

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/257935435>

Tecnologías OLAP en la gestión de servicios de salud: su aplicación en el estudio del gasto en medicamentos.

Conference Paper · December 2001

CITATIONS

0

READS

499

4 authors:



Alan David March

10 PUBLICATIONS 41 CITATIONS

[SEE PROFILE](#)



Adrián Gomez

Hospital Italiano de Buenos Aires

92 PUBLICATIONS 326 CITATIONS

[SEE PROFILE](#)



Daniel Luna

Hospital Italiano de Buenos Aires

269 PUBLICATIONS 1,467 CITATIONS

[SEE PROFILE](#)



Fernán González Bernaldo de Quirós

Hospital Italiano de Buenos Aires

295 PUBLICATIONS 1,928 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Registro Institucional de Hiponatremia [View project](#)



Proyecto Itálica - Hospital Italiano de Buenos Aires [View project](#)

Tecnologías OLAP en la gestión de servicios de salud: su aplicación en el estudio del gasto en medicamentos.*

Dr. Alan D. March* **, Dr. Jorge L. Leguiza*, Ing. Eitel J.M. Lauría*, Lic. Adrian Gomez**, Dr. Fernán G. Bernaldo de Quirós**, Dr. Daniel Luna**.

*Facultad de Ciencias de la Administración, Universidad del Salvador, Buenos Aires, Argentina;

**Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

Introducción.

En su polémico libro "The Trouble with Computers", Landauer ha sostenido que las computadoras han fracasado como herramientas potenciadoras de la productividad¹. Su argumento central es que la informática ha sido exitosamente aplicada para la automatización de tareas repetitivas, rutinarias y bien definidas, en tanto que aún no han demostrado un desempeño similar en lo concerniente a su rol como herramientas para el soporte de la toma de decisiones. Si bien el caso de las bases de datos transaccionales (BDT) no fue específicamente mencionado por el citado autor, el problema planteado por estas se ajusta en forma sorprendente a sus argumentaciones. En la mayor parte de las organizaciones humanas, la tecnología de bases de datos relacionales ha sido una verdadera solución para aumentar la eficiencia de los procedimientos vinculados a transacciones que involucran datos referidos a una situación específica dentro de un circuito predeterminado de tareas. Actividades tales como la facturación, los procesos vinculados al mantenimiento de *stocks*, etc., han aumentado notablemente su eficiencia gracias a la tecnología informática. Sin embargo, es posible decir, siguiendo a Landauer, que todo el provecho potencial encerrado en esos grandes cúmulos de datos aun no ha sido explotado en toda su extensión. Si bien estos son adecuadamente utilizados en cuanto a su función transaccional, no puede decirse lo mismo de su uso como materia prima de análisis. En efecto, las BDT encierran una historia detallada de los procesos ejecutados por una organización, los cuales debidamente analizados permitirían reconocer patrones y tendencias de comportamiento sumamente útiles para las decisiones gerenciales vinculadas al planeamiento y la definición de estrategias. La aplicación de los datos transaccionales a este tipo de tareas indudablemente aumentaría el retorno sobre la inversión en tecnologías de la información, ya que los datos transaccionales serían "reutilizados" con fines que trascienden lo meramente operativo. Las razones por las cuales esto no se ha dado en ambientes corporativos se encuentra relacionado con la naturaleza relacional de las BDT utilizadas en la mayor parte de las organizaciones. Como se analizará en la sección siguiente, el tipo de estructura que impone el modelo relacional determina que estas bases de datos se encuentren poco adaptadas a los requerimientos de análisis, tornándose, como ha sostenido Zornes, en verdaderas "cárceles de datos" (*data jailhouses*)².

Las organizaciones dedicadas a la atención de la salud, como muchas de sus pares en otras áreas de la economía, asisten a un proceso de creciente informatización. La mayor parte de las aplicaciones aún se vinculan con procesos netamente administrativo-contables, pero el grado de informatización de datos estrictamente médicos es cada vez mayor³. Las BDT propias de las organizaciones médicas no escapan a los problemas que afectan a las organizaciones de los otros sectores, y los analistas se enfrentan a los mismos problemas de "encarcelamiento" de los datos.

Las nuevas tecnologías de *Data Warehousing*, *OLAP* y *Data Mining* representan una innovadora oportunidad para "liberar" los datos de las BDT y ponerlos a disposición de los encargados de tomar decisiones en las organizaciones. Se trate de tres tecnologías complementarias: el *Data Warehouse*, en el que se consolidan los datos operativos de la organización provenientes de múltiples fuentes en un único

El presente trabajo ha sido realizado en el marco del convenio de colaboración entre la Facultad de Ciencias de la Administración de la Universidad del Salvador e IBM Argentina S.A. Los autores desean agradecer a los Sres. Jorge Teodoro, Cristina Prat y Paola Rodríguez, de la citada empresa, por la colaboración prestada.

gran repositorio de información; la tecnología *OLAP* (procesamiento analítico on-line), que permite al usuario elaborar complejas consultas basadas en el análisis de la información desde la perspectiva de las múltiples dimensiones del negocio; y el *Data Mining* (minería de datos), permitiendo efectuar análisis exploratorios de los datos a efectos de reconocer tendencias y/o patrones ocultos de comportamiento⁴. El objetivo último es concretar el sueño del decision-maker: alinear los procesos de negocio con la estrategia corporativa y transformar la información generada por los mismos en una base de conocimiento que permita conducir con menor riesgo y mayores ventajas los destinos de la organización a su cargo (ver Figura 1).



Figura 1. Una base de conocimiento para pilotear la organización

A la fecha existen numerosas implementaciones en las áreas del comercio mayorista, servicios personales de distinta índole (transporte, supermercados, etc). Su aplicación en el área médica es aún minoritaria, pero ya existen algunos ejemplos exitosos. En lo que sigue se presenta un ejemplo de aplicación de OLAP al análisis del gasto en medicamentos, realizado en un hospital privado de la ciudad de Buenos Aires.

Las bases de datos estadísticas: antecedentes inmediatos de las tecnologías de OLAP y Data Warehousing.

Las motivaciones que han dado origen al OLAP y al Data Warehousing (DW) en modo alguno pueden ser consideradas recientes. Sus antecedentes históricos pueden encontrarse en las llamadas bases de datos estadísticas (BDE), que aún hoy siguen siendo utilizadas y que poseen muchos componentes comunes a sus pares de OLAP y DW⁵. Una breve descripción de las BDE brindará una introducción familiar para el usuario poco familiarizado con las nuevas tecnologías.

Desde el punto de vista de su objetivo final, todas estas tecnologías se basan en una motivación estadística, es decir, la intención de obtener medidas sumarias y efectuar predicciones en base a grandes volúmenes de datos. Resulta difícil decir que las modernas herramientas OLAP son herramientas estadísticas, ya que muchas de ellas incorporan funciones que no pueden ser debidamente consideradas como parte de la estadística clásica. Tal el caso de la lógica difusa o las redes neuronales. Por ello hemos optado por la denominación "motivación estadística". En todos los casos, el objetivo de la sumaria o consolidación de los datos se realiza en base a lo que se ha dado en llamar una *función sumaria* o *función de resumen*. Dicha función es fácilmente comprensible por parte del usuario de software estadístico, quien característicamente efectúa extensas operaciones de suma sobre grandes cantidades de registros individuales para luego determinar medidas de tendencia central (media, moda o mediana) y de dispersión (varianza, desvío estándar). Estas funciones sumarias se ejecutan sobre un rasgo particular de una muestra, que se denomina *medida sumaria* o *medida de resumen*, también comunes a todas estas tecnologías. Son ejemplos de medidas de resumen las ventas totales de una organización, el gasto total de una empresa, la cantidad de afectados por una determinada enfermedad, etc. Estas medidas son alcanzadas mediante la ejecución de la función sumaria antes mencionada, típicamente una suma en los ejemplos anteriores.

Una vez determinada la medida sumaria, el analista se interesa por determinar la distribución de esta a través de diferentes rasgos de la población bajo análisis. Estos rasgos o dimensiones son una vez mas una

característica común y son de especial relevancia en las tecnologías de OLAP y DW. Son ejemplos de dimensiones, tomando como base el ejemplo de las ventas de una empresa citado mas arriba, las filiales, departamentos, regiones geográficas, características demográficas de los clientes, etc. Estas dimensiones permiten desglosar el volumen total de ventas según el rasgo seleccionado. Una característica adicional de las dimensiones es la posibilidad de su agrupación jerárquica, según la dependencia conceptual entre las mismas. Así, "regiones geográficas", "filiales" y "departamentos" son pasos sucesivos de una desagregación cada vez mayor basada en un concepto espacial.

Estas grandes similitudes entre las BDE y las herramientas OLAP y de DW, permiten una comprensión global de las segundas partiendo de las características comunes.

Las limitaciones de las BDT y el advenimiento de las tecnologías OLAP y Data Warehousing.

OLAP y DW son dos tecnologías que comparten una gran cantidad de similitudes. Las razones básicas de la distinción entre ambas son sus orígenes como tecnologías independientes, y el mayor énfasis de la primera sobre los aspectos del análisis de los datos y de la segunda sobre las estructuras en las cuales los mismos son almacenados. Los rasgos que en conjunto permiten diferenciarlas de las BDE se relacionan con ciertas limitaciones que representan un problema de larga data para sus usuarios. La mayor parte de los análisis estadísticos se realizan en una de dos formas: prospectiva y retrospectiva. En los estudios prospectivos, se diseña una hipótesis básica y se sale a recabar los datos *de novo*, mediante el diseño de encuestas y formularios *ad-hoc*. Este es una característica saliente de las investigaciones científicas, donde se pretende tener un control muy estricto sobre la calidad de los datos*. No obstante lo expuesto, esta estrategia no siempre puede cumplirse, ya sea porque un programa para obtener datos prospectivos es caro y laborioso, o porque la urgencia de contar con la información no permite el diseño y ejecución de un estudio prospectivo. Finalmente, debe considerarse la conveniencia de utilizar datos históricos para efectuar proyecciones cuando su calidad es lo suficientemente confiable.

Uno de los grandes impedimentos para la ejecución de este tipo de análisis, que ha significado un obstáculo tanto para estadísticos e investigadores como para gerentes y administradores, es la manera en que estos datos son usualmente almacenados. Entre los numerosos diseños propuestos para el armado de bases de datos, el modelo relacional es el que ha prevalecido en el mercado. En parte por sus ventajas intrínsecas para el soporte de transacciones individuales sobre registros únicos, y en parte por el eco obtenido en el mercado como un estándar *de facto*, la mayor parte de las BDT utilizan este modelo y allí es donde se encuentran grandes volúmenes de datos históricos propios de las organizaciones. Las limitaciones que esta tecnología plantea a los analistas son de varios tipos, todas ellas fundadas en el tipo de fragmentación de los datos que el modelo relacional impone a lo que se considera un registro individual. Las mismas pueden ser divididas en problemas de recuperación y problemas de visualización.

El problema de la recuperación.

Desde la propia concepción del modelo relacional por Codd⁶, el objetivo primordial de las BDT ha sido la optimización de la recuperación de registros individuales o de grupos pequeños de los mismos. Una típica transacción sobre una BDT es la localización de un registro en base a un dato suministrado por un usuario, por comparación de ese dato frente a los miles a millones de candidatos que existen. Para ello las BDT son optimizadas mediante índices que imponen ordenamientos particulares a los datos o generan ordenamientos separados del dato en cuestión. Similarmente, la introducción de un nuevo dato significa situarlo en el lugar correspondiente al índice correspondiente y a menudo involucra la reconstrucción del indexado de la base de datos. Dado que los datos de los cuales parte una consulta a una base de datos son generalmente bien conocidos (por ejemplo, un apellido para localizar a una persona o un número para localizar una factura) y que un indexado excesivo de una base de datos torna lentos a los procesos de actualización, los diseñadores

En las ciencias médicas, por ejemplo, suele distinguirse entre los estudios prospectivos y retrospectivos en cuanto a su calidad, ya que los primeros suelen considerarse mejores que los segundos. La diferencia radica en que los estudios retrospectivos se realizan tomando datos que han sido recogidos con un propósito diferente al del análisis (por ejemplo, anotaciones en una historia clínica en el curso de un tratamiento médico. Se parte del supuesto, generalmente correcto, de que como la recopilación de datos original no se realizó con fines de investigación, pueden no haberse efectuados controles de calidad sobre los mismos.

restringen los índices a un conjunto limitado de datos (campos) de una base de datos. Así, en los ejemplos antes citados, es poco probable que un domicilio sea indexado para los fines de localización de una persona o de una factura. La búsqueda de un registro en base a un dato no indexado puede no ser requerida en las transacciones cotidianas de una organización. Sin embargo, un analista puede tener ese requerimiento (por ejemplo, localizar a todas las personas que viven en una determinada localidad). El problema aquí planteado será el de la lentitud intrínseca que significa la búsqueda de un dato no indexado. Asimismo, si el analista desea buscar datos dentro de un determinado rango (por ejemplo, edades de las personas), el problema puede empeorar al tener que efectuarse un cálculo comparativo, o, como suele suceder con el caso del ejemplo en el cual el dato "edad" no suele asentarse como dato en si mismo sino que es el producto de un cálculo realizado sobre la fecha de nacimiento.

Un segundo problema que plantea el modelo relacional tiene que ver con la fragmentación de los datos para evitar las redundancias. Como es bien sabido, muchos registros vinculados a un solo objeto conceptual implican redundancia. Así, el encabezado de una cuenta corriente (datos demográficos de la persona o la empresa) se relaciona lógicamente con cada asiento realizado en dicha cuenta. De no mediar el modelo relacional, cada asiento en una cuenta corriente debería acompañarse de la totalidad de los datos de encabezado si todos los datos deben ser recuperados en el futuro. El modelo relacional propone la separación entre los datos repetitivos y los datos variables a través de la construcción de tablas lógicamente vinculadas. Esta separación persigue fines de optimización y en un típico diseño de BDT aquello que constituye un solo registro desde el punto de vista lógico en realidad se transforma en diferentes cantidades de registros en un número a veces grande de tablas separadas. Las uniones lógicas entre dichas tablas requieren de índices para ser ejecutadas en forma óptima.

La complicación central de la recuperación de datos es la performance, tanto la experimentada por el analista como el impacto sobre los recursos de hardware. Las bases de datos típicas de las organizaciones modernas se encuentran implementadas sobre un servidor o grupo de servidores limitado y es utilizada por una gran cantidad de usuarios. Una consulta compleja absorberá una gran cantidad de los recursos de hardware del procesador si la misma trabaja sobre tablas no indexadas por no haberse previsto dicha necesidad. El analista observará tiempos de ejecución muy prolongados y los demás usuarios se verán penalizados por el exceso de recursos utilizado por un solo usuario. En algunos casos la complejidad de la consulta podría saturar la capacidad del procesador y provocar la caída (standstill) del sistema completo. Por estas razones es bastante típico que los administradores de las BDT limiten el tipo de consultas permitidas a los usuarios, provocando esa caracterización ya mencionada de "cárceles de datos".

El problema de la visualización.

Como ya se ha mencionado al considerar las BDE, las dimensiones de análisis que una medida sumaria admite son numerosas. Estas dimensiones son habitualmente combinadas por el analista en busca de datos específicos. Las herramientas de software habitualmente utilizadas en estas tareas son típicas matrices bidimensionales (por ejemplo, planillas de cálculo). Dichas herramientas resuelven perfectamente los "entrecruzamientos" bidimensionales mediante la metáfora de la fila y la columna. A la hora de utilizar mas dimensiones, la visualización se torna más compleja. La solución de compromiso por la cual se opta generalmente es el agregado secuencial de filas y columnas de encabezado en las típicas presentaciones bidimensionales, algo frecuentemente observado en presentaciones estadísticas. En estos casos, una dimensión jerárquicamente superior abarca numerosas dimensiones jerárquicamente subordinadas. En tanto esta es una solución adecuada para un número limitado de dimensiones, la exposición de un número considerable de las mismas implicaría la producción de matrices bidimensionales demasiado extensas como para ser fácilmente visualizables. Desde el área de las planillas de cálculo se ha suministrado una solución parcial a través de las tablas dinámicas, pero estas suelen presentar limitaciones para la carga de datos.

OLAP y Data Warehousing: aspectos principales.

Como ya se ha mencionado anteriormente, OLAP Y DW comparten características. Las razones fundamentales de la distinción entre ambas tiene motivos históricos y de énfasis. En cuanto a la primera razón, solo cabe destacar sus comienzos como emprendimientos independientes apuntando a la solución de

los mismo problemas arriba citados. A los fines de una mejor exposición, serán descriptas independientemente y luego se ofrecerá una integración.

Data Warehousing.

Un *data warehouse* (almacén de datos) podría definirse como "una colección no volátil de datos integrados, temáticamente orientada y creciente con el tiempo, cuyo propósito principal es la provisión de información para la toma de decisiones organizacionales"⁷. Del análisis de esta definición pueden extraerse sus características principales. Su condición de "colección de datos" constituye la característica vinculante con las BDT: ambas coleccionan datos y solo difieren en la manera en que dichos datos son obtenidos y almacenados. Las BDT transaccionales obtienen sus datos a partir del proceso de *data-entry* por parte de los usuarios del sistema en su trabajo cotidiano. Un DW obtiene sus datos de las BDT a través de procesos que serán descritos mas adelante. El atributo de la "no-volatilidad" refiere a un proceso que a menudo diferencia las BDT de los DW. En el proceso de mantenimiento de las primeras, es bastante común que determinados conjuntos de datos sean retirados de las mismas por no ser requeridos en las transacciones cotidianas. Así, por ejemplo, un archivo maestro de personas que constituye el padrón de afiliados de un sistema de medicina prepaga es a menudo depurado de los datos atinentes a personas que han cesado su afiliación. Si bien es posible "marcar" el registro de una persona en dicha base de datos de manera tal de señalar su condición de inactivo, en el largo plazo la recuperación del registro de una persona activa (como puede ser necesario para generar una factura) se tornaría lenta dada la necesidad de recorrer registros de personas "inactivas" cuyo número necesariamente aumenta durante el ciclo de vida de una organización en la cual los clientes constantemente entran y salen. Dicha estrategia, si bien aumenta la velocidad de recuperación de registros activos, inhabilita el análisis de las tendencias históricas del negocio por eliminar a quienes ya no forman parte de el. Las penalidades de un análisis sobre estas BDT "depuradas" son claras: en su expresión mas sencilla, la organización no podría conocer las razones por las cuales determinados clientes han dejado de tomar sus servicios por haberlos eliminado totalmente del sistema. Si en el proceso de depuración estos datos se han eliminado la tarea de análisis será imposible. Si han sido almacenados en una base de datos "histórica", el análisis solo será posible mediante la fusión (*merging*) de los dos padrones (el de "bajas" y el de "activos"). Aún así, si no se han tomado los recaudos debidos, aún podrían surgir problemas de consistencia en el caso de personas que han cancelado su afiliación en el pasado para luego volver a afiliarse y revestir una condición de "activo" al momento del análisis. Un DW almacenará la totalidad de la información histórica de una organización e implementará estrategias para controlar cambios históricos que puedan "contaminar" la calidad de los datos. Así, por ejemplo, en el caso de re-afiliaciones, una BDT solo reconocerá a una persona en su calidad de activo o inactivo. En cambio en un DW, la historia de afiliaciones y desafiliaciones de una persona determinada deberá ser adecuadamente resuelta al momento de diseño, y la condición de afiliación no solo "marcada" como tal sino registrada históricamente en sus sucesivas condiciones de "actividad" o "inactividad". Del ejemplo anterior se deduce que desde el punto de vista estructural, un DW no será un sencillo espejo de la BDT en la cual se originan sus datos. Por ello, un componente esencial de estas tecnologías son las "herramientas de extracción y transformación" que permiten trasladar datos desde una BDT hacia un DW.

La "orientación temática" que distingue a un DW merece un examen mas detenido. Toda colección de datos es organizada sobre un principio rector. Como ya se ha mencionado anteriormente, uno de los principales criterios que inspiran una BDT es la performance transaccional. Esto a menudo llega a la fragmentación de datos que en una situación analítica podría no ser deseable. En el caso del archivo maestro de personas del ejemplo anterior, es perfectamente posible que el diseñador de una BDT separe en dos tablas los atributos de esa persona que fueran considerados "fijos" y "variables" o "únicos" y "múltiples". Tal el caso del sexo o la fecha de nacimiento para el primer elemento de ambos pares, y el domicilio para el caso del segundo. En tanto una persona posee una sola fecha de nacimiento posible, su domicilio puede cambiar o puede poseer varios domicilios (personal, laboral, postal, de facturación, etc.). El diseñador de una BDT apelará a la tercera forma normal de Codd y establecerá una relación 1→N entre las tablas de "personas" y "domicilios", en tanto que el responsable de un DW probablemente prefiera una sola "tabla" que contenga un registro completo de los datos de la persona aun cuando incurra en redundancias. En el caso de la BDT se ha optado por la performance. En el caso del DW se ha optado por el "tema": los diversos lugares geográficos en donde puede localizarse a una persona

La condición de "crecimiento con el tiempo" también requiere de explicación. Si bien es cierto que tanto una BDT como un DW crecen en la medida que se agregan datos, la primera puede ser sometida a procesos de depuración de datos no considerados críticos (las personas "inactivas" del archivo maestro de personas). Es por ello posible decir que, en tanto ambas colecciones pueden crecer con el tiempo, un DW siempre lo hace. Una BDT hasta puede decrecer con el tiempo, ya que en el caso del archivo maestro de personas el volumen puede disminuir si a partir de un cierto punto en el tiempo las bajas de afiliados superan en número a las altas. Asimismo, la redundancia señalada en el párrafo anterior es un motivo adicional para las diferencias de volumen a favor de un DW. Lo expuesto no es una simple consideración teórica, dado que en ambientes que cuentan con DW plenamente instalados, las relaciones de volumen con las BDT son de decenas de terabytes a cientos de megabytes.

Un DW constituye el repositorio de datos sobre los cuales se ejecutan las consultas orientadas a obtener información para la toma de decisiones. La modelización de un DW es a menudo compleja. Esta condición es doblemente impuesta por su condición de repositorio histórico (que impone requerimientos de representación del tiempo a menudo complejos) y por la variedad de "temas" alrededor de los cuales pueden organizarse colecciones de datos. La dificultad de elaborar estos modelos para la totalidad de una organización ha determinado la producción de expresiones mas pequeñas de DW conocidas como *data marts* (DM). Estos no son mas que DW mas restringidos y por ello menos voluminosos. Desde el punto de vista estructural no difieren de un DW. Su única diferencia radica en que constituyen modelos parciales de una organización.

El principio fundamental de organización de los datos en un DW es la *multidimensionalidad*. La característica estructural principal de los datos almacenados multidimensionalmente es la generación de grandes agrupamientos parciales que generalmente comprenden una serie de dimensiones vinculadas jerárquicamente. Dichas estructuras representan un enfoque radicalmente diferente al del modelo relacional por el elevado grado de redundancia de datos que se encuentra en un DW.

Existen dos tipos de tecnologías para la construcción de un DW. Sus nombres representan una demostración del grado de solapamiento que existe actualmente entre DW y OLAP. Por un lado, existen DW basados en sistemas de base de datos relacionales tradicionales, que implementan extensiones al lenguaje de consulta que permiten efectuar las típicas funciones de OLAP. Estas implementaciones reciben el nombre de ROLAP, donde la "R" significa "relacional". El modelo alternativo almacena los datos en forma nativamente multidimensional (por ejemplo, como *arrays*) y recibe el nombre de MOLAP, significando la "M" del acrónimo "multidimensional".

Como resumen de lo hasta aquí expuesto, las características fundamentales de un DW pueden definirse, en comparación con las BDT, conforme a lo expuesto en la Tabla 1.

Rasgos de los datos	BDT	DW
Detalle	detallados	resumidos
Temporalidad	actuales	series temporales
Uso	administrativo	gerencial
Requerimientos informativos	conocidos y previstos	imperfectamente conocidos
Organización general	orientada al programa aplicativo	orientada al tema
Redundancia	no redundantes	redundantes

Tabla 1. Comparación entre los datos de una BDT y un DW (Tomado, con modificaciones, de Thomsen)

Online Analytical Processing (OLAP).

Introducido por Codd⁸, el OLAP puede ser definido como una arquitectura disponible para proveer a los usuarios la habilidad de realizar análisis dinámicos de datos. Una vez más queda demostrado el solapamiento entre ambas tecnologías por la existencia de productos de DW que proveen "servicios OLAP".

Utilizando herramientas OLAP, los usuarios pueden acceder al DW, brindando a los responsables de la toma de decisiones en las organizaciones el potencial de mejorar su comprensión del negocio y los cambios que lo afectan, de incrementar su habilidad para identificar o generar soluciones posibles a problemas de decisión, y de efectuar oportunamente formulaciones tácticas o estratégicas alineadas con los objetivos de la organización. Esta característica de analizar y sintetizar información a partir del OLAP surge de la elaboración de múltiples escenarios especulativos que contestan preguntas tales como “qué pasaría si ..” (“what if”) o “por qué”. En estos verdaderos modelos de simulación, y a partir de la modificación de ciertas variables clave, se analiza el comportamiento del resto de las variables⁹.

En 1994 Codd y Codd¹⁰ introdujeron 12 reglas sobre el modelo OLAP, pensadas para eliminar interpretaciones erróneas sobre las características de esta tecnología. En forma resumida, las 12 reglas de Codd para los sistemas OLAP son las siguientes:

1. *Vistas multidimensionales.* Manejo y organización conceptual y física de la información en forma multidimensional.
2. *Transparencia.* Capacidad para acceder a datos de otras fuentes (por ejemplo planillas de cálculo), de manera sencilla y transparente.
3. *Accesibilidad.* Habilidad para obtener información completa y estructurada de fuentes externas de datos tales como bases de datos relacionales, archivos planos, etc.
4. *Desempeño y consistencia.* El número de dimensiones utilizadas en el sistema no debe degradar el desempeño del sistema, ni tampoco afectar la consistencia de la información.
5. *Cliente/servidor.* Las herramientas deben poder operar en ambientes cliente/servidor.
6. *Dimensionalidad genérica.* Cada dimensión deberá ser tratada de igual manera.
7. *Uso eficiente del almacenamiento.* Manejo eficiente de la porosidad¹¹ (sparseness) de la base multidimensional, para ocupar la mínima cantidad de espacio. Por "porosidad" se entiende la manera en la que herramienta maneja el espacio requerido para almacenar la información multidimensional; este punto es muy importante ya que, debido a la estructura de los datos en las bases multidimensionales se cuenta con muchos "celdas" o campos vacíos. Un buen manejo de la porosidad implica que la herramienta es capaz de detectar las celdas vacías, y hacer eficiente el espacio que éstos requieren.
8. *Soporte a múltiples usuarios.* Permitir el acceso de múltiples usuarios al mismo tiempo al mismo modelo.
9. *Operaciones entre dimensiones sin límite.* Capacidad para realizar operaciones entre varias dimensiones sin ningún tipo de restricción.
10. *Manipulación intuitiva de datos.* Capacidad de navegación a través de los datos, dimensiones y jerarquías de la base mediante una interfaz de usuario fácil de usar.
11. *Producción flexible de reportes.* Utilitarios para la creación rápida de reportes, consultas y gráficos.
12. *Capacidad ilimitada para dimensiones y relaciones (jerarquías).* Capacidad para manejar un número ilimitado de jerarquías, relaciones y dimensiones de los datos.

El concepto de OLAP es poderoso. Sin embargo, su aplicación práctica a través de distintas herramientas comerciales ha dado lugar a implementaciones que cubren parcial o totalmente el enunciado de Codd. Las empresas desarrolladoras de software han, inclusive, agregado “reglas” adicionales que permiten mejorar el rendimiento de los procesos analíticos de consulta y actualización¹².

Estudiando las características de OLAP se concluye en primera instancia que el modelo MOLAP está mejor preparado que la alternativa ROLAP para desplegar perspectivas multidimensionales con un rendimiento adecuado. De hecho, el modelo MOLAP presenta ventajas de performance cuando se trata de procesar juegos de datos densos –es decir, estructuras poco porosas-. Codd, al introducir el concepto de OLAP, enfatizó el hecho de que los sistemas relacionales carecían de la versatilidad de sus competidores multidimensionales en términos de funcionalidad analítica. Sin embargo, desde entonces han aparecido múltiples defensores de la tecnología ROLAP, y este enfoque es naturalmente defendido por los fabricantes de bases de datos relacionales¹³. Las ventajas y desventajas de utilizar alguno de estos dos modelos están ampliamente documentadas en la bibliografía especializada y quedan fuera del alcance de este trabajo^{14 15}. De todos modos, el modelo a utilizar es transparente para el decision-maker. Lo que este último necesita y que la tecnología OLAP debe proveer se resume en lo siguiente:

1. Acceso a la información residiendo en el DW
2. Estructuración de la información en categorías de datos organizadas en jerarquías de consolidación (dimensiones), definidas en base a los requerimientos del usuario.
3. Disponibilidad de una interfaz adecuada para visualizar la información desde múltiples perspectivas de análisis.

Visualización multidimensional

La posibilidad de visualizar la información desde la perspectiva de múltiples dimensiones de análisis es una característica clave de la tecnología OLAP. En general una dimensión es una categoría de datos (por ejemplo cuenta, tiempo, región). Esta categoría posee una o más valores característicos que son los miembros de la dimensión (ver figura 2: en la dimensión región, “Argentina”, “Bolivia”, etc).

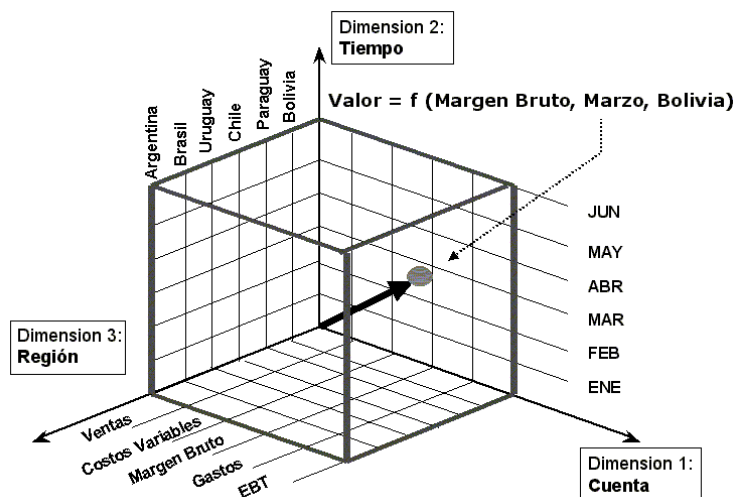


Figura 2. Las dimensiones de un modelo OLAP basado en información contable

En términos geométricos podemos caracterizarlos como cada uno de los valores en cada una de las aristas del cubo, asimilando cada arista o eje del sistema de coordenadas a una dimensión de análisis y el “cubo” al modelo de datos OLAP en si mismo. Así un valor almacenado en una celda del cubo corresponde a un punto cuya ubicación es identificada por los valores de los ejes coordenados (Margen Bruto en Bolivia y en Marzo, para el ejemplo de la figura 2). La analogía es válida para comprender el concepto desde un punto de vista gráfico, pero introduce una simplificación insalvable pues limita el modelo a tres dimensiones de análisis. En la práctica el número de dimensiones de un modelo OLAP puede ser considerablemente superior a tres. La extrapolación del modelo geométrico a más de tres dimensiones da lugar a la denominación de “hipercubo” entendiéndose, por tal al espacio de nodos definidos por el sistema de coordenadas n-dimensionales.

Dimensión Tiempo

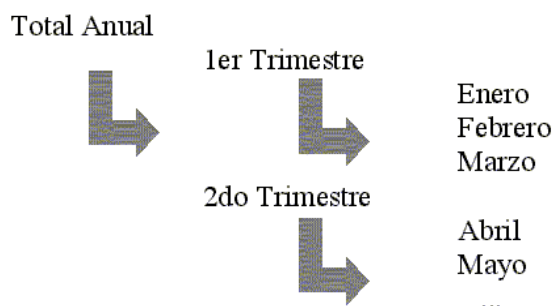


Figura 3. Estructuras jerárquicas dimensionales

La condición intrínsecamente jerárquica de la estructura de las dimensiones nos aleja aún más de la simplificación del modelo geométrico. En efecto, otra de las virtudes de la tecnología OLAP consiste no solamente en poder analizar la información desde múltiples dimensiones, sino disponer en las mismas de estructuras de agregación o consolidación que permitan totalizar los resultados del análisis en forma parcial o absoluta. Así por ejemplo la dimensión *tiempo* podrá contener como miembros a cada uno de los meses del año y su vez contar con niveles de agregación trimestrales y anual (ver figura 3).

Es importante destacar en el ejemplo de la figura 2 la existencia de una dimensión especial (la dimensión cuenta) que contiene los miembros (variables de medida) sobre los que se efectúa la medición y el análisis de datos. En efecto, dicha dimensión de medidas esta siempre presente en todo modelo de datos OLAP y es quien le da sentido al análisis multidimensional pues contiene al conjunto de variables que son objeto de análisis desde múltiples perspectivas.

Podemos formalizar un ejemplo de visualización multidimensional a través de una planilla de cálculo. Supongamos para ello que un usuario necesita analizar información contable a través del tiempo en distintas regiones. Retrotrayéndonos al ejemplo de la figura 2, las dimensiones a considerar son *cuenta*, *tiempo* y *región*. En este caso una vista multidimensional desplegada en una planilla de cálculo para analizar las cuentas en un grupo de países para un mes determinado tendrá la forma:

Enero				
	<i>Argentina</i>	<i>Brasil</i>	<i>Uruguay</i>	...
<i>Ventas</i>				
<i>Costos Variables</i>				
<i>Margen Bruto</i>				
...				

Figura 4. Vista tridimensional

El agregado de una cuarta dimensión *escenario* que permita comparar la información real (*Real*) contra los valores presupuestados (*Plan*) da lugar a una vista con el siguiente formato:

Enero	Argentina		Brasil		Uruguay		
	Real	Plan	Real	Plan	Real	Plan	Real	Plan
Ventas								
Costos Variables								
Margen Bruto								
...								

Figura 5. Vista cuatridimensional

Es importante comprender que, en términos de la tecnología OLAP, estas vistas son configurables por el usuario. Las dimensiones no se encuentran estáticamente dispuestas en filas o columnas. El usuario puede modificar la forma y el nivel de agregación de los datos desplegados apelando a las funciones OLAP *drill-down*, *roll-up*, *slice*, *dice* y *nesting*. Así, podrá producir una vista que totalice las cuentas anualmente en lugar de considerar solamente el mes de Enero, o bien desplegar los resultados de las ventas en cada país y para cada mes del año.

Drill-down y Roll-up:

Como ya se ha mencionado anteriormente, los datos que componen una dimensión pueden ser desglosados conforme a una subdivisión en categorías sucesivas organizadas jerárquicamente. Los procesos de agregación o desagregación de datos para cada una de las categorías de la jerarquía de una dimensión se conocen, respectivamente, como *roll-up* y *drill-down*. El *drill-down* lleva a un nivel mayor de detalle, en tanto que el *roll-up* cumple la función opuesta. Constituyen el equivalente conceptual de las funciones gráficas de *zoom-in* y *zoom-out*. *Roll-up*, consolidación y agregación son sinónimos para la terminología OLAP¹⁶. En el caso de la figura 5, la aplicación de la función *roll-up* sobre la dimensión *tiempo* generaría una vista con resultados totalizados anualmente. Aplicando *drill-down* sobre este resultado podríamos obtener la misma vista pero con el detalle del mes seleccionado.

Slicing, dicing y nesting

Son operaciones utilizadas para visualizar la información en base a una configuración dimensional definida por el usuario. La operación *slice* implica “cortar una rebanada” bidimensional del hipercubo. Así, por ejemplo, la figura 4 despliega las cuentas de las distintas regiones haciendo un corte (*slice*) para el mes de Enero. *Dice* (en inglés, lanzar los dados) es la rotación del hipercubo a efectos de desplegar un nuevo *slice* o corte bidimensional. Por ejemplo en la figura 6, la función *dice* genera un nuevo corte que despliega las ventas presupuestadas, para todas las regiones y meses del año.

Ventas				
Plan				
	Enero	Febrero	Marzo	...
Argentina				
Uruguay				
Brasil				
...				

Figura 6. Nuevo *slice* generado a partir de una función *dice*

Nesting (anidado) es la acción que permite desplegar una dimensión desde adentro (es decir para cada uno de los miembros) de otra dimensión. Aplicando *nesting* sobre la dimensión *escenario* en la figura 6, se obtiene la figura 7, en la que la dimensión *escenario* queda anidada dentro de la dimensión *tiempo*.

Ventas	Enero		Febrero		Marzo		
	Real	Plan	Real	Plan	Real	Plan	Real	Plan
Argentina								
Uruguay								
Brasil								
...								

Figura 7. Nuevo *slice* generado a partir de una función *nesting*

Es importante considerar que en el estado actual del arte en materia de desarrollo de interfaces gráficas, estas operaciones no necesitan ser realizadas a través de complejos comandos. El usuario selecciona la(s) dimension(es) afectada(s) y la(s) desplaza, la(s) rota o la(s) re-configura, produciendo implícitamente la ejecución de las funciones antes mencionadas.

Los requerimientos analíticos en las organizaciones médicas.

La atención de la salud constituye una actividad económica de gran relevancia en el contexto de las economías nacionales. En la Argentina el gasto en salud representa un 7.9% del PBI, en tanto que en economías centrales como la de EEUU alcanza el 14%. De particular importancia es la tendencia constantemente ascendente exhibida por este gasto, lo cual ha llevado a numerosos analistas a señalar la importancia de explorar estrategias para controlar dicho gasto a fin de evitar la quiebra del sistema. Inspirados en lo que ocurre en otros sectores de la actividad económica, desde hace mas de una década los administradores de las organizaciones proveedoras de servicios de salud han vuelto su mirada sobre la información histórica de sus instituciones, con el fin de efectuar evaluaciones económicas de sus actividades. Este tipo de análisis invariablemente presupone como condición inicial contar con un estudio descriptivo de los costos, lo cual habilita la posterior evaluación económica integral, fundamentalmente basada en análisis de costo-efectividad, costo-beneficio y costo-utilidad¹⁷.

La determinación precisa de los costos en la industria de servicios resulta en general mas dificultosa en comparación con lo que sucede en la industria manufacturera. Esta situación es particularmente crítica en los servicios de salud, por motivos que van desde la dificultad para definir el concepto de producto hasta la enorme cantidad y variedad de factores que involucran procesos tales como una internación hospitalaria. Debido a esto, las técnicas de costeo de alta precisión tales como el microcosteo no han encontrado gran eco en las instituciones de salud y en su lugar se utilizan métodos menos precisos tales como los grupos de case-mix, o los costos por período, ya sea generales o basados en enfermedades específicas.

No obstante lo expuesto, la creciente informatización de las organizaciones de salud dirigida a la automatización de los procesos administrativos ha generado un subproducto de grandes bases de datos transaccionales (BDT) que encierran el detalle de la composición de numerosos procesos de atención. Dado que la mayor parte de estos sistemas se han empleado para propósitos administrativo-contables tales como la facturación o el control de stock, el potencial informativo de estas BDT para estudios de costos es considerable.

En la Argentina, el mercado médico suma al contexto inflacionario global del costo de la salud las condiciones locales de una economía recesiva, cuyo efecto mas directo es la progresiva migración de personas desde la medicina privada a la medicina pública. Esto impone importantes penalidades tanto al sistema de medicina privada como a su par estatal y público. En el caso del primero, el efecto mas directo es la disminución del volumen del negocio. Este hecho es particularmente grave para las firmas ya que el modelo de negocios mas extendido es el del seguro, donde los individuos aportan una prima dirigida a cubrir los riesgos de salud y una disminución del número de cápitats atenta contra la posibilidad de

licuación del riesgo basada en el volumen¹⁸. El impacto sobre el sistema público de salud es efecto directo de la traslación de gran número de personas desde el sistema privado a este, que por las condiciones recesivas y los esfuerzos de contención del gasto social que realiza el estado en la actualidad, no se encuentra preparado para enfrentar el aumento de la demanda de servicios..

Por las razones arriba expuestas, diversos analistas han coincidido en señalar que la clave para la gestión exitosa de organizaciones de atención de la salud es el manejo de la información, solo posible mediante la implantación de sistema informáticos adecuados. Desde el punto de vista transaccional, es posible decir que dichos sistemas ya se encuentran emplazados en mayor o menor medida. Sin embargo, la transformación de la información sobre procesos en materia prima para el análisis constituye una deuda pendiente. Las modernas tecnologías de Data Warehousing y OLAP, por la propia filosofía que las anima, constituyen la herramienta habilitante fundamental para lograr el objetivo analítico.

Material y métodos:

Se trabajó sobre la base de datos transaccional del Plan de Salud (PS) del Hospital Italiano de Buenos Aires (HIBA). El PS es un sistema de atención médica prepaga que cuenta con un padrón de afiliados de aproximadamente 80.000 personas, la mayor parte de las cuales son atendidas en el HIBA o en uno de sus centros de atención ambulatoria periféricos. Estos últimos se encuentran localizados en el área suburbana de la provincia de Buenos Aires. Se elaboró un datamart a partir de 980.000 registros correspondientes a los consumos de fármacos por parte de pacientes ambulatorios afiliados a los diferentes planes de cobertura del Plan de Salud, entre el 1° de Enero de 1998 y el 1° de Octubre de 1999. Los datos utilizados fueron los siguientes:

- Identificación del afiliado:
 - Número de afiliación;
 - Fecha de nacimiento;
 - Sexo;
 - Fecha de ingreso al sistema prepago
- Identificación del evento de consumo:
 - Fecha de compra del medicamento;
 - Numero de matrícula del profesional que prescribió el medicamento;
 - Plan de cobertura del afiliado al momento del evento de consumo;
 - Fecha de ingreso al sistema prepago;
 - Tiempo de permanencia en el plan al momento del consumo;
 - Tiempo transcurrido entre el consumo y la baja cuando esta se produjo dentro del período estudiado
- Identificación del medicamento consumido:
 - Nombre del medicamento;
 - Costo del medicamento

El análisis multidimensional se realizó utilizando las siguientes herramientas:

- como servidor OLAP, IBM DB2 OLAP Server 1.1, con Hyperion Application Manager como interfaz de administración y diseño, sobre un servidor IBM Netfinity 3000 (128 MB RAM, disco rígido de 9 GB) utilizando Windows NT Server 4.0 como sistema operativo. El cubo fue construido sobre un modelo ROLAP con DB2 versión 5 como back-end relacional;
- como cliente OLAP, empleamos Cognos Powerplay 6.0 para generar reportes on-line, Powerplay Connect 6.0 para acceder al servidor OLAP desde el cliente, y Hyperion Application Manager como interfase de diseño y administración remota, sobre una PC Compaq Deskpro (RAM 64 MB, disco rígido de 6 GB) utilizando Windows NT Workstation 4.0 como sistema operativo.

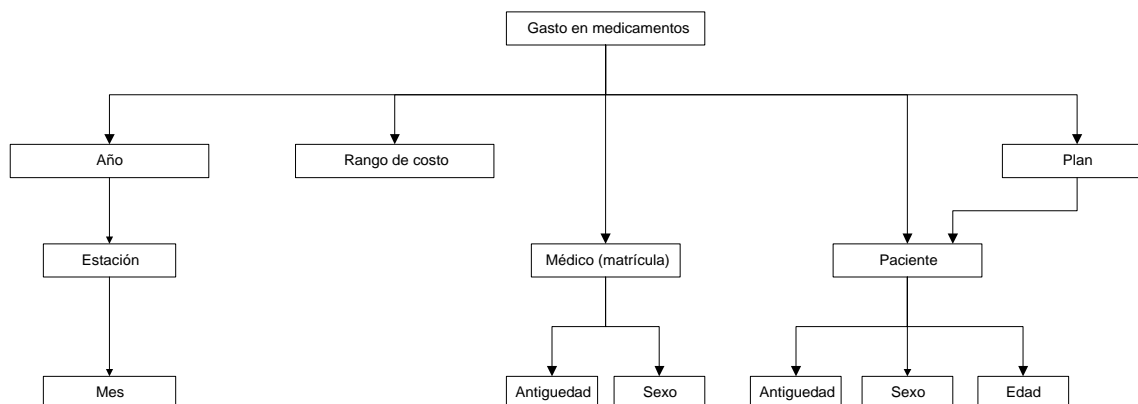


Figura 8. Grafo de dimensiones

El modelo de datos utilizado para la construcción del cubo utilizó como función la adición del precio de los medicamentos consumidos durante el período estudiado, obteniéndose como medida sumaria el gasto total en medicamentos. Las dimensiones analizadas se ilustran como grafo en la Figura 8. La dimensión temporal utilizó la jerarquía Año>Estación>Mes. Se optó por el agregado trimestral basado en estaciones para permitir la detección de los patrones de consumo asociados a las enfermedades de incidencia estacional. La dimensión de rango de costos de los medicamentos se utilizó para estudiar la incidencia diferencial de medicamentos de costo alto, medio y bajo en el consumo total.

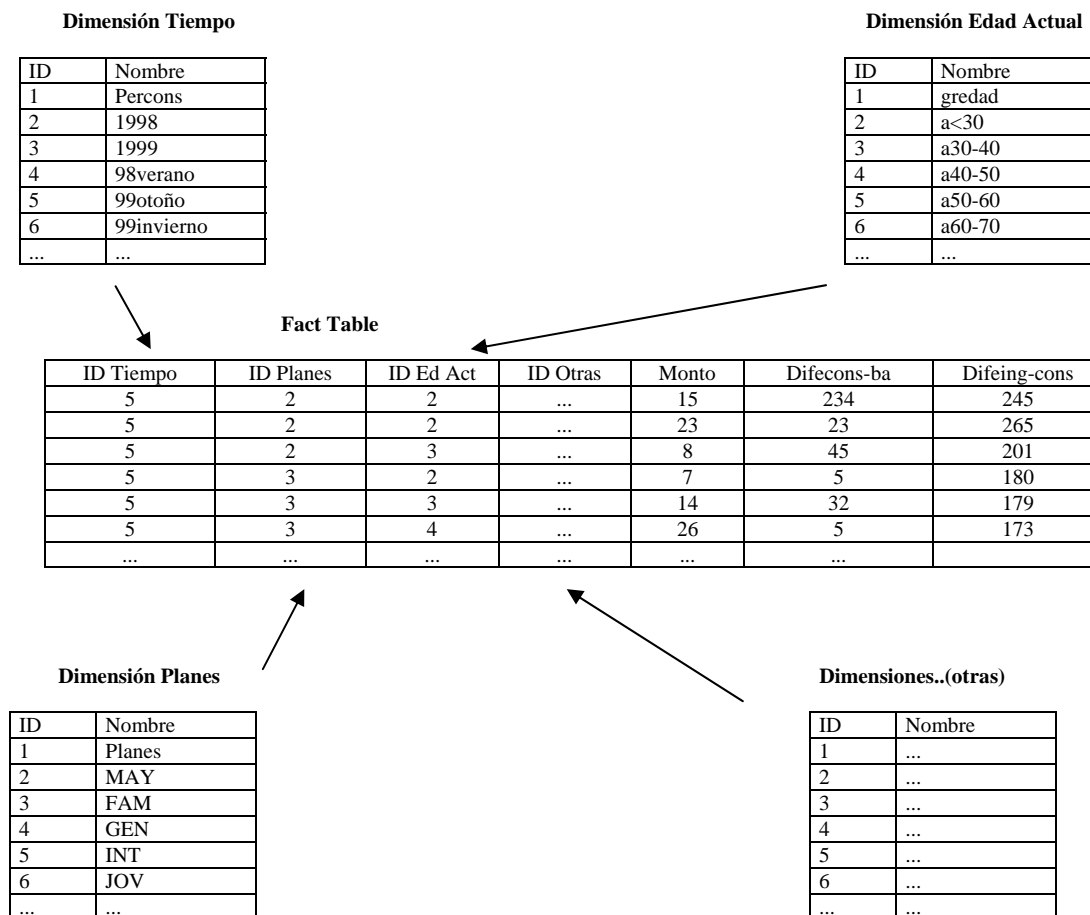


Figura 9. Modelo de datos - (Ejemplo).

Los rangos fueron determinados empíricamente. La dimensión del profesional prescriptor utilizó las variables de sexo y antigüedad en la profesión para medir la posible incidencia de dichos rasgos en la conducta prescriptora. Al no contar con el dato de la fecha de graduación de los médicos se utilizó el número de matrícula como dato subrogante. La dimensión paciente empleó las variables de sexo, edad y antigüedad en el sistema prepago. La dimensión de tipo de plan, que refiere a los diferentes contratos de cobertura ofrecidos por el sistema, fue estudiada como dimensión independiente y en relación a las características ya citadas de la dimensión paciente. (ver figura 9).

Partiendo del modelo descrito, se procedió a obtener los datos de partir de la BDT. Se obtuvieron dos sets de tablas planas conteniendo información referente al consumo de medicamentos y al padrón de afiliados, respectivamente. Ambos archivos fueron de formato ASCII, separados por *tabs* y dotados de delimitadores de texto (Figura 10). Su tamaño fue de aproximadamente 85MB. Para producir el consolidado necesario para la creación del cubo se utilizaron herramientas de extracción provistas por la BDT. Las tablas obtenidas contenían información agrupada de los consumos individuales, incluyendo tanto los datos demográficos previstos para el análisis (fecha de nacimiento del afiliado, sexo, etc.) surgidos del evento de adquisición del medicamento (fecha, precio, etc) como los valores calculados (edad al ingreso al sistema, edad al momento del consumo, etc.).

[illegible]

Figura 10. Formato original de los datos

A continuación, se empleó al Application Manager para crear la aplicación "farma" con el set de datos "farma1", generándose así un *outline* básico de las dimensiones. (figura 11).

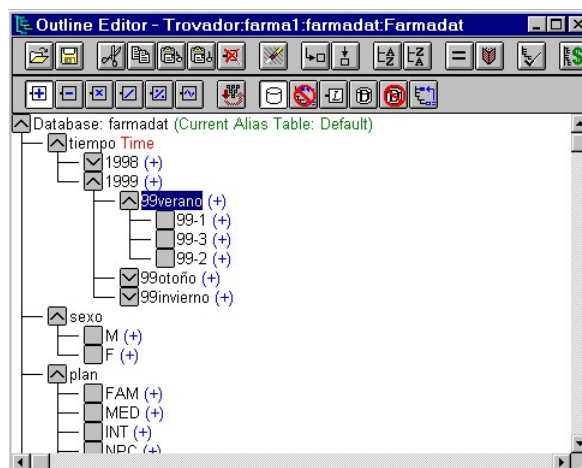


Figura 11. Diseño dimensional (Outline)

Se creó una regla para producir el resto de la estructura del cubo en forma automática, generando las dimensiones y miembros necesarios desde la misma, a partir de los datos fuente. Los nuevos miembros fueron obtenidos mediante el sistema de clasificación jerárquica denominado "generación" a partir de los valores paternos, utilizando los nombres de campo existentes en la primera fila del archivo ASCII. Se agruparon los valores incluidos en cada columna como sub-miembros del paterno correspondiente (ej. la dimensión "precio" incluyó "alto", "medio" y "bajo" como sub-miembros. Con la nueva regla se procedió a modificar el *outline* del set de datos..

The screenshot shows the 'Data Prep Editor' window with a rule definition at the top and a data preview table below it.

Rule Definition:

```

1 "afiliado"■"sexo"■"edadact"■"gredadact"■"edadcons"■"greda
2 1"■"M"■90■"a80-90"■87■"c80-90"■69■"i60-70"■89■"b80-90"■"IN
3 1"■"M"■90■"a80-90"■87■"c80-90"■69■"i60-70"■89■"b80-90"■"IN
4 1"■"M"■90■"a80-90"■88■"c80-90"■69■"i60-70"■89■"b80-90"■"IN
5 1"■"M"■90■"a80-90"■88■"c80-90"■69■"i60-70"■89■"b80-90"■"IN
6 1"■"M"■90■"a80-90"■88■"c80-90"■69■"i60-70"■89■"b80-90"■"IN
7 1"■"M"■90■"a80-90"■88■"c80-90"■69■"i60-70"■89■"b80-90"■"IN
8 1"■"M"■90■"a80-90"■88■"c80-90"■69■"i60-70"■89■"b80-90"■"IN
9 1"■"M"■90■"a80-90"■88■"c80-90"■69■"i60-70"■89■"b80-90"■"IN

```

Data Preview Table:

	tiempo	field 15	tiempo	field 17	medico
1	percons	estacons	mes	fechacons	medico
2	1998	98verano	98-2	2/25/98 0:0	30483
3	1998	98verano	98-2	2/25/98 0:0	30483
4	1998	98invierno	98-8	8/4/98 0:00	43280
5	1998	98primavera	98-12	12/1/98 0:0	30483
6	1998	98primavera	98-12	12/1/98 0:0	30483
7	1998	98primavera	98-12	12/1/98 0:0	30483

Figura 12. Definición de reglas de carga

Luego de validar el outline asignando los *tags* de "time" y "account" a las dimensiones de "tiempo" y "medida", y de definir al tiempo como dimensión "*anchor*", se procedió a crear una nueva regla para cargar el cubo con la totalidad de los datos del archivo fuente (figura 12). En dicha regla se operó específicamente para adaptar la carga de los campos con valores nulos o no existentes en la jerarquía del modelo, a los efectos de minimizar los errores de inclusión o exclusión.

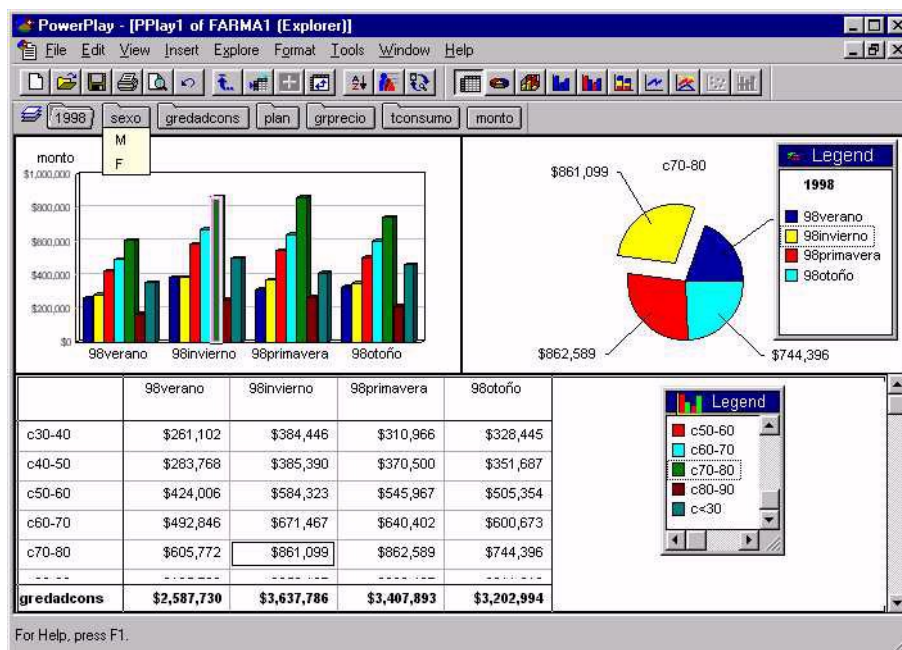


Figura 13. Consumos según grupos etarios y estación

Dado que en el modelo desarrollado cada fila/campo del archivo fuente incluía montos individuales para cada intersección de dimensiones, se definió la columna "monto" como "data field", reemplazando los espacios en blanco (afiliados registrados pero sin consumo) con valores "0".¹⁹

Finalmente, se comprobó la validez de la regla mediante la carga de la base a partir de los archivos fuente. Los errores detectados fueron corregidos y se recalculó la base para completar las intersecciones de agrupamiento de valores paternos no existentes en el archivo fuente a partir de los valores de sus respectivos sub-miembros.²⁰

A fin de estudiar la composición detallada del gasto en medicamentos para el período estudiado, se emplearon las funciones OLAP para desagregar dicho gasto conforme a las dimensiones predeterminadas. En particular, se buscó establecer los perfiles de consumo en base a los agrupamientos fundados en datos demográficos, temporales, de afiliación al sistema (distintos planes de cobertura), de costo unitario de los fármacos consumidos (para determinar la incidencia comparativa de los medicamentos de diferente rango de costo -alto, mediano y bajo-), etc. (figura 13).

Sobre un total de 67.500 afiliados activos al fin del período estudiado, el gasto en medicamentos rondó los 20 millones de unidades económicas*. Del apareamiento y comparación de las diferentes dimensiones planteadas se obtuvieron numerosas conclusiones, de las cuales se citan las más relevantes.

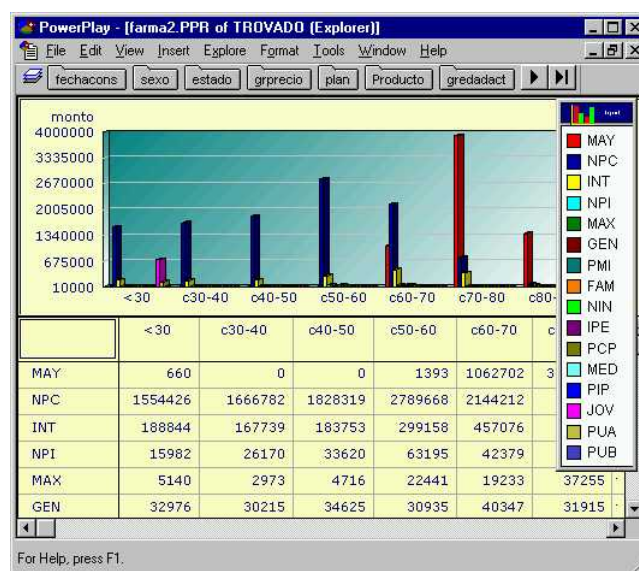


Figura 14. Consumos según plan y grupo etario

El 65% del consumo correspondió al sexo femenino. En cuanto al rango etario, 25% del consumo fue efectuado por afiliados entre 70 y 80 años de edad, seguido por un 19% para el grupo 60-70 y 16% al grupo 50-60. El grupo 30-40 consumió menos del 10% del gasto total en medicamentos.

La tendencia al consumo de medicamentos de bajo costo fue liderada por el segmento etario menor de 30 años, donde el 25% del gasto se ubicó en ese rango. El consumo de medicación más costosa fue liderado conjuntamente por los grupos de 70-80 y 30-40 años. (figuras 14 y 15)

En lo referente a los consumos desagregados por tipo de cobertura, el plan NPC absorbió mas de la mitad del gasto, con un 60% empleado en medicación de alto costo.

La variación estacional del gasto en medicamentos fue escasa. En invierno se consumió el 28% del total de gasto, en tanto que en el verano se consumió el 20%, y en primavera y otoño el 40% (20% y 20%).

* La cifra real en moneda constituye un dato confidencial, por lo cual se emplea una medida ficticia.

El 25% del gasto se originó en 22 médicos, mayormente por indicación de medicamentos de alto costo (60 al 92% de sus prescripciones correspondieron a este grupo). El set de datos con el cual se trabajó no contenía información acerca de la especialidad del profesional prescriptor, por lo cual esta dimensión no pudo estudiarse.

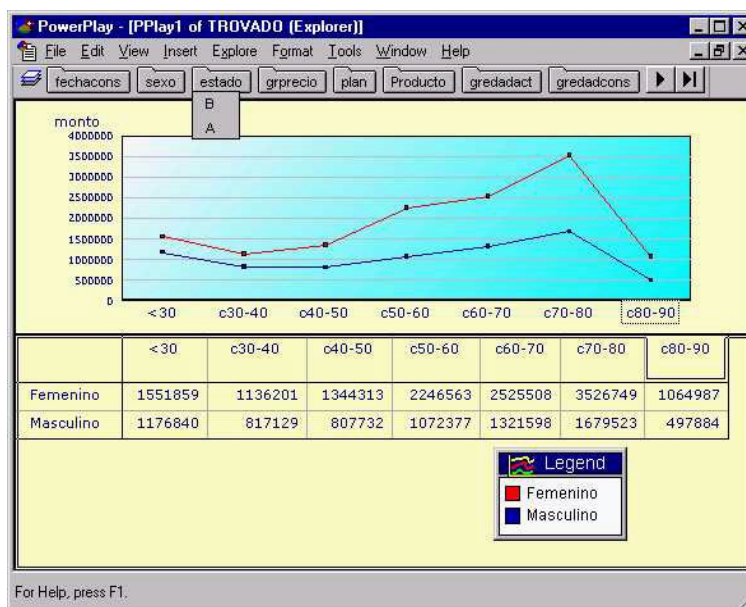


Figura 15. Variación de consumo por grupo etario y sexo

Discusión:

El propósito del presente trabajo no fue la evaluación de un sistema de atención organizada de la salud cual es un sistema de medicina prepaga, sino una "prueba de concepto" sobre la factibilidad de aplicar herramientas OLAP al estudio de dichas organizaciones. Por ello se tomo una muestra limitada de la totalidad del contenido de las BDT operativas en una organización activa. Las dimensiones estudiadas corresponden a variables cuya incidencia es bien conocida a través de estudios estadísticos prospectivos. Se intentó reproducir dichos hallazgos a través de una herramienta OLAP. Los resultados obtenidos fueron altamente satisfactorios, ya que tendencias conocidas tales como la relación directa entre volúmenes de gasto y edad, o el agrupamiento del mayor gasto alrededor de un grupo reducido de profesionales prescriptores son hechos bien conocidos²¹. La herramienta OLAP confirmó plenamente dichas tendencias clásicas. Asimismo, se descubrieron tendencias menos usuales que merecen un mayor análisis y que podrían ser autóctonas de la población bajo estudio y efecto probable del case-mix u otros factores menos conocidos. En cualquier caso, la capacidad de detectar patrones específicos de consumo en una área de incidencia importante sobre el gasto total cual es la de los medicamentos, habilitaría a gerentes y directores a acometer estudios mas detallados tendientes a contener el gasto allí donde ello fuera posible.

Una importante restricción al presente análisis es la falta de una dimensión que permita desagregar los medicamentos consumidos según su función terapéutica (ej: antiarrítmicos, antiinflamatorios, etc.). La razón por la cual dicho análisis no fue acometido se relaciona con la fuente original de los datos, constituida por los listados enviados al PS por la red de farmacias contratada para la provisión de medicamentos a los afiliados al sistema. Los mismos son altamente granulares ya que proveen el detalle del medicamento adquirido por cada afiliado. Sin embargo, para la representación del mismo utilizan su denominación comercial, lo cual implica problemas tales como múltiples denominaciones para un solo tipo genérico de medicamento (las diferentes "marcas" del mismo principio activo), la imposibilidad de efectuar agrupamientos en base a grupos y subgrupos terapéuticos conforme a las taxonomías clásicas (ej: drogas

cardíacas > antiarrítmicos > propranolol), etc. En la actualidad los autores trabajan en la incorporación de un sistema de clasificación de medicamentos que permite la asignación de los mismos a diferentes grupos terapéuticos según un sistema de tres ejes: anatómico (sitio de acción del medicamento), terapéutico (acción del medicamento) y químico (denominación científica del medicamento)^{22,23}. Dicha clasificación ha demostrado importantes beneficios para el análisis automatizado de bases de datos de medicamentos²⁴⁻²⁵⁻²⁶.

Tampoco se ha intentado vincular los consumos de medicamentos a los diagnósticos emitidos durante el proceso de atención médica, dado que la BDT no contaba con vínculos directos entre el evento de prescripción y la adquisición del medicamento en las farmacias. En la actualidad se trabaja en la elaboración de un sistema que permita detectar dichos vínculos, basado en la nomenclatura controlada SNOMED (Systematized Nomenclature of Medicine)²⁷.

Conclusión:

Los resultados obtenidos demuestran que un sistema OLAP adecuadamente implementado permite detectar tendencias conocidas en el consumo de medicamentos, previamente demostrados en la literatura de economía de la salud. Asimismo, permiten detectar tendencias locales al sistema de salud en el cual se las implementa, lo cual habilitaría al planeamiento de estrategias que favorezcan un uso racional de los medicamentos. El agregado de esquemas de metadatos que permitan la agregación por grupos anatómicos, terapéuticos y químicos, así como la utilización de listas de diagnósticos de granularidad adecuada, permitirá un mejor análisis de las bases de datos transaccionales compiladas durante el proceso de atención de la salud en un sistema de medicina prepaga.

Bibliografía

- ¹ Landauer TK: **The Trouble with Computers. Usefulness, Usability and Productivity**. MIT Press, Cambridge: 1995.
- ² Citado en Thomsen E: **OLAP Solutions. Building Multidimensional Information Systems**. Wiley, NY: 1997.
- ³ Drazen EL, Metzger JB, Ritter JL, Schneider MK: **Patient Care Information Systems. Successful Design and Implementation**. Springer, NY: 1994.
- ⁴ A. Berson, S.J. Smith: **Data Warehousing, Data Mining and OLAP**, McGraw-Hill, 1997
- ⁵ Shoshani A: **OLAP and Statistical Databases: Similarities and Differences**. Proc. ACM PODS '97, 185-196, 1997.
- ⁶ E.F. Codd: **A Relational Model for Large Shared Data Banks**, CACM, vol. 13, #6, June 1970
- ⁷ Chaudhuri S, Dayal U: **An Overview of Data Warehousing and OLAP Technology**. ACM SIGMOD Record 26(1), March 1997.
- ⁸ E.F. Codd, S.B. Codd, C.T. Salley, **Providing On-line Analytical Processing to User-Analysts: An IT Mandate**, E.F. Codd and Associates, 1993.
- ⁹ Nikitas-Spiros Koutsoukis, Gautman Mitra., Cormac Lucas: **Adapting on-line analytical processing for decision modeling: the interaction of information and decision technologies**, Elsevier, Decision Support Systems 26 (1999) 1- 30
- ¹⁰ E.F. Codd, S.B. Codd, **OLAP with TM/1**, E.F. Codd and Associates, 1994
- ¹¹ Marisol Gonzalez Lozano, **Introducción a los Sistemas de Información Ejecutiva y la Evaluación de Herramientas para su Desarrollo**, Soluciones avanzadas Nro 35, Mexico 1996
- ¹² Frank A. Buytendijk, **OLAP: Playing for Keeps**. <http://www.xs4all.nl/~fab/olapkeep.html>
- ¹³ **MicroStrategy, The Case of Relational OLAP**, MicroStrategy, 1995
- ¹⁴ R. Finkelstein, **Understanding the need for on-line analytical servers**, Arbor Software , 1994
- ¹⁵ Arbor Software, **Relational OLAP: Expectations and Reality**, Arbor Software, 1995
- ¹⁶ **OLAP Council, OLAP Glossary**, <http://www.olapcouncil.org/research/glossary.htm>, OLAP Council, 1995
- ¹⁷ Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW: **Methods for the Economic Evaluation of Health Care Programs**. Oxford University Press, Oxford: 1997.
- ¹⁸ Cherney A: **The Capitation and Risk Sharing Guidebook**. Irwin, Chicago: 1996
- ¹⁹ M and M Whitehorn, **Business Intelligence: The IBM Solution. Datawarehousing y OLAP**, Springer-Verlag, 1999.

- ²⁰ **IBM DB2 OLAP Server - Arbor Essbase database Administrator**, IBM, 1997.
- ²¹ Folland S, Goodman AC, Stano M: **The Economics of Health and Healthcare**. Prentice-Hall, NY: 1996
- ²² WHO Collaborating Centre for Drug Statistics Methodology: **ATC classification index with DDDs**. WHOCC, Oslo: 2000
- ²³ WHO Collaborating Centre for Drug Statistics Methodology: **About the ATC/DDD system**.
<http://www.whocc.nmd.no/atcsystem.htm>
- ²⁴ Miller GC, Britt H: **A new drug classification for computer systems: the ATC extension code**. Int J Biomed Comput 40:121-4, 1995.
- ²⁵ Gronroos PE, Irjala KM, Huupponen RK, Scheinin H, Forsstrom J, Forsstrom JJ: **A medication database--a tool for detecting drug interactions in hospital**. Eur J Clin Pharmacol 53:13-7, 1997.
- ²⁶ Forsstrom JJ, Gronroos P, Irjala K, Heiskanen J, Torniainen K: **Linking patient medication data with laboratory information system**. Int J Biomed Comput 42:111-6, 1996.
- ²⁷ Lussier YA, Rothwell DJ, Cote RA: **The SNOMED model: a knowledge source for the controlled terminology of the computerized patient record**. Methods Inf Med: 37:161-4, 1998