



El rol de las bases de datos multidimensionales en una solución de Data Warehousing

Gabriela Lopez Uhalde

Silvana Manganiello

Eduardo Martinez

<p>TES 99/14 DIF-02086 SALA</p>	<p> UNIVERSIDAD NACIONAL DE LA PLATA FACULTAD DE INFORMATICA Biblioteca 50 y 120 La Plata catalogo.info.unlp.edu.ar biblioteca@info.unlp.edu.ar</p> <p> DIF-02086</p>
---	---

DONACION.....
S.....
Fecha..... 3-10-05
Inv. E..... Inv. B..... 2086

TES
99/14



Tema	Página
Introducción	1
Data Warehousing	1
Data Warehouse	3
Finalidad del Data Warehouse	3
Características claves de una solución de Data Warehouse	4
Datos en un Data Warehouse	5
Característica de los datos	5
Componentes de un Data Warehouse (Arquitectura de Referencia)	6
Uso del Data Warehouse	11
Diferencias entre Data Warehouse y bases de datos operacionales	12
Técnicas para utilizar el Data Warehouse	13
Procesamiento informático	14
Procesamiento analítico	15
Análisis Multidimensional	16
Procesamiento analítico en línea (OLAP)	17
Arquitectura OLAP	17
Arquitectura Lógica	18
Arquitectura Funcional	18
Arquitectura Física	19
Bases de datos multidimensionales	
Introducción	21
Ventajas de performance	24
Beneficios	25
Cuando la tecnología multidimensional no es apropiada	25
Características de las bases de datos multidimensionales	28
Rotación	28
Ranging	31
Jerarquías, Roll-Ups y Drill Downs	32
Queries	34
Vistas de datos multidimensionales	35
Cálculos multidimensionales	35
Dimensión Tiempo	36



Datos esparcidos	37
Integración de datos	37
Estructuras relacional, multidimensional y spreadsheet	39
Características de las distintas tecnologías	40
OFA (Oracle Financial Analyst)	
Elementos de OFA	41
Tipos de Workstations	41
Tipos de bases de datos	41
Diseño de arquitectura por filas	42
Workstation super administradora	42
Workstation administradora	43
Workstation task processor	44
Workstation budget	44
Workstation analyst	45
Ejemplificación de los conceptos utilizados en OFA	46
División consultoría	47
División producción	48
División finanzas	48
Comparación entre los distintos tipos de workstations	49
Configuraciones de OFA típicas	51
Dimensiones y valores de dimensión	52
Mantenimiento de dimensiones y valores de dimensión	53
Tipos de dimensiones	54
Items de datos financieros	54
Mantenimiento de datos financieros	55
Copia de datos financieros	55
Jerarquías	55
Jerarquías múltiples	56
Jerarquías con múltiples raíces	57
Mantenimiento de jerarquías	58
Atributos	58
Tipos de atributos	59
Mantenimiento de atributos	61
Modelos	61



Resolución de datos	62
Transferencia de datos y estructuras	63
Procesamiento de tareas	69
Reportes gráficos y worksheets	70

Sistema de Control de Presupuestos

Descripción del problema	71
Diagrama de contexto	73
Concepción y arquitectura general del sistema	74
Descripción del flujo de datos	76
Estructura de datos	79
Reporting y análisis	85



Introducción

Las empresas acumulan grandes cantidades de datos corporativos, que constituyen un gran potencial en cuanto a información histórica relacionada con el desempeño, la competencia, los clientes, los productos, la experiencia, los presupuestos y demás. La idea es lograr que toda esta información pueda ser investigada y analizada para obtener ventajas estratégicas, pero existen muchas barreras para que este conjunto de fragmentos de información pueda ser considerado como una unidad, especialmente teniendo en cuenta que los sistemas en la empresa se desarrollan en distintos departamentos en diferentes ubicaciones.

Aún con sistemas integrados, la complejidad de los datos hace difícil su localización, comprensión y uso. Descubrir el dato correcto es frecuentemente la tarea más difícil y la que más tiempo consume. Es necesario proveer mecanismos para permitir al usuario localizar y entender rápidamente los datos.

*La estrategia de **Data Warehousing** reconoce la necesidad de consolidar y almacenar los datos en sistemas de información para ayudar al usuario a tomar decisiones efectivas. El objetivo es mejorar la performance de la empresa, proveyendo a los usuarios finales acceso a información crítica con un tiempo de respuesta razonable.*

Un Data Warehouse organiza y orienta los datos desde la perspectiva del usuario final. Contiene datos históricos y datos resumidos, junto con datos detallados extraídos de los sistemas operacionales. Proveer datos resumidos simplifica el procesamiento de las consultas del usuario final, facilita el análisis y la toma de decisiones.

La componente principal de un Data Warehouse es el análisis, mediante el cual el usuario final, analizando grandes volúmenes de información, puede planificar, pronosticar y tomar decisiones.

Datawarehouses enormes que integran datos operacionales con información de mercado, clientes y proveedores, muchas veces resultan en una explosión de información que se vuelve muy difícil de utilizar.

*Para alcanzar conjuntamente los objetivos de almacenamiento de grandes volúmenes de información, planificación y análisis en el Data Warehouse, aparece la idea de utilizar la **tecnología OLAP** (On Line Analytical Processing), diseñada para proveer la funcionalidad y performance necesarias.*

Esta nueva y sofisticada tecnología provee a los usuarios poderosas funciones para análisis, síntesis y consolidación de datos, con un mínimo conocimiento de la estructura de los mismos pero con un máximo control sobre la información a la cual deben acceder de acuerdo al nivel de decisión definido dentro de la organización.

Es necesario diferenciar entre las tecnologías OLAP y OLTP (On Line Transaction Processing). En las aplicaciones OLTP, los datos se reúnen y almacenan para operación y control mientras que en las OLAP se depositan datos para consultas, análisis y divulgación.

Las aplicaciones OLTP están caracterizadas por varios usuarios creando, actualizando o recuperando registros individuales. Son utilizadas por los gerentes y analistas que frecuentemente requieren altos grados de agregación de los datos desde distintas perspectivas.

Las aplicaciones OLAP son normalmente actualizadas en forma batch, en general, desde distintas fuentes de datos y proveen un poderoso back-end analítico para múltiples aplicaciones de usuarios.

Las aplicaciones OLAP comparten un conjunto de usuarios y requerimientos funcionales que no pueden ser alcanzados aplicando herramientas de query o de productividad personal directamente contra los datos históricos mantenidos en la base de datos relacional del warehouse.

La tecnología OLAP logra su máxima flexibilidad y poder utilizando la tecnología de bases de datos multidimensionales.

Una base de datos multidimensional es una base de datos diseñada con el propósito de almacenar y recuperar grandes volúmenes de información en forma eficiente.

Muchas veces los datos empresariales son claramente de naturaleza multidimensional, con lo cual es natural almacenarlos en una base de datos de este tipo. El hecho de almacenar los datos de acuerdo a su naturaleza, ofrece ventajas como facilidad de presentación y navegación de datos, facilidad de mantenimiento y mejor performance.

En este marco, analizamos la problemática del Data Warehouse, la tecnología OLAP para alcanzar la funcionalidad y eficiencia requeridas y estudiamos las características de las bases de datos multidimensionales como una alternativa para el almacenamiento de datos.

Finalmente presentamos una herramienta usada actualmente en el mercado para resolver problemas de este tipo, mostrando los conceptos de las bases de datos multidimensionales y las distintas formas de efectuar consultas y análisis, en un problema real de control de presupuestos.

Data Warehousing





Data Warehousing

La estrategia de Data Warehousing reconoce la necesidad de consolidar y almacenar los datos manejados en los distintos sistemas de información de la compañía facilitando la toma de decisiones y asegurando respuestas más rápidas y efectivas.

Data Warehouse

Un Data Warehouse es un conjunto de datos integrados que varían con el tiempo, que no son transitorios y soportan el proceso de toma de decisiones.

Es un repositorio de datos con información proveniente de diversos orígenes, principalmente de bases de datos transaccionales.

La información que contiene puede ser de distintos tipos:

- ✓ *información puramente transaccional*
- ✓ *información resultante de procesar datos transaccionales*
- ✓ *información estadística preelaborada que permite optimizar el tiempo de respuesta*
- ✓ *datos elaborados en función de los objetivos del negocio*

Los objetivos de un Data Warehouse son:

- ✓ *proveer acceso inmediato a datos corporativos con una alta performance*
- ✓ *asegurar la consistencia de los datos que contiene*
- ✓ *los datos almacenados pueden estar combinados y separados según los objetivos del negocio*
- ✓ *los datos se almacenan de modo de facilitar las consultas, análisis y presentación de información*
- ✓ *los datos provenientes de distintas fuentes son cuidadosamente ensamblados, depurados y publicados sólo si son apropiados para su uso*

Finalidad del Data Warehouse

La finalidad del Data Warehouse consiste en auxiliar a la empresa a comprender el pasado y planear el futuro, la idea es obtener la respuesta adecuada a tiempo para tomar mejores decisiones.

Los datos operacionales correctos se encuentran almacenados en el Data Warehouse, el desafío es sacar esos datos del Data Warehouse y convertirlos en información que ayude a la empresa a realizar elecciones más adecuadas, lo que implicará tomar mejores decisiones y obtener una ventaja significativa respecto de aquellos que no cuentan con la información .

Los datos históricos operativos reunidos en el Data Warehouse , estarán organizados para derivar un valor estratégico y no solamente como un conjunto abundante de datos.

La utilización del Data Warehouse evoluciona en distintas fases. En la primera, se usa para reportes y consultas predefinidos. A continuación, se utiliza para analizar datos de resumen y de detalle, presentando los resultados en forma de reportes y gráficos.

Estas primeras fases se usan también para verificar la calidad y suficiencia de los datos y capacitar a los usuarios de la empresa. A medida que éstos se sienten más seguros comienzan a usar el Data Warehouse con fines estratégicos, con análisis multidimensionales y operaciones sofisticadas.

Es necesario tener en cuenta que aún cuando el Data Warehouse contiene información estratégica, derivar de él un valor empresarial es un esfuerzo complejo.

Las empresas desean usar los datos del Data Warehouse, no sólo para reportar el pasado, sino para ayudar a planear el futuro.

Características claves de una solución de Data Warehouse

✓ Política

La determinación de las reglas del negocio, tales como objetivos y alcance de los datos, tópicos operacionales, tales como carga y frecuencia de mantenimiento y tópicos organizacionales como seguridad de la información, son aspectos tan importantes como la selección de la tecnología y productos a utilizar para asegurarse el éxito en la implementación de un Data Warehouse.

✓ Transformación

Los datos pueden ser almacenados en el Data Warehouse sin ningún tratamiento previo o luego de un proceso de transformación y refinamiento. La elección de la forma de almacenamiento depende de quién será el usuario que accederá a la información.

Esta transformación puede involucrar la reestructuración, redefinición, filtrado, combinación, recálculo y sumarización de datos provenientes de distintos sistemas transaccionales.

✓ Almacenamiento

Los datos almacenados en el Data Warehouse provienen de distintos sistemas transaccionales, debido a ésto el sistema de almacenamiento debe ser muy flexible para permitir la incorporación de datos nuevos y el acceso total a la información. En el Data Warehouse se almacenan datos históricos, transaccionales y los que se denominan metadatos (información sobre el origen, formato y ubicación de los datos). La tecnología más utilizada para el almacenamiento de los datos históricos responde a un modelo relacional.

✓ Análisis

Es la componente de la solución de Data Warehouse encargada de comprender cual es el rumbo del negocio, por lo tanto debe soportar los queries que necesita el usuario final para planificar, pronosticar y poder tomar mediciones del estado del negocio analizando grandes volúmenes de información. En el manejo de esta información de gestión la tecnología más adecuada es un servidor OLAP (On Line Analytical

Processing) a diferencia de la componente de almacenamiento donde el modelo transaccional era el más adecuado.

✓ *Presentación y acceso*

Debe permitir al usuario seleccionar, consultar y manipular los datos almacenados en el Data Warehouse, para esto debe proveer herramientas de navegación.

Datos en un Data Warehouse

Un Data Warehouse contiene datos históricos, datos resumidos y datos detallados extraídos de los sistemas operacionales.

*Proveer **datos resumidos** simplifica el procesamiento de las consultas del usuario final, facilita el análisis y la toma de decisiones.*

*Los **datos históricos** juegan un papel muy importante en el análisis de gestión, ya que nos permiten observar el comportamiento de la empresa a lo largo del tiempo. De esta forma se pueden ver tendencias y desvíos que serán de gran ayuda en la elaboración de nuevas políticas empresariales. Estos datos generalmente provienen de fuentes de información diversas, las cuales pueden involucrar formatos complejos como: archivos multiregistros, grupos repetitivos y registros de tamaño variable. Por este motivo es necesario tener en el Data Warehouse lo que se denomina metadatos.*

*Los **metadatos** contienen la información necesaria para acceder y comprender las distintas fuentes de datos almacenadas en un Data Warehouse.*

Características de los datos

✓ *Granularidad de los datos*

Los datos pueden estar en distintos niveles de detalle: actuales, anteriores, poco resumidos, muy resumidos, agregados, etc. Las necesidades de la empresa determinan cuáles datos se necesitan y el nivel de granularidad requerido.

Las empresas podrían almacenar datos con distintos niveles de granularidad combinando Data Warehouses y Data Marts, con información detallada para consultas en el Data Warehouse e información resumida en el Data Mart para una recuperación frecuente y rápida.

✓ *Información procesada con anticipación*

Un Data Warehouse contiene información procesada y almacenada con anticipación para acceso y utilización rápidos. De este modo se cuenta con información de rutina, como los reportes mensuales y trimestrales y reportes y gráficos de resumen de

comercialización. En ocasiones también se cargan y ofrecen bases de datos multidimensionales incorporadas con anticipación para un fácil análisis interactivo.

✓ *Selección y procesamiento de información por solicitud*

Para la información no rutinaria, existen consultas y reportes predefinidos de forma tal que el usuario pueda seleccionar el reporte o consulta y procesarlo cuando lo necesita. Los reportes, los gráficos y la carga de bases de datos multidimensionales están disponibles cuando se solicitan.

✓ *Acceso y procesamiento Ad-hoc*

Para cubrir las necesidades no establecidas de los usuarios (ya que no es posible preparar las posibles necesidades de todos los usuarios con anticipación) se requiere de un acceso y procesamiento ad-hoc.

✓ *Calidad de datos de la información*

La calidad y confiabilidad de los datos es la base de la confianza en el Data Warehouse. El Data warehouse debe contener información sobre la calidad de los datos, de forma tal que las decisiones se tomen con un absoluto conocimiento de causa y se manejen los riesgos en forma adecuada.

Componentes de un Data Warehouse

Presentamos una arquitectura de referencia que permite analizar las componentes de un Data Warehouse. Esta arquitectura divide las componentes en bloques y capas.

Los bloques se relacionan con la funcionalidad específica del Data Warehouse y son los siguientes:

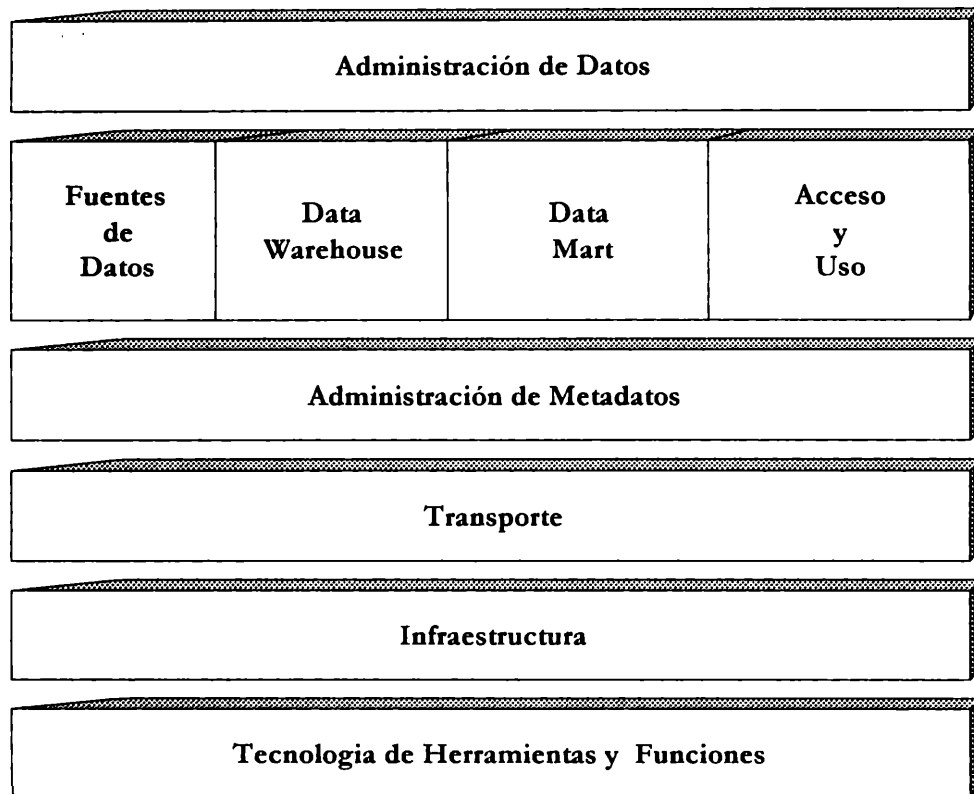
- ✓ *Fuentes de Datos*
- ✓ *Data Warehouse*
- ✓ *Data Mart*
- ✓ *bloque de Acceso y Uso*

Las capas representan el ambiente necesario para implementar los bloques y son:

- ✓ *Administración de Datos*
- ✓ *Administración de Metadatos*
- ✓ *Transporte*
- ✓ *Infraestructura*

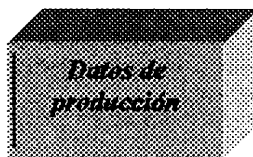
Las capas de Administración de Datos y Administración de Metadatos corresponden a las actividades que se relacionan con la extracción, carga y actualización de datos que se requieren para mantener la información del Data Warehouse.

Arquitectura de Referencia

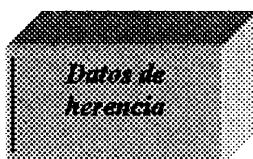


Bloques

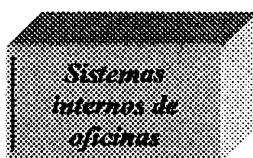
- **Bloque de fuentes de datos:** *este bloque se divide a su vez en las siguientes componentes:*



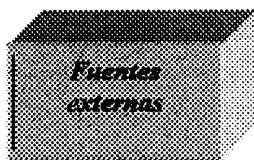
- *Se refieren a las bases operacionales de datos que contienen la información recopilada de las aplicaciones con las que opera la empresa. Estas bases de datos operacionales pueden ser relacionales, no relacionales o basadas en archivos.*



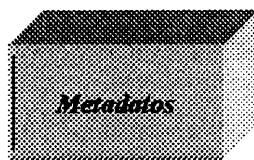
- *Los datos de herencia están fuera de línea, porque ya no son necesarios para apoyar aplicaciones operacionales actuales. Sin embargo, esta información tiene un gran valor histórico y es de gran utilidad para el análisis de tendencias.*



- *Existen fuentes de datos que no están almacenadas en una base de datos operacional, ni son usados por una aplicación operacional. En general estos datos son los siguientes: reportes, documentos de procesadores de texto, planillas de cálculo, reportes anuales o archivos.*



- *Las fuentes externas de información no las controla, posee, ni opera la empresa. Pueden ser electrónicas (resúmenes de análisis competitivos de mercado), o no electrónicas como los artículos en revistas o diarios especializados.*

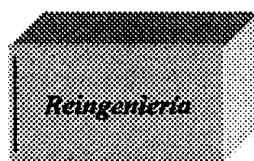


- *Los metadatos representan la información de definición acerca de los datos. Los metadatos contienen el nombre del dato capturado, la definición del contenido de los datos (campo), la fecha en que fueron creados y la fuente de origen.*

- **Bloque de construcción del Data Warehouse:** este bloque se divide a su vez en las siguientes componentes:



- *La componente de refinamiento es responsable de estandarizar, limpiar, filtrar, confrontar y agregar fecha de origen de la información a los datos seleccionados. En esta componente también se crean y capturan metadatos adicionales para los nombres y definiciones de datos estándar.*



- *La componente de reingeniería es responsable de preparar los datos para que sean congruentes con las necesidades del usuario final. La reingeniería en el contexto del Data Warehouse, es distinta de la reingeniería de las aplicaciones o procesos empresariales.*

La reingeniería implica las siguientes actividades:

- ✓ *La integración de diferentes tipos de datos de sistemas múltiples para crear datos nuevos.*
- ✓ *El cálculo previo de la información resumida y derivación de otra información que requiera el usuario empresarial.*
- ✓ *La traducción y formato de datos de diferentes fuentes para poder combinarlos de manera uniforme y consistente.*
- ✓ *La transformación y reubicación de los datos almacenados en las fuentes originales de información para permitir la actualización constante de la nueva información derivada.*

Los metadatos se crean en la componente de reingeniería. Estos son una parte esencial del Data Warehouse.

Los metadatos incluyen:

- ✓ *La estructura misma del Data Warehouse*
- ✓ *Descripción de la base de datos*
- ✓ *Las rutas y las reglas de navegación del Data Warehouse*
- ✓ *El glosario empresarial*

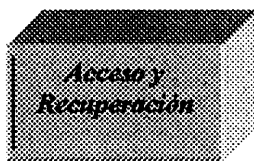
- ✓ *Las listas de consultas y reportes diseñados y predefinidos*

Los metadatos guían al usuario en un contexto apropiado para la navegación y visualización de los datos. Debido a que se condensa la información antigua y detallada, se guardan los metadatos de estas acciones.

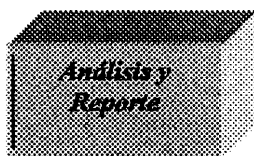
▪ **Bloque de construcción del Data Mart:** *el Data Mart tiene componentes similares a los del Data Warehouse. La principal diferencia entre estos dos es el enfoque del usuario final. El Data Mart es opcional para la mayoría de las organizaciones que ya tienen un Data Warehouse. Por otro lado algunas organizaciones implementan solo varios Data Mart, sin un Data Warehouse para toda la organización.*

El bloque de creación del Data Mart posee los mismos componentes de los del bloque de creación del Data Warehouse. La diferencia radica en el enfoque de cada uno de los componentes. Por ejemplo, el bloque de modelado en el Data Mart se concentra en las necesidades del departamento individual que posee el Data Mart, mientras que el bloque de modelado del Data Warehouse debe contemplar las necesidades de todos los departamentos de la organización.

▪ **Bloque de acceso y uso del Data Warehouse:** *el tercer gran bloque de la arquitectura del Data Warehouse tiene dos componentes:*



▪ *Esta componente es responsable de transformar los datos recuperados en vistas multidimensionales o almacenarlos en una base multidimensional para su posterior análisis. Otra función de este bloque es permitir el acceso al usuario para que examine y navegue por los metadatos, de esta manera podrá entender qué es el Data Warehouse, el contexto de los datos, las transformaciones que se hicieron, la ubicación de los datos, verificar la confiabilidad de los mismos, etc.*



▪ *Esta componente es la responsable de brindar las herramientas y aplicaciones necesarias para explotar todas las posibilidades que brindan el Data Warehouse y el Data Mart. Estas se clasifican de la siguiente manera :*

- ✓ *herramientas de reportes*
- ✓ *herramientas de análisis y soporte de decisiones*
- ✓ *herramientas de modelado empresarial y procesamiento analítico*
- ✓ *herramientas de Data Mining*

Capas

▪ **Capa de administración de datos:** *en esta capa se realizan las tareas de extracción, carga, actualización, seguridad y restauración del Data Warehouse. El interés particular en esta capa reside en:*

- ✓ *Extraer los datos apropiados a fin de seleccionar las fuentes de información para el refinamiento posterior, la reingeniería y la incorporación en el Data Warehouse*
- ✓ *Seguir y llenar las solicitudes de datos nuevos por parte de fuentes de datos nuevas o actuales*
- ✓ *Capturar los datos modificados en las fuentes de datos operacionales y después actualizar el Data Warehouse.*

▪ **Capa de administración de metadatos:** *la arquitectura del Data Warehouse se basa en el concepto de definiciones de datos, o metadatos. Los metadatos cumplen un rol muy importante en el Data Warehouse ya que están presentes en todas las actividades del mismo. Las fuentes de información se caracterizan por la definición de los datos que llegan. Registrar la fecha requiere definir metadatos adicionales relacionados con tal proceso. Lo mismo ocurre con el registro de las fuentes de origen.*

La capa de administración de metadatos proporciona guías y punteros hacia la información del Data Warehouse, esta capa se divide en las siguientes componentes:

- ✓ *Administración del Data Warehouse, Data Mart y glosario*
- ✓ *Administración de la extracción, creación, depósito y actualización de los metadatos*
- ✓ *Administración de las consultas predefinidas, reportes, índices y perfiles*
- ✓ *Administración de la actualización y duplicación*
- ✓ *Administración de conexiones, generación de archivos, restauración y depuración*

En esta capa es donde se administran las bases de datos, con su complejidad en las áreas de índices múltiples, compactación de datos, claves compuestas y las versiones de los datos.

▪ **Capa de transporte:** *Es la responsable de transportar los datos entre los distintos bloques de la arquitectura. Esta capa utiliza tecnología de actualización y duplicación, red para transferencia y entrega de datos y componentes de middleware. La capa de transporte esta compuesta por las siguientes componentes:*



- *Esta componente contiene los siguientes tipos de sistemas: protocolos de red, diagramas de administración de redes, sistemas operativos de red y tipos de redes.*



- *La componente de agente cliente/servidor y middleware contiene los siguientes subsistemas: accesos a la base de datos, middleware orientado a mensajes e intermediarios de solicitudes objeto.*



- La componente de duplicación contiene sistemas de duplicación y propagación.

- **Capa de infraestructura:** esta capa se encarga de administrar la integración y ejecución ordenada de las aplicaciones para extraer, actualizar, duplicar, agregar y resumir el Data Warehouse

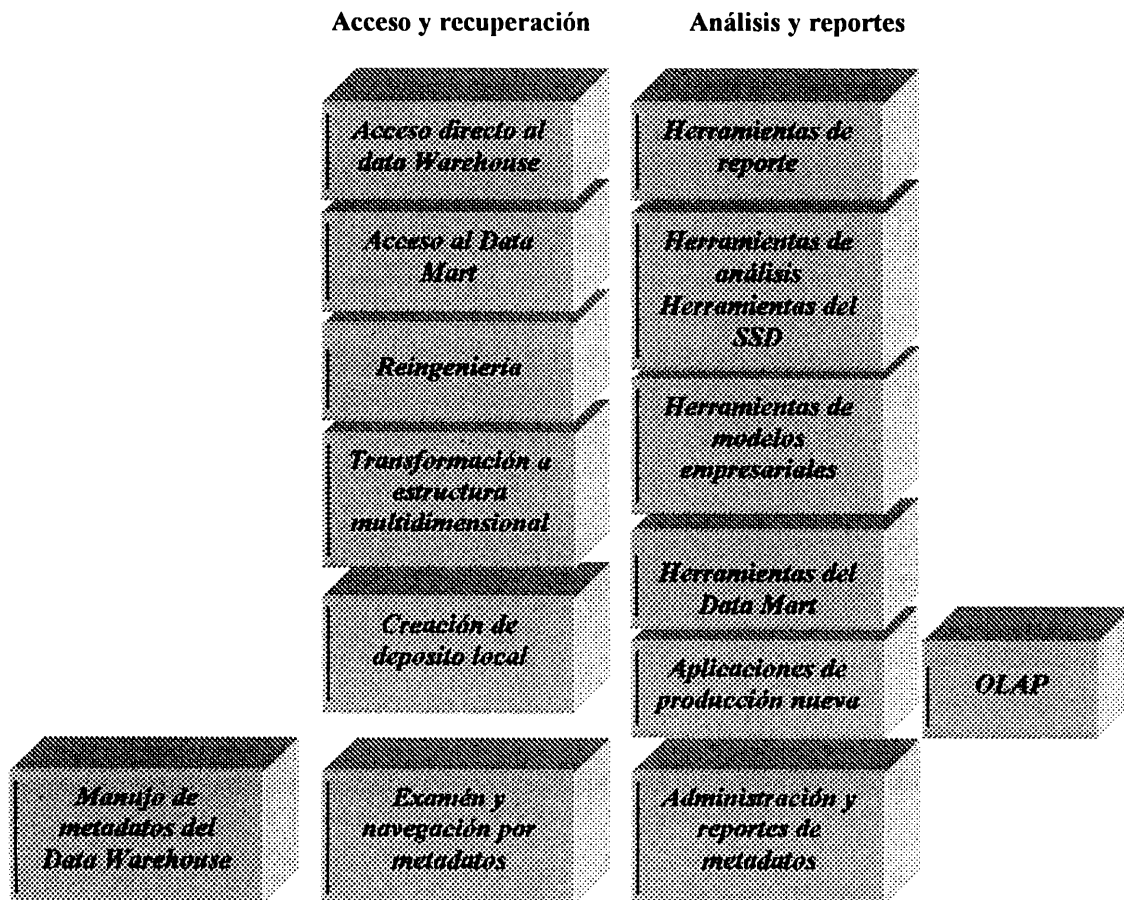
Uso del Data Warehouse

Desde una perspectiva empresarial el Data Warehouse se construye para proporcionar datos e información de una manera y en un formato utilizables para los usuarios de la empresa y para fortalecerlos con diversas herramientas y características para analizar los datos y tomar decisiones.

El Data Warehouse debe proporcionar un acceso flexible a los datos a partir de una amplia variedad de herramientas.

El uso de un Data Warehouse consta de dos actividades principales: acceso y recuperación y análisis y reportes.

Estas actividades están encuadradas en el bloque de Acceso y Uso de la arquitectura de referencia, el cual contiene ambas componentes.



El sub-bloque de Acceso y Recuperación proporciona acceso directo tanto al Data Warehouse como al Data Mart. Ofrece servicios como reingeniería de los datos recuperados para almacenarlos correctamente en bases de datos multidimensionales, o preparación de vistas multidimensionales para análisis.

Una característica importante de este sub-bloque es el manejo para acceso y exámen de metadatos.

Los usuarios aprovechan estas capacidades de exámen y navegación por los metadatos para realizar las siguientes acciones:

- ✓ Localizar y entender qué existe en el almacenamiento de datos y en el Data Mart.*
- ✓ Ver los datos en su contexto apropiado para convertirlos en información.*
- ✓ Conocer la ubicación de los datos y como acceder a ellos.*
- ✓ Interpretar la calidad de la fuente de datos, la validez de las transformaciones realizadas en los datos desde el momento de su extracción hasta su almacenamiento, y juzgar la confiabilidad y calidad de los datos.*

El sub-bloque de Análisis y Reportes proporciona un conjunto de herramientas y aplicaciones para aprovechar el Data Warehouse.

Tenemos distintos tipos de herramientas tales como las de reporte para estadísticas y análisis de los datos, las de modelos empresariales, las de proceso analítico y las de minería de datos.

Las herramientas adicionales para navegación y reportes sobre los metadatos, proporcionan el mapa esencial del Data Warehouse y del Data Mart. Este mapa lleva al usuario a comprender los datos en el contexto adecuado, con información sobre el nivel de granularidad, el nivel de resumen, la fuente de los datos, la fecha de origen y las reglas empresariales aplicadas a los datos.

Diferencias entre Data Warehouse y bases de datos operacionales

- ✓ El almacenamiento de datos en un Data Warehouse es diferente del almacenamiento en las bases de datos operacionales que soportan las aplicaciones de Procesamiento de Transacciones en Línea (OLTP), ya que organiza y orienta los datos desde la perspectiva del usuario final. En cambio, los sistemas operacionales organizan sus datos desde el punto de vista de la aplicación, de modo de realizar el acceso a los mismos con la mayor eficiencia posible. Con frecuencia los datos organizados de esta manera no permiten realizar las consultas empresariales necesarias para la toma de decisiones en los niveles gerenciales.*
- ✓ Administra grandes volúmenes de información: la mayoría de los Data Warehouse contienen información histórica que se retira con frecuencia de los sistemas operativos porque ya no es necesaria para las aplicaciones operacionales y de producción. Por el volumen de información que un Data Warehouse debe manejar, debe ofrecer opciones para la adición y la condensación de datos. Un Data Warehouse maneja información a diferentes niveles de granularidad. Debido a la necesidad de administrar toda la información histórica junto con los datos actuales, la dimensión de un Data Warehouse es mucho mayor que la de las bases de datos operacionales.*
- ✓ Debido a los grandes volúmenes de información que debe manejar, un Data Warehouse guarda información en diferentes medios de almacenamiento.*

- ✓ *Comprende múltiples versiones de un esquema de base de datos: el Data Warehouse tiene que guardar y administrar información histórica que ha sido manejada en distintos momentos por diferentes versiones de esquemas de base de datos.*
- ✓ *Condensa y agrega información: con frecuencia es muy alto el nivel de detalle de la información guardada por las bases de datos operacionales para cualquier toma de decisiones sensata. Un Data Warehouse condensa y agrega la información para presentarla en forma comprensible a los usuarios.*
- ✓ *Integra y asocia información de diversas fuentes de información: debido a que las organizaciones han administrado históricamente sus operaciones utilizando numerosas aplicaciones de software y múltiples bases de datos, se requiere de un Data Warehouse para recopilar y organizar en un solo lugar la información que estas aplicaciones han acumulado con el paso de los años. Esta es una tarea desafiante por la diversidad de tecnologías de almacenamiento, de técnicas de administración de base de datos y de la semántica de los datos.*

Técnicas para utilizar el Data Warehouse

Las aplicaciones y herramientas del Data Warehouse que se emplean para recuperar, manipular y analizar los datos, y presentar después los resultados, se denominan genéricamente herramientas de soporte de decisiones. Se usan en dos modalidades: verificación y descubrimiento.

En la modalidad de verificación, el usuario crea una hipótesis e intenta confirmarla accediendo a los datos del Data Warehouse. Las herramientas que implementan ésta modalidad son de consulta, de sistemas de reporte y de análisis multidimensional.

La modalidad de verificación es también conocida como modalidad conducida por el usuario

En la modalidad de descubrimiento, las herramientas intentan descubrir características en los datos no conocidas ni sospechadas, por ejemplo: patrones de compra, asociación entre la adquisición de artículos diferentes. Una herramienta de la modalidad de descubrimiento es por ejemplo la herramienta de minería de datos (data mining).

La modalidad de descubrimiento es también llamada modalidad conducida por los datos.

Desde la perspectiva de disponibilidad de herramientas, las dos modalidades de verificación y descubrimiento se clasifican en tres enfoques: procesamiento informático, procesamiento analítico y minería de datos.

El procesamiento informático y el procesamiento analítico apoyan la modalidad de verificación mientras que el data mining apoya la modalidad de descubrimiento.

Las herramientas de data mining recorren los datos transaccionales detallados para descubrir patrones y asociaciones ocultos.

Procesamiento analítico

El procesamiento informático apoya la modalidad verificación del soporte de decisiones. Comprende técnicas como análisis estadísticos básicos y de datos, consultas para acceder y recuperar los datos del Data Warehouse y la presentación de los análisis realizados en forma de reportes, gráficos, tablas, etc. Los datos que se acceden y procesan pueden ser históricos o recientes y pueden estar poco o muy resumidos.

El procesamiento informático asiste a los usuarios en la búsqueda de respuestas a cuestiones tales como:

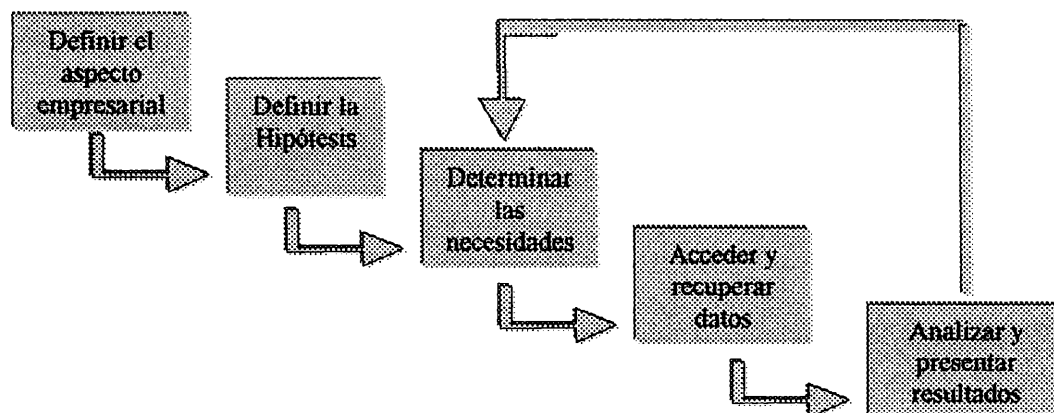
- ✓ ¿Están los artículos deportivos sujetos a tendencias por temporadas? ¿Cuáles elementos se afectan y cuando?
- ✓ ¿Cuántos automóviles se vendieron el mes pasado? ¿Cómo se compara esta cantidad con el mismo mes en los últimos cinco años?
- ✓ ¿Quiénes son los diez mejores Vendedores en la región sur?
- ✓ ¿Cuáles son los diez artículos de menor rentabilidad en el catálogo de ventas?

El procesamiento informático consta de tres componentes distintos:

- ✓ Consultas para acceder y recuperar datos del Data Warehouse
- ✓ El cálculo, análisis y manipulación de los datos
- ✓ Presentación del análisis en forma de reportes tabulares y matriciales, tablas, cuadros y gráficos.

Desde la perspectiva de la arquitectura de referencia, el procesamiento informático está relacionado con el bloque de Acceso y Uso.

Los pasos que comprenden el procesamiento informático son en esencia los mismos pasos de análisis de decisiones que un usuario empresarial experimentado aplica normalmente



El usuario debe primero definir el problema empresarial a analizar. A continuación deben definirse las hipótesis o cuestiones que requieren una respuesta. Es muy importante poder transformar el problema en una hipótesis y en cuestiones que pueda resolver el Data Warehouse.

A continuación, el usuario determina si se cuenta con los datos apropiados y si los datos son suficientes para resolver el problema. Se emplean ahora las herramientas de procesamiento informático para acceder al Data Warehouse y recuperar los datos.

Por último se analizan los datos y se presentan los resultados como recomendaciones factibles.

En muchas ocasiones, se requiere un análisis iterativo. El usuario vuelve a definir la hipótesis original o accede a datos adicionales para verificar dicha hipótesis.

Para satisfacer las necesidades antes descritas el ambiente del procesamiento informático debe manejar un amplio rango de: tipos de consultas y reportes, capacidad de análisis y funciones de acceso y recuperación del Data Warehouse.

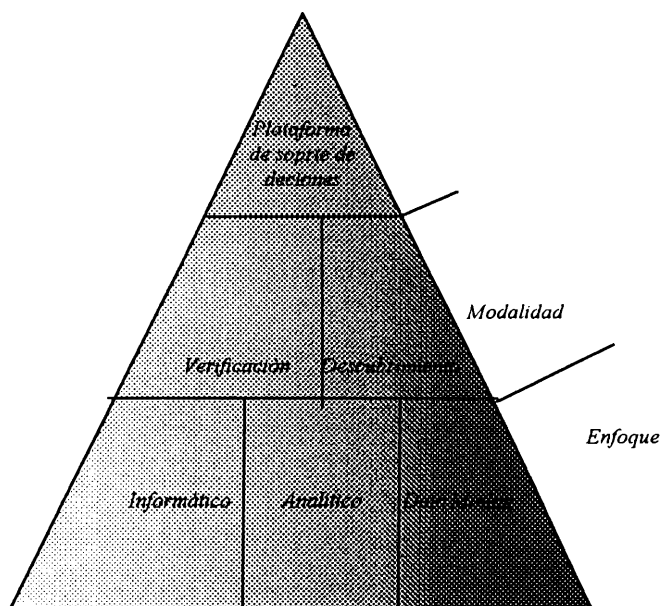
Procesamiento analítico

Los usuarios desean extraer los datos adecuados de forma intuitiva con una mínima inversión de tiempo y frustración. Una vez extraídos los datos correctos, los analizan, sintetizan y consolidan en información que utilizan para la toma de decisiones.

El procesamiento analítico se circunscribe a la modalidad de verificación: el usuario crea una hipótesis y accede a los datos para verificarla o confirmarla. Por lo general el análisis de los datos es interactivo, una hipótesis conduce a otra hasta obtener un conjunto claro de alternativas y recomendaciones potenciales.

Durante el análisis iterativo, los usuarios descubren muchas veces relaciones insospechadas entre parámetros empresariales, por ejemplo la relación de las ventas con la edad y el sexo de los clientes. Es decir, el procesamiento analítico no está limitado únicamente a la "verificación" sino también al "descubrimiento".

Existen varias formas de extraer y analizar información de valor del Data Warehouse como se puede ver en esta figura:



Visión de los datos del usuario

Muchos de los datos de la empresa son, de hecho, eminentemente multidimensionales. Se encuentran relacionados y son regularmente jerárquicos. Por ejemplo los datos de ventas, los de inventario y los de presupuesto están totalmente interrelacionados y dependen entre sí.

En la práctica, para predecir las ventas de un nuevo producto, habrá que estudiar los patrones de compras anteriores, la adopción de nuevos productos, las preferencias regionales y otros factores.

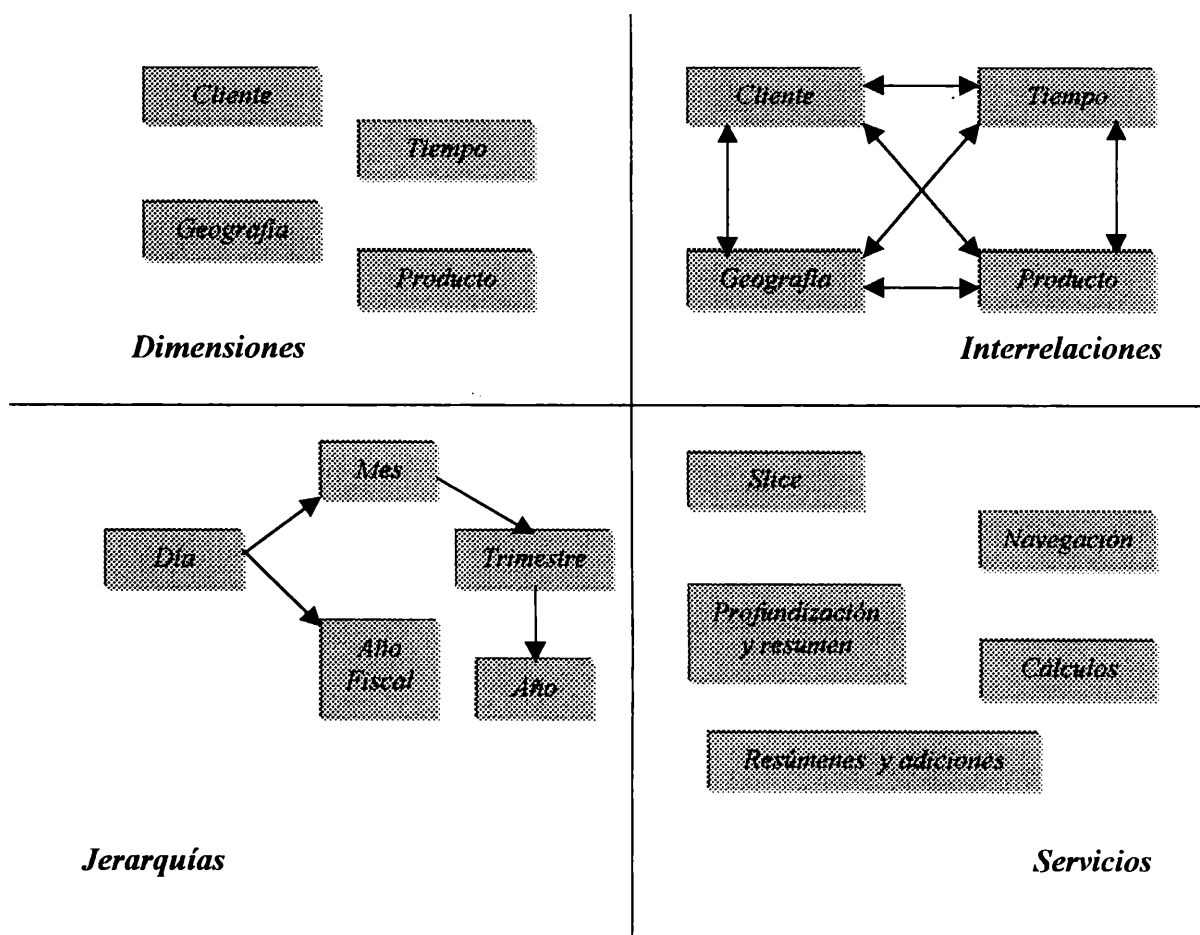
Las empresas necesitan buscar mercados donde sus productos y sus servicios tengan ventajas competitivas y sean diferentes. Para lograr esto es un requisito el análisis multidimensional.

Análisis Multidimensional

Tanto para la eficiencia operativa como para la planeación a futuro es necesario analizar datos de la empresa interrelacionados. Esta necesidad se aborda mediante el "procesamiento analítico". En éste el enfoque está en el análisis de los datos especialmente en el análisis multidimensional.

En el análisis multidimensional los datos se representan mediante "dimensiones" como producto, territorio y cliente. Por lo general las dimensiones se relacionan mediante "jerarquías" por ejemplo: ciudad, estado, región, país y continente o estado, territorio y región. El tiempo es otra dimensión con su propia jerarquía: día, semana, mes, trimestre y año o día y año calendario.

Para facilitar el análisis complejo, el procesamiento analítico o análisis multidimensional presenta una visión sencilla de los datos.



Procesamiento analítico en línea (OLAP)

En un Data Warehouse, se depositan datos para consulta, análisis y divulgación, a diferencia del procesamiento de transacciones en línea (OLTP), en donde los datos se almacenan para operación y control. OLAP es una tecnología de procesamiento analítica que crea nueva información de análisis de gestión a partir de los datos existentes. El procesamiento analítico en línea provee a los usuarios con poderosas funciones para el análisis, síntesis y consolidación de datos con un mínimo conocimiento de la estructura de los mismos.

- ✓ *Presenta una visión multidimensional lógica de los datos en el Data Warehouse. La visión es independiente de cómo se almacenan los datos*
- ✓ *Comprende siempre la consulta interactiva y el análisis de los datos. Por lo general la interacción es de varias pasadas, lo cual incluye la profundización en niveles cada vez mas detallados o el ascenso a niveles superiores de resumen y adición.*
- ✓ *Ofrece opciones de modelado analítico, incluyendo un motor de cálculo para obtener proporciones, desviaciones, etc, que comprende medición de datos numéricos a través de muchas dimensiones*
- ✓ *Crea resúmenes, adiciones (también conocidas como consolidaciones) y jerarquías*
- ✓ *Maneja modelos funcionales de pronósticos, análisis de tendencias y análisis estadísticos.*
- ✓ *Recupera y exhibe datos tabulares en dos o tres dimensiones, cuadros y gráficas, posibilitando la vista de los mismos desde diferentes perspectivas.*
- ✓ *Responde con rapidez a las consultas, de modo que el proceso de análisis no se interrumpe y la información no se desactualiza.*
- ✓ *Tiene un motor de depósito de datos multidimensional, donde las dimensiones son una representación lógica de las dimensiones empresariales.*

La tecnología OLAP se aplica en muchas áreas funcionales de la empresa, tales como: producción, ventas, análisis de rentabilidad, análisis de logística, consolidaciones financieras, presupuestos, pronósticos y contabilidad de costos.

OLAP es una opción de análisis y reporte, muy importante dentro del bloque de Acceso y Uso del Data Warehouse.

Arquitectura OLAP

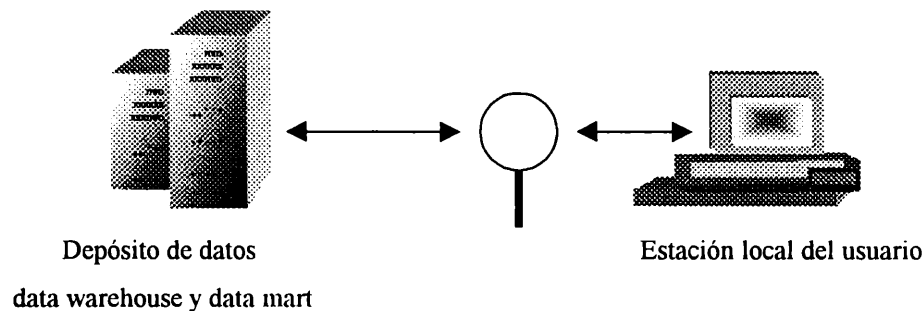
La arquitectura OLAP ofrece las siguientes opciones:

- *Acceder los datos directamente del Data Warehouse o del Data Mart, transformarlos en una estructura multidimensional y almacenarlos localmente en una estación de trabajo*
- *Acceder a los datos del Data Warehouse, transformarlos en una estructura multidimensional y almacenarlos con esta estructura en el Data Mart*
- *Acceder los datos directamente del Data Warehouse o del Data Mart y transformarlos en una visión multidimensional para su posterior análisis y reporte en la estación de trabajo local*

Arquitectura lógica

La arquitectura lógica consta de dos partes :

- **Visión OLAP:** *la visión OLAP es la presentación multidimensional y lógica de los datos del Data Warehouse o del Data Mart al usuario, sin importar cómo y dónde están almacenados.*
- **Tecnología de depósito de datos:** *el depósito de los datos puede realizarse bajo una estructura de base de datos relacional o multidimensional.*



Tecnología de depósito de datos

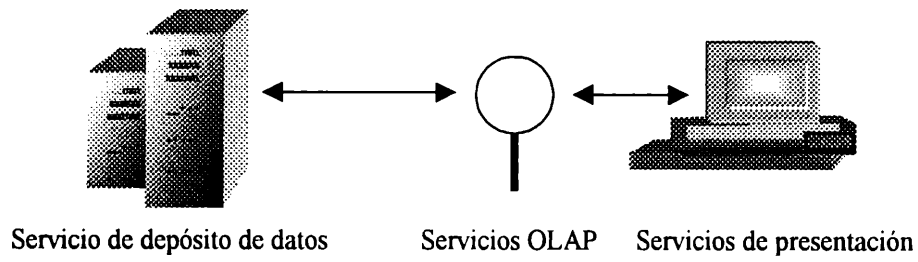
Visión OLAP

Al usuario sólo le interesa la visión multidimensional de los datos, mientras que al administrador del sistema le interesa dónde y cómo se almacenan para poder asegurar un desempeño eficaz.

Arquitectura funcional

La arquitectura funcional OLAP consta de tres componentes de servicio:

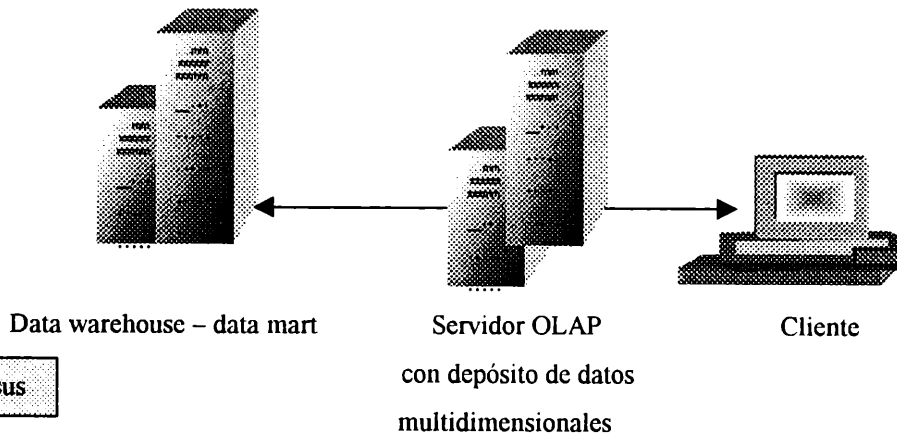
- *Servicio de depósito de datos*
- *Servicios OLAP*
- *Servicios de presentación al usuario*



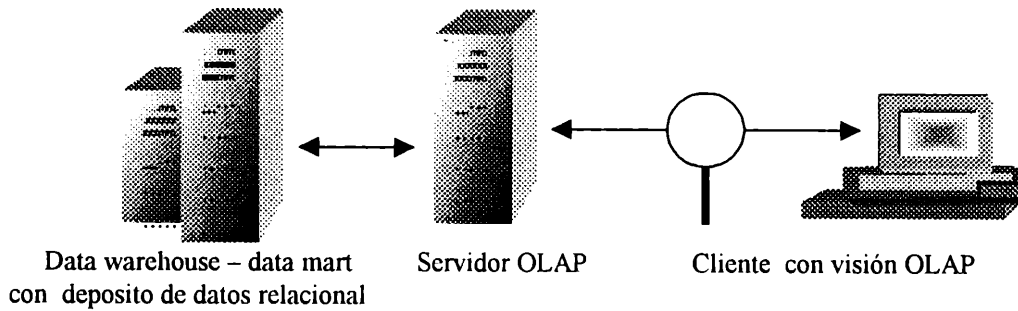
Arquitectura Física

La arquitectura física consta de dos grandes categorías basadas en tecnologías de depósito de datos multidimensional y depósito de datos relacional.

Multidimensiona



Versus



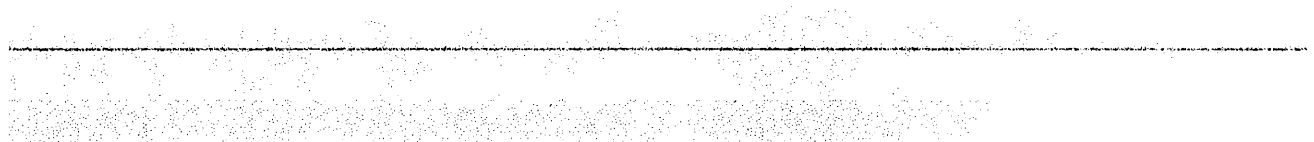
Relacional

Las bases de datos relacionales almacenan los datos en tablas como registros con claves y se acceden mediante un lenguaje común, el SQL. Por otra parte los depósitos de datos multidimensionales, almacenan los datos en arreglos.

Con un depósito de datos relacional, el Data Warehouse o el Data Mart pueden tener una dimensión muy grande; el tamaño aumenta por el uso de índices y técnicas de recuperación de datos para lograr un desempeño aceptable de las consultas multidimensionales.

Con el depósito multidimensional, el tamaño es limitado, pero utiliza tecnología de compresión de matriz dispersa para almacenar más datos en menos espacio.

Base de datos multidimensionales



Bases de datos multidimensionales

Una base de datos multidimensional es una base de datos diseñada con el propósito de almacenar y recuperar grandes volúmenes de información en forma eficiente. Esta información estará representada por un conjunto de datos íntimamente relacionados, vistos y analizados desde diferentes perspectivas. Estas perspectivas son llamadas **dimensiones**.

Cuando el usuario necesita utilizar los datos almacenados con el propósito de tomar decisiones importantes de su negocio, identificar tendencias o formular estrategias efectivas requieren ver los datos desde distintos puntos de vista.

Por ejemplo: En una agencia de automóviles un punto importante sería examinar los datos de ventas realizadas. La evaluación requeriría ver datos históricos de ventas desde diferentes perspectivas:

- ✓ Ventas por Modelo
- ✓ Ventas por Color
- ✓ Ventas a través del tiempo
- ✓ Ventas por Vendedor

En realidad las bases de datos relacionales podrían responder estas consultas, pero los queries deberían darnos resultados de una manera significativa (no sólo un listado de datos) con un buen tiempo de respuesta. Los usuarios finales cuando necesitan acceso interactivo a grandes cantidades de datos almacenados en un entorno relacional se ven muchas veces frustrados por los pobres tiempos de respuesta y la falta de flexibilidad ofrecida por la tecnología relacional y las herramientas de construcción de queries SQL.

En una base de datos relacional representaríamos los datos del ejemplo para un Vendedor con una tabla de este tipo:

Modelo	Color	Ventas
Sedan	Azul	4
Sedan	Rojo	3
Sedan	Blanco	2
4 x 4	Azul	3
4 x 4	Rojo	5
4 x 4	Blanco	5
Utilitario	Azul	6
Utilitario	Rojo	5
Utilitario	Blanco	4

La columna Ventas contiene el dato que queremos analizar. Las columnas Color y Modelo contienen las perspectivas desde dónde queremos analizar los datos. Por ejemplo el primer registro en la tabla dice que se han vendido 4 Sedan azules, se puede ver que hay 3 valores posibles para la columna Modelo: Sedan, 4x4, Utilitarios, la segunda columna Color tiene también 3 valores posibles Azul, Blanco y Rojo.

Una manera alternativa de representar estos datos es mediante una matriz. Así podemos representar las ventas en una matriz de 3x3 que es un ejemplo de un arreglo de 2 dimensiones. Un arreglo es la componente fundamental de una base de datos multidimensional.

M o d e l o	Sedan	4	3	2
	4 x 4	3	5	5
	Utilitario	6	5	4
		Azul	Rojo	Blanco
		Color		

En un arreglo cada eje es llamado **dimensión** (representa una de las perspectivas de nuestros datos), y cada elemento dentro de una dimensión es llamado **posición**. En este ejemplo las dimensiones son Modelo con 3 posiciones: 4x4, Sedan, Utilitario; y Color también con 3 posiciones: Azul, Blanco y Rojo. El valor dado por la intersección de Modelo y Color (fila y columna) representa la cantidad de unidades vendidas para ese Color y Modelo. Por ejemplo se vendieron 4 Utilitarios Blancos.

Hay varias razones por las que un arreglo es una forma más eficiente y efectiva de representar las ventas que en la tabla relacional.

1. El usuario detecta a simple vista que se tienen 2 dimensiones de 3 posiciones cada una, es mucho más difícil llegar a esta conclusión examinando la tabla relacional.
2. El arreglo agrupa convenientemente la información en filas y columnas, por ejemplo todas las ventas de Sedan están en una misma fila, así es muy fácil obtener rápidamente un total para Sedan o se pueden comparar rápidamente las ventas de ese Modelo para distintos colores. Por ejemplo se puede inferir cuál es el Color de Sedan más vendido. De la misma manera las ventas para cada Color aparecen en cada columna.

Es decir, el arreglo multidimensional representa un nivel más alto de organización que la tabla relacional. La estructura, al considerar las relaciones entre los datos, contiene en sí misma más "inteligencia" (información significativa), debido a que nuestras perspectivas están embebidas directamente en la estructura como dimensiones, en oposición a la representación mediante columnas en la tabla relacional.

La estructura de tabla relacional nos dice solamente que hay dos campos: Modelo y Color, pero no nos da ninguna información acerca de los posibles contenidos de esos campos. La estructura de arreglo, por otro lado nos dice no sólo que hay dos dimensiones Color y Modelo, sino que también presenta todos los valores posibles de cada dimensión como posiciones a lo largo de la dimensión. Debido a esta presentación estructurada, todas las posibles combinaciones de perspectivas conteniendo un atributo específico (el Color azul por ejemplo) están alineadas a lo

largo de las posiciones de la dimensión para ese atributo. Esto hace muy fácil no sólo el acceso y recuperación de datos, de acuerdo al tipo de datos específicos en el que estamos interesados, sino que hace más intuitivo el browsing y manipulación de datos al usuario final. Esta estructura de arreglo inteligente facilita el análisis de datos y ofrece ventajas en la performance. Un query consiste en seleccionar y organizar datos de una manera particular. Por ej. obtener las ventas para todos los Sedanes de Color azul. En la tabla relacional habría que recorrer cada registro individual y ver si corresponde al criterio de selección, mientras que con un arreglo multidimensional al tener pre-organizados los datos se reduce notablemente la búsqueda.

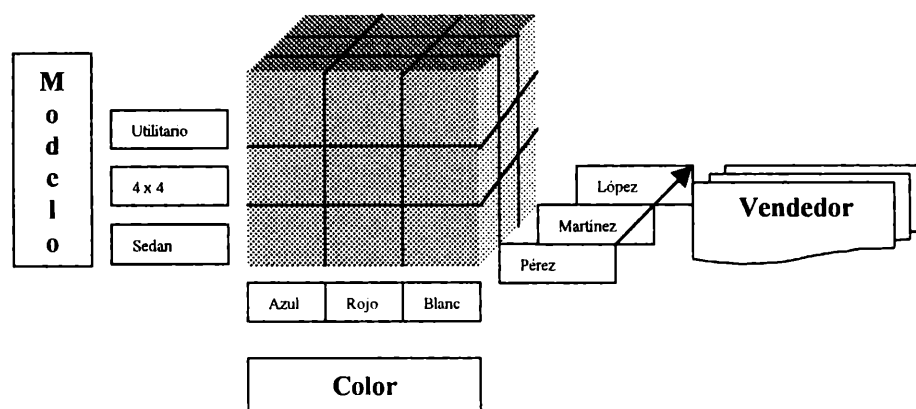
Consideramos el mismo ejemplo agregando una tercera dimensión Vendedor con tres valores posibles:

La estructura relacional quedaria representada en esta tabla:

<i>Modelo</i>	<i>Color</i>	<i>Vendedor</i>	<i>Ventas</i>
<i>Sedan</i>	<i>Azul</i>	<i>Pérez</i>	<i>1</i>
<i>Sedan</i>	<i>Azul</i>	<i>López</i>	<i>1</i>
<i>Sedan</i>	<i>Azul</i>	<i>Martínez</i>	<i>2</i>
<i>Sedan</i>	<i>Rojo</i>	<i>Pérez</i>	<i>1</i>
<i>Sedan</i>	<i>Rojo</i>	<i>López</i>	<i>1</i>
<i>Sedan</i>	<i>Rojo</i>	<i>Martínez</i>	<i>1</i>
<i>Sedan</i>	<i>Blanco</i>	<i>Pérez</i>	<i>2</i>
<i>Sedan</i>	<i>Blanco</i>	<i>López</i>	<i>1</i>
<i>Sedan</i>	<i>Blanco</i>	<i>Martínez</i>	<i>1</i>
<i>4 x 4</i>	<i>Azul</i>	<i>Pérez</i>	<i>3</i>
<i>4 x 4</i>	<i>Azul</i>	<i>López</i>	<i>2</i>
<i>4 x 4</i>	<i>Azul</i>	<i>Martínez</i>	<i>1</i>
<i>4 x 4</i>	<i>Rojo</i>	<i>Pérez</i>	<i>2</i>
<i>4 x 4</i>	<i>Rojo</i>	<i>López</i>	<i>1</i>
<i>4 x 4</i>	<i>Rojo</i>	<i>Martínez</i>	<i>3</i>
<i>4 x 4</i>	<i>Blanco</i>	<i>Pérez</i>	<i>4</i>
<i>4 x 4</i>	<i>Blanco</i>	<i>López</i>	<i>1</i>
<i>4 x 4</i>	<i>Blanco</i>	<i>Martínez</i>	<i>1</i>
<i>Utilitario</i>	<i>Azul</i>	<i>Pérez</i>	<i>2</i>
<i>Utilitario</i>	<i>Azul</i>	<i>López</i>	<i>3</i>
<i>Utilitario</i>	<i>Azul</i>	<i>Martínez</i>	<i>1</i>
<i>Utilitario</i>	<i>Rojo</i>	<i>Pérez</i>	<i>2</i>
<i>Utilitario</i>	<i>Rojo</i>	<i>López</i>	<i>3</i>
<i>Utilitario</i>	<i>Rojo</i>	<i>Martínez</i>	<i>1</i>
<i>Utilitario</i>	<i>Blanco</i>	<i>Pérez</i>	<i>2</i>
<i>Utilitario</i>	<i>Blanco</i>	<i>López</i>	<i>2</i>
<i>Utilitario</i>	<i>Blanco</i>	<i>Martínez</i>	<i>1</i>

En una estructura multidimensional agregar el campo Vendedor se traduce simplemente en agregar una tercera dimensión con tres posiciones. Ahora tenemos un arreglo de 3x3x3

conteniendo 27 celdas. Una vez más, el usuario puede visualizar más fácilmente los datos en las dimensiones pre-organizadas que en la tabla relacional.



Si la cantidad de posiciones en lugar de 3 es n , el arreglo se vería como un cubo de $n \times n \times n$ posiciones y en el formato relacional sería una tabla de $n \times n \times n$ registros

Ventajas de performance

Usando un arreglo de $10 \times 10 \times 10$ encontrar las ventas para Modelo = Sedan, Color = Azul y Vendedor = Pérez, en el modelo relacional requeriría recorrer 1000 registros para encontrar el correcto. En contraste, al tener la estructura multidimensional más "conocimiento" sobre los datos, solamente tiene que buscar a lo largo de tres dimensiones de 10 posiciones cada una para encontrar el "registro" (celda) correcto. Necesitamos como máximo 10 accesos a la dimensión Modelo para encontrar Sedan + 10 accesos a la dimensión Color para encontrar azul + 10 accesos a la dimensión Vendedor para encontrar a Perez, es decir, un total de 30 accesos como máximo. Vemos así una significativa ventaja en performance al comparar los 30 accesos en el peor de los casos en la estructura multidimensional versus los 1000 en la estructura relacional.

Veamos otro ejemplo: queremos encontrar el total de ventas para el Modelo Sedan en la tabla relacional tendríamos que recorrer los 1000 registros para extraer aquellos en los cuales el Modelo = Sedan. En el arreglo tridimensional tenemos simplemente que sumar los contenidos de un Slice de 10×10 del arreglo

Agreguemos la dimensión Tiempo para indicar el mes del año en que se hizo la venta.

Visualizar una cuarta dimensión es más difícil que visualizar 3: se puede ver como 12 cubos de 3 dimensiones cada uno, uno para cada mes. Las celdas del cubo Enero contiene ventas efectuadas en Enero.



Beneficios

Cuando los datos son de naturaleza multidimensional, la performance alcanzada con una base de datos de este tipo es mejor que la alcanzada con una relacional. El tiempo de respuesta de un query en un entorno multidimensional es mucho más rápido que en uno relacional, particularmente para queries que generan vistas cross tab de datos.

Es importante destacar que cualquier acción de datos posible con una base de datos multidimensional siempre puede realizarse usando tecnología relacional.

Las bases de datos multidimensionales ofrecen varias ventajas:

√ *Facilidad de presentación y navegación de datos*

Las salidas de datos de las bases de datos multidimensionales son tipo spreadsheet. Obtener una salida del mismo tipo en el mundo relacional requiere escribir un query SQL complejo o usar un generador SQL para presentar los datos de una forma más intuitiva. Más aún algunos formatos de salida no pueden realizarse en SQL (por ejemplo reportes de ranking)

√ *Facilidad de mantenimiento*

Las bases de datos multidimensionales se caracterizan por la facilidad extrema de mantenimiento. Debido a que los datos están almacenados en la forma en que son vistos, no se requiere ninguna traducción adicional para convertir queries del usuario en requerimientos de datos. Las bases de datos relacionales utilizan índices y joins sofisticados para proveer este mismo nivel de intuitividad y ésto requiere mantenimiento y almacenamiento significativo.

√ *Performance*

Las bases de datos multidimensionales alcanzan niveles de performance difíciles de obtener en un entorno relacional. Para poder mejorar la performance en el relacional, habría que agregar índices y claves, pero no es posible optimizar la base de datos para todos los queries ad hoc que puedan presentarse.

Cuando la tecnología multidimensional no es apropiada

Cualquier conjunto de datos puede almacenarse en un arreglo multidimensional, sin embargo, por razones de performance y naturaleza de los datos, no siempre es conveniente este tipo de almacenamiento. Veamos el siguiente ejemplo de datos de empleados:

Apellido	Número de empleado	Edad
Pérez	01	21
García	12	19
Sánchez	31	63
Mendoza	14	31
Medero	54	27
Naya	03	56
Martínez	41	45
López	33	41
Fontanari	23	19

Podríamos representar estos datos en un arreglo multidimensional tomando Apellido y Número de Empleado como dos dimensiones, la edad estaría en las intersecciones de las posiciones de las dos dimensiones. Es fácil observar que no se repiten apellidos ni números de empleados. Así cada una de las dimensiones tiene 9 posiciones en el arreglo de 9 x 9, con 81 celdas, donde únicamente 9 tienen un valor.

Edad de empleados

A p e l l i d o s	Pérez				21					
	García								19	
	Sánchez	63								
	Mendoza					31				
	Medero						27			
	Naya							56		
	Martínez		45							
	López									41
	Fontanari			19						
		131/1	141/2	230/0	011/8	140/7	541/7	103/0	125/5	330/8
Números de Empleado										

Comparemos este ejemplo con el anterior:

Modelo	Color	Ventas
Sedan	Azul	4
Sedan	Rojo	3
Sedan	Blanco	2
4 x 4	Azul	3
4 x 4	Rojo	5
4 x 4	Blanco	5
Utilitario	Azul	6
Utilitario	Rojo	5
Utilitario	Blanco	4

Este modelo también con 9 registros tiene solamente 3 valores posibles para cada campo, cayendo solamente tres posiciones a lo largo de tres dimensiones para un arreglo de 3x3. Así un arreglo tridimensional tiene nueve celdas conteniendo cada una un valor.

Volúmen de ventas

M o d e l o	Sedan	4	3	2
	4 x 4	3	5	5
	Utilitario	6	5	4
		Azul	Rojo	Blanco
		Color		

El problema es que el ejemplo de la base de datos de empleados no es multidimensional ya que no hay interrelación inherente entre los elementos de los diferentes registros: ningún apellido corresponde a más de un número de empleado y ningún número de empleado corresponde a más de un apellido. Esto se representa en un arreglo multidimensional de 9 x 9 pobremente poblado. En oposición, en el ejemplo original, los datos son netamente multidimensionales, almacenados en un arreglo de 3x3 completamente poblado.

En la forma relacional se necesita un máximo de 9 búsquedas para localizar un registro, mientras que en la forma multidimensional se necesitan 18 (9 búsquedas en cada dimensión), con lo cual la performance disminuye. En contraste en el ejemplo de ventas, en la forma relacional se requiere 9 búsquedas, contra solamente 6 en la forma multidimensional.

Las ventajas de performance de almacenar datos multidimensionales en un arreglo multidimensional aumentan a medida que se incrementa el tamaño del conjunto de datos. De la misma manera, las desventajas de la performance de almacenar datos no multidimensionales en un arreglo multidimensional aumentan a medida que se incrementa el tamaño del conjunto de datos.

Teniendo en cuenta que las bases de datos multidimensionales son diseñadas para facilitar la manipulación y el análisis de estructuras de datos complejos muy interrelacionados, no presenta ninguna ventaja almacenar datos no - multidimensionales en una base de datos multidimensional.

En el ejemplo de los empleados es altamente improbable querer obtener la suma de las edades de los empleados o ver la tendencia de la edad de los empleados a medida que se incrementa el número de empleados.

Hay un número muy limitado de relaciones significativas entre los datos de los empleados, la información reside en el contenido individual de los elementos de datos, este ejemplo no es un buen caso para la tecnología de base de datos multidimensionales.

En contraste, el ejemplo de ventas tiene un gran número de interrelaciones significativas.

Frecuentemente el usuario querrá comparar ventas de un modelo frente a otro, total de ventas para cada color o ver la tendencia de las ventas a lo largo del tiempo para un modelo y color específicos. Las interrelaciones del conjunto de datos de ventas son más importantes que los elementos de datos en sí mismos. Estas son las características de un conjunto de datos que debería ser almacenado en una base de datos multidimensional.

Algunas aplicaciones donde la tecnología multidimensional es apropiada son:

- ✓ Análisis financiero*
- ✓ Análisis de presupuestos*
- ✓ Control de calidad*
- ✓ Análisis de utilidades por producto*

Características de las bases de datos multidimensionales

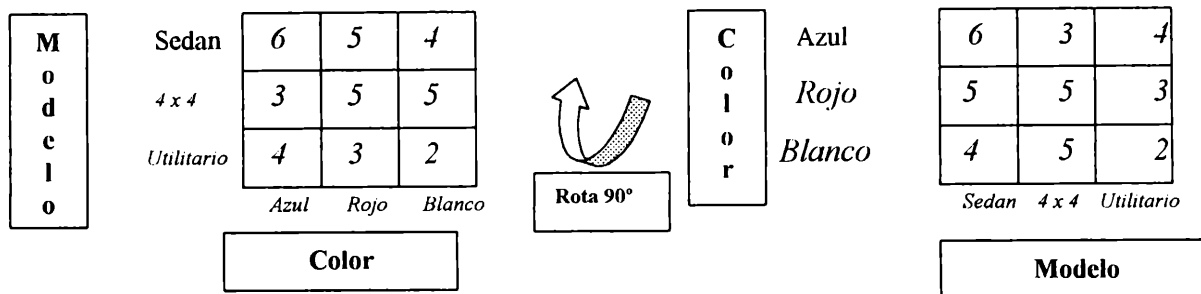
Rotación (data slicing)

El almacenamiento de los datos en un arreglo multidimensional tiene utilidades inherentes. La visualización del arreglo muestra explícitamente dimensiones, posiciones a lo largo de las dimensiones, intersecciones y valores. De hecho, la visualización de un arreglo multidimensional es similar a la de un spreadsheet, que es lo que generalmente requiere el usuario.

Las bases de datos multidimensionales permiten el browse directo de los arreglos tal como el usuario final haría un browse sobre un spreadsheet.

En el entorno relacional este tipo de visualización estructurada debe ser construido a través de un query complejo. Si después de tener una primer visualización el usuario quiere ver otra desde un punto de vista diferente, se debe construir un nuevo query complejo. En el entorno multidimensional esta nueva visión es llamada rotación, la visualización es rotada 90 grados sin la necesidad de reordenar los datos. La facilidad y velocidad con se hace la rotación es otro ejemplo de las ventajas inherentes de la manipulación de datos en un arreglo multidimensional.

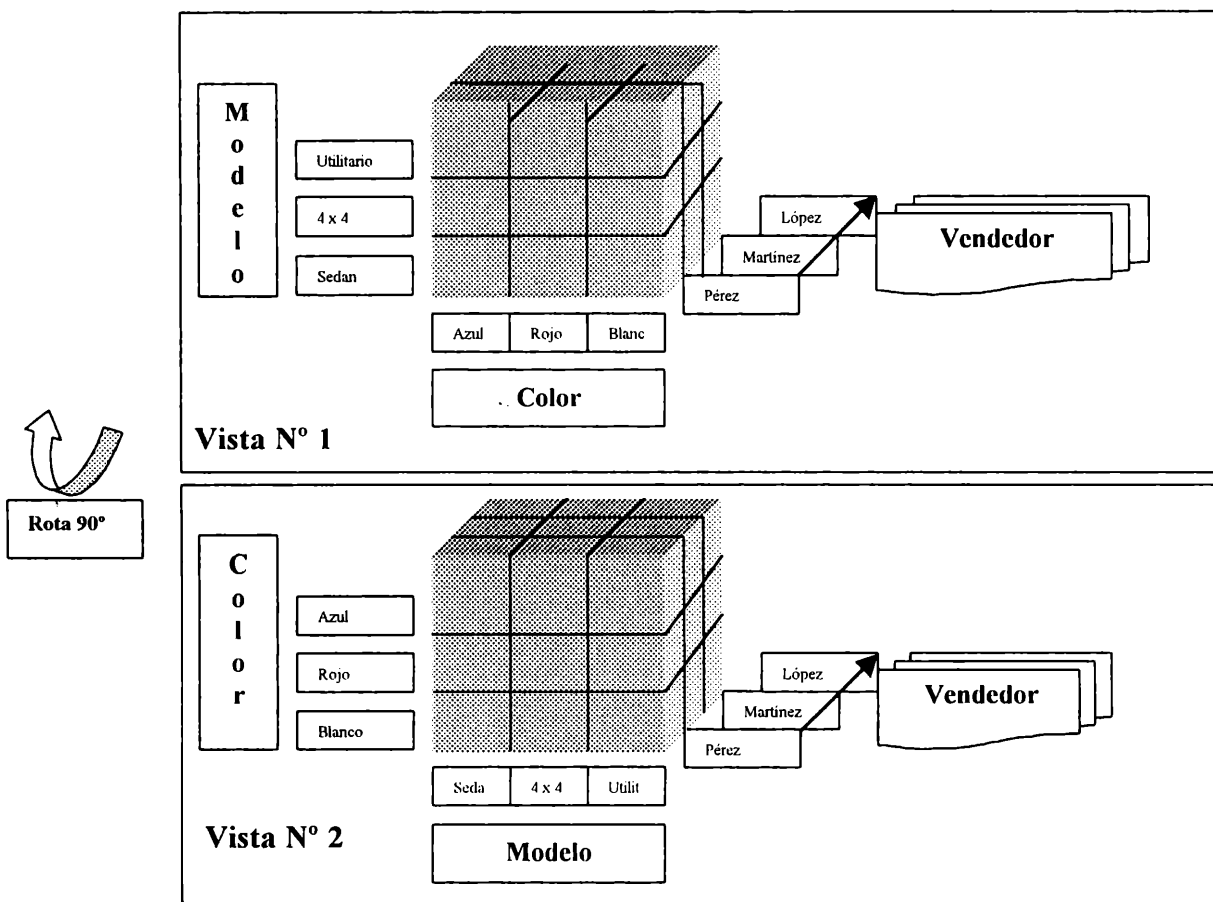
Volúmen de ventas

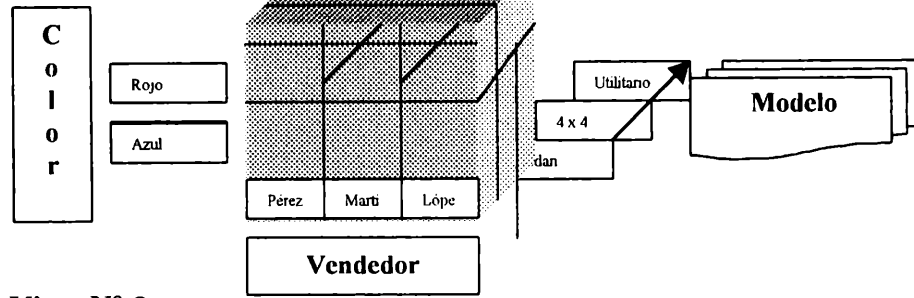


En el ejemplo hay 2 visualizaciones posibles Modelo por Color o Color por Modelo. En ambos casos en la intersección de las dos dimensiones aparecen los valores de ventas.

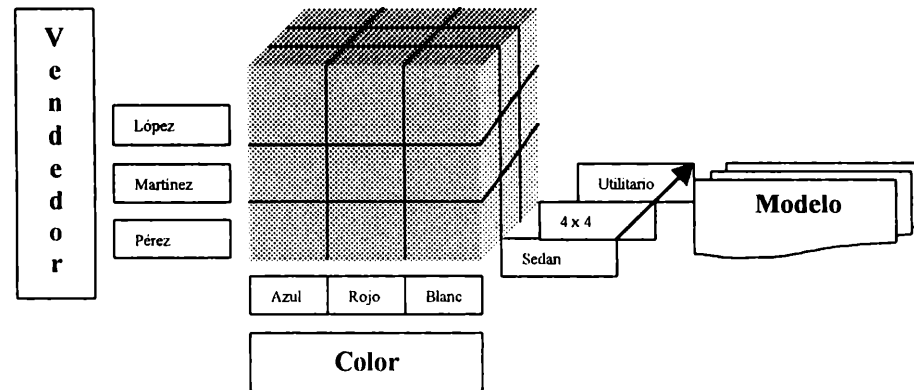
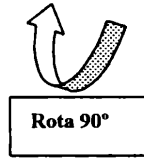
En el ejemplo anterior con tres dimensiones Modelo, Color y Vendedor se tienen seis vistas posibles:

- 1- Modelo x Color (con Vendedor en profundidad)
- 2- Color x Modelo (con Vendedor en profundidad)
- 3- Color x Vendedor (con Modelo en profundidad)
- 4- Vendedor x Color (con Modelo en profundidad)
- 5- Vendedor x Modelo (con Color en profundidad)
- 6- Modelo x Vendedor (con Color en profundidad)

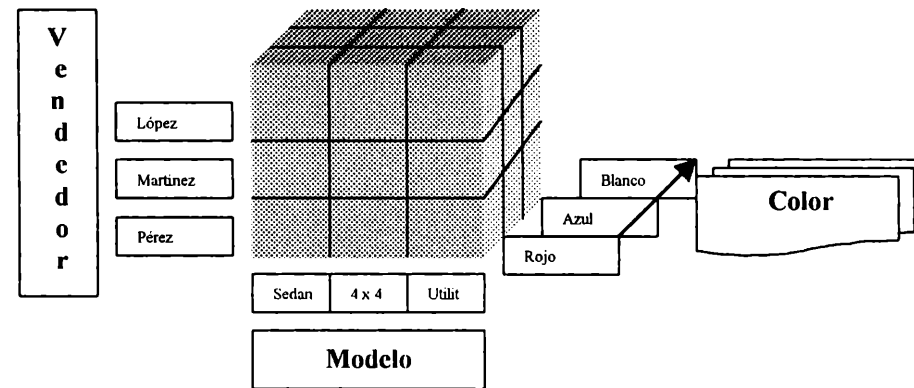
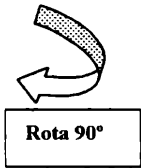




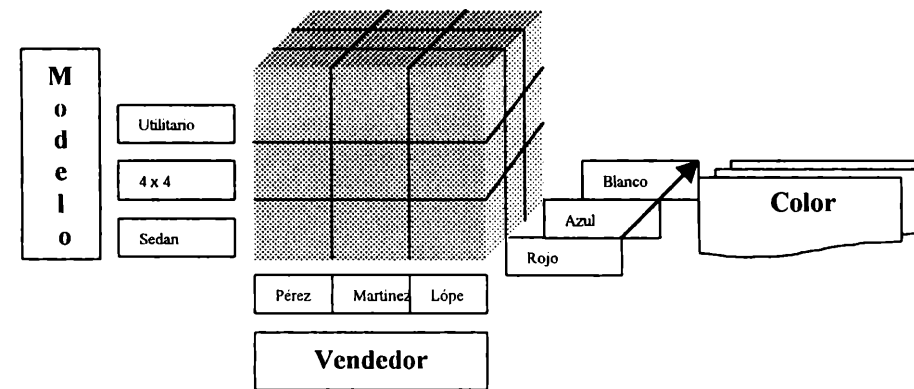
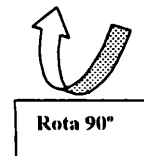
Vista N° 3



Vista N° 4



Vista N° 5



Vista N° 6

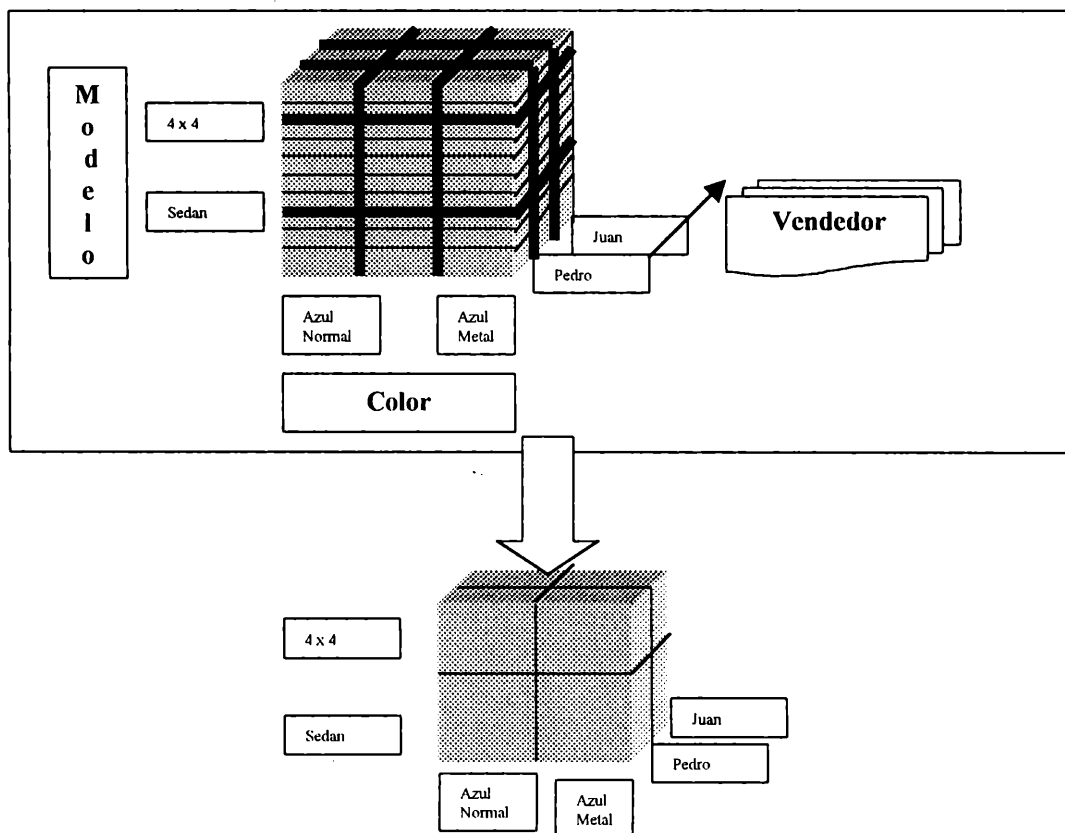
Cualquiera de las 6 vistas puede ser accedida a través de una rotación simple y sin necesidad de reordenamiento de datos. La rotación es también denominada “data slicing” ya que cada rotación muestra un “slice” diferente o tabla bidimensional de datos.

El número de posibles vistas (o slice) se incrementa exponencialmente con el número de dimensiones: un arreglo bidimensional tiene 2 vistas, uno tridimensional tiene 6 vistas, uno cuatridimensional tiene 24 vistas.

Ranging (data dicing)

Supongamos en nuestro ejemplo que el usuario quiere comparar los volúmenes de ventas de los nuevos colores Azul Metálico y Azul Francia. Se sabe que los únicos Modelos que se fabrican en estos colores son 4X4 y Sedan y por otro lado sólo dos Vendedores (Pedro y Juan) los han recibido. A través de la operación llamada “ranging” el usuario selecciona las posiciones deseadas en cada dimensión:

- ✓ En la dimensión Modelo: 4X4 y Sedan
- ✓ En la dimensión Color: Azul Metálico y Azul Francia
- ✓ En la dimensión Vendedor: Pedro y Juan



Esta selección determina un “arreglo reducido” que puede ser rotado y usado de la misma forma que el arreglo original (llamado padre). Realizar una operación similar al “ranging” en el entorno relacional involucra un query sumamente complejo. El “ranging” es también llamado “data dicing”. La performance de este tipo de operaciones en un arreglo multidimensional es mucho mejor que la del query análogo en el entorno relacional. Una vez más vemos que para obtener las vistas deseadas en el entorno multidimensional se requieren búsquedas con escaso consumo de recursos.

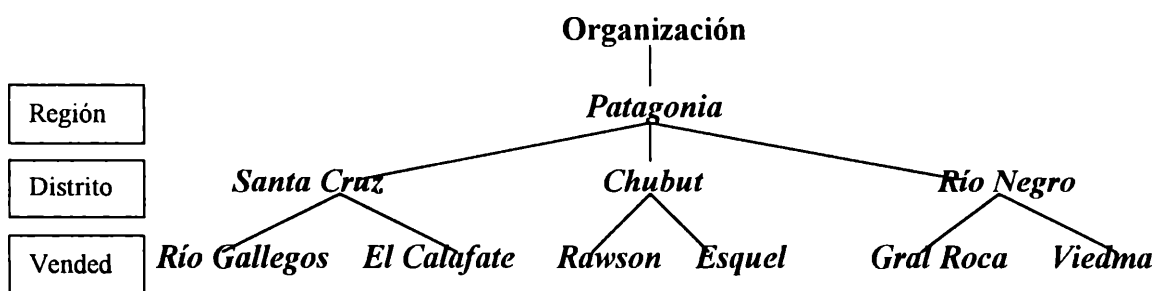
Jerarquías, Roll-Ups y Drill Downs

Consideremos, además, distritos de ventas, donde cada vendedor pertenece a un distrito. El usuario puede querer ver los volúmenes de venta a nivel distrito. Si los volúmenes en un distrito determinado no alcanzan las expectativas, el usuario se puede mover a nivel vendedor para determinar cuales vendedores no están alcanzando los objetivos previstos.

Hay una relación natural entre los volúmenes de ventas a nivel Vendedor y los volúmenes de ventas a nivel distrito. La suma de los volúmenes de ventas de todos los vendedores pertenecientes a un distrito determina el volumen de ventas del mismo.

La tecnología de bases de datos multidimensional está diseñada especialmente para facilitar estas relaciones naturales. Si bien es posible crear dimensiones independientes y separadas para Vendedor y Distrito, una solución más poderosa y eficiente es definir dos agregaciones relacionadas dentro de la misma dimensión. Una agregación es Vendedor y la otra es Distrito. Se percibe que estas dos agregaciones forman parte de una misma dimensión. En nuestro ejemplo todos los vendedores pertenecen a un distrito de ventas. Así, Distrito es una agregación de nivel superior dentro de una dimensión general que podríamos llamar Organización.

Las bases de datos multidimensionales permiten la creación de jerarquías de este tipo con múltiples niveles. Por ejemplo, el nivel Vendedor está relacionado al nivel Distrito, que a su vez puede estar relacionado al nivel Región.



Es una jerarquía de tres niveles Región, Distrito y Vendedor dentro de la dimensión Organización.

La alternativa en el entorno relacional es crear un campo para cada nivel en la jerarquía o bien crear tablas separadas que definan las relaciones. Esta aproximación es menos eficiente y

elegante que la provista por el entorno multidimensional, que maneja las relaciones jerárquicas entre diferentes agregaciones en forma natural.

La habilidad de definir jerarquías permite rapidez en la manipulación y análisis detallado de datos a lo largo de diferente niveles dentro de las dimensiones de un arreglo multidimensional. Moverse hacia arriba y hacia abajo en los distintos niveles de la jerarquía se denomina “roll-up” y “drill-down”.

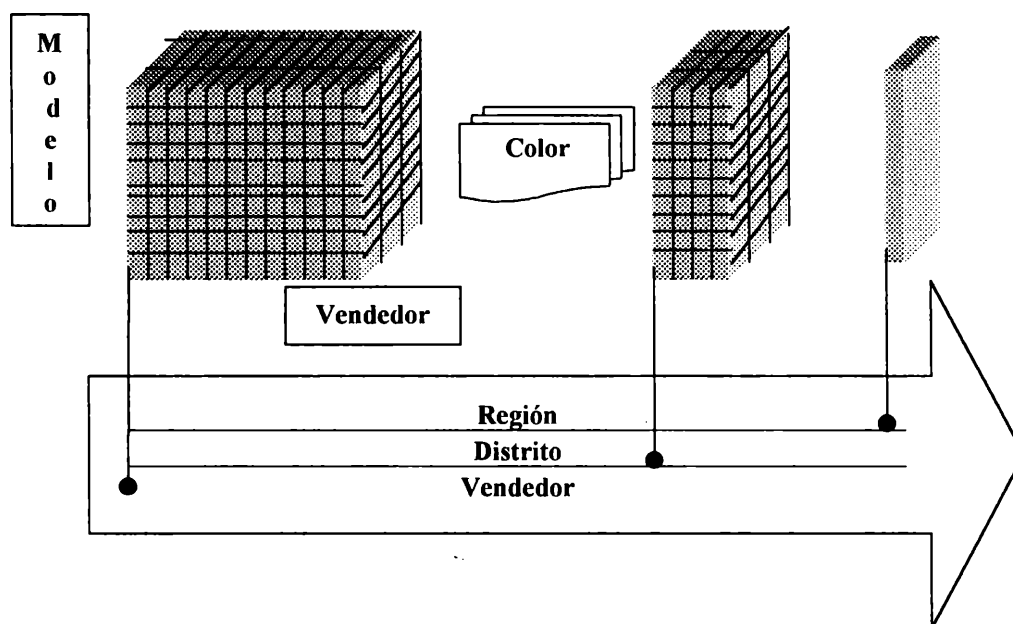
En el ejemplo anterior, el usuario puede hacer roll-up y drill-down dentro de la dimensión simple Organización. También debería poder hacer estas operaciones en varias dimensiones simultáneamente.

Así el usuario podría mantener constantes ciertas posiciones como por ejemplo manager (en Personal), distrito (en Organización) y producto (en Productos) y hacer roll-up y drill-down en las ventas sobre la dimensión Tiempo.

Una vista de este tipo puede ser rotada y “rangueada” como fue previamente descrito.

Los análogos a roll-up y drill-down en el ambiente relacional requieren sucesivas corridas de un query para cada nueva vista.

Nuevamente el entorno multidimensional presenta ventajas relativas a la performance.



Otro concepto importante es el de jerarquías simultáneas múltiples dentro de la misma dimensión. El usuario puede querer hacer roll-up dentro de la dimensión Organización en dos formas diferentes:

1. ¿ Cuáles son los volúmenes de ventas por Región ?
2. ¿ Cuáles son los volúmenes de ventas por Punto de Distribución ?

No todos los vendedores que pertenecen a una región pertenecen a un mismo punto de distribución. Por lo tanto tendremos dos jerarquías diferentes pero simultáneas para poder hacer roll-up desde Vendedor a Región y desde Vendedor a Punto de Distribución.

Jerarquías de ventas

<i>Organización de ventas</i>	<i>Organización de distribución</i>	<i>Modelo</i>	<i>Tipo de color</i>
<i>Región</i>	<i>Importador</i>	<i>Talle</i>	<i>Pintura base</i>
<i>Distrito</i>	<i>Distribuidor</i>	<i>Tipo</i>	<i>Color</i>
<i>Vendedor</i>	<i>Vendedor</i>	<i>Modelo</i>	

Queries

El alto grado de estructura en un arreglo multidimensional hace que el lenguaje de queries sea más simple y eficiente. No solamente el lenguaje es más intuitivo sino que la salida es directamente útil para el usuario final.

Volvamos al ejemplo de 3 dimensiones con 3 posiciones cada una Modelo, Color y Vendedor. Si quiere ver las ventas de cada Vendedor para un Modelo particular, es decir, la suma de las ventas por Color para cada Modelo dentro de cada Vendedor, el query será:

PRINT TOTAL (VENTAS KEEP MODELO VENDEDOR)

La salida del query es la siguiente tabla bidimensional:

M o d e l o	Vendedor		
	López	Martínez	Rodríguez
4 x 4	7	5	6
Utilitario	4	6	8
Sedan	3	8	12

La salida del query corresponde a la manera en la que los datos son almacenados en el arreglo. La salida es de valor inmediato, se visualizan fácilmente tendencias y comparaciones. Comparemos el query multidimensional con el query SQL contra los mismos datos en una base de datos relacional.

```
SELECT MODELO, VENDEDOR, SUM(VENTAS)
FROM VENTAS
GROUP BY MODELO, VENDEDOR
ORDER BY MODELO, VENDEDOR
```

Tienen los mismos datos y la misma consulta, pero el query en un ambiente relacional resulta más complejo y menos intuitivo. La salida de este query será:

<i>Modelo</i>	<i>Vendedor</i>	<i>Ventas</i>
<i>Sedan</i>	<i>Pérez</i>	<i>12</i>
<i>Sedan</i>	<i>López</i>	<i>3</i>
<i>Sedan</i>	<i>Martínez</i>	<i>8</i>
<i>4 x 4</i>	<i>Pérez</i>	<i>6</i>
<i>4 x 4</i>	<i>López</i>	<i>7</i>
<i>4 x 4</i>	<i>Martínez</i>	<i>5</i>
<i>Utilitario</i>	<i>López</i>	<i>4</i>
<i>Utilitario</i>	<i>Martínez</i>	<i>6</i>
<i>Utilitario</i>	<i>Pérez</i>	<i>8</i>

Obviamente los datos resultantes son los mismos, pero su formato es menos útil e intuitivo que el anterior.

Para obtener una salida similar a la del query multidimensional el usuario deberá escribir rutinas en algún lenguaje de programación con comandos SQL embebidos. En resumen la facilidad en el acceso a los datos y la utilidad directa de la salida, son aumentadas por el poder del lenguaje de queries multidimensional.

Vistas de datos multidimensionales

Una de las ventajas clave de las bases de datos multidimensional es permitir a múltiples usuarios el acceso y visualización de los mismos datos. Cada usuario puede realizar operaciones de rotación, rangueo, drill-down y roll-up a través de la base de datos sin producir impactos en las vistas de otros usuarios. La base de datos multidimensional debe también soportar múltiples usuarios, utilizando herramientas front-end para desarrollo de aplicaciones, browsing de datos y análisis, sin degradar la performance.

Cálculos multidimensionales

Los datos almacenados en un arreglo multidimensional ofrecen un alto grado de organización que facilita el análisis.

Por ejemplo, en nuestro arreglo bidimensional obtener el total de Ventas = Azul se reduce a sumar los datos de una fila. Esta misma operación en el modelo relacional requiere recorrer los registros uno a uno sumando aquellos para los que el Color sea igual a Azul.

Por supuesto las operaciones matemáticas generalmente son más sofisticadas, las bases de datos de datos multidimensionales están bien equipadas para el manejo de funciones matemáticas.

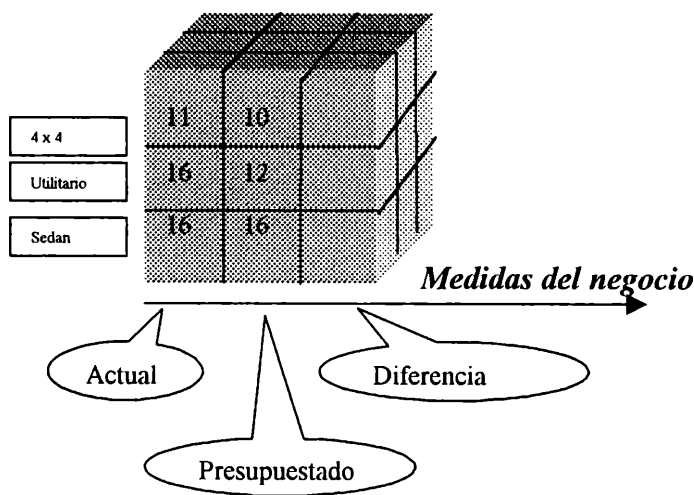
Las aplicaciones basadas en tecnología de base de datos multidimensional generalmente tienen una dimensión formada por los parámetros utilizados para medir la evolución del negocio. En esta dimensión vamos a encontrar parámetros como Volúmenes de Ventas, Presupuestos, Ganancias, Ingresos brutos, etc.

Tomemos un ejemplo tridimensional donde las primeras dos dimensiones son Modelo y Color (que conforman una matriz bidimensional), y una tercera dimensión con los parámetros de Medición del Negocio (actual, presupuestado, etc.). Es poco probable que un usuario esté interesado en realizar sumas a través de las filas o columnas dentro del arreglo bidimensional. Frecuentemente querrá realizar operaciones aritméticas a través de slices enteros, comparando por ejemplo la Varianza entre el arreglo bidimensional VENTAS ACTUALES y el arreglo bidimensional PRESUPUESTADO.

En nuestro arreglo multidimensional esta operación consiste en dividir el arreglo ACTUAL por el arreglo PRESUPUESTADO. En las bases de datos multidimensionales este tipo de operación se realiza eficientemente ya que los arreglos, en nuestro caso ACTUAL y PRESUPUESTADO, pueden ser tratados como celdas individuales, realizando las operaciones matemáticas muy rápidamente aún en el caso de estar trabajando con arreglos muy grandes. Esto puede realizarse aunque los arreglos sean de tamaño y forma diferentes.

Las bases de datos multidimensionales poderosas pueden conformar un arreglo en otro ignorando inteligentemente las celdas donde no hay datos. En el entorno relacional operaciones tales como la de realizar una división entre dos tablas (ACTUAL y PRESUPUESTADO) requiere utilizar una gran cantidad de recursos ya que implica realizar un join entre dos grandes tablas o realizar un join individual para cada celda o registro. En grandes conjuntos de datos, realizar este query tomaría horas.

Volúmenes de ventas



Concluimos que una base de datos multidimensional además de proveer un acceso a los datos más eficiente, provee herramientas de cálculo muy relacionadas a la estructura de la base de datos.

Dimensión Tiempo

Generalmente, muchos de los análisis requeridos para la toma de decisiones involucran análisis de datos a través del tiempo. El mundo relacional no provee herramientas para un manejo efectivo del tiempo. Muchas bases de datos multidimensionales tienen una dimensión Tiempo

predefinida, con jerarquías de rolling- up y drilling-down a través de días, semanas, meses, años, año fiscal, etc.

Este tratamiento de la dimensión Tiempo como una dimensión predefinida tiene dos ventajas significativas:

- ✓ *Elimina el esfuerzo requerido para construir jerarquías sofisticadas cada vez que la base de datos es instalada.*
- ✓ *Debido a que la dimensión tiempo esta codificada en el hard, tenemos ventajas extras de performance para aquellas aplicaciones que realizan análisis a través del tiempo como demoras, seguimiento de proyectos, etc.*

Datos esparcidos

Los arreglos esparcidos son conjuntos de datos que tienen gran cantidad de celdas vacías. El ejemplo de personal almacenado en un arreglo multidimensional es un ejemplo extremo de lo que denominamos "arreglos esparcidos". Hemos visto otros ejemplos en los que mostramos conjuntos de datos conteniendo valores en todas sus celdas, éste es el otro extremo que llamamos arreglos completamente poblados. En la práctica la mayoría de los conjuntos de datos multidimensionales se ubican en alguna posición entre estos dos extremos. Por ejemplo un vendedor puede estar autorizado a vender autos Sedan azules y rojos en todos los distritos pero blancos sólo en dos de ellos. En una tabla relacional sólo habrá registros para Sedan blanco en esos dos distritos; en una estructura multidimensional existirá una celda para cada distrito aún en el caso en que no le corresponda un valor de ventas. La celda estará vacía conteniendo un valor nulo necesitando los mismos recursos de procesamiento y almacenamiento que una celda con valor.

En grandes arreglos con muchas dimensiones, aunque el conjunto de datos sea eminentemente multidimensional, se tendrán muchas celdas con valores nulos.

En algunos casos, el número de celdas con valores nulos excede el 90%, ésto es un arreglo esparcido. Por razones de optimización de performance y minimización de espacio de almacenamiento, es importante tratar de forma especial este tipo de arreglos. El motor de base de datos detecta bloques de valores nulos, los remueve y comprime el arreglo multidimensional. Este tratamiento especial de los arreglos esparcidos incrementa fuertemente la performance a la vez que reduce los requerimientos de almacenamiento.

Integración de datos

La noción de separar las aplicaciones OLTP de las aplicaciones OLAP cada vez está cobrando más fuerza.

La carga de trabajo en OLAP es discontinua, presenta ocasionalmente periodos de intensa actividad cuando los usuarios lanzan queries complejos y luego el sistema permanece ocioso mientras los datos son analizados y manipulados en la workstation. (figura a)

En contraste, la carga de trabajo en aplicaciones OLTP es más constante (figura b).

Si se mezclan las cargas de trabajo, las aplicaciones OLTP se sobrecargan con los picos momentáneos generados por las aplicaciones OLAP.

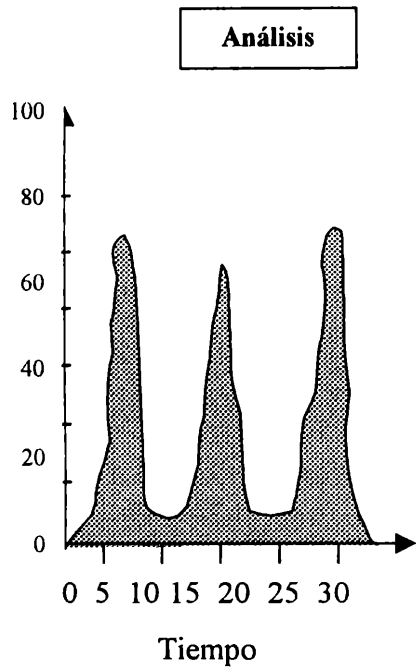


Figura A

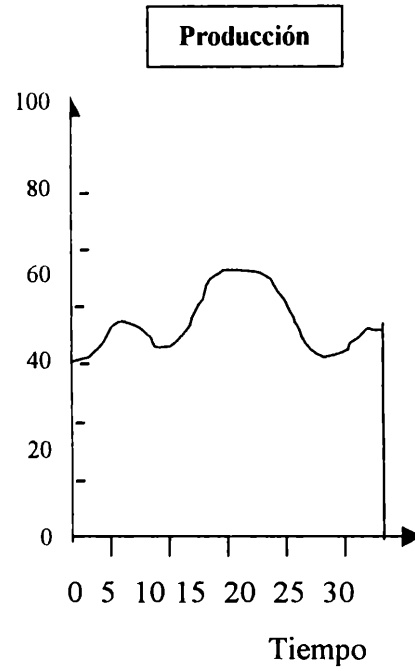


Figura B

**Análisis
Producción**

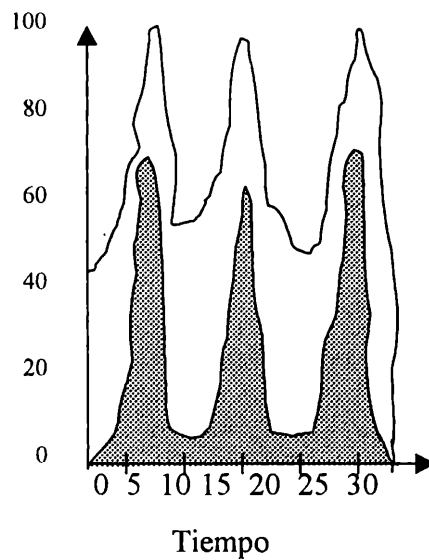


Figura C

Es deseable separar las aplicaciones operacionales de las de análisis. Una base de datos multidimensional bien diseñada soporta esta separación proveyendo herramientas para automatizar la extracción e integración de datos y manteniendo un ambiente separado para análisis de datos y reportes.

Estructuras relacional, multidimensional y spreadsheet

La tecnología de base de datos multidimensional tiene el objetivo de optimizar la performance en el acceso, manipulación y análisis de datos. Esta tecnología no está diseñada para reemplazar el modelo relacional ni otras tecnologías de bases de datos empleadas para aplicaciones transaccionales.

La tecnología multidimensional se puede ver como un complemento de la relacional optimizando el análisis, manipulación e integración de datos.

Históricamente las bases de datos relacionales han sido usadas para aplicaciones que registran y almacenan grandes volúmenes de transacciones. Sin embargo, no han sido especialmente diseñadas para el análisis de grandes volúmenes de datos. Para este tipo de aplicaciones, la base de datos multidimensional es más apropiada.

Las estructuras de datos multidimensionales intentan complementar los sistemas transaccionales que incluyen tecnología relacional proveyendo:

- ✓ *una base para la integración de los datos extraídos de los sistemas de almacenamiento ya existentes en la compañía*
- ✓ *una estructura de datos más natural para el usuario final*
- ✓ *un entorno optimizado para aplicaciones de análisis intensivo*

Las spreadsheets son otro ejemplo de estructuras de datos utilizadas hoy en día.

Proveen poderosas capacidades para el análisis de datos pero son limitadas en cuanto a la modelización de datos. Además, carecen de habilidad para visualizar los datos desde diferentes perspectivas. Hay en el mercado algunas herramientas de spreadsheet multidimensionales que dirigen el acceso del usuario final al modelo de datos multidimensional. Pero presentan limitaciones en:

- ✓ *extracción e integración de datos desde fuentes dispersas en la compañía*
- ✓ *analizar grandes volúmenes de datos*
- ✓ *proveer capacidades de rotación dinámica de datos, drill down y roll up*
- ✓ *Soportar múltiples usuarios concurrentes y vistas de los mismos modelos de datos.*

Características de las distintas tecnologías

Relacional

- ✓ *Standard establecido*
- ✓ *Procesamiento de transacciones eficiente*

Multidimensional

- ✓ *Soporte consultas del usuario final*
- ✓ *Eficiencia en el análisis comparativo*
- ✓ *Manejo de grandes volúmenes de datos*
- ✓ *Facilidad en el mantenimiento*

Spreadsheets Multidimensionales

- ✓ *Extiende las capacidades del spreadsheet a manipulación de rotaciones*
- ✓ *Es familiar a los usuarios finales*
- ✓ *Aplicación de escritorio estándar*
- ✓ *Facilidades de soporte*

OFA

OFA

Es una aplicación de software que se utiliza para:

- ✓ *Realizar análisis y reporting de los datos financieros de la compañía, crear presupuestos financieros, forecast y planificaciones.*
- ✓ *Crear y ejecutar modelos financieros.*
- ✓ *Configurar un sistema financiero eficiente para la empresa.*

Elementos de Ofa

Para realizar las funciones anteriores, usa tres tipos de elementos: reportes, gráficos y worksheets. Los reportes permiten ver los datos financieros desde diferentes perspectivas y preparar presentaciones basadas en los análisis realizados. Los gráficos se usan para representar visualmente los datos financieros. Las worksheets se usan para ingresar y manipular datos financieros.

OFA corre sobre una red de computadoras usando una LAN (local area network) o sobre una combinación de computadoras personales y servers en red.

Módulos de software conocidos como workstations se instalan en servers y computadoras personales y pueden acceder a bases de datos comunes llamadas "bases de datos shared".

Tipos de workstations

OFA incluye dos tipos básicos de workstations:

- ✓ *Administradoras*
 1. *Workstation super administradora*
 2. *Workstation administradora*
 3. *Workstation Task Processor*
- ✓ *Users*
 1. *Workstation budget*
 2. *Workstation analyst*

Tipos de bases de datos

- ✓ *Personales*

Se usan para almacenar objetos y en algunos tipos de workstations datos financieros.

Estos objetos pueden ser: dimensiones, items de datos financieros, atributos y jerarquías y también reportes, worksheets y gráficos.

✓ *Shared*

Contienen datos financieros y objetos que son compartidos por todos los usuarios. Son mantenidas por los administradores.

Diseño de arquitectura “por filas”

Usando la workstation super administradora como punto de partida, se puede construir una arquitectura por filas creando workstations administradoras subordinadas que, a su vez, pueden ser usadas para crear otras workstations administradoras subordinadas.

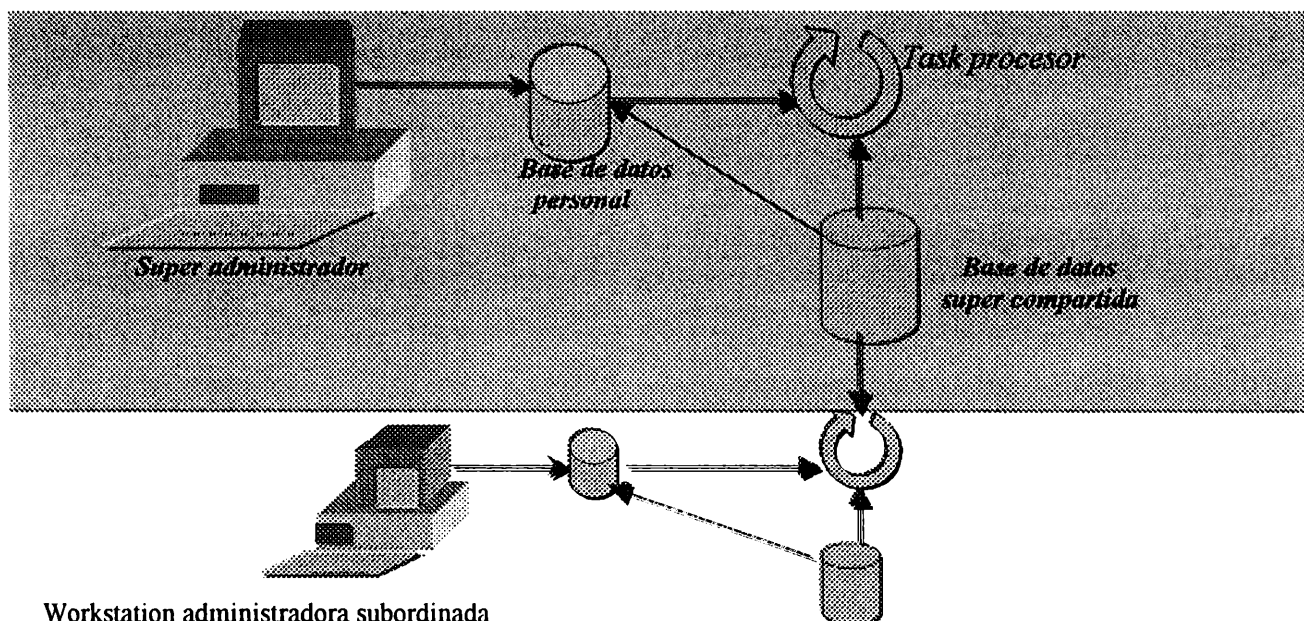
Esta aproximación por filas reduce el tamaño y la complejidad de las estructuras top level y permite ganar autonomía y control en los niveles más bajos y sitios remotos.

Workstation super administradora

Es una herramienta de acceso de datos y manejo de sistemas que permite realizar todas las tareas que otras workstations administradoras pueden desarrollar. Es única y está ubicada en el tope de la jerarquía.

Es la única workstation que permite setear el tiempo fiscal para todo el sistema.

La workstation super administradora está asociada a una base de datos Super compartida, a una base de datos personal y a un Task Processor.



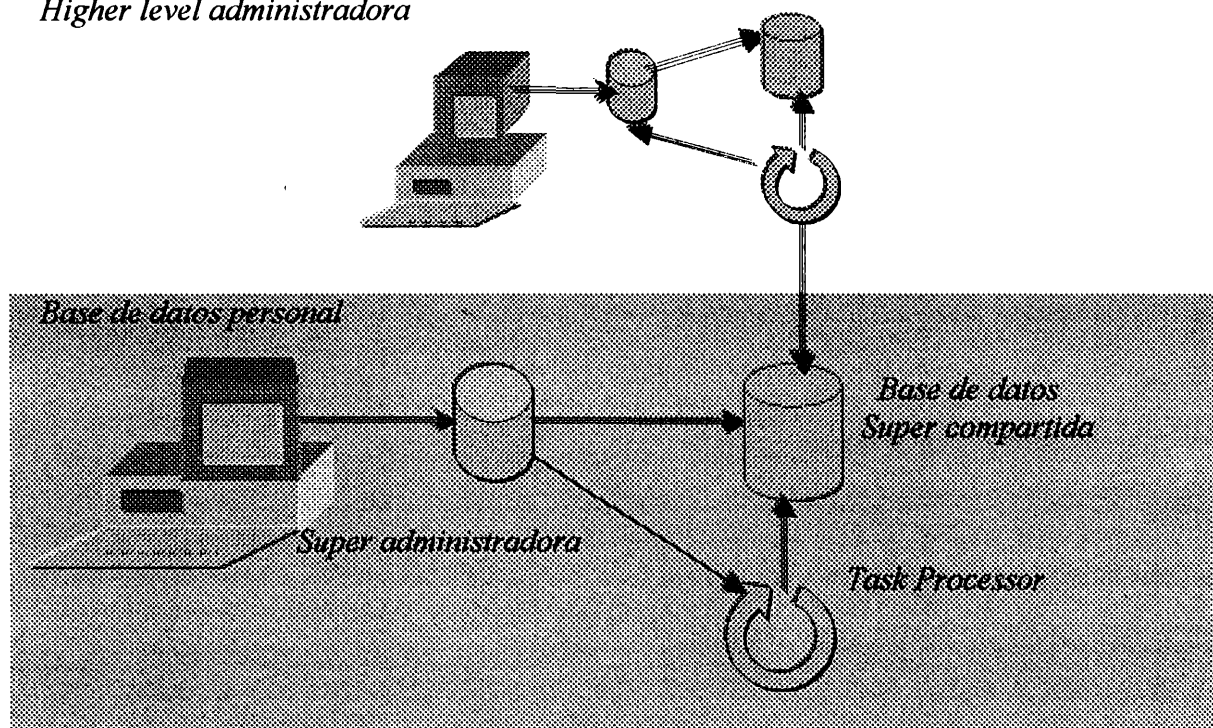
Workstation administradora

Es una herramienta de acceso de datos y manejo de sistemas que permite realizar las siguientes tareas:

- ✓ Definir y mantener usuarios de workstations budget, analyst y otras administradoras subordinadas.
- ✓ Especificar la configuración default para workstations budget, analyst y otras administradoras subordinadas, cuando son creadas.
- ✓ Restringir los cambios a datos compartidos controlando el acceso de los usuarios a los datos financieros en la base de datos compartida.
- ✓ Definir estructuras de bases de datos y distribuirlas a la base de datos compartida, workstations administradoras subordinadas, workstations budget y workstations analyst. Estas estructuras son dimensiones, valores de dimensión, atributos, jerarquías, modelos, reportes, gráficos, worksheets y folders.
- ✓ Distribuir slices de datos financieros a los usuarios y la base de datos compartida.
- ✓ Refrescar la base de datos compartida propia con los datos enviados por el administrador.
- ✓ Submitir datos financieros desde la base de datos compartida propia a la base de datos compartida del administrador.

En la estructura por filas, pueden existir múltiples workstations administradoras en distintos niveles. Cada workstation administradora tiene asociada una base de datos compartida propia y un Task Processor.

Higher level administradora



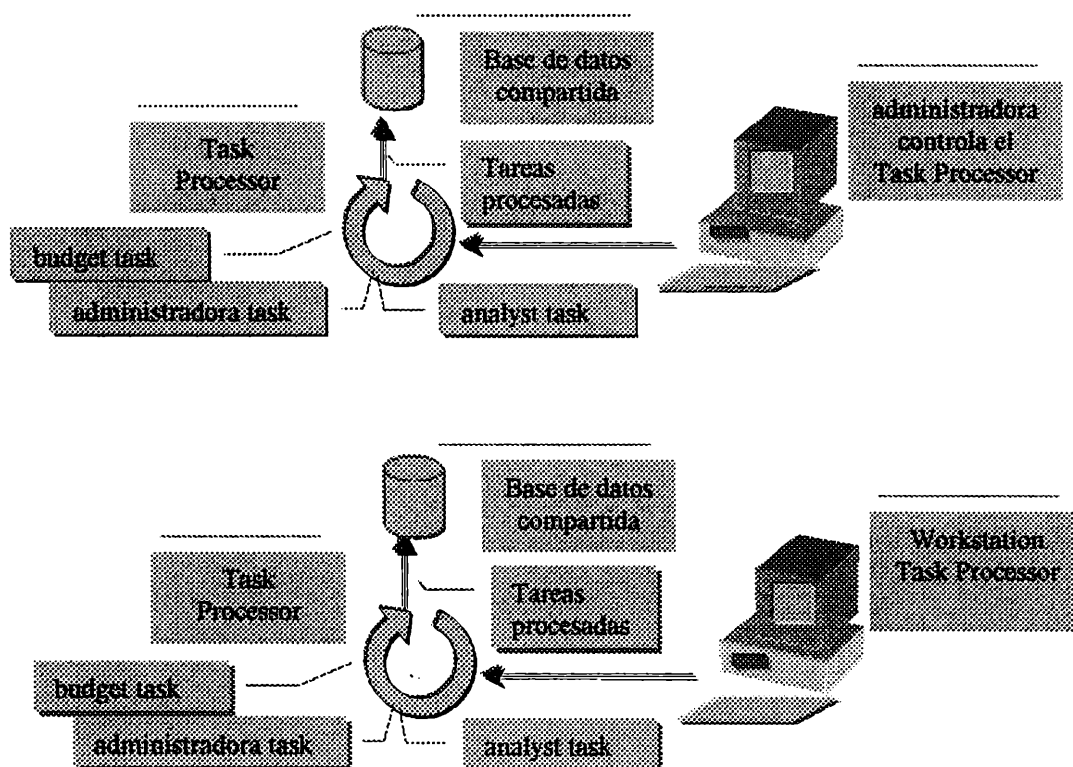
Workstation Task Processor

Es una herramienta de manejo de colas con las siguientes características:

- ✓ En un sistema OFA está asociada cada workstation administradora.
- ✓ Permite a los administradores monitorear y controlar la transferencia de datos entre workstations de usuarios y sus bases de datos compartidas asociadas.
- ✓ Permite a los administradores distribuir estructuras a las workstations subordinadas.

El Task Processor puede instalarse en la misma PC que la workstation administradora o en otra separada dedicada solamente al task processing (stand alone).

En el primer caso, el control del Task Processor lo ejerce el administradora; en el segundo, controla la workstation Task Processor, con la ventaja de que las tareas pueden procesarse sin interrupción.



Workstation Budget

Es una herramienta de acceso de datos que permite realizar las siguientes tareas:

- ✓ Define y mantiene los objetos de la base de datos personal: dimensiones, atributos, jerarquías, modelos, solve definitions .
- ✓ Crear y mantener reportes, gráficos, worksheets y folders personales.
- ✓ Usar objetos de base de datos y documentos definidos por el administrador.

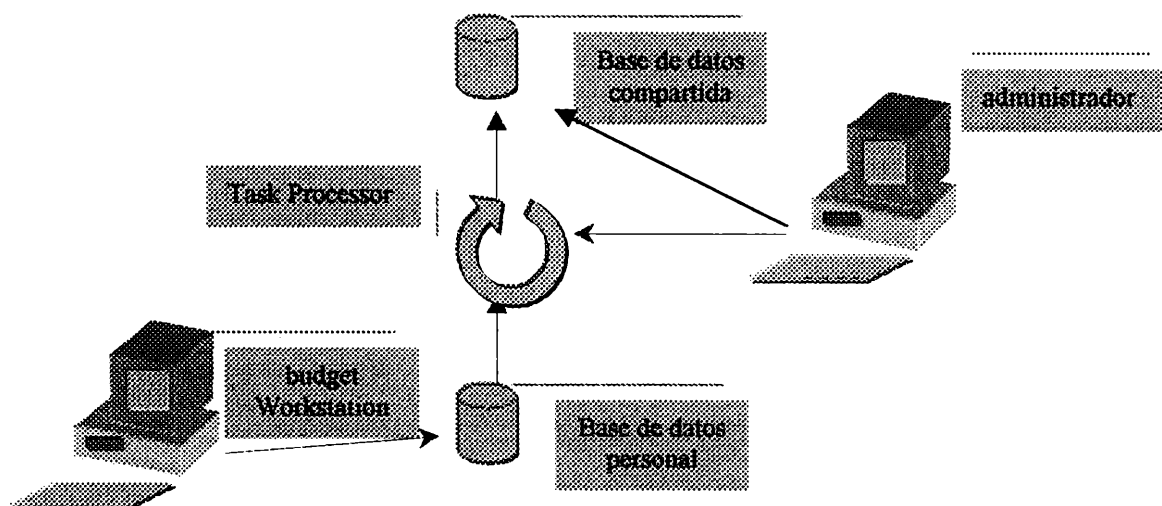
- ✓ *Submitir datos financieros desde la base de datos personal a la base de datos compartida del administrador.*
- ✓ *Refrescar la base de datos personal con los datos de la base de datos compartida de la workstation administradora.*

En la estructura por filas, la workstation budget está ubicada debajo de la workstation administradora que la creó y tiene asociada una base de datos personal que contiene estructuras de bases de datos, seteos y datos financieros.

Pueden existir múltiples workstations budget debajo de la misma workstation administradora.

En una estructura por filas, pueden existir múltiples workstations budget en distintos niveles de la configuración.

La siguiente ilustración muestra como la workstation budget se comunica con la base de datos compartida a través del Task Processor.



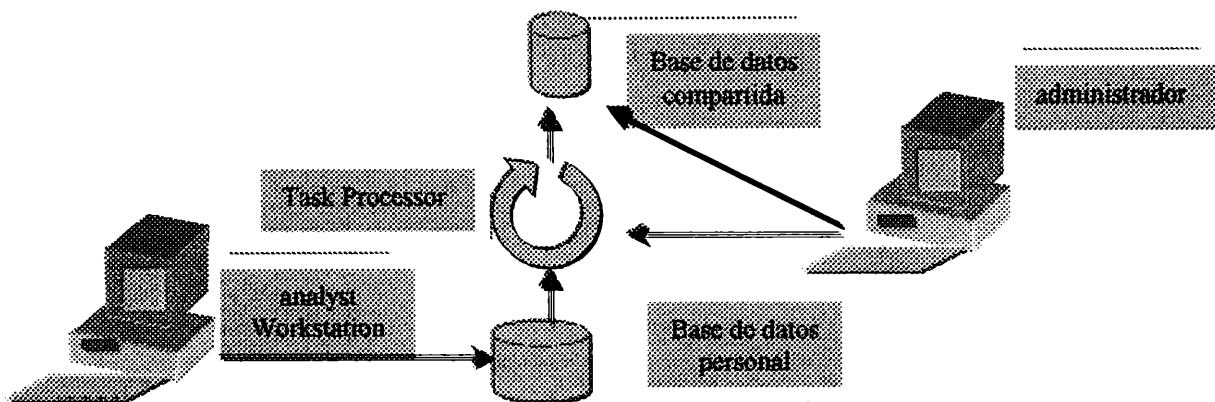
Workstation analyst

Es una herramienta de acceso de datos que permite realizar las siguientes tareas:

- ✓ *Refrescar las estructuras distribuidas por la workstation administradora.*
- ✓ *Definir y usar reportes, gráficos, worksheets y folders personales o usar documentos definidos por el administrador.*
- ✓ *Acceder y manipular datos asociados con dimensiones, valores de dimensión, tiempo, atributos, jerarquías y modelos en la base de datos compartida.*

En una estructura por filas, una workstation analyst está ubicada debajo de la workstation administradora que la creó y está asociada con una base de datos personal que contiene estructuras de base de datos y seteos pero no una copia de los datos financieros.

La siguiente ilustración muestra como la workstation analyst se comunica con la base de datos compartida a través del Task Processor.



Para ver los datos, la workstation analyst se comunica directamente con la base de datos compartida, sin pasar por el Task Processor.

Ejemplificación de los conceptos utilizados por OFA

Consideraremos una compañía llamada XX Informática que se dedica a la fabricación de computadoras y equipos relacionados para ilustrar los conceptos propios de OFA:

La compañía está organizada de la siguiente forma:

- ✓ Existen 7 líneas de productos.
- ✓ Los productos están divididos en 14 categorías distintas.
- ✓ El sector de Ventas de la compañía está dividido en 4 divisiones:
 - Gerencia de Ventas
 - Ventas Comerciales
 - Ventas al Estado
 - Integración y distribución de Ventas.

La división más importante en términos de beneficios es la de Ventas Comerciales.

La compañía mantiene un año calendario y desarrolla análisis de varianzas mensuales comparando valores reales y presupuestados.

En una corporación típica, hay varias formas diferentes de configurar y utilizar OFA. Veremos diferentes divisiones de la compañía y cómo cada una de ellas usa esta herramienta.

✓ División Consultoría

En esta división existe un gerente de Consultoría y un subgerente por cada región que reporta al mismo. Los subgerentes regionales de Consultoría usan workstation budget que reportan en la workstation super administradora del gerente de la división.

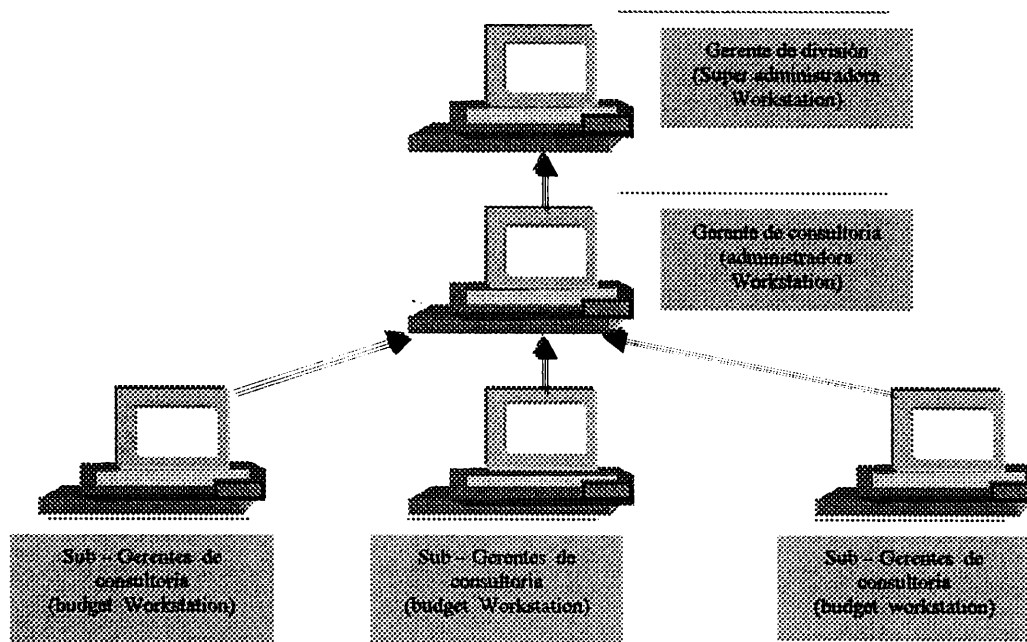
Los subgerentes de cada región pueden:

- 1. Editar los gastos e ingresos de su región*
- 2. Ver los gastos e ingresos de los subgerentes de otras regiones*
- 3. Ingresar sus valores presupuestados y proyectados (forecast)*
- 4. Desarrollar análisis What If para evaluar la performance potencial en su región*
- 5. Aumentar sus bases de datos personales, agregando datos personales y financieros a aquellos recibidos de la workstation administradora*

Una vez que un subgerente finaliza sus budgets y forecasts, submite sus números finales al gerente de Consultoría, quien los revisa y pide modificaciones si es necesario. También puede editar los datos compartidos que se convierten en el budget final de la división.

Luego, las workstations budget de menor nivel (las de los subgerentes regionales) refrescan sus bases de datos personales para obtener los valores finales.

El gerente de la división submite un resumen final de sus valores a la workstation super administradora corrida por el gerente financiero.



✓ División Producción

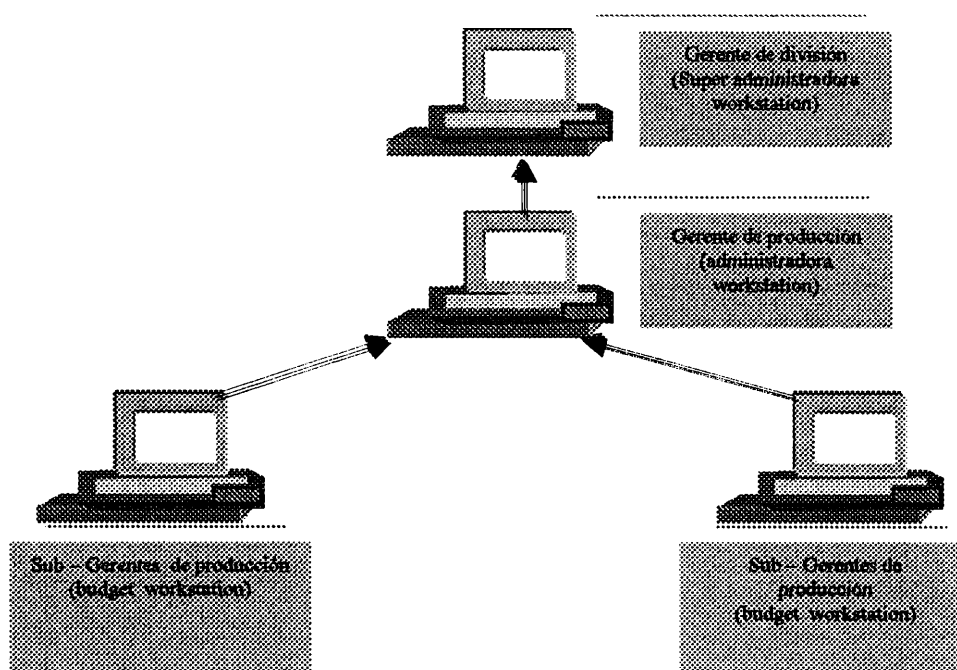
La gerencia de Producción tiene una organización similar a la de la gerencia de Consultoría, pero lo que cada subgerente ve y edita está determinado por la responsabilidad que tiene sobre el producto y no por las regiones geográficas.

Los subgerentes de Producción utilizan workstations budget para presupuestar, proyectar y realizar análisis de utilidades del producto, beneficios del cliente, análisis de tendencia, análisis what if y forecast de producción. Submiten presupuestos y proyecciones al gerente de Producción o directamente al Analista Financiero.

Muchos de ellos están provistos de workstations analyst para mantenerse actualizados en los datos de performance de la compañía y soportar la toma de decisiones en las líneas de producción.

Los usuarios de las workstations analyst pueden ver datos históricos, budgets y los datos de forecast hasta la fecha para que los budgets y forecast puedan ser planificados con más exactitud. Su visión de los datos está también muy orientada al producto.

Usan datos corporativos concluidos para la toma de decisiones actuales y planificar, por lo tanto la funcionalidad de una workstation analyst se adapta a sus necesidades.



✓ División Finanzas

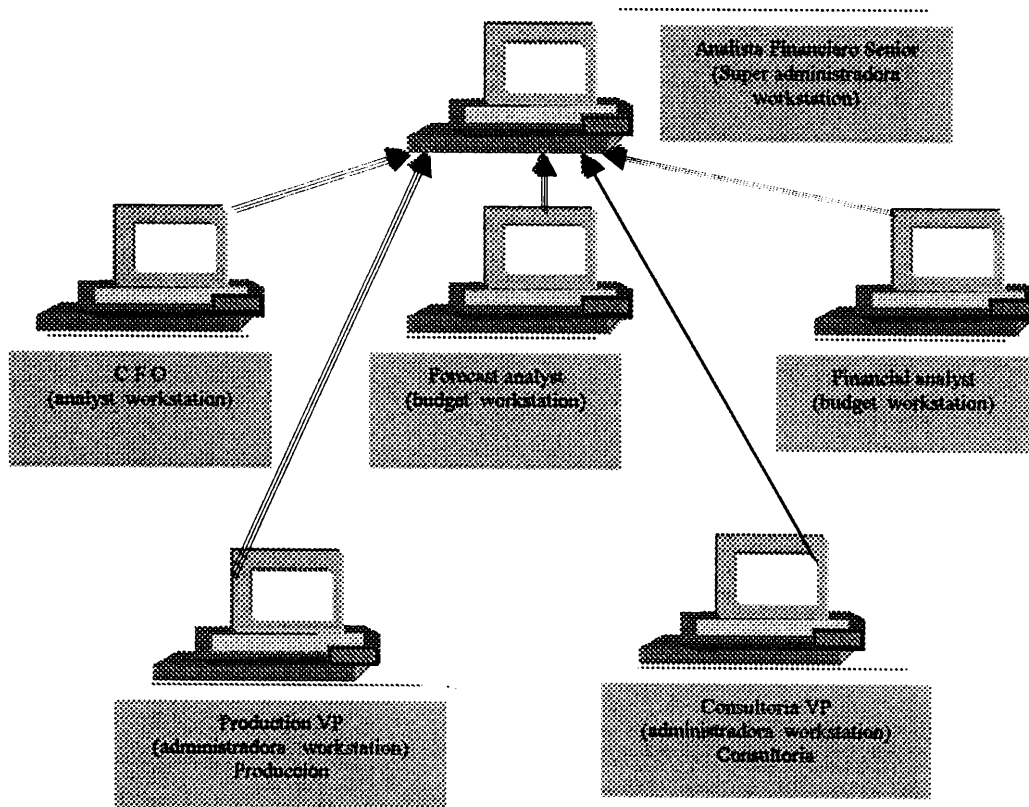
Todas las workstations de las divisiones de Producción y Consultoría reportan directamente al super administrador en la División Finanzas.

Otros usuarios de esta división usan *workstations budget* y *analyst* que también reportan al super administrador.

Puede haber uno o varios analistas financieros senior que son los responsables de recolectar los datos de *budgets* y *forecasts* de las distintas divisiones de la organización. Si hubiera varios analistas, actuarían como *workstation administradora* que reportan a la *workstation Super administradora*.

Los analistas requieren capacidades de análisis y modelización para *rever* y posiblemente *modificar* la información recibida desde otras áreas de la organización.

El Super administrador reside en la División Finanzas y el número de *workstations Sub administradora* que ellos soportan depende de si los datos de otras áreas son *submitidos* a una o varias *workstations* dentro de esta división.



Comparación entre los distintos tipos de workstations

- ✓ Tareas de manejo de sistema

Una workstation super administradora, puede:



1. *Configurar workstations*
2. *Controlar el acceso de los usuarios a los objetos de base de datos*
3. *Definir y mantener usuarios*
4. *Monitorear y controlar transferencia de datos*
5. *Setear el año fiscal*
6. *Especificar configuraciones default para workstations budgets y analysts*

Una workstation administradora, puede:

1. *Configurar workstations excepto de tipo super administrador*
2. *Controlar el acceso de los usuarios a los objetos de base de datos*
3. *Definir y mantener usuarios, excepto de tipo super administrador*
4. *Monitorear y controlar transferencia de datos*
5. *Especificar configuraciones default para workstations budgets y analysts*

El Task Processor, puede:

1. *Monitorear y controlar transferencia de datos*

Las workstations budgets y analysts no pueden realizar tareas de este tipo.

✓ *Tareas de manejo de datos*

Una workstation super administradora, puede:

1. *Crear estructuras de datos*
2. *Crear y mantener documentos personales*
3. *Distribuir datos financieros*
4. *Distribuir estructuras*
5. *Mantener datos personales*
6. *Submitir datos a la base de datos compartida a través del Task Processor*
7. *Ver el log de tareas*
8. *Usar la facilidad de copia de datos*

Una workstation administradora, puede:

1. *Crear estructuras de datos*
2. *Crear y mantener documentos personales*
3. *Distribuir datos financieros*
4. *Distribuir estructuras a todas excepto al Super administrador*
5. *Mantener datos personales*
6. *Submitir datos a la base de datos compartida a través del Task Processor*
7. *Ver el log de tareas*
8. *Refrescar los datos recibidos desde administradores de nivel superior*
9. *Usar la facilidad de copia de datos*

El Task Processor, puede:

1. *Procesar tareas en la base de datos compartida*
2. *Ver el log de tareas*

Una workstation budget, pueden:

1. *Crear estructuras de datos personales únicamente*
2. *Crear y mantener documentos personales*
3. *Mantener datos personales*
4. *Submitir datos a la base de datos compartida a través del Task Processor*
5. *Ver el log de tareas*
6. *Refrescar los datos recibidos desde administradores de nivel superior*
7. *Usar la facilidad de copia de datos*

Una workstation analyst, pueden:

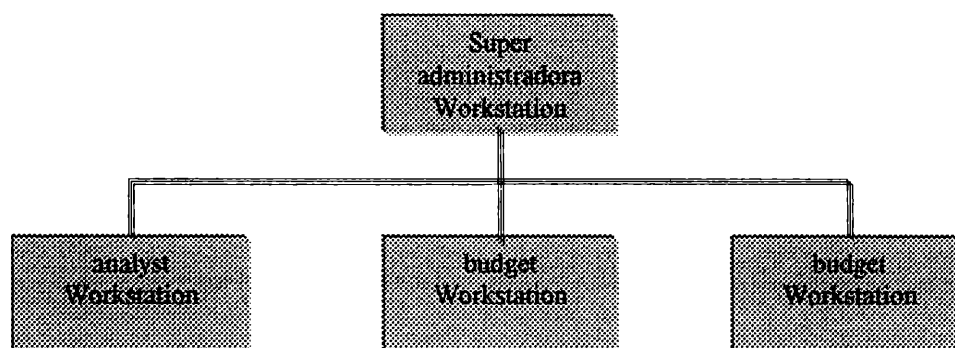
1. *Crear y mantener documentos personales*
2. *Submitir datos a la base de datos compartida a través del Task Processor*
3. *Ver el log de tareas*
4. *Refrescar los datos recibidos desde administradores de nivel superior*

Configuraciones de OFA típicas

- ✓ *Configuraciones simples*

La configuración más simple está compuesta por:

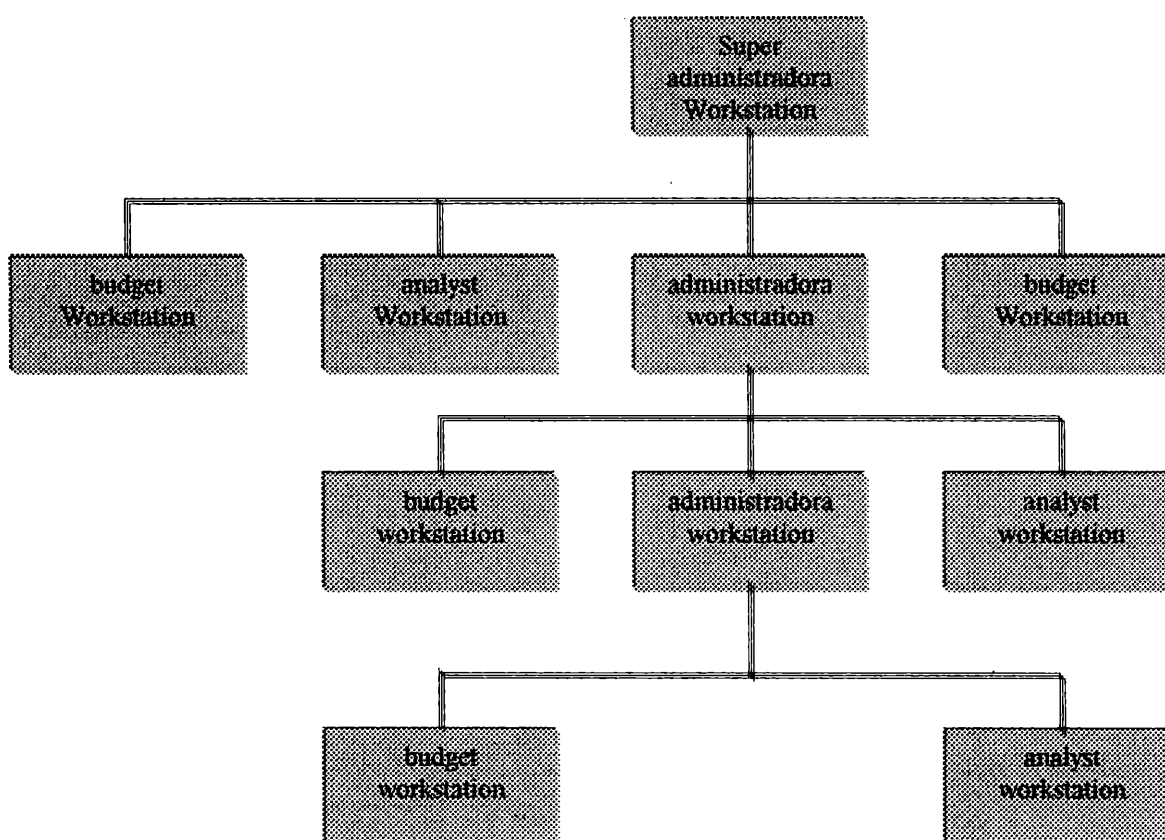
1. *Una única workstation super administradora que incluye una base de datos super compartida y un Task Processor.*
2. *Una o más workstations usuarios.*



✓ *Configuraciones complejas*

Una configuración más compleja está dada por una estructura con múltiples filas donde cada nivel o fila incluye:

- 1. Una única workstation super administradora con una base de datos super compartida y un Task Processor.*
- 2. Una o más workstations administradoras cada una de las cuales incluye una base de datos super compartida y un Task Processor.*
- 3. Una o más workstations usuarios que se comunican con la workstation administradora de la fila inmediata superior*



Dimensiones y Valores de dimensión

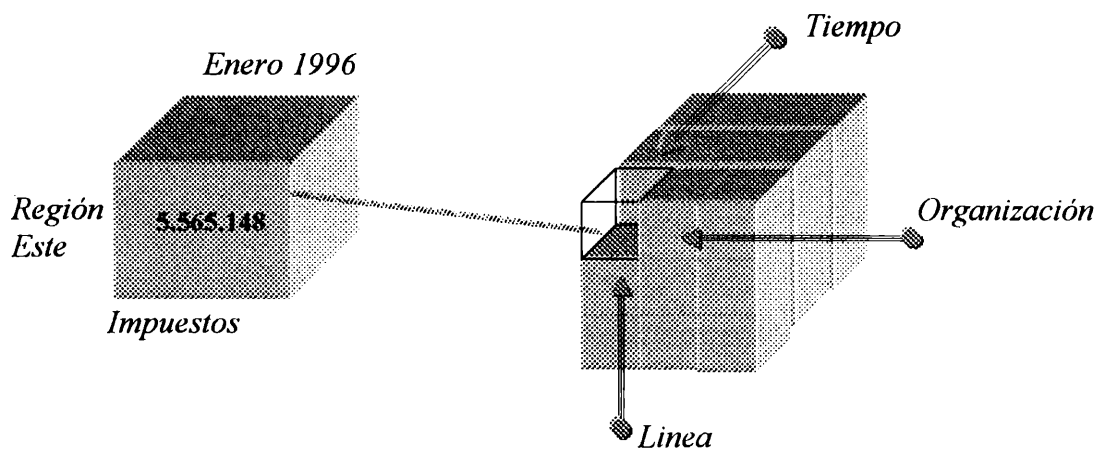
Los datos se organizan en Ofa usando tres tipos de estructuras:

✓ *Dimensiones*

Son objetos de bases de datos que organizan los datos contenidos en los ítems de datos financieros.

Permiten seleccionar y trabajar con un subconjunto específico de datos.

- ✓ *Valores de dimensión*
Son los elementos que integran una dimensión. Se usan para organizar los valores de los ítems de datos financieros.
- ✓ *Ítems de datos financieros*
Son los datos propiamente dichos.



Esto muestra cómo la intersección de los valores de dimensión Impuesto, Región Este y Enero 1998 apuntan al valor de ítem de dato financiero 5.565.146 en la base de datos.

Mantenimiento de dimensiones y valores de dimensión

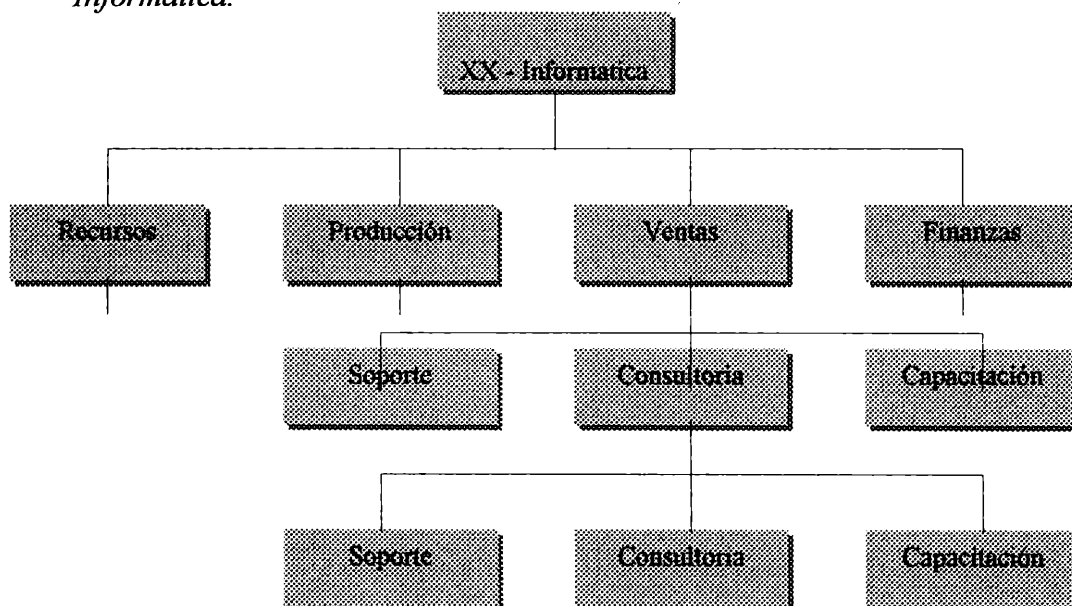
Workstation administradora

Es responsabilidad del administrador mantener valores de dimensión y dimensiones en su base de datos compartida. Si modifica una dimensión o valor de dimensión y quiere que otros usuarios puedan ver esta modificación, deberá distribuirla a la base de datos compartida y a otros usuarios.

Workstation budget

El usuario de una workstation budget puede crear y modificar dimensiones y valores de dimensión para su uso personal, pero no podrán ser submitidos a la base de datos compartida. Podrá modificar dimensiones y valores de dimensión creados por el administrador pero no podrá salvar las modificaciones a no ser que le asigne un nuevo nombre.

- ✓ *Los datos para Soporte, Capacitación y Consultoría harán roll-up en los datos de Ventas. Recursos, Producción, Ventas y Finanzas harán roll-up en los datos de XX Informática.*

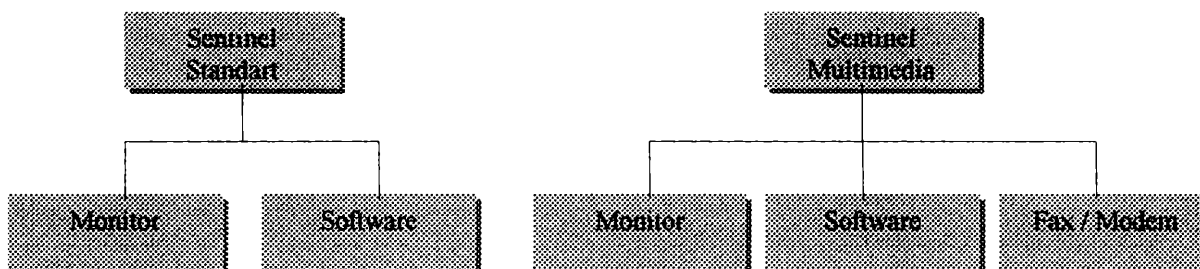


Jerarquías múltiples

OFA soporta jerarquías múltiples, se pueden especificar cualquier número de jerarquías para hacer roll-up de datos.

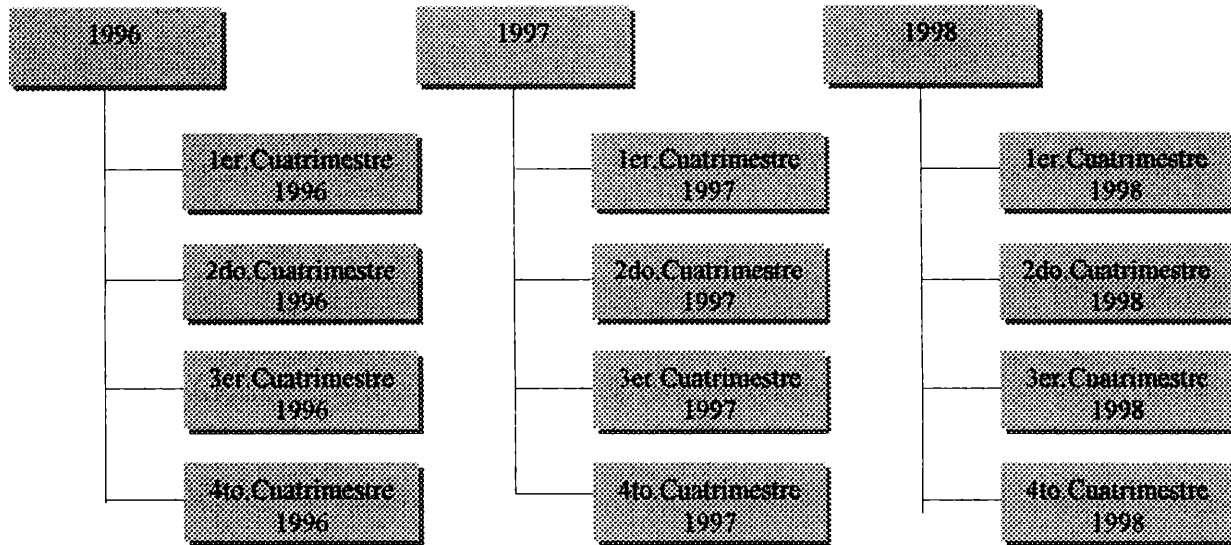
Por ejemplo, se podrían definir jerarquías separadas para examinar datos gerenciales y funcionales. En una jerarquía cualquier valor de dimensión que no está en nivel más alto puede tener un solo padre. Sin embargo, si el valor de la dimensión pertenece a más de una jerarquía, puede tener un padre diferente en cada una.

Otro ejemplo, dos jerarquías representando tipos diferentes de equipos: Sentinel Standard y Sentinel Multimedia: las dos con los valores de dimensión: Monitor y Software. La segunda contiene otro valor de dimensión Fax/Modem:



Jerarquías con múltiples raíces

OFA soporta también jerarquías con más de una raíz, esto es útil para las dimensiones relativas al tiempo:



1996, 1997 y 1998 aparecen como raíces en la jerarquía Tiempo. Estos valores son hermanos y no tienen padre en la jerarquía.

Mantenimiento de jerarquías

Es responsabilidad del administrador mantener las jerarquías en la base de datos compartida y distribuir las modificaciones a los otros usuarios.

Los usuarios de workstations budget pueden crear y modificar sus propias jerarquías, sólo para su uso personal.

Atributos

Son objetos de la base de datos que establecen relaciones entre los valores de dos dimensiones de tal forma que los valores de una dimensión (dimensión grouping) pueden ser usados para agrupar valores de otra (dimensión base). Cada valor de la dimensión grouping será una característica o atributo que se aplica a múltiples valores de la dimensión base.

Se usan para:

- ✓ *facilitar el proceso de selección de datos a usar en reportes, gráficos y worksheets*

- ✓ *facilitar la selección de datos relacionados ya que se seleccionan un conjunto de valores de la dimensión base especificando solamente un valor de la dimensión grouping*
- ✓ *eliminar la necesidad de especificar cada valor de la dimensión base en operaciones separadas cuando se seleccionan datos*

Tipos de atributos

✓ Uno a muchos

Establece una relación entre los valores de dos dimensiones donde cada valor de la dimensión grouping está relacionado con múltiples valores de la dimensión base.

Simplifica el proceso de selección de datos permitiendo especificar un único valor de la dimensión grouping para seleccionar muchos valores de la dimensión base.

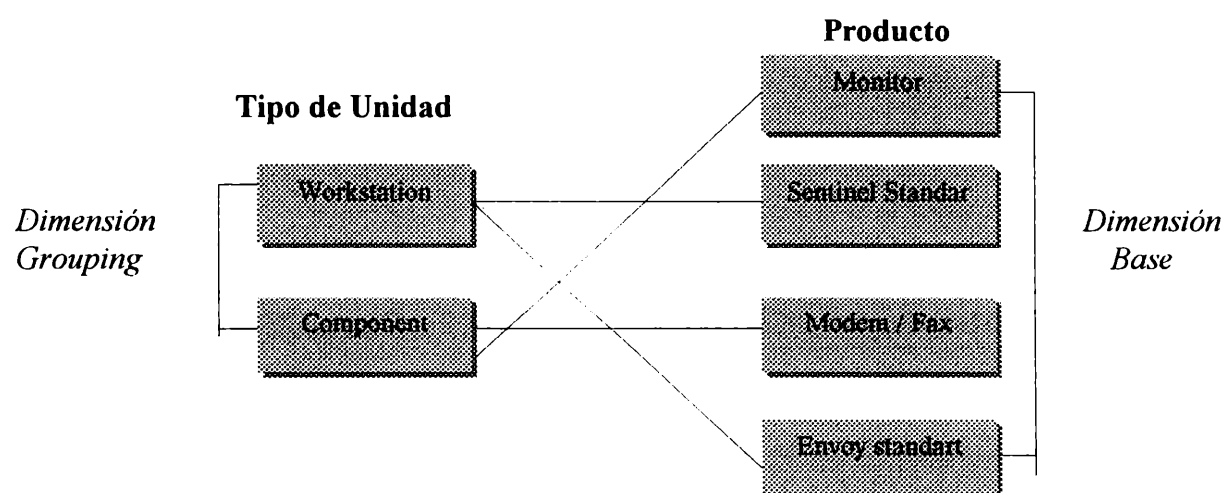
Ejemplo:

Llamamos Tipo de Unidad a la dimensión grouping con dos valores posibles: workstation y Componentes. La dimensión base es Producto y tiene cuatro valores posibles: Monitor, Sentinel Standard, Modem/Fax y Envoy Standard.

El valor de dimensión grouping workstation está relacionado a dos valores de la dimensión base: Sentinel Standard y Envoy Standard.

El valor de dimensión grouping Componente está relacionado a dos valores de la dimensión base: Monitor y Modem/Fax.

Se podría seleccionar un subconjunto de Producto especificando workstation o Componente.



✓ Muchos a muchos

Establece una relación entre los valores de dos dimensiones donde múltiples valores de la dimensión grouping están relacionados con múltiples valores de la dimensión base, permitiendo ver los mismos datos en múltiples contextos.

Simplifica el proceso de selección de datos permitiendo especificar valores simples de la dimensión grouping para seleccionar muchos valores de la dimensión base, al igual que los atributos uno-a-muchos.

Proporciona una característica adicional para refinar la selección de datos ya que, debido a que los valores de la dimensión base pueden estar relacionados a múltiples valores de la dimensión grouping, los valores de la dimensión grouping proveen un contexto para la selección de datos.

Ejemplo:

La dimensión grouping es TiposDeMonitores con dos valores posibles: VGA y Super VGA. La dimensión base es Producto con los valores: Sentinel Standard, Sentinel Finacial, Sentinel Multimedia y Envoy Standard.

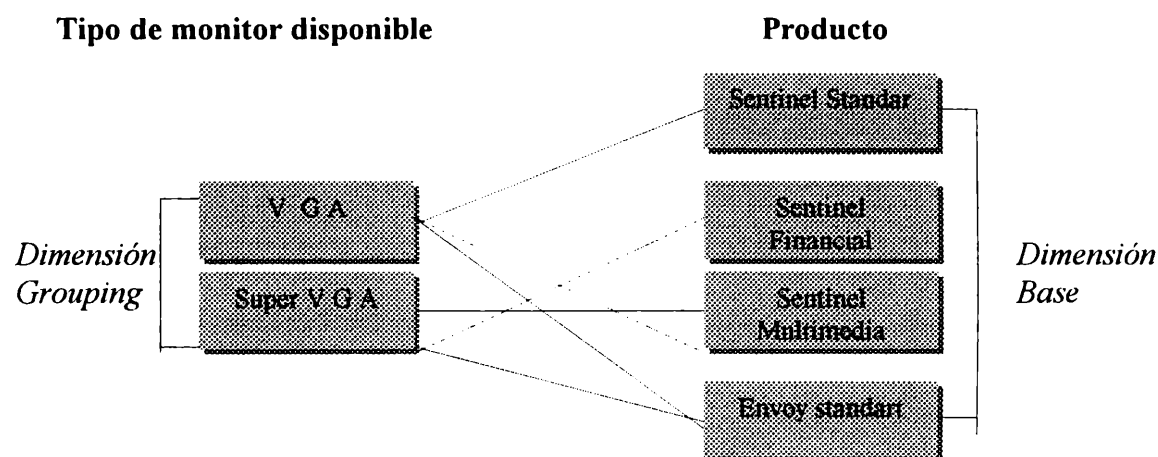
Múltiples Productos están relacionados con los TiposDeMonitores disponibles para ellos.

El valor de dimensión grouping VGA está relacionado con los valores de la dimensión base: Sentinel Standard, Sentinel Multimedia y Envoy Standard.

Super VGA está relacionado con los valores de la dimensión base: Sentinel Finacial, Sentinel Multimedia y Envoy Standard.

Los Productos Sentinel Multimedia y Envoy Standard se relacionan a los dos TiposDeMonitor disponibles.

Se puede seleccionar el valor Sentinel Multimedia dentro del contexto en que uno quiera analizarlo: como parte de las workstation VGA o dentro de las workstation Super VGA.



Mantenimiento de atributos

Los atributos se usan para facilitar la selección de datos relacionados. Podemos usar atributos para agrupar datos relacionados en diferentes formas: por tipo de producto, por período de tiempo, por usuario, etc.

Por ejemplo, un administrador podría crear un atributo `budgetEste` que incluye todos los usuarios de workstation `budget` en la división `Este` de la compañía. El administrador podrá entonces distribuir datos específicos a todos los usuarios de workstation `budget` en la división `Este` sin necesidad de seleccionarlos individualmente.

Es responsabilidad del administrador mantener los atributos en la base de datos compartida y distribuir las modificaciones a los otros usuarios.

Los usuarios de workstations `budget` pueden crear y modificar atributos para su uso personal.

Modelos

Son objetos de la base de datos formados por ecuaciones interrelacionadas que pueden usarse para calcular nuevos datos. También pueden ser usados en worksheets para realizar análisis `WhatIf`.

Los datos calculados se mantienen separados de los datos financieros.

El mismo modelo puede ser usado para calcular datos para múltiples datos financieros siempre y cuando tengan las mismas dimensiones.

Los modelos permiten realizar diferentes funciones de manejo de datos financieros, tales como:

- ✓ *establecer relaciones simples entre valores de dimensión*
- ✓ *establecer relaciones complejas involucrando datos almacenados en cubos separados*
- ✓ *ecuaciones simultáneas donde es esencial la secuencia de múltiples cálculos para llegar a los resultados deseados.*

Ejemplo:

Un modelo para generar valores de ingresos y gastos incluye:

- ✓ *Costo de las mercaderías vendidas*
- ✓ *Ventas brutas*
- ✓ *Ganancia bruta*

$$\text{CostoDeMercaderíasVendidas} = \text{CantUnidades} * \text{CostoUnitario}$$

$$\text{VentasBrutas} = \text{CantUnidades} * \text{PrecioUnitario}$$

$$\text{IngresoVentasNeto} = \text{VentasBrutas} - (\text{DescuentosVentas} + \text{RetornosVentas})$$

$$\text{GananciaBrutaVendida} = \text{IngresoVentasNeto} - \text{CostoDeMercaderíasVendidas}$$

Los valores *CantUnidades*, *PrecioUnitario*, *CostoUnitario*, *DescuentosVentas* y *RetornosVentas* son valores de entrada.

La combinación de éstas y otras ecuaciones se llama modelo. Al aplicar el modelo a los datos financieros se calculan las fórmulas y se obtienen resultados que servirán para el análisis posterior.

Otro ejemplo:

Modelización del beneficio de la división Consultoría considerando cantidad de empleados y cargos que ocupan:

$$\text{BeneficioConsultoría} = (\text{BeneficioMensual de empl A} * \text{cant empl A}) + ((\text{BeneficioMensual de empl B} * \text{cant empl B}) + \dots)$$

En este ejemplo los analistas pueden variar la cantidad de empleados de cada tipo (Director, Senior, Junior, Consultor), aplican el modelo y ven así el impacto que ese cambio genera en los beneficios totales de la división.

Es responsabilidad del administrador mantener los modelos en la base de datos compartida y distribuir las modificaciones a los otros usuarios.

Los usuarios workstation budget pueden crear y modificar modelos para su uso personal.

Resolución de datos

Si hubo cambios en datos relacionados a valores de dimensión en la base de datos, se deben "resolver" los datos para asegurar que los cambios sean corrientes. Cuando se resuelven datos asociados a un ítem de dato financiero particular, OFA realiza los cálculos, y agrega los datos a lo largo de las jerarquías de acuerdo a los seteos especificados usando modelos.

Es necesario resolver datos en un ítem de dato financiero luego de:

- ✓ cambiar o agregar una ecuación en un modelo
- ✓ cambiar el tipo de agregación
- ✓ cambiar o agregar una jerarquía para una dimensión asociada con el ítem de dato financiero
- ✓ agregar nuevos periodos de tiempo a la base de datos
- ✓ agregar nuevas dimensiones a la base de datos
- ✓ agregar nuevos ítems de datos financieros a la base de datos
- ✓ cambiar el período de tiempo a la base de datos
- ✓ refrescar el ítem de dato financiero luego de agregar nuevos datos
- ✓ entrar o cambiar datos para el ítem de dato financiero

Es importante notar la diferencia entre recalcular datos y resolver datos. Cuando se recalculan datos, el recálculo se realiza sólo en las dimensiones activas en el worksheets. Mientras que cuando se resuelven datos, el recálculo se realiza para los datos financieros del modelo completo.

El mecanismo básico de OFA para realizar estas resoluciones de datos es llamado "solve definition".

Se trata de una herramienta que permite definir un contexto para calcular nuevos datos para un ítem de dato financiero.

Transferencia de datos y estructuras

Con respecto a este tema, los administradores harán procesos de transferencia de datos, distribución de estructuras, distribución de datos, submisión de datos, refresh de datos.

Los usuarios de workstation budget harán procesos de transferencia de datos, submisión de datos, refresh de datos.

Un usuario de una workstation analista hará procesos de transferencia de datos y salvado de los datos.

Métodos de transferencia de información

- ✓ *Distribución de Estructuras*
Permite a los administradores (incluyendo el super administrador del sistema) transferir objetos de la base de datos a sus usuarios.
- ✓ *Distribución de Datos*
Permite a los administradores (incluyendo el super administrador del sistema) transferir datos a sus usuarios.
- ✓ *Refresh de Estructuras*
Permite a los usuarios incorporar estructuras que les son enviadas por sus administradores.
- ✓ *Refresh de Datos*
Permite a los usuarios actualizar los datos con los que trabajan con la versión más actualizada desde la base de datos compartida.
- ✓ *Submit de Datos*
Permite a los administradores y a los usuarios de workstation budget actualizar con sus datos la base de datos compartida.

Cómo se transfiere la información

- ✓ *El super administrador del sistema crea un conjunto de objetos de base de datos (dimensiones, valores de dimensión, jerarquías, modelos, worksheets, reportes y gráficos)*

- ✓ *El super administrador distribuye objetos de la base de datos (estructuras) a los usuarios que soporta.*
- ✓ *Los usuarios de workstations budget en el nivel inmediatamente inferior al del super administrador incorpora aquellas estructuras en su base de datos personal. (refresh de estructuras de datos)*
- ✓ *Los usuarios de workstations budget realizan entrada de datos y corren cálculos sobre ellos para crear nuevos datos. (ingreso de datos, modificación de datos en los ítems de datos financieros, creación de modelos)*
- ✓ *Los usuarios de workstations budget submiten datos para actualizar la base de datos compartida del super administrador con los datos que ellos han creado.*
- ✓ *Una vez que los datos están en la base de datos compartida quedan disponibles para ser usados por usuarios analistas, otros usuarios budget y usuarios administradores a los que el super administrador les haya distribuido los objetos apropiados.*
- ✓ *Estos usuarios ejecutan el proceso de Refresh Data para acceder a los datos. (refresh de estructuras de datos, refresh de datos)*

Flujo de información

- ✓ *Flujo de datos*

En la estructura por filas usada por Ofa los datos fluyen hacia abajo y hacia arriba a través de las distintas filas en una forma específica y controlada que permite a los administradores asegurar la integridad de los datos en cada nivel y saber que los usuarios que soportan tienen acceso a los datos apropiados.

El administrador debe dar a los usuarios a quienes quiere transferir los datos, acceso a los objetos en la base de datos compartida que estructuran esos datos.

Cuando se distribuyen objetos a un usuario, por default se le debe dar permisos de lectura y escritura a los datos referenciados por esos objetos.

- ✓ *Flujo de estructuras*

En la estructura por filas los objetos de la base de datos (estructuras) fluyen sólo hacia abajo a través de los distintos niveles.

El super administrador, que ocupa el nivel más alto de la estructura, crea objetos de la base de datos para representar las necesidades de la organización y los deja disponibles a los usuarios del próximo nivel de la estructura.

Realiza ésto distribuyendo los objetos a la base de datos compartida que mantiene y a los usuarios que soporta.

Flujo de información en una estructura por filas

Los datos financieros son submitidos y distribuidos hacia arriba y hacia abajo a través de la estructura por filas. La estructuras de base de datos son distribuidas desde el administrador hacia los usuarios, por lo tanto el flujo es sólo hacia abajo en la estructura por filas.

Este proceso es manejado por los administradores y permite a los datos originados en un nivel de la estructura ser pasados a otros niveles.

Transferencia de datos desde y hacia la base de datos compartida

Los usuarios budget y analistas envían sus datos a la base de datos compartida vía el Task processor.

Luego otros usuarios asociados a la base de datos compartida pueden refrescar su visión de la base para ver los nuevos datos, pueden trabajar con ellos y generar datos nuevos basados en ellos.

Luego pueden usar el proceso de submisión de datos para enviar los nuevos datos que han creado a la base de datos compartida de su administrador, haciéndolos de esta forma, disponibles a otros usuarios de su mismo nivel en la estructura.

De la misma manera, su administrador puede usar el proceso de submisión para enviar datos que tengan significado para ambos niveles, a la base de datos compartida de su administrador.

Los procesos de transferencia de datos son controlados por el administrador correspondiente a cada nivel, quien mantiene la base de datos compartida para usuarios del nivel inmediatamente inferior.

Administrador de tareas de distribución

Se llama distribución al proceso de proveer acceso a los objetos en la base de datos compartida.

Esta tarea es realizada por el administrador quien distribuye objetos de la base de datos y datos a los usuarios que soporta.

Los objetos que pueden distribuirse son dimensiones, valores de dimensión, ítems de datos financieros, jerarquías, documentos (reportes, gráficos y worksheets), modelos, atributos, folders y datos asociados a los objetos.

El administrador distribuye objetos y datos a workstation analistas, budget y administradoras que están en la fila inmediata inferior o desde su propia base personal a la base de datos compartida que mantiene.

Es responsabilidad del administrador distribuir los objetos y datos apropiados a cada usuario basados en los slices de datos que cada usuario necesita acceder para poder realizar su tarea.

El administrador debe realizar una distribución cuando:

- ✓ *ha creado un nuevo objeto de base de datos o modificado un existente, para uno o más de los usuarios que soporta*
- ✓ *necesita borrar un objeto de la base de datos compartida y remover el acceso de uno o más usuarios al mismo*
- ✓ *nuevos usuarios necesitan acceder a objetos en la base de datos compartida*
- ✓ *necesita proveer, a uno o más usuarios, acceso a objetos adicionales a los cuales no tenía acceso hasta el momento*
- ✓ *ha creado o modificado datos asociados a estructuras que el/los usuarios ya tenían acceso*

Distribución de datos

Los administradores distribuyen slices de datos desde su base de datos personal tanto a la base de datos compartida como a las workstation subordinadas.

Una vez especificados los datos a distribuir, se debe correr el Task Processor para completar la distribución. Así cada usuario de workstation budget deberá refrescar su base de datos personal para incorporar los datos modificados.

Los usuarios de workstations analistas deberán refrescar sus vistas de la base de datos compartida para visualizar los nuevos datos.

Distribución de estructuras

El administrador necesita distribuir las estructuras de la base de datos a los usuarios periódicamente para asegurarse que tienen acceso apropiado a los datos en la base de datos compartida.

Distribuir estructuras implica seleccionar los ítems que se quieren distribuir y especificar qué usuarios han de recibirlos. Además el administrador deberá especificar si los ítems distribuidos van a ser agregados, modificados o borrados en la base de datos destino.

Después de especificadas las estructuras a distribuir y los receptores, el administrador deberá correr el Task Processor para completar la distribución. Luego cada usuario deberá refrescar su base de datos personal para incorporar las estructuras modificadas.

Task Processor

El Task Processor mantiene una cola de tareas que procesa en orden para que los objetos sean distribuidos.

El Task Processor puede estar seteado para correr continuamente como una workstation separada o ser corrido manualmente por el administrador.

Submit de datos desde una workstation administrador

Una vez que los usuarios han submitido datos a la base de datos compartida el administrador de la misma debe a su vez submitirlos a la base de datos compartida de su administrador.

Debido a que la workstation super administrador está en el lugar más alto de la estructura la base de datos que mantiene contiene el nivel más alto de datos consolidados.

Tanto el administrador como el super administrador debe submitir los datos creados o modificados en su base de datos personal a la base compartida que mantiene.

El proceso de submitir datos es el mismo si se submiten datos desde la base de datos personal a la compartida o si se submiten datos de la base de datos compartida a la base de datos compartida del administrador.

Submitiendo datos desde una workstation budget

Una vez preparado un budget o un forecast usando un worksheets, el usuario puede submitir su trabajo a la base de datos compartida donde puede ser accedido por otros usuarios.

El usuario puede manipular los datos en su base de datos personal usando una worksheets y salvar el trabajo sin afectar la base de datos compartida. Para que otros usuarios puedan acceder al trabajo realizado, el usuario deberá submitirlo desde la base de datos personal a la compartida.

Antes de hacerlo, el usuario deberá especificar los slices de datos que desea submitir. Un slice de datos está definido por un ítem de dato financiero almacenado (Actuales, budget, Forecast) quien a su vez está definido por un conjunto de dimensiones (Organización, Línea, Tiempo). Por lo tanto especificar un slice de datos implica seleccionar tanto el ítem de dato financiero almacenado como un subconjunto de sus valores de dimensión (un conjunto de organizaciones, un conjunto de ítems de línea, un conjunto de periodos de tiempo).

Salvado de datos desde la workstation analista

Un usuario de una workstation analista no mantiene copias de la base de datos compartida si no que accede directamente a ella cuando abre una worksheet. Por lo tanto necesitará sólo salvar datos en la worksheet para submitir su trabajo a la base de datos compartida.

Refreshing

Para asegurarse que se está trabajando con la información más actualizada de la base de datos compartida, el usuario deberá refrescar las estructuras de datos de su workstation (dimensiones, jerarquías, modelos y documentos).

El refresh puede hacerse manualmente o automáticamente. En el segundo caso, el refresh de datos y estructuras se hace cada vez que el usuario comienza una nueva sesión de trabajo.

Refrescando en una workstation budget

Si el administrador decide que el refresh sea automático, el software refresca las estructuras de datos automáticamente cada vez que el usuario de la workstation budget inicia una nueva sesión de trabajo.

De lo contrario el usuario debe refrescar manualmente sus estructuras.

Debido a que el usuario trabaja con una copia personal de los datos de la base de datos compartida, debe hacer un refresh separado para refrescar sus datos.

Cuando se refrescan datos el usuario debe especificar cuales slices de datos se quieren actualizar.

Si el administrador ha distribuido nuevas estructuras o estructuras modificadas, el usuario budget debe refrescar las nuevas estructuras antes de refrescar nuevos datos asociados con ellas.

Refrescando en una workstation administradora

El administrador necesita refrescar periódicamente sus bases de datos compartida y personal. Puede refrescar estructuras y datos automáticamente o manualmente.

El administrador puede refrescar los datos en su base de datos personal desde la base de datos compartida que él maneja, o los datos en la base de datos compartida que administra desde la base de datos.

Refresh de estructuras automático

La workstation en el nivel más alto de la jerarquía determina si una workstation tendrá refresh automático o no. Así el super administrador controla el refresh automático para sus administradores subordinados y a su vez, cada administrador controla el refresh automático para las workstation subordinadas a él.

El super administrador no necesita refrescar estructuras ya que no existe ningún administrador arriba de él que se las distribuya, y las estructuras que distribuye de la base de datos personal a la compartida son aplicadas directamente a la base de datos compartida por el Task Processor.

El administrador es el responsable de refrescar los datos en la base de datos compartida que mantiene con cualquier dato que se haya creado o modificado en la base de datos compartida en el nivel inmediatamente anterior en la estructura.

Si el administrador o super administrador trabaja con datos en una base de datos personal, debe refrescar los datos en la misma desde la base de datos compartida que mantiene.

Cuando se refrescan datos se debe especificar el slice de datos que se quiere refrescar

Si un administrador le ha distribuido estructuras nuevas o actualizadas a otro, éste último debe refrescar las nuevas estructuras antes de refrescar los datos asociados con ellas.

El super administrador puede refrescar datos en su base de datos personal desde la base de datos compartida que maneja, pero no puede refrescar los datos en la base de datos compartida porque no existe otra base de datos compartida en un nivel superior en la jerarquía

Si un administrador distribuye estructuras desde la base de datos personal a la base de datos compartida que mantiene, cuando corre el Task Processor para procesar la distribución, las estructuras son automáticamente incorporadas en la base de datos compartida y no hay necesidad de refrescar las estructuras explícitamente en la base de datos compartida.

Procesamiento de tareas

El administrador es el encargado de monitorear y correr el Task Processor.

OFA procesa todas las tareas usando el Task Processor y lleva el registro de ellas usando el Task Queue.

El Task Processor permite a los administradores mantener la base de datos compartida. Las distribuciones del administrador, los procesos de resolución y las submisiones de datos a la base de datos compartida, son ruteados al Task Processor, donde esperan en una cola hasta poder ser procesados.

Este proceso permite a múltiples usuarios submitir tareas simultáneamente y da al administrador la oportunidad de monitorear todas las tareas de procesamiento en la base de datos.

Cuando se procesan tareas usando el Task Processor se están enviando a la base de datos compartida, la cual hace disponible toda la información a los usuarios apropiados.

Debido a que todas las tareas en el Task Processor son procesadas cronológicamente, las tareas pueden sobre escribirse unas a otras dependiendo el orden en el que fueron submitidas. El administrador puede reordenar las tareas, tal que la información más crítica no sea inadvertidamente sobre escrita.

Cuando las tareas son submitidas y procesadas por el Task Processor aparecen en el Task Queue, quien mantiene información acerca de cada tarea, tal como:

- ✓ *el usuario que la inició*
- ✓ *el tipo de tarea (por ejemplo, distribución)*
- ✓ *si la tarea está pendiente o ha sido procesada*
- ✓ *cuándo la tarea fue submitida para procesamiento*
- ✓ *si la tarea está pendiente, cuándo va a ser procesada y cuál es su posición en la cola (prioridad)*
- ✓ *si la tarea ya fue procesada, cuándo se completó el procesamiento*

Usar el Task Queue para monitorear el procesamiento de tareas permite controlar todos los eventos del procesamiento, ayudando a manejar la actividades de la base de datos más efectivamente.

El administrador puede ver información sobre las tareas en la cola de tareas que él maneja. Puede ver la historia de cualquier tarea, y ver los detalles de las tareas de submisión de datos.

Reportes, gráficos y worksheets

Son herramientas poderosas e interactivas que permiten ver los datos financieros desde diferentes perspectivas y preparar presentaciones efectivas basadas en los análisis realizados.

Se debe especificar el slices de datos que se quiere examinar, luego se puede manipular el layout para verlo desde diferentes maneras.

Ejemplo - Sistema de control de presupuestos

Gabriela Lopez Uhalde

Silvana Manganiello

Eduardo Martinez

Ejemplo – Sistema de Control de Presupuestos

Presentamos un problema real, para mostrar los conceptos de bases de datos multidimensionales utilizando los elementos provistos por OFA:

Descripción del problema

El objetivo del sistema es brindar las herramientas necesarias para el Control y Análisis de Presupuesto en una empresa automotriz.

El sistema deberá emitir dos niveles de reportes:

- ✓ *Costos primarios por Gerencia – Departamento – Sección*
- ✓ *Consolidado total de la Compañía*

El responsable de cada Gerencia – Departamento – Sección ingresará las unidades mínimas de presupuesto con las que se elaborará el Presupuesto de Costos Primarios.

Una vez obtenidos los presupuestos de costos primarios de cada responsable presupuestario se arma el presupuesto a nivel Compañía. Para obtener este reporte se consolidan en la base de datos compartida los datos ingresados por los responsables presupuestarios.

El Sistema incluye la posibilidad de:

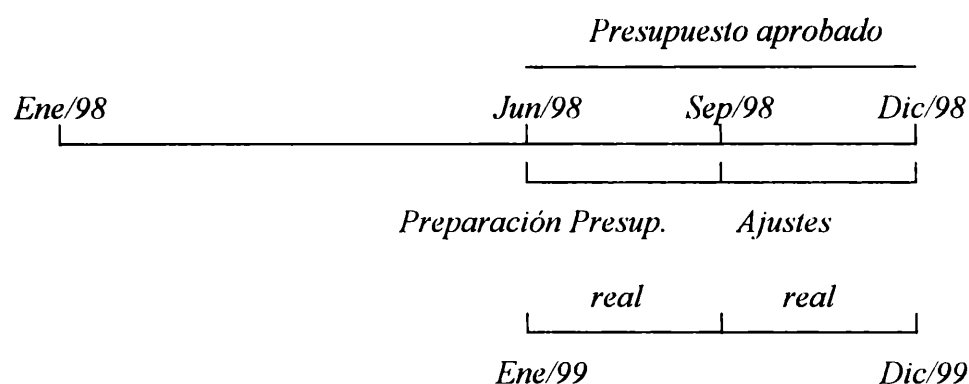
- ✓ *Comparación del costo real (interfase con el sistema contable), el forecast (input de presupuesto mensual re-estimado) y el presupuesto aprobado*
- ✓ *Mantener el presupuesto aprobado y los gastos reales de los años anteriores para poder realizar comparaciones y estadísticas*
- ✓ *Emitir reportes de realización con respecto al presupuesto aprobado y / o el forecast*
- ✓ *Llevar a cabo simulaciones requeridas eventuales y programadas, tales como:*
 - *Ingreso de información a niveles intermedios (por ej total por gerencia, sin detallar la naturaleza del gasto, agrupados por naturaleza de gastos de acuerdo al tema, etc) a los efectos de simular el presupuesto desde un enfoque global de la Compañía (no detallado)*
 - *Cambio de cotización en un determinado periodo de tiempo y moneda*
 - *Variación del presupuesto en un coeficiente especificado*

Preparación del Presupuesto

A mediados de cada año, durante Junio y Julio, se realiza en la compañía la preparación del Presupuesto para el año siguiente.

Los responsables presupuestarios de cada gerencia, en Junio y Julio de 1998 determinan valores para cada mes de 1999, para cada departamento y sección dentro de la gerencia y al detalle de

naturaleza de gasto. Al revisar los valores preparados, junto con los gerentes y más tarde con los socios de la empresa, surgen ajustes que se aplican para llegar al fin del año con un presupuesto final cerrado.

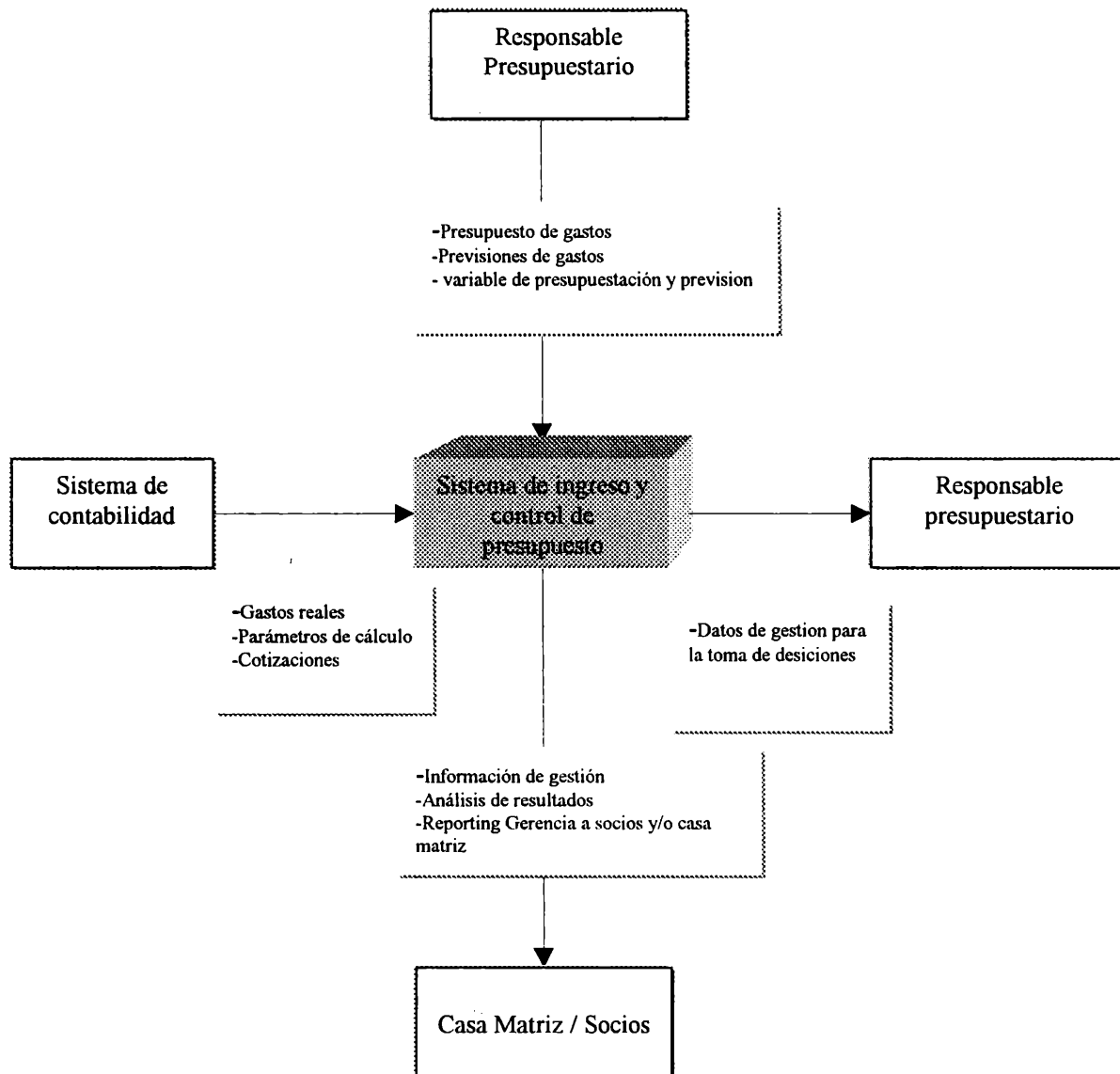


Durante el año siguiente, se realizan previsiones (forecast) o reestimaciones teniendo en cuenta los gastos reales. Al comenzar el año, los valores de forecast son iguales a los valores presupuestados pero van a ir variando a medida que avanza el año. Por ejemplo, para un determinado mes de 1999 se presupuestaron 15.000 \$ y en realidad se gastaron 10.000 \$ debido a la suspensión de un proyecto que pensaba hacerse ese mes o por el contrario se presupuestaron 15.000 \$ y en realidad se gastaron 20.000 \$ debido a gastos imprevistos generados por un accidente ocurrido dicho mes. Este tipo de diferencias se ven reflejadas en los valores de forecast para los meses posteriores a la ocurrencia sin modificación en los anteriores.

Una característica del forecast es que los valores de los meses transcurridos quedan fijos y sólo se modifican los de los meses siguientes, en caso de ser necesario.

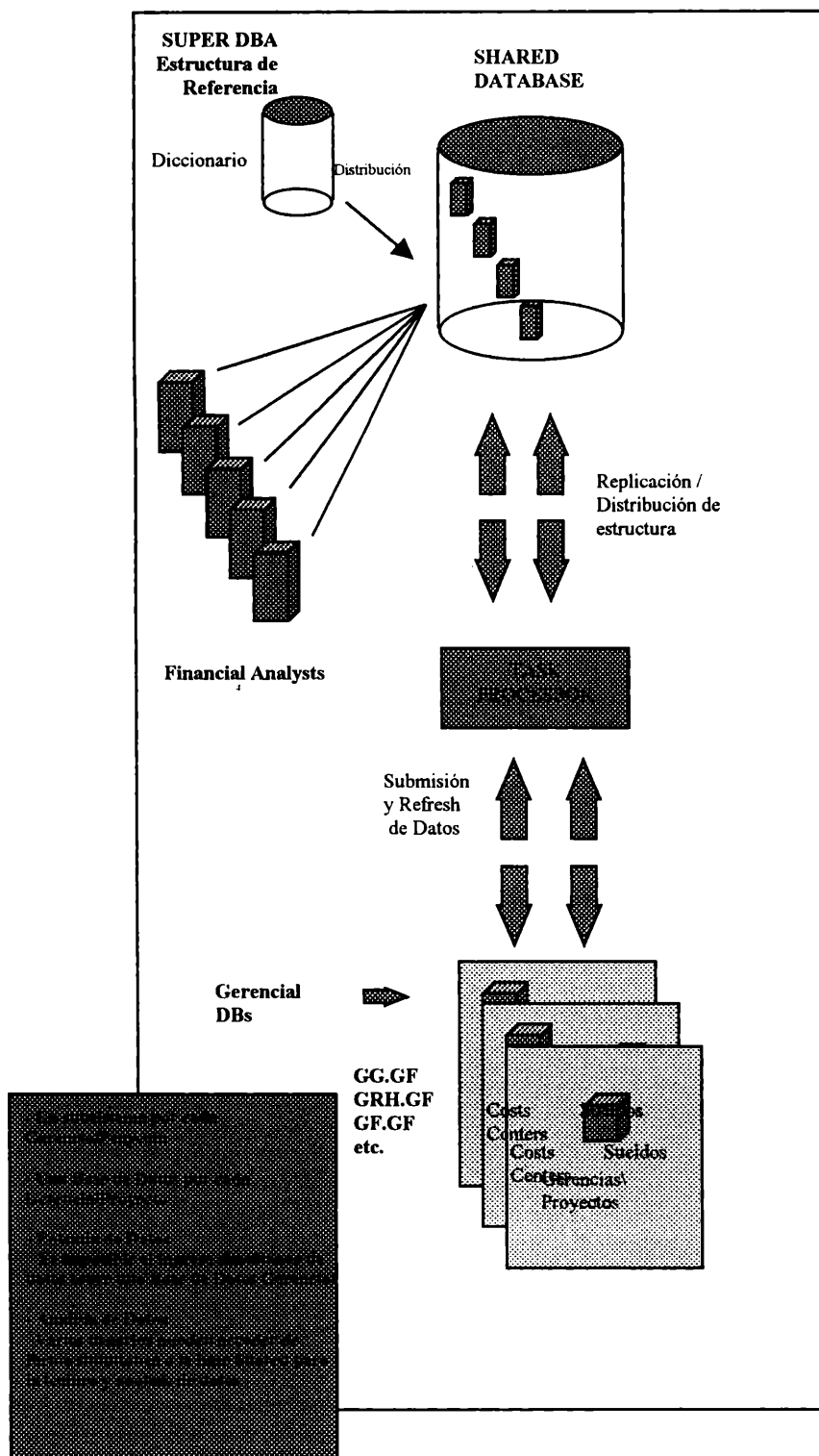
La idea de una buena presupuestación es que al final del año, al comparar los valores presupuestados contra los valores realmente gastados difieran en no más del 5 al 10 %.

Diagrama de contexto



Concepción y arquitectura general del sistema

ORACLE FINANCIAL ANALYZER Sistema de Control del Budget - Concepción y Arquitectura General



*Se tendrán varias **workstations budget**, una por Gerencia, donde el responsable presupuestario (Cost Controller) ingresa sus valores presupuestados (budget) y los reestimados (forecast), al nivel de detalle requerido por la Gerencia.*

Cada workstation budget está asociada a una base de datos personal que contiene estructuras de datos, seteos y datos financieros propios de la Gerencia.

En una workstation de este tipo, el Cost Controller puede definir sus propios objetos de base de datos tales como dimensiones, jerarquías, modelos. Por otro lado podrá usar los objetos de base de datos definidos por el administrador.

Se crearán reportes y gráficos para la Gerencia y worksheets para ingreso de datos y análisis.

Luego de ingresados los datos de la Gerencia serán submitidos desde la base de datos personal a la base de datos compartida de la workstation administradora.

Los datos submitidos tendrán operaciones de refresh, actualizando la base de datos personal con datos de la base de datos compartida de la workstations administradora.

La comunicación entre la base de datos personal y la compartida se realiza a través del Task Processor.

*Se tendrá una **workstation administradora** (punto inicial de la arquitectura) que define y mantiene los usuarios de las workstations budget y otros tantos usuarios de workstations analistas que tendrán acceso a la base de datos compartida.*

Justamente es responsabilidad del administrador controlar el acceso a los datos en la base de datos compartida, además de la definición y la distribución de estructuras de datos comunes (dimensiones, atributos, modelos) al resto de las workstations, además de la distribución de slices de datos a los usuarios.

La base de datos compartida que mantiene esta workstation administradora, contiene datos consolidados correspondientes a la totalidad de las gerencias y serán consultados por las distintas workstations analistas.

*Las **workstations analistas** refrescan las estructuras distribuidas por la workstation administradora y definen y usan sus propios reportes, gráficos y worksheets para realizar tareas de análisis.*

Por ejemplo, la Gerencia de Sistemas de la compañía usa workstations budget que reportan a la workstation administradora.

Los usuarios de la Gerencia de Sistemas pueden :

- ✦ *Editar los gastos de la Gerencia, lo cual facilita la determinación de los valores de presupuesto y forecast basados en las tendencias o en valores reales históricos para la Gerencia.*
- ✦ *Realizar análisis What If considerando factores que pudieran ocurrir.*



- √ *Agregar a la base de datos personal datos propios, no presentes en el conjunto de datos que se distribuyen desde la base de datos del administrador, que permitan realizar análisis más completos.*

Cuando se completa el ingreso de los datos de budget y forecast, se submiten a la base de datos compartida del administrador, previa consolidación.

En la compañía automotriz tendremos una workstation budget para cada una de las gerencias: Recursos Humanos, Finanzas, Comercial, Producción, Sistemas, Auditoría y Análisis. En cada gerencia habrá un responsable presupuestario o Cost Controller encargado de determinar e ingresar al sistema los valores presupuestados durante la preparación del presupuesto así como los valores reestimados (forecast) para dicha gerencia, a nivel departamento y sección y por naturaleza de gasto, para cada mes del año.

Nota: una gerencia está formada por varios departamentos y cada departamento puede estar formado por varias secciones.

También estarán encargados de comunicar al administrador novedades en cuanto a los valores que se manejan, por ejemplo cuando se agrega una nueva clase de gasto (naturaleza) o se crea una nueva sección o se elimina (no tan frecuente como el caso anterior).

Los analistas acceden directamente a la base de datos compartida, desde workstations analistas, a los datos consolidados a un nivel más alto de detalle y pueden realizar sobre ellos diferentes tipos de análisis, reporting y simulaciones .

Los valores correspondientes a los gastos reales se cargarán al sistema mediante una interfase con el Sistema de Contabilidad directamente sobre la base de datos compartida.

Descripción del Flujo de Datos

1 - ENVIAR NOVEDADES AL ADMINISTRADOR		
INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
Cost Controller	<ol style="list-style-type: none"> 1. <i>Chequea las novedades. Por ejemplo: nuevos valores de gerencias, departamentos o secciones, si los hubiera, nuevas naturalezas de gastos, atributos, etc.</i> 2. <i>Envía la nueva estructura al Administrador.</i> 	MAIL

2 - ACTUALIZAR ESTRUCTURA DE DATOS

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
Administrador	<ol style="list-style-type: none"> 1. Recibe el mail 2. Actualiza la estructura con los nuevos valores. 3. Submite los datos a la base de datos shared y a las bases personales correspondientes. 4. Informa de los cambios a los usuarios. 	OFA
Cost Controller	<ol style="list-style-type: none"> 5. Refresca la nueva estructura . 	OFA

3 - INGRESAR VALORES DE PRESUPUESTO

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
Cost Controller	<ol style="list-style-type: none"> 1. Ingreso de los valores de Presupuesto 	<p>OFA: Worksheet</p> <p>Se ingresan los datos por Gerencia/Departamento/Sección y por Naturaleza de Gasto, para el mes indicado.</p>

4 - CONSOLIDAR VALORES DE PRESUPUESTO

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
Cost Controller	<ol style="list-style-type: none"> 1. Consolidar los datos de Presupuesto por Gerencia/Departamento/Sección. <p>Se consolida el cubo InformaciónAnalíticaGastos, por Gerencia (a lo largo de la jerarquía).</p> <p>Se obtienen totales para cada Naturaleza y para cada mes, además el total del año para la Gerencia/Departamento/Sección.</p>	<p>OFA: Solve</p> <p>Ejecuta Rollups sobre el cubo InformaciónAnalíticaGastos .</p>

5 - GENERAR REPORTES DE CONTROL

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
<i>Cost Controller</i>	1. <i>Generar reportes de control de los valores de Presupuesto ingresados.</i>	<i>OFA: Report Se genera un reporte de valores mes a mes para cada Gerencia</i>

6 - SUBMITIR DATOS

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
<i>Cost Controller</i>	1. <i>Submite los valores de Presupuestos consolidados a la base de datos compartida. (cubo InformaciónAnalíticaGastos)</i>	<i>OFA</i>

7 - CHEQUEAR SUBMISIONES

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
<i>Administrador</i>	1. <i>Controla que todas las bases personales hayan submitido los valores de Presupuesto a la fecha especificada.</i> 2. <i>Chequea la actualización en la base de datos Shared.</i>	<i>OFA: Task Processor</i>

8 - ROLLUP CUBO INFORMACIÓNANALÍTICAGASTOS

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
<i>Administrador</i>	1. <i>Rollups sobre la base de datos Shared del cubo InformaciónAnalíticaGastos, consolidando los valores submitidos desde las bases de datos personales.</i> <i>(consolidación a nivel Compañía)</i>	<i>OFA: Solve</i>

9 - EMISIÓN DE REPORTES DE CONTROL

INTERVIENE	ACCIÓN	PROGRAMA/HERRAMIENTA
<i>Analista</i>	1. <i>Genera distintos tipos de reportes finales de Control.</i>	<i>OFA: reports.</i>

Para el ingreso de los datos presupuestados se usará una worksheet OFA como la que sigue. Para una Gerencia/Departamento/Sección fijas, se ingresan los valores para cada mes y cada naturaleza de gasto.

	Ene	Feb	Mar	Abr	Dic	Total
<i>Total</i>	40	40	40	40	40	480
<i>RRHH/ Adm. Personal</i>	10	10	10	10	10	120
<i>Sueldos</i>	10	10	10	10	10	120
<i>Aportes</i>	10	10	10	10	10	120
<i>Indemnizaciones</i>	10	10	10	10			10	120
.....								

Luego se consolidan los datos obteniendo los totales de cada mes y de cada tipo de gasto y un total general para la Gerencia/Departamento/Sección .

Estructura de datos

Los datos están organizados usando tres tipos de estructuras:

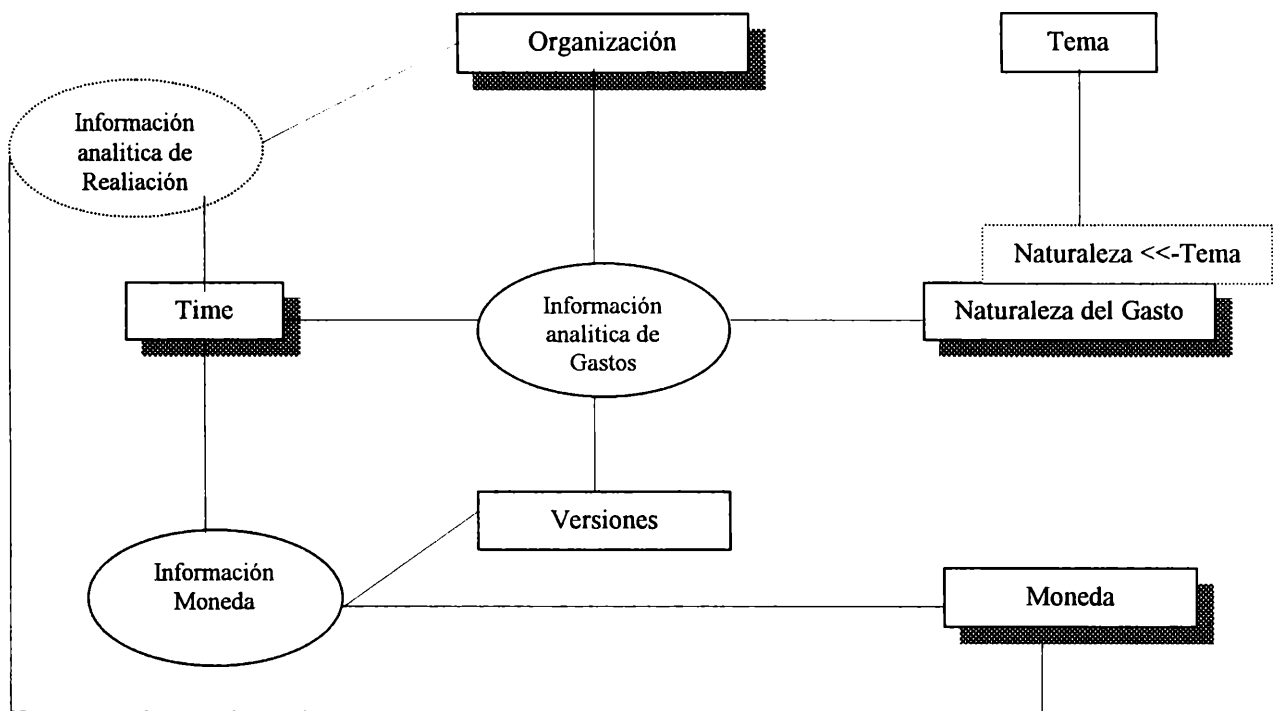
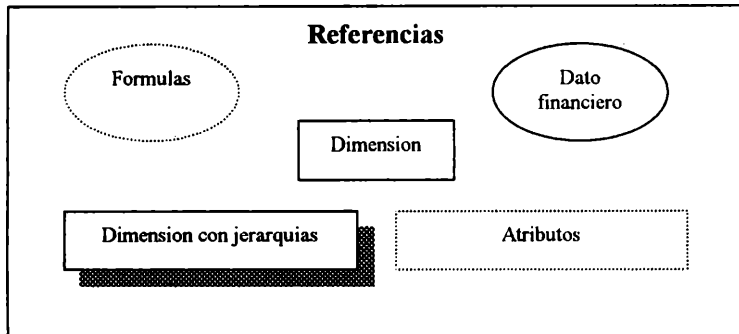
- ✓ Dimensiones
- ✓ Valores de dimensión
- ✓ Items de datos financieros

Las dimensiones son objetos de la base de datos que organizan los datos financieros, permitiendo seleccionar subconjuntos específicos de datos.

Los valores de dimensión son los elementos que componen una dimensión.

En nuestro ejemplo tenemos seis dimensiones:

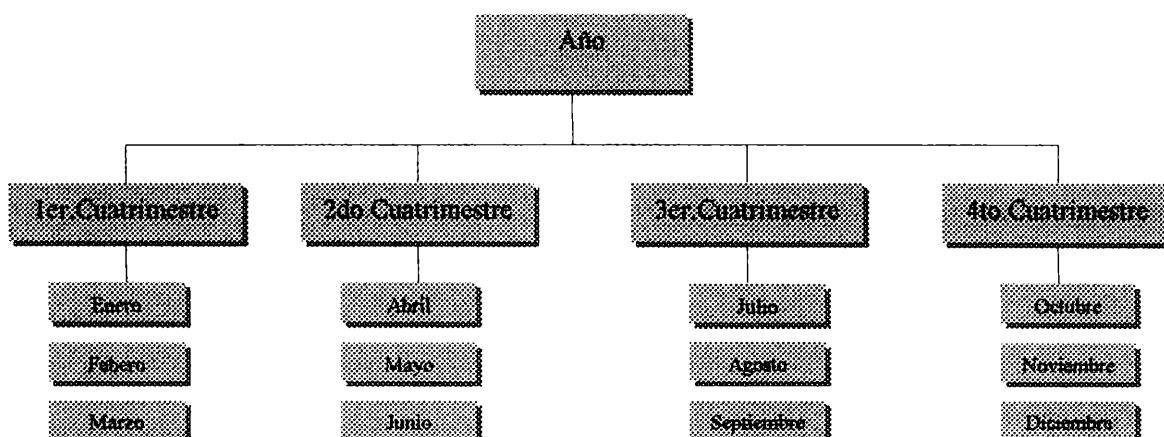
- ✓ Tiempo
- ✓ Organización
- ✓ Naturaleza del Gasto
- ✓ Tema
- ✓ Versiones
- ✓ Moneda



Dimensión Tiempo

Es una dimensión cuyos valores representan períodos de tiempo, un periodo de tiempo puede ser un mes, un cuatrimestre o un año.

Esta dimensión es una dimensión jerárquica, cuya jerarquía esta representada por el siguiente esquema:



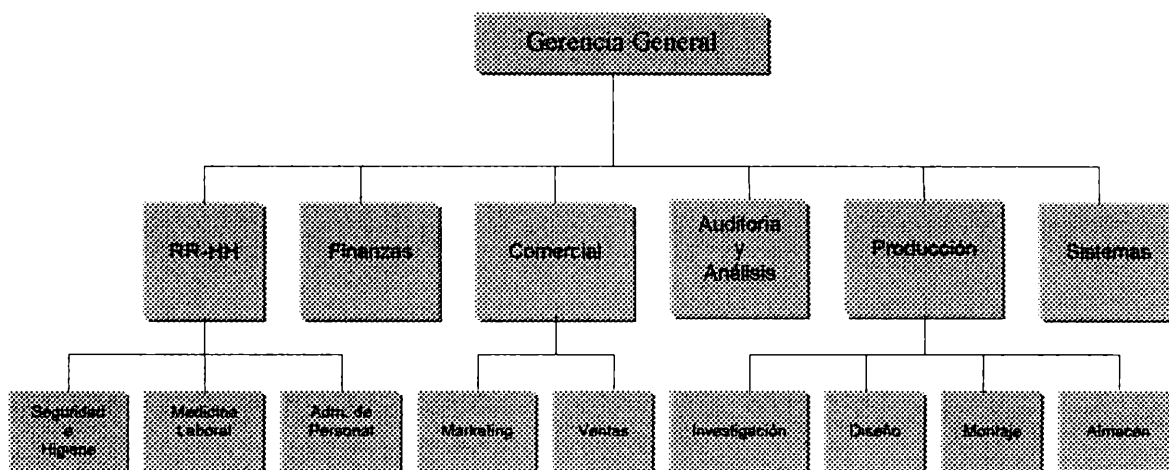
Como el esquema se repite para todos los años se trata de una jerarquía con mas de una raíz, es lo que se denomina jerarquía múltiple.

Notar que los valores de dimensión año (1996,1997,1998,etc) son hermanos sin ningún valor de nivel superior a ellos en la jerarquía.

Dimensión Organización

Es una dimensión cuyos valores representan la organización funcional de la empresa.

Esta dimensión es una dimensión jerárquica, cuya jerarquía esta representada por el siguiente esquema:



Como se ve en el diagrama el valor más alto en la jerarquía esta dado por la Gerencia General, de la que dependen seis gerencias:

Recursos Humanos, con tres departamentos:

- ✓ Seguridad e Higiene
- ✓ Medicina Laboral
- ✓ Administración de personal

Finanzas

Comercial, con dos departamentos:

- ✓ Marketing
- ✓ Ventas

Producción, con tres departamentos:

- ✓ Diseño e Investigación
- ✓ Montaje
- ✓ Almacén

Sistemas, con dos departamentos:

- ✓ Soporte Técnico
- ✓ Análisis y Desarrollo

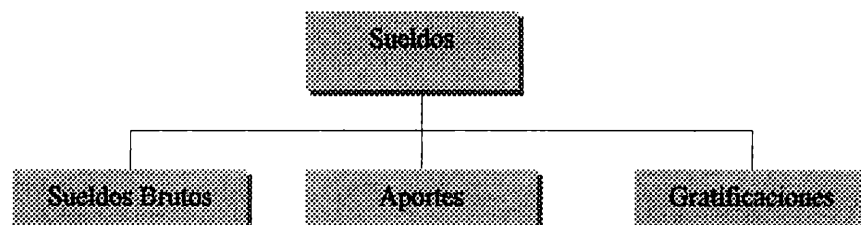
Auditoria y Análisis

Dimensión Naturaleza del gasto

Es una dimensión jerárquica cuyos valores representan los conceptos que originan los gastos.

Algunos ejemplos de valores de dimensión son: Capacitación, Elementos de seguridad, Gastos legales, Campañas de prevención, Medicamentos, Chequeos Médicos periódicos, Preocupacionales, Sueldos brutos, Aportes patronales, Gratificaciones, Indemnizaciones, Elementos de librería, Maquinarias, Transporte, Viáticos, Impuestos aduaneros, Auto partes, Cubiertas, Llantas, Instrumental, Pintura, Servicios de mantenimiento, etc.

Un ejemplo de jerarquía para esta dimensión es “Sueldos” que esta formado por: Sueldos brutos, Aportes patronales y Gratificaciones.



Dimensión Temas

Es una dimensión agrupamiento donde cada valor de dimensión está formado por un conjunto de valores de la dimensión base.

Cuando se define la base de datos se pueden crear objetos de base de datos denominados **atributos**. Los atributos establecen relaciones entre los valores de dos dimensiones de tal forma que los valores de una dimensión (dimensión agrupamiento) pueden ser usados para agrupar múltiples valores de otra dimensión (dimensión base).

Cada valor de la dimensión agrupamiento es una característica o atributo que se aplica a múltiples valores de la dimensión base.

Para una misma dimensión se pueden definir múltiples atributos.

Utilizar atributos simplifica el proceso de selección de datos relacionados, debido a que se pueden seleccionar múltiples valores de la dimensión base especificando un único valor de la dimensión agrupamiento.

El uso de atributos para seleccionar datos, elimina la necesidad de especificar cada valor de la dimensión base en una operación separada.

La dimensión Temas (dimensión agrupamiento) agrupa valores de la dimensión naturaleza del gasto (dimensión base) mediante el uso de atributos.

Los valores de la dimensión temas son:

Insumos de producción

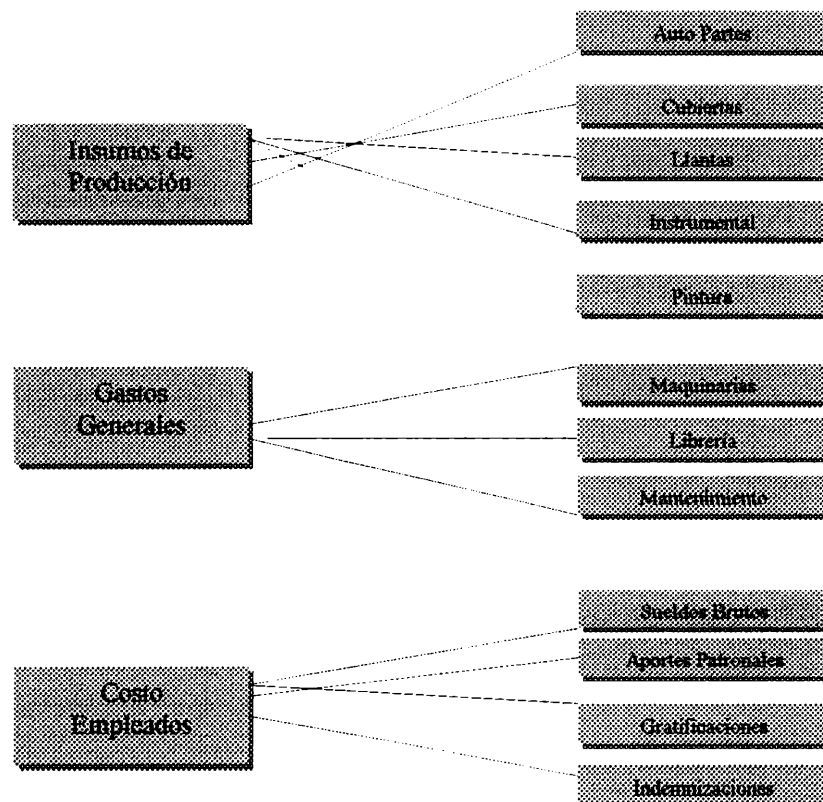
Gastos generales

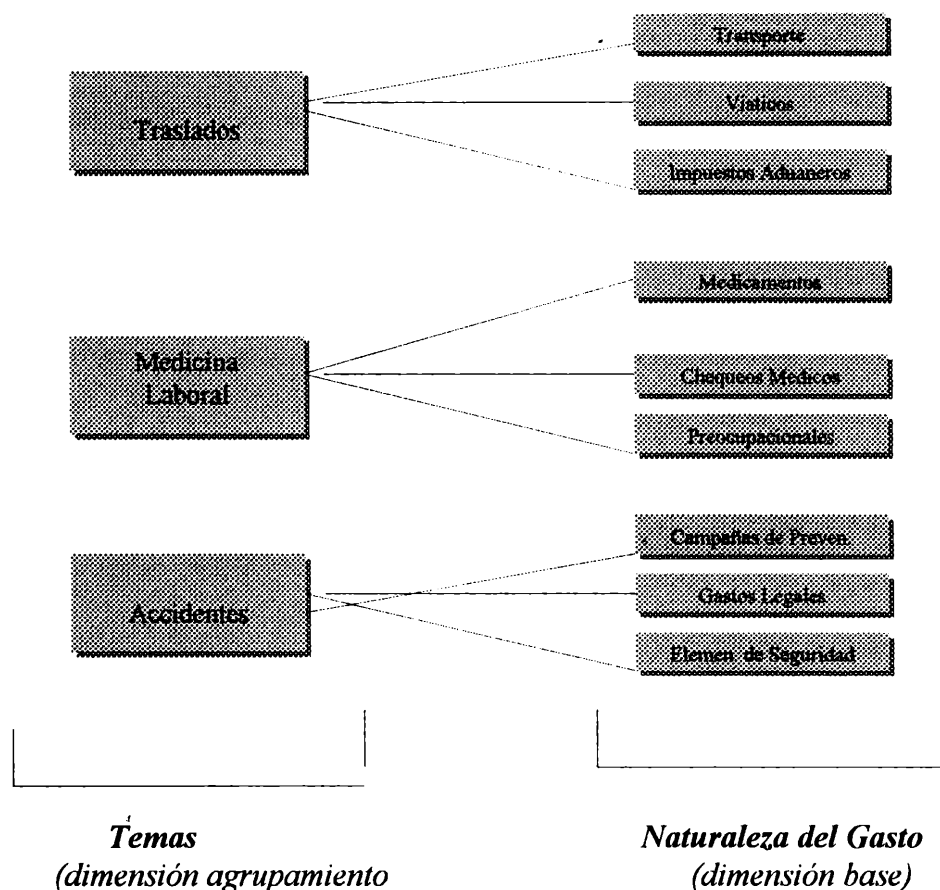
Costo empleado

Traslados

Accidentes

Para la dimensión temas definimos el atributo **Clasificación del gasto** de la siguiente forma:





Como se ve en el esquema el atributo Clasificación del Gasto define la relación entre los valores de la dimensión Temas y los valores de la dimensión Naturaleza del gasto

Dimensión Versiones

Es una dimensión cuyos valores representan los valores de Budget, Forecast y Gastos Reales.

- *Budget*: los datos financieros correspondientes a este valor de dimensión representan el presupuesto de los gastos planeados para el año siguiente.
- *Forecast*: los datos financieros del Forecast representan valores presupuestados re-estimados
- *Gastos Reales*: representan los gastos realmente ocurridos en el transcurso del año.

Dimensión Moneda

Es una dimensión cuyos valores representan la moneda en que están calculados los datos. Los valores posibles de esta dimensión son: Pesos, Dólares, Francos, Marcos, etc.

Cubo Información Analítica Gastos

Es un cubo de cinco dimensiones cuyo contenido son los datos financieros propiamente dichos.

Se encuentra dimensionado por : Versiones, Naturaleza del Gasto, Moneda , Organización y Tiempo.

Cubo Información Moneda

Es un cubo cuyo contenido representa los valores de conversión de la Moneda a lo largo del tiempo para los distintos presupuestos.

Está dimensionado por : Versiones, Tiempo y Moneda.

Cubo Información Analítica Realización

Es un cubo cuyo contenido representa el porcentaje de realización, donde

$$\%Realización = Valor Real / Valor Presupuestado.$$

Este ítem de dato financiero es de tipo fórmula, cada valor se obtiene de realizar el cálculo entre los datos financieros almacenados Real y Presupuestado. Estos datos no quedan almacenados permanentemente en la base sino que son calculados cada vez.

Está dimensionado por : Tiempo, Organización, Naturaleza del Gasto, y Moneda.

Reporting y análisis

Sobre los datos en la base de datos compartida los usuarios analistas podrán hacer distintos tipos de análisis, reportes y simulaciones. Worksheets, Reports y Gráficos son herramientas de OFA que permiten visualizar y analizar un slice de datos financieros desde diferentes perspectivas y preparar presentaciones efectivas basadas en ese análisis.

Utilizando estas herramientas podríamos realizar los siguientes estudios:

Comparación entre los valores Reales, Presupuestados y Forecast, lo cual permite analizar las desviaciones ocurridas.

Realizar comparaciones y estadísticas considerando el Presupuesto y Gastos Reales de años anteriores.

Emitir reportes de realización con respecto al Presupuesto utilizando el cubo Información Analítica Realización (de tipo fórmula).

Simulación de cambio de cotización: cuando se hicieron los cálculos del Presupuesto con una cierta cotización puede requerirse un recálculo utilizando una cotización diferente.

Variación del Presupuesto en un coeficiente especificado: analizar qué efecto tendría en los gastos un aumento del 2% en el departamento de Marketing.

Analizar sobre los valores presupuestados qué porcentaje del total de los gastos de la Compañía, son originados por cierta Gerencia.

Podrían generarse gráficos que muestren de un vistazo el porcentaje de los gastos de cada Gerencia en el total.

Lo mismo podría analizarse considerando naturaleza de gastos.

Averiguar dentro de una Gerencia cuál es el Departamento más costoso.

Controlar que los Presupuestos obtenidos para los Departamentos de una Gerencia no superen un determinado monto.

Los mismos análisis podrían hacerse sobre los gastos reales. Por ejemplo cuánto se gastó en Telefonía en la Gerencia de Recursos Humanos.



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.

Bibliografía

- ✓ *W.H. Immon, Building the Data Warehouse, John Wiley & Sons, Inc. New York*
- ✓ *Ralph Kimball, The Data Warehouse Toolkit, John Wiley & Sons, Inc. New York*
- ✓ *Vidette Poe, Building a Data Warehouse for decision Support, prentice Hall, New Jersey*
- ✓ *W.H. Immon, Claudia Imhoff, Greg Battas, Building the operational Data Store, John Wiley & Sons, Inc.,New York*
- ✓ *Datamation Magazine (1997)*
- ✓ *Olap In Data Warehousing Arbor Software*
- ✓ *Introducción Al Procesamiento Analítico On-Line (Olap) David Friend , Chairman- Pilot Software*
- ✓ *Personal Express User's Guide, Oracle*



BIBLIOTECA
FAC. DE INFORMÁTICA
U.N.L.P.

DONACION..... TES
\$..... 99/14
Fecha..... 3-10-05
Inv. E..... 2086

TES
99/14
DIF-02086
SALA



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE INFORMATICA
Biblioteca
50 y 120 La Plata
catálogo.info.unlp.edu.ar
biblioteca@info.unlp.edu.ar



DIF-02086