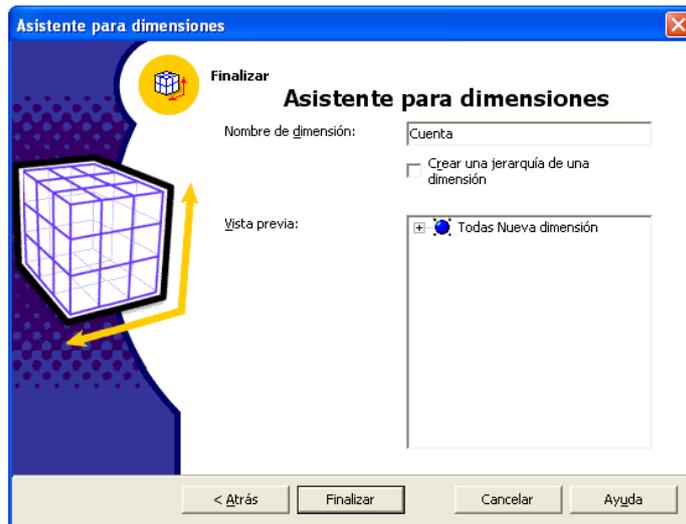
La siguiente pantalla indica que el método de almacenamiento será ROLAP.



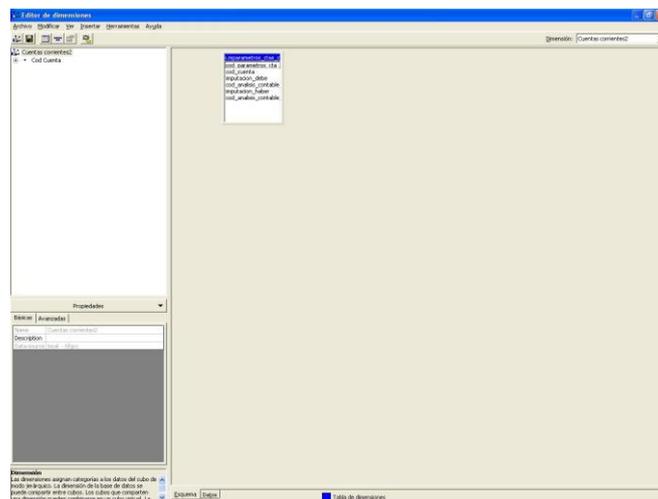
El nombre y la clave del miembro corresponden al mismo valor cod_cuenta, por lo cual en la siguiente ventana no se hace necesario especificar si los miembros se encuentran repetidos en la dimensión ya que estos serán siempre únicos.



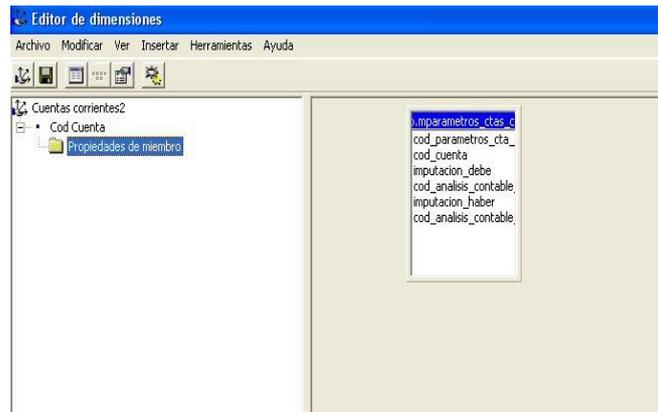
En la siguiente ventana se ingresa el nombre de la dimensión (Cuenta).



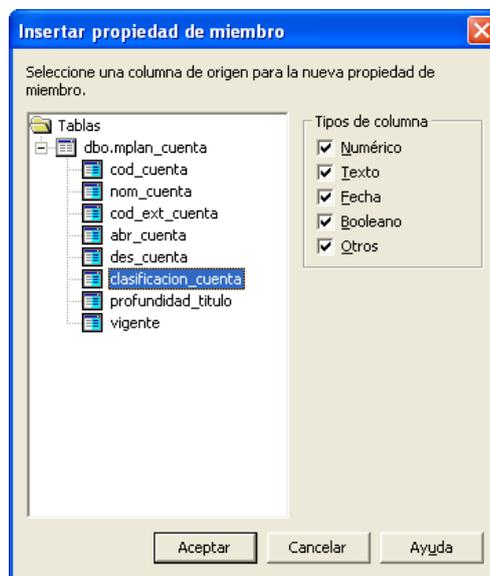
Luego se abrirá un menú que nos permitirá configurar la dimensión de acuerdo a nuestros requerimientos.



Para realizar los cálculos de los saldos a una determinada fecha es necesario saber las clasificaciones de las cuentas, es por ello que se deben agregar a la dimensión las propiedades de las cuentas. Esto se hace seleccionando con el botón derecho sobre la carpeta Propiedades de miembro.



Seleccionar `clasificacion_cuenta`



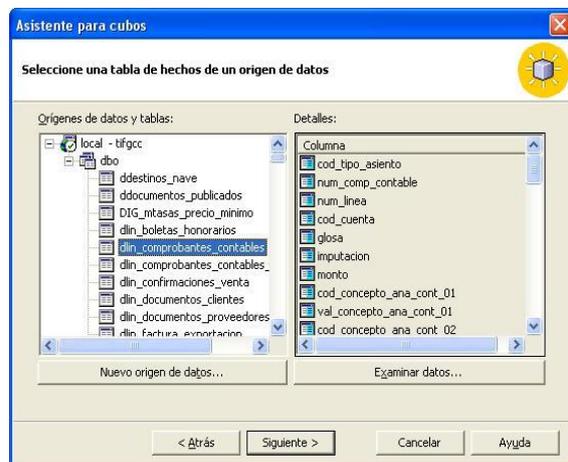
Posteriormente guardar y procesar.

1.1.2 Creación del cubo “Análisis Cuenta”

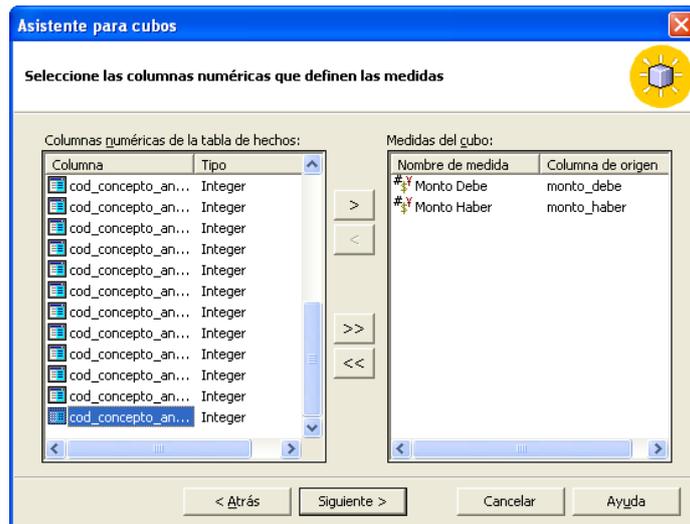
Con el botón derecho sobre la opción cubos seleccione “Nuevo cubo”, “Asistente”.



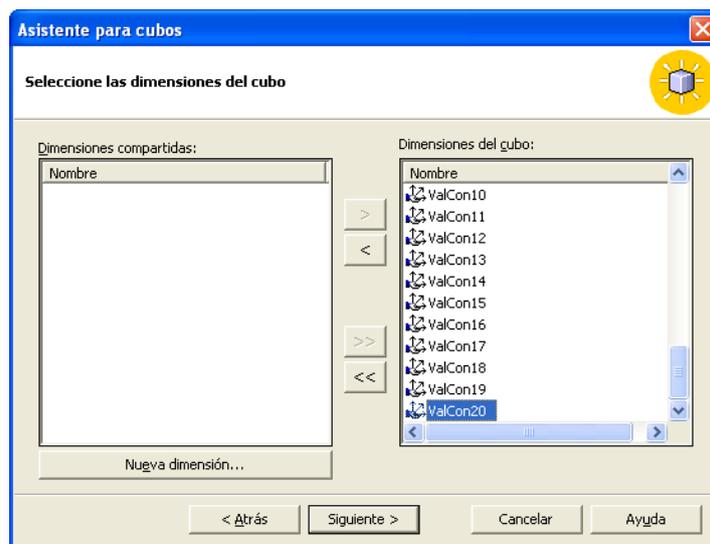
Luego debe seleccionar la tabla de hechos, para este caso dlin_comprobantes_contables.



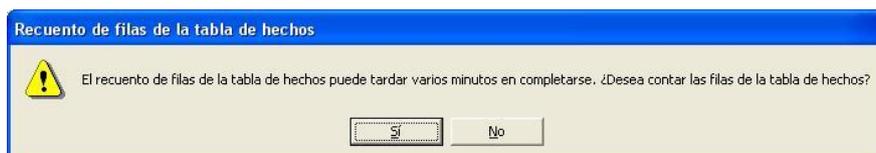
En el siguiente paso se debe seleccionar las columnas de montos de las cuales se quiere sacar información. Para nuestro caso las columnas “monto_debe” y “monto_haber”.



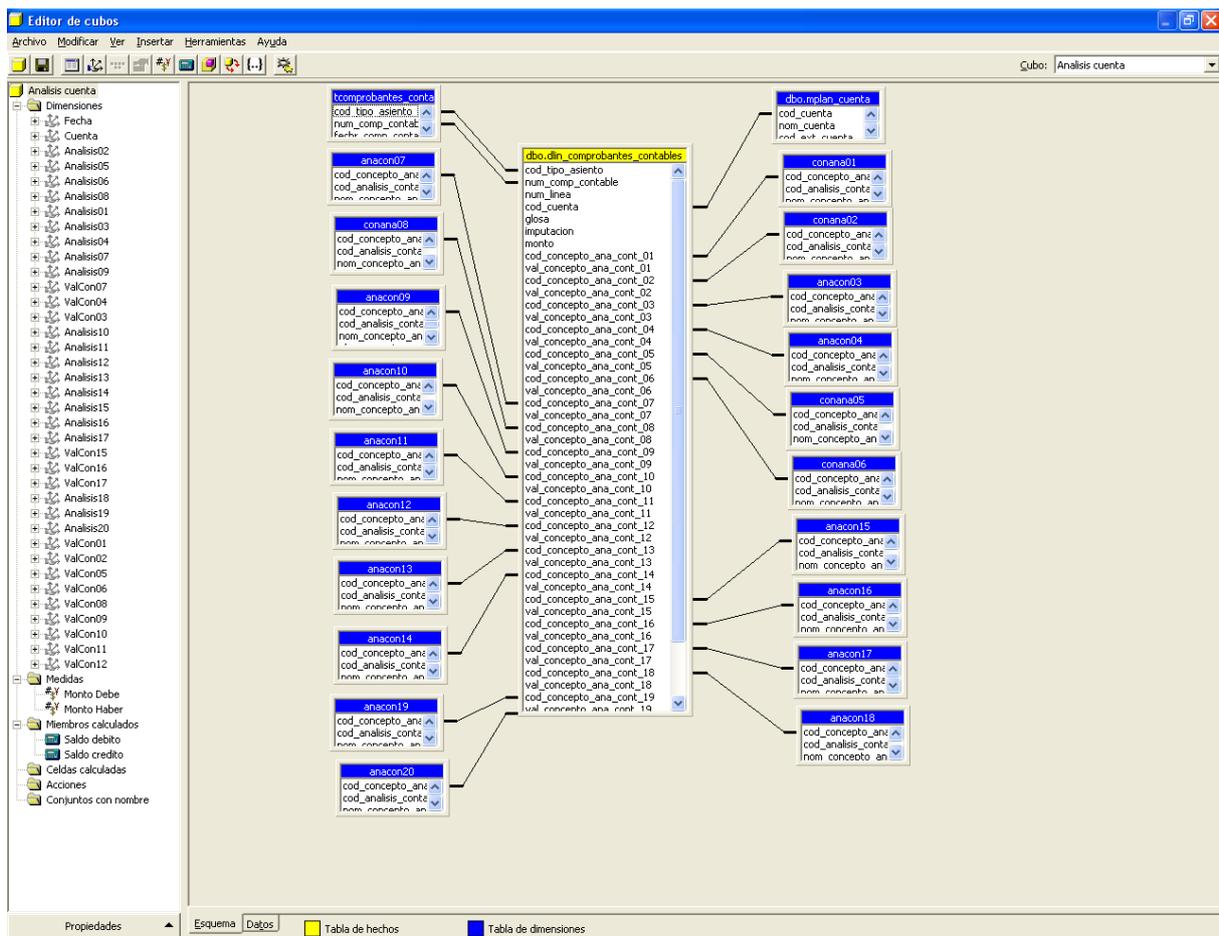
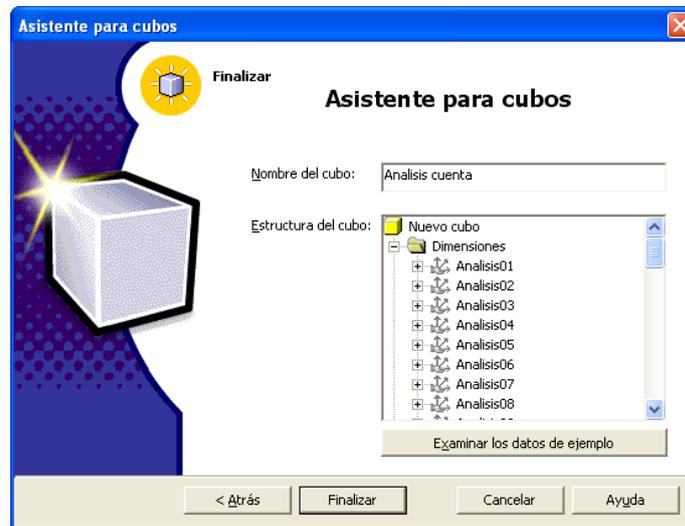
Luego se seleccionan todas las dimensiones que se utilizarán para calcular las distintas agregaciones. Estas corresponden a todos las dimensiones de análisis y valores creados anteriormente, además de las columna Fecha y Cuenta.



Luego aparecerá el siguiente mensaje en el que se debe seleccionar "Sí".

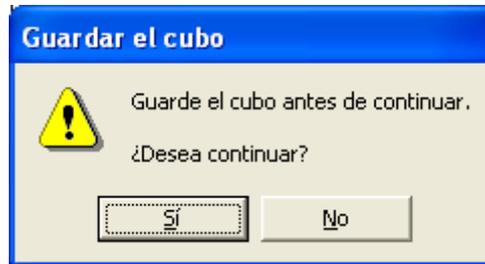


Posteriormente se debe ingresar el nombre del cubo (“Análisis cuenta”)

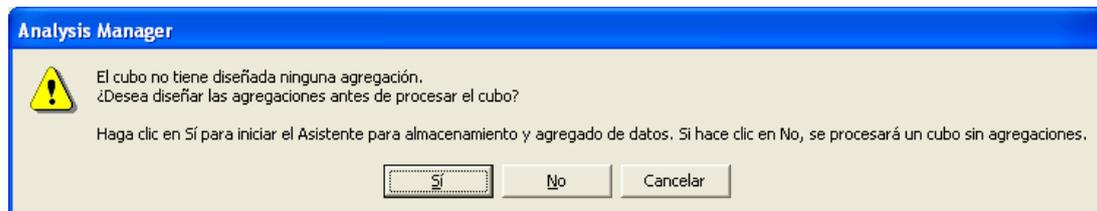


La siguiente etapa es procesar el cubo seleccionando la opción del menú “Herramientas”, “Procesar cubo”.

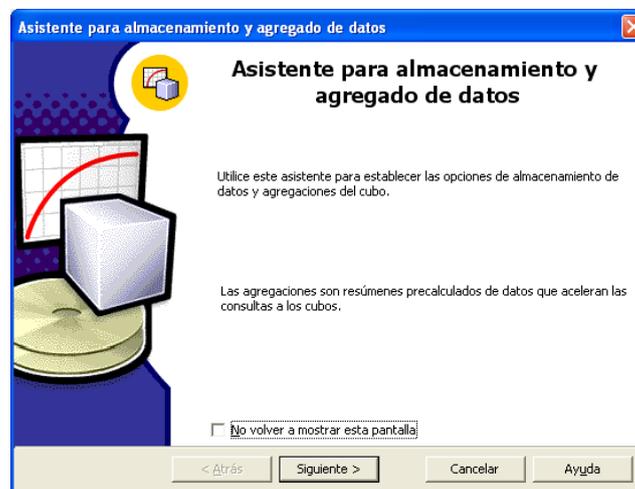
Al mensaje de guardar, seleccionar “Sí”



Al siguiente mensaje seleccionar “Sí”, esto hará un recuento de los datos de la tabla de detalle del comprobante contable además de evitar posibles errores al hacer consultas a datos que no estén calculados.

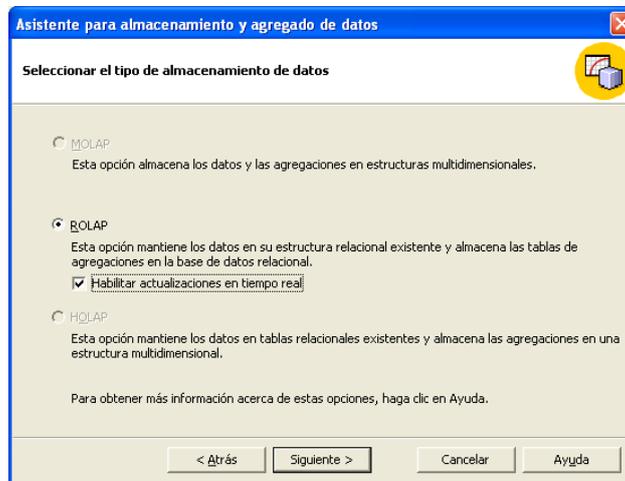


Seleccionar “Siguiente”

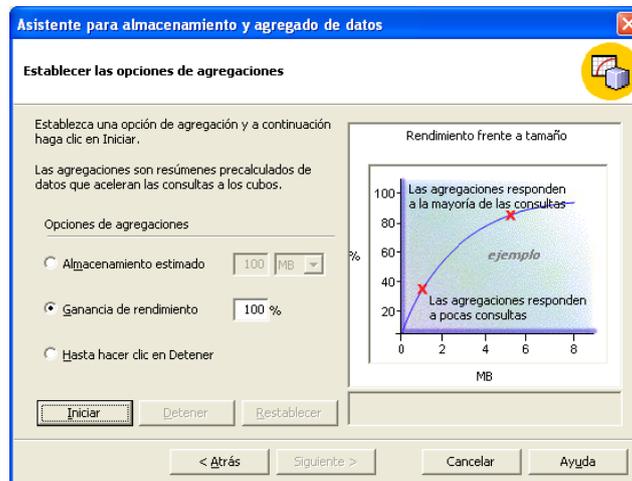


Como se seleccionó al crear las dimensiones, el modo de almacenamiento de los datos será ROLAP, además debido a que cada vez que se realicen actualizaciones en el modelo

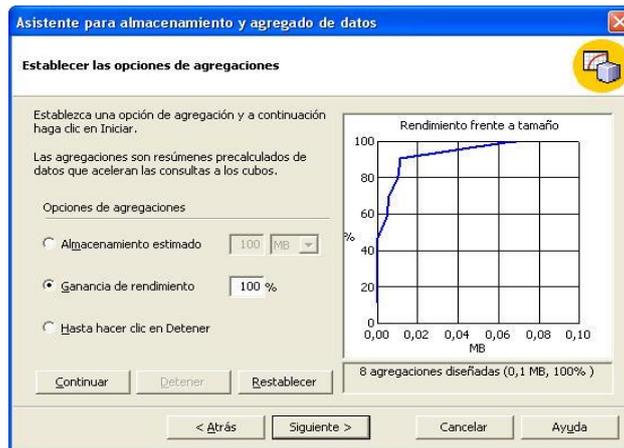
relacional es necesario actualizar los datos del cubo se debe marcar “**Habilitar actualizaciones automáticas en tiempo real**”.



Con respecto a las opciones de agregaciones se le da mayor importancia al rendimiento, desechando la opción de almacenamiento estimado en tamaño. Dado lo anterior se debe ingresar un rendimiento de 100% y presionar aceptar.



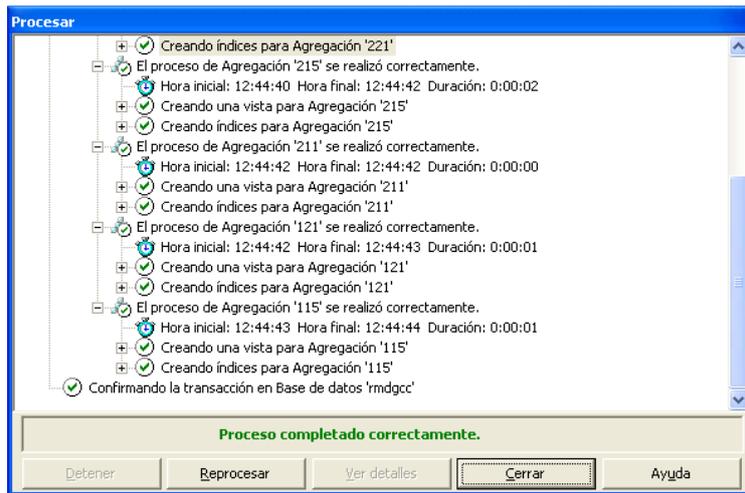
Una vez calculadas a las agregaciones aparecerá un gráfico que indicará el rendimiento del cubo frente al tamaño.



Una vez terminada la etapa de procesamiento se debe seleccionar **“Procesar ahora”** y finalizar.



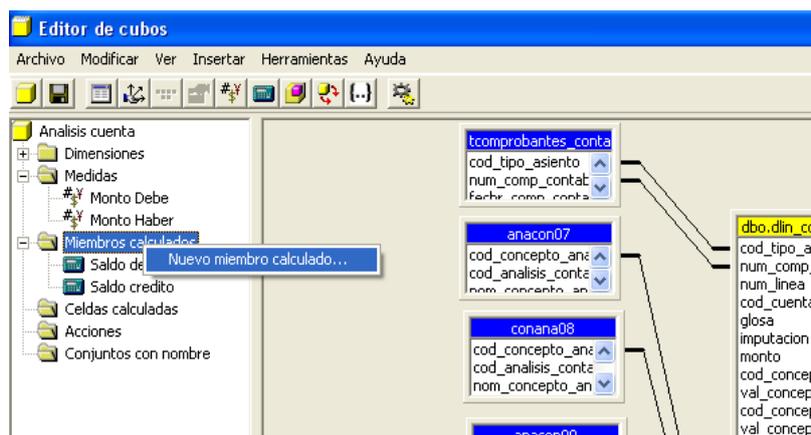
Para verificar que se realizó correctamente el procesamiento de los datos verificar que la siguiente ventana no contenga errores.



1.1.2.1 Creación del miembro calculado “Saldo débito”

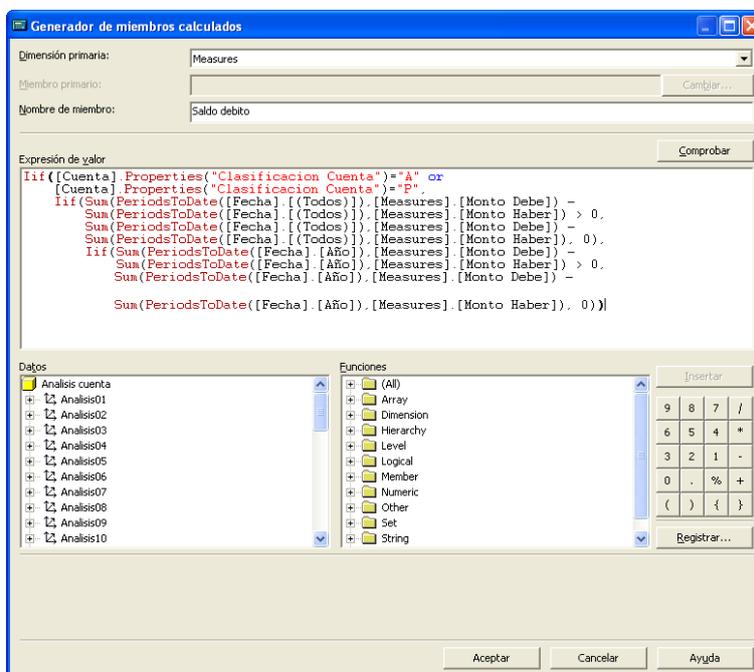
Los miembros calculados que se crearán para este cubo corresponden a los saldos débitos y créditos de las cuentas para los distintos análisis.

Sobre el nivel Miembros calculados presionar con el botón derecho y seleccionar “**Nuevo miembro calculado**”.



En la sección primaria seleccionar “**Measures**”, esto indica que el miembro que creemos quedará incorporado como subnivel de la dimensión de medidas, esto permite tener todos los montos ordenados en una misma dimensión, facilitando la forma de ver los datos.

El nombre del primer miembro calculado será “**Saldo debito**” (Sin acento ya que causa conflictos al ejecutar las consultas).



La fórmula o expresión de valor para este miembro es la siguiente:

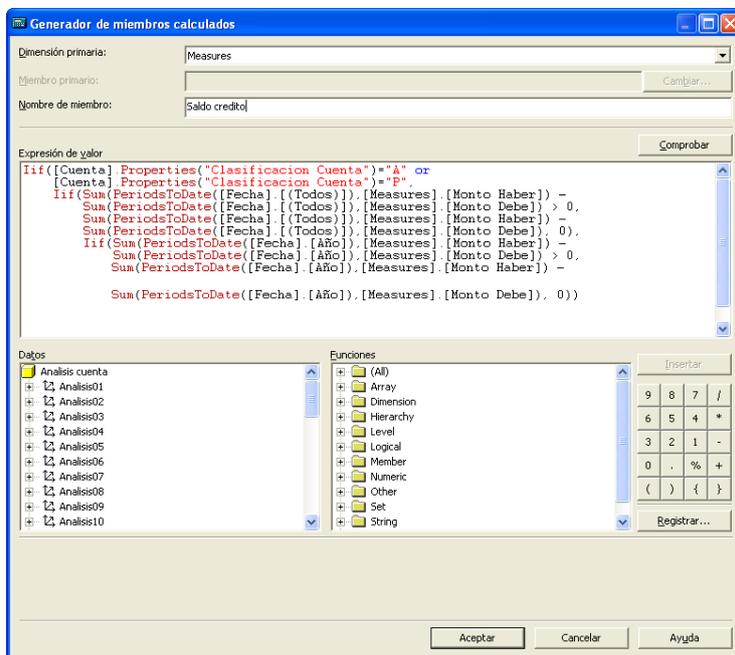
```
Iif([Cuenta].Properties("Clasificacion Cuenta")="A" or
[Cuenta].Properties("Clasificacion Cuenta")="P",
Iif(Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Debe]) -
Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Haber]) > 0,
Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Debe]) -
Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Haber]), 0),
Iif(Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Debe]) -
Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Haber]) > 0,
Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Debe]) -
Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Haber]), 0))
```

1.1.2.2 Creación del miembro calculado “Saldo crédito”

Seguir los pasos anteriores para crear un nuevo miembro calculado.

Al igual que en la creación del miembro anterior en la sección primaria seleccionar “Measures”.

El nombre del primer miembro calculado será “Saldo crédito” (Sin acento ya que causa conflictos al ejecutar las consultas).



La fórmula o expresión de valor para este miembro es la siguiente:

```
Iif([Cuenta].Properties("Clasificacion Cuenta")="A" or
[Cuenta].Properties("Clasificacion Cuenta")="P",
Iif(Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Haber]) -
Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Debe]) > 0,
Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Haber]) -
Sum(PeriodsToDate([Fecha].[Todos]),[Measures].[Monto Debe]), 0),
Iif(Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Haber]) -
Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Debe]) > 0,
Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Haber]) -
Sum(PeriodsToDate([Fecha].[Año]),[Measures].[Monto Debe]), 0))
```

Finalmente presionar aceptar.

Anexo 3

Glosario

Anexo 3 - Glosario

- **Asientos:** Un asiento no es más que la forma de representar un hecho contable. En el momento del cobro del dinero, debemos reflejar ese hecho a través de un asiento.

Por ejemplo, supongamos que el cliente Juan paga \$ 5.000 que debía, las cuales fueron cobrados por caja. El asiento sería el siguiente:

	Fecha		
5.000	CAJA	CLIENTES	5.000
a			
Cobro de Juan, Factura 1			

- **Conjuntos:** Colección ordenada de tuplas, donde todas deben tener la misma dimensionalidad y el mismo orden.
- **Cuenta:** La cuenta es el elemento básico de la contabilidad. Es como una ficha individual donde se anotan todos los movimientos contables como: Caja, clientes, bancos, etc. La cuenta se representa en forma de T y tiene la siguiente estructura:

Debe	CAJA	Haber
1000		

- **Elaboración presupuestaria:** Método que permite la preparación, el cálculo y el seguimiento del presupuesto de la empresa.
- **Extracción:** Proceso de obtención de datos clave para la toma de decisiones de las bases de datos operacionales.

- **Hechos contables:** Los hechos contables son aquellos sucesos puntuales que hacen variar las cuentas de la empresa. Un hecho contable puede ser, por ejemplo, una venta, un pago, una compra, una devolución, etc.

La contabilización de los hechos contables se realiza basándose en el método de la partida doble e que dice: "**todo hecho contable afecta como mínimo a dos cuentas**". Los hechos contables se reflejan en la empresa en forma de asientos.

Debe	CLIENTE	Haber
10.000	3.000	
7.000		

- **Indicador:** Es un dato significativo que permite tener una cierta visión del estado de la actividad de la empresa. Puede ser el volumen de negocios, el número de clientes o de empleados. Los indicadores son analizados por dimensión.
- **Jerarquía:** Reagrupación de varios niveles dentro de una dimensión. Por ejemplo, la dimensión ‘tiempo’ puede tener como jerarquía: Año > Mes > Día, etc. El usuario podrá obtener datos de forma detallada.
- **MDX** es el lenguaje tipo SQL (Structured Query Lenguaje) que se usa en la programación de los servidores de base de datos, pero aplicado en este caso a la extracción y consulta de la información interna de un cubo de datos.
- **Metadatos:** ‘Dato sobre dato’, es decir, la información enlazada con el dato: origen, fecha de actualización, fecha de importación a la base, persona que le ha modificado últimamente, fecha de validez, formato, etc. Los metadatos son indispensables para una buena comprensión del sistema de información. Se almacenan en el referencial.
- **Multidimensional:** Dícese de una petición o una base de datos dentro de la cual cada indicador se analiza en función de varios criterios, las dimensiones.

- **OLAP:** Modelo conceptual de bases de datos multidimensionales, inventado por el Dr. Codd. Este modelo conoce algunas variaciones: MOLAP, ROLAP, DOLAP, HOLAP, etc., que no son más que variantes de un mismo modelo inicial.
- **OLTP:** On-Line Transactional Processing. Modelo conceptual de bases de datos relacionales.
- **Petición o query:** Solicitud hecha a la base de datos en forma de campos condicionales, realizada de forma frecuente en lenguaje SQL.
- **Referencial:** Bases de datos en la que se almacenan metadatos.
- **Relacional:** Base de datos dentro de la cual toda información es almacenada en forma de tabla en dos dimensiones. Las relaciones entre estas tablas se crean en forma de joins.
- **SQL:** Structured Query Language. Lenguaje con el que se realizan las peticiones, que permite ‘interrogar’ a todas las bases de datos compatibles utilizando las mismas peticiones.
- **Tabla:** Los datos se almacenan en las bases de datos relacionales en forma de tablas, es decir, hojas de dos dimensiones (líneas y columnas).
- **Transacción:** Diálogo informático entre una aplicación y un usuario que termina en la realización de una instrucción o una orden, la edición de un documento.
- **Transformación:** Proceso de homogeneización de los datos clave. Esta fase es esencial para que los datos que hay que analizar sean fiables sobre todo cuando provienen de bases de datos distintas.
- **Tuplas:** Combinación de miembros de diferentes dimensiones en la que puede incluirse a los sumo un miembro de cada dimensión.

Bibliografía

- **Valle, J.** (2005). **Bodega de datos.** Disponible en <http://www.monografias.com/trabajos24/bodega-de-datos/bodega-de-datos.shtml>. Leído el 15 de enero de 2006
- **Bohorquez, M.** (1999). **Aproximación metodológica de un Spatial Data WareHouse.** Colombia: Universidad Nacional. Disponible en http://gis.esri.com/library/userconf/latinproc00/colombia/spatial_data.pdf. Leído el 20 diciembre 2005.
- **Borja R.,** (2002). **La empresa multidimensional: OLAP.** Disponible en www.businessobjects.com. Leído el 12 de enero 2006.
- **Fernández A.,** (2000). **Data WareHouse.** Colombia: Universidad Nacional. Disponible en <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosLezca no.pdf>. Leído el 15 de diciembre.
- **Almonacid F.,** (2000). **Data WareHouse.** Argentina. Universidad de Buenos Aires. Disponible en <http://www.monografias.com/trabajos6/dawa/dawa.html>. Leído el 16 de enero de 2006.
- **Inmon H.,** (1992). **La Tecnología Datawarehousing.** Second edition, Wiley. Disponible en <http://www.alfa-redi.com/rdi-articulo.shtml?x=638>. Leído el 5 de diciembre de 2005.
- **Osterfeldt, M.** (1993). **Bodega de Datos.** Argentina. Facultad de Ciencias Exactas y Naturales y Agrimensura. Disponible en <http://www.inf.udec.cl/revista/ediciones/edicion3/cwolff.PDF>. Leído el 30 de enero de 2006.
- **Microsoft,** (2000). **Libros en línea, SQL 2000.** Disponible en <http://msdn2.microsoft.com/es-es/library/default.aspx>. Leído el 12 de septiembre de 2006.

- **Microsoft**, (2000), **Libros en pantalla, Analysis Service SQL Server**.
[http://msdn2.microsoft.com/es-es/library/ms175609\(SQL.90\).aspx](http://msdn2.microsoft.com/es-es/library/ms175609(SQL.90).aspx). Leído el 12 de septiembre de 2006.
- **Nader J.**, (1998). **Sistema de Apoyo Gerencial Universitario**. Argentina. Instituto tecnológico de Buenos Aires. Disponible en <http://www.itba.edu.ar/capis/webcapis/tesisdemagister/nader-tesisdemagister.pdf>. Leído el 15 de enero de 2006.
- **PradoLand Consultores**, (2000). **Cubos Conceptos**. México. Disponible en http://www.prado.com.mx/Cubos/Cubos_Conceptos/cubos-definiciones-y-conceptos-15.htm. Leído el 25 de enero de 2006.