



**UNIVERSIDAD NACIONAL  
AUTONOMA DE MEXICO**

**FACULTAD DE ESTUDIOS SUPERIORES  
ACATLÁN**

**“LA METODOLOGÍA BUSINESS INTELLIGENCE PARA LA  
CREACIÓN DE CUBOS PARA LA TOMA DE DECISIONES  
EMPRESARIALES. CASO PRÁCTICO CREACIÓN DE UN CUBO  
EMPLEANDO LAS HERRAMIENTAS DE COGNOS”**

**TESINA**

**QUE PARA OBTENER EL TITULO DE**

**Licenciado en Matemáticas Aplicadas y Computación**

**PRESENTA**

**Rubén Salas Cristóbal**

**Asesor: Profesora Teresa Carrillo Ramírez**

**Septiembre del 2008**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**“LA METODOLOGÍA *BUSINESS INTELLIGENCE* PARA LA  
CREACIÓN DE CUBOS PARA LA TOMA DE DECISIONES  
EMPRESARIALES. CASO PRÁCTICO CREACIÓN DE UN CUBO  
EMPLEANDO LAS HERRAMIENTAS DE COGNOS”**

*Algo peor que no tener información disponible es tener mucha información y no  
saber qué hacer con ella*

## INDICE

<b>Introducción:</b> .....	<b>3</b>
<b>CAPITULO 1. Business intelligence</b> .....	<b>6</b>
1.1 Definición .....	6
1.1.1 Tomar mejores decisiones rápidamente .....	6
1.1.2 Convertir los datos en información. ....	7
1.1.3 Utilizar un método razonable para gestión empresarial. ....	8
1.2 Ventajas .....	9
1.3 Antecedentes del Business Intelligence.....	12
1.4 El ciclo de Business Intelligence .....	14
1.4.1 Análisis.....	15
1.4.2 Acción .....	16
1.4.3 Medición.....	16
<b>CAPITULO 2. Datawarehouse</b> .....	<b>18</b>
2.1 Definición de Datawarehouse.....	18
2.2 Características de un Datawarehouse .....	19
2.3 Definición de OLTP.....	21
2.4 Datawarehouse vs. OLTP.....	23
2.5 Proceso de Preparación de los datos (ETL) .....	24
2.6 Modelo Estrella .....	28
2.6.1 Tabla Hechos o Transaccional .....	29
2.6.2 Tablas Estructurales o Dimensionales .....	29
2.7 Datawarehouse y OLAP .....	30
<b>CAPITULO 3. El Procesamiento Analítico en Línea (OLAP)</b> .....	<b>32</b>
3.1 Definición .....	32
3.2 Sistemas MOLAP .....	34
3.3 Sistemas ROLAP.....	35
3.4 Cubo Multidimensional .....	36
3.4.1 Medidas o Métricas.....	37
3.4.2 Dimensión .....	37
3.4.3 Niveles .....	38
3.4.4 Categorías .....	39
3.4.5 Análisis Multidimensional .....	40
<b>CAPITULO 4.- Caso Práctico</b> .....	<b>41</b>
4.1 Características de Powerplay .....	41
4.2 Metodología.....	43
4.3 Caso Práctico.....	44
4.3.2 Modelo y Creación del Cubo (en Power Play Transformer) .....	51
4.3.3 Herramientas Front- End y Análisis de la información (Reportes) .....	54
<b>CONCLUSIÓN</b> .....	<b>60</b>
<b>GLOSARIO</b> .....	<b>62</b>
<b>BIBLIOGRAFÍA</b> .....	<b>64</b>

## **Introducción:**

En la actualidad el mercado tiene un ritmo acelerado, así que las compañías que puedan crear ventajas competitivas y sean creativas en estrategias superiores de negocio dejarán lentamente y con astucia al resto de las compañías. Aunque vivimos en la era de la información, donde crecientes volúmenes de información están a nuestra disposición, frecuentemente luchamos para comprender el significado de los datos. Mientras el progreso en la potencia de las nuevas tecnologías nunca termina, necesitamos reconocer que la tecnología por sí sola no puede resolver los problemas de los negocios. Entonces, para conducir el negocio del futuro, las compañías necesitan gente que lleven el timón y que tomen decisiones efectivas.

Algo interesante ha sucedido en el universo de los negocios: los resultados de las empresas más grandes dependen ahora no sólo de factores como productos, servicios o ubicación, sino que consideran el conocimiento. Es decir, conocimiento acerca de clientes, productos y mercados; conocimiento que revela oportunidades prometedoras e identificación de amenazas potenciales; conocimiento que levanta los ingresos y reduce los costos; conocimiento que soporta una mejor administración, amplia y más efectiva. Tal conocimiento basado en información comprensible detallada y relevante es crucial para lograr y sostener una ventaja competitiva.

De esta manera, cualquier buena decisión de negocios requiere hechos y cifras (datos). El nivel de confianza en una decisión es directamente proporcional al nivel de confianza de los hechos y cifras sobre los cuales está basada. Así como la competencia crece y las decisiones deben ser tomadas en menor tiempo, además de que dichas decisiones se hacen más complejas, por lo que se vuelve sumamente importante la capacidad de obtener la mayor cantidad de información relevante posible en menos tiempo. Indudablemente, para analizar grandes árboles de decisión se necesita más tiempo, y es por este motivo que se requieren herramientas de análisis automatizadas que ayudan a reducir el tiempo de analizar la información con velocidad y precisión. Lógicamente, con el apoyo y sustento en hechos y cifras que den seguridad en las decisiones.

Por otro lado, para mejorar su desempeño las compañías están dispuestas a entender mejor a sus clientes y prospectos, a desarrollar y expandir su relación con ellos, impulsar el servicio a clientes, mejorar la rentabilidad de sus productos y servicios, crear nuevas ofertas muy imaginativas y reducir los gastos. A través de diferentes procesos como clasificación de clientes y análisis de utilización de productos o

servicios, las empresas finalmente empezaran a utilizar la información de sus clientes como un activo estratégico.

El poseer la “inteligencia” en el negocio significa tener respuestas a preguntas claves tales como: ¿Cuáles de nuestros clientes son más rentables, y cómo podemos expandir la relación con ellos?, ¿Qué productos y servicios pueden ser vendidos de forma cruzada más efectivamente y a quiénes?, ¿Cómo podemos mejorar los resultados de las campañas de mercadotecnia?, ¿Qué canales de ventas son más efectivos y para qué productos? , ¿Cómo podemos mejorar la imagen que tienen los clientes de nosotros?.

La mayoría de las empresas tienen los datos que contestan estas preguntas ya que los sistemas operacionales generan una gran cantidad de datos de productos, clientes y mercados desde los puntos de venta, sistemas de reservas, servicio a clientes, sistemas contables y otros sistemas. Sin embargo, sólo una pequeña fracción de los datos es analizada para propósitos estratégicos por las compañías en todo el mundo; el resto de los datos que no son explotados, frecuentemente combinados con fuentes de datos externos a la empresa como reportes de gobierno, asociaciones comerciales, Internet e información adquirida de fuentes especializadas, forman una mina de oro que está a la espera de ser explotada, refinada y transformada en conocimiento vital y valioso para las empresas.

La información resultante puede ser utilizada de muchas maneras, desde gráficos que muestren los resultados generales de la empresa hasta una comunicación personalizada con los clientes a través de centros de llamadas, emisión de facturas, y otros medios de contacto que facilitan la mercadotecnia uno-a-uno en una escala sin precedentes.

A la tecnología que nos permite contar con la información suficiente, así como con la capacidad de analizar la información con el fin de ayudar al proceso de toma de decisiones se le conoce como “*Business Intelligence*”, esta tecnología se basa en la información almacenada en *Datawarehouse* que a su vez se carga con los datos que genera la empresa (transacciones).

*Business Intelligence* ofrece una posibilidad real de manejar eficientemente a la información, la gente y a la tecnología y llegar, de esta manera, a la gerencia exitosa de una compañía u organización.

En esta tesina proporcionaremos los conceptos y técnicas necesarias para un buen entendimiento de lo que involucra *Business Intelligence* así como la forma de utilizar las herramientas computacionales que

nos permiten generar los análisis deseados con el fin de hacer más sencilla y efectiva la toma de decisiones. Con este propósito los capítulos están organizados de la siguiente forma.

#### **CAPÍTULO 1.- Business intelligence**

En este capítulo se presenta un panorama completo de *Bussiness intelligence*, así como el impacto que representa en una empresa el tener esta tecnología como una herramienta para el soporte de la toma de decisiones. Además se incluye la manera de identificar si una empresa es un buen candidato a beneficiarse con el *Business Intelligence*.

#### **CAPÍTULO 2.- Datawarehouse**

En este capítulo se define lo que es un Datawarehouse, la diferencia que existe con una base de datos operacional, y las características de las tablas que componen dicho Datawarehouse.

#### **CAPÍTULO 3.- El Procesamiento Analítico en Línea (OLAP )**

En este capítulo se definen las características principales de la tecnología del Procesamiento Analítico en Línea (OLAP), los sistemas basados en esta tecnología (MOLAP Y ROLAP), cómo esta relacionado el MOLAP con la creación de cubos, así como la definición y elementos que componen a un cubo.

#### **CAPÍTULO 4.- Caso Práctico**

En este capítulo se presenta un caso práctico con el fin de hacer uso de los conceptos expuestos en los primeros capítulos utilizando PowerPlay de COGNOS. Se hace uso de esta herramienta por que es una de las herramientas mas completas y que más se utiliza actualmente. Además de que cuento con la posibilidad de utilizarlo porque donde actualmente laboro me dieron la facilidad de utilizarla.

# CAPITULO 1. Business Intelligence

## **1.1 Definición**

Cuando hablamos de *Business Intelligence* nos estamos refiriendo al conjunto de metodologías y tecnologías orientadas a potencializar la gestión inteligente de la empresa, que permiten a los equipos de directivos controlar los negocios. O dicho de otra forma, *Business Intelligence* es un esfuerzo por captar y analizar los datos de un negocio para comprender los mercados, clientes o proveedores con mayor claridad, para mejorar los procesos de negocio y competir con más efectividad.

El término *Business Intelligence*, también conocido como BI, es relativamente nuevo y es empleado por diferentes expertos y fabricantes de software para distinguir un amplio rango de tecnología, plataformas de software, aplicaciones específicas o procesos, que permiten convertir los datos en información valiosa. De ahí que se trata de un concepto multifacético el cual examinaremos éste desde tres diferentes perspectivas:

- Tomar mejores decisiones rápidamente
- Convertir los datos en información
- Utilizar un método razonable para la gestión empresarial.

### **1.1.1 Tomar mejores decisiones rápidamente**

El objetivo primario de *Business Intelligence* es ayudar a la gente a tomar decisiones que mejoren el funcionamiento de la compañía e impulsen su ventaja competitiva en el mercado. En forma breve, *Business Intelligence* faculta a las organizaciones a tomar mejores decisiones rápidamente.

En el mejor de todos los mundos, los gerentes, desde el liderazgo mas bajo hasta el gerente general, toman decisiones tomando en cuenta sus experiencias, sus entendimientos del negocio, sus planes de negocio y sobre todo la información. Frecuentemente las experiencias, entendimientos y estrategias que acompañan a la toma de decisiones son bástate rígidas; es decir, ellas cambian muy lentamente. La información, sin embargo, es siempre nueva, lo que significa que cambian con frecuencia muy rápido y de una forma significativa. A menudo es difícil tener un manejo de estos cambios y comprender su significado. Tomar mejores decisiones significa mejorar alguna o todas las partes del proceso: esto

también significa un menor número de decisiones erróneas y un mayor número de decisiones acertadas. Mejores decisiones dan como resultado un perfeccionamiento de los objetivos corporativos, tales como mejorar el valor del accionista.

*Business Intelligence* ayuda a tomar mejores decisiones analizando si las acciones tomadas verdaderamente están dando como resultados el mejoramiento los objetivos de la compañía. Decidir cuál es la mejor decisión para una organización, es más fácil si se lleva a cabo con un definido y claro conjunto de objetivos y un plan para conseguirlos. Esta conexión entre el plan global de la compañía y *Business Intelligence* no es una vía de una sola dirección, ya que no se trata simplemente de recibir el plan y utilizar éste como la escala de medición de la calidad de las decisiones. *Business Intelligence* tiene un papel más importante en la creación de las estrategias y planes. Para lo cual es indispensable contar con todos los datos y manipularlos de tal manera que permitan tomar decisiones acertadas.

La necesidad de ser veloz también se aplica a la ganancia de reacciones dentro de una organización. Si un acceso y procesamiento rápido de la información no son posibles, las decisiones son tomadas sin información o con información antigua o discontinuada. Las consecuencias negativas de contar con información antigua o discontinuada pueden ser enormes, tal como perder un cliente clave o continuar produciendo un producto que los clientes no desean. Aunque disponer de información perfecta para apoyar todas las decisiones es un objetivo inalcanzable, no hay duda de que tomar de forma continua decisiones adecuadas en un plazo corto proporcionarían ventajas competitivas.

### ***1.1.2 Convertir los datos en información.***

Para tomar mejores decisiones y más rápidamente, los directivos y gerentes necesitan información relevante y útil al alcance de la mano. Pero es común una larga brecha entre las informaciones que los responsables en la toma de decisiones requieren y las grandes cantidades de datos que las organizaciones recopilan cada día. Nosotros llamamos a esto la brecha del análisis (analysis gap<sup>1</sup>).

Para saltar esta brecha, las organizaciones hacen significativas inversiones para desarrollar sistemas *Business Intelligence* para convertir los datos originales o crudos en información de utilidad. Los sistemas de *Business Intelligence* más efectivos tienen acceso a inmensas cantidades de datos (medidas

---

<sup>1</sup> Es un sistema que identifica las carencias entre "en donde estoy" y "dónde debo estar".

en gigabytes incluso terabytes) para posteriormente entregar a los responsables en la toma de decisiones información expresada de una forma que ellos pueden asimilar y analizar fácilmente. Esto se conoce como análisis a la velocidad del pensamiento, es decir, la capacidad de obtener una respuesta a una pregunta tan rápido como la pregunta es formulada. Esto permite hacer un gran salto en la calidad del análisis que puede ser ejecutado, trayendo como consecuencia un mejor entendimiento de negocio.

Algunas personas consideran que la contribución de la tecnología para proporcionar la información de utilidad es la definición exacta de *Business Intelligence*, esto es, el conjunto de todos los sistemas, aplicaciones, procesos y procedimientos que recopilan y convierten grandes cantidades de datos en información útil para administrar y controlar actividades del negocio dentro de cada uno de los departamentos, divisiones y unidades de negocio.

### ***1.1.3 Utilizar un método razonable para gestión empresarial.***

*Business Intelligence* también puede ser definido como un método para la gestión empresarial, una forma de pensamiento organizacional, una filosofía de gestión; en pocas palabras, un interés hacia la inteligencia del negocio. Tanto las personas como las organizaciones se interesan en el BI porque creen que el uso de un enfoque racional y basado en hechos a la hora de tomar decisiones resulta positivo en la medida que sea posible.

El interés por adoptar el *Business Intelligence* tiene las siguientes características:

- Buscar hechos (datos) que se pueden medir cuantitativamente acerca del negocio.
- Usar métodos organizados y tecnologías para analizar los hechos.
- Inventar o compartir modelos que expliquen las relaciones de causas y efectos entre decisiones operativas y los efectos que éstas tienen en alcanzar los objetivos de negocios.
- Experimentar con métodos alternos y supervisar con retroalimentación sobre los resultados.
- Comprender que las personas no siempre son seres racionales.
- Basar la gestión de la empresa (decisiones e iniciativas) en todas estas características.

La información resultante puede ser utilizada de muchas maneras, desde gráficos que muestren los resultados generales de la empresa hasta una comunicación personalizada con los clientes a través de

centros de llamadas, emisión de facturas y otros medios de contacto que facilitan la mercadotecnia uno-a-uno en una escala sin precedentes.

Este proceso de explotación, refinamiento y transformación de la mina de oro en información es el corazón de *Business Intelligence* (BI).

Después de todo lo antes expuesto, *Business Intelligence* se define como:

*La habilidad de consolidar información y analizarla con la suficiente velocidad y precisión para tomar las mejores decisiones de negocios.*<sup>2</sup>

## **1.2 Ventajas**

En un medio en que la fidelidad del cliente está siempre en duda y la presión de la competencia se hace más intensa que nunca, las relaciones con los clientes y el servicio que se presta son esenciales para la supervivencia de un negocio en una economía cada vez más difícil.

Es evidente que las buenas decisiones repercuten directamente en el crecimiento de un negocio. Por eso es importante contar con todo lo necesario para hacer un verdadero análisis de la información y así ganarle a los competidores. Entonces la metodología *Business Intelligence* ofrece las siguientes ventajas:

- Evita que el proceso de toma de decisiones sea muy tardado y lo convierte en algo sencillo y rápido. Toda empresa que desarrolle un sistema *Business Intelligence* cuenta con un elemento primordial para seguir creciendo y para siempre tomar la delantera a la competencia.
- Pretende colocar todos los datos al alcance de toda la empresa, proporcionando las herramientas para extraerlos de las aplicaciones, conferirles un formato estándar, para posteriormente almacenarlos en un repositorio optimizado para una entrega de la información rápida y resumida que haga posible una consulta o análisis más detallado.
- Trabaja impulsada por un motor incansable: la necesidad de las empresas de aumentar su competitividad. Las tecnologías analíticas destacan lo relevante sobre los datos, permitiendo dedicar recursos en los frentes que exigen las acciones inmediatas para ganar competitividad.

---

<sup>2</sup> Elizabeth Bitt, Micheñ Luckevich, Stacia Misner ; “BUSINESS INTELLIGENCE Técnicas de análisis para la toma de decisiones estratégicas”, Editorial Mc Graw Hill 2003

En el siguiente punto es importante enfatizar que todo aquel responsable de la toma de decisiones en algún negocio y que conteste afirmativamente algunas de las siguientes preguntas es un usuario potencial de *Business Intelligence*:

- a) ¿Pasa más tiempo recolectando y preparando información que analizándola?
- b) En ocasiones, ¿le frustra el no poder encontrar información que usted está seguro que existe dentro de la empresa?
- c) ¿Pasa mucho tiempo tratando de hacer que los reportes en Excel luzcan bien?
- d) ¿No sabe qué hacer con tanta información que tiene disponible en la empresa?
- e) ¿Quiere saber qué productos fueron los más rentables durante un periodo determinado?
- f) ¿No sabe cuáles son los patrones de compra de sus clientes dependiendo de las zonas?
- g) ¿Ha perdido oportunidades de negocio por recibir información retrasada?
- h) ¿Trabaja horas extras el fin de un determinado periodo (año, trimestre, mes, semana, etc.) para procesar documentos o reportes?
- i) ¿No sabe con certeza si su gente está alcanzando los objetivos planeados?
- j) ¿No sabe si mantiene una comunicación estrecha entre las diversas áreas de su empresa hacia una estrategia común?
- k) ¿No tiene idea de por qué sus clientes le regresan mercancía?

En este sentido, las empresas necesitan mejorar su gestión empresarial, hacerla más moderna y competitiva. Una de las formas de hacerlo consiste en aplicar el análisis cuantitativo y cualitativo de los datos, conseguir mecanismos de visualización de los informes, y hacerlo en un plazo lo más corto posible. Esa es la misión del *Business Intelligence*.



Figura 1. La información al alcance de todos

El constante desarrollo tecnológico ha permitido que los directivos accedan a mejor y mas información. Lo logrado hasta ahora por el *Business Intelligence* va de la mano con este hecho, ya que su progreso sería inimaginable si la tecnología no hubiera alcanzado los actuales niveles de calidad y no sólo en la presentación y la entrega de información personalizada sino también en las capacidades de almacenamiento y consulta de grandes bases de datos. La industria ha ido progresando en la homogeneización de los datos, lo que permite a múltiples usuarios utilizar un lenguaje común, aprovechando la consistencia, la calidad y los contenidos de los datos para todas las aplicaciones que funcionan en una misma empresa.

El objetivo de *Business Intelligence* es disponer de la información y analizarla desde múltiples perspectivas a fin de que se puedan elegir las soluciones más válidas para la empresa. No hay nada mejor que tener el control total de la información y con esto asegurar que se dispone del conocimiento necesario para tomar las decisiones que hagan que una empresa esté a la vanguardia y que además se esté seguro que las decisiones no son tomadas al azar sino que tiene un gran respaldo.

En la Figura 2 se ilustra la jerarquía que existe en una base de datos entre dato, información y conocimiento. Se observa igualmente el volumen que representa en cada nivel y el valor que los responsables de las decisiones le dan en esa jerarquía. El área interna dentro del triángulo representa los objetivos que se han propuesto. La separación del triángulo representa la unión estrecha entre dato e información, no así entre la información y el conocimiento.

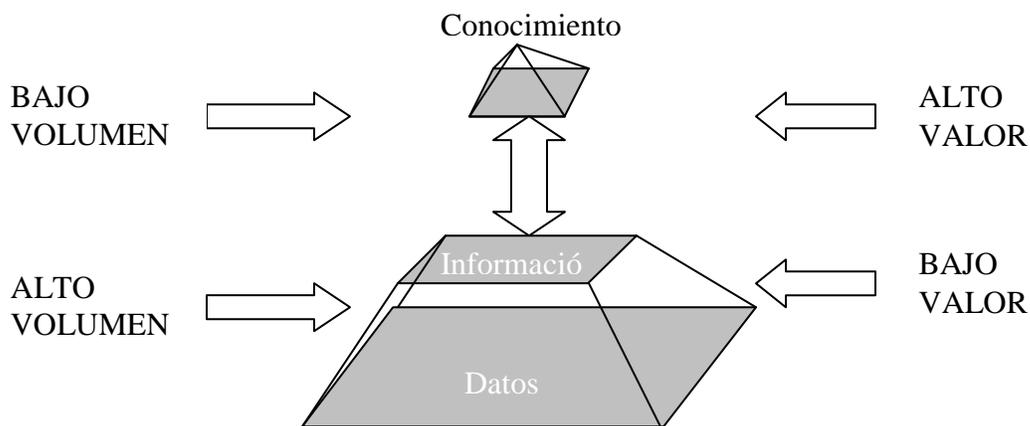


Figura 2. Relación entre Valor y Volumen

### ***1.3 Antecedentes del Business Intelligence***

Los orígenes de los sistemas de apoyo para la toma de decisiones se remontan a finales de los años cuarenta y principios de los cincuenta, mucho antes de que las computadoras existieran. La idea básica era, y por supuesto sigue siendo, recolectar datos operacionales del negocio y reducirlos a una forma que pudieran ser usados para analizar el comportamiento del mismo y modificarlos de una manera inteligente.

A finales de los años sesenta y a principio de los setenta, los investigadores de la Universidad de Harvard y del Instituto Tecnológico de Massachussets comenzaron a promover el uso de computadoras para que ayudaran en el proceso de toma de decisiones, al principio, dicho uso estuvo limitado a la automatización de las tareas de generación de informes, aunque en ocasiones también se proporcionaron herramientas analíticas algo rudimentarias. Esos primeros sistemas de cómputo fueron conocidos inicialmente como Sistemas de Decisiones para la Administración, y posteriormente también se les llamo Sistemas de Información para la Administración. Los cuales pueden considerarse “Sistema de Apoyo para la Toma de Decisiones” o “*Business Intelligence*”.

Los años setenta también vieron el desarrollo de varios lenguajes de consulta y alrededor de dichos lenguajes se construyeron varios sistemas personalizados de apoyo para la toma de decisiones desarrollados por las propias empresas. También fueron implementados generadores de informes, tales como el RPG, o productos para la recuperación de datos, tales como Focus, Datatrieve y NOMAD. Estos sistemas fueron los primeros que permitieron a los usuarios finales, debidamente entrenados, acceder directamente a los almacenes de datos de la computadora, es decir, permitieron que los usuarios formularan consultas relacionadas con el negocio. Sin embargo, las implementaciones que requerían eran muy costosas.

Por otro lado, para generar tales informes era necesario procesar los datos por lotes, lo que no garantizaba que la información que se desprendía fuera precisamente actualizada. Además, las propias limitaciones de almacenamiento obligaban a borrar datos de los sistemas, lo que eliminaba la posibilidad de contemplar la evolución del negocio (lo histórico).

En los ochenta se introdujeron los almacenes de datos para agilizar y simplificar el acceso a los datos. A partir de ellos se creaban repositorios de dimensiones más modestas sobre la base, de los cuales se generaba el análisis, lo que redujo el periodo de tiempo que se necesitaba entre la recopilación y el análisis en sí.

El siguiente gran paso viene de la mano de la tecnología del Procesamiento Analítico en Línea, *OLAP*, especialmente en conjunción con tecnología de almacenamiento de datos, las cuales se convirtieron en la manera de exportar de una forma útil los datos, la información y el conocimiento. El *OLAP* se utiliza para proporcionar conocimiento estratégico, elaborar informes y análisis, o para entender los datos cuando el acceso a ellos no se encuentra estructurado.

En cualquier caso, en los años noventa el *Business Intelligence* estaba circunscrito a las grandes corporaciones por el costo de hardware y aplicaciones, la complejidad de la integración de datos de sistemas cerrados, además del alto costo del mantenimiento y de administración de estos sistemas. Sin embargo, en la actualidad el valor del *Business Intelligence* es accesible para la mayoría de las empresas. El hardware necesario para su implementación se ha reducido sustancialmente; ya que se puede tener la base de datos multidimensional en una computadora personal y la capacidad de almacenamiento y de proceso también ha mejorado mucho.

*Business Intelligence* ha evolucionado y madurado en los últimos años gracias a las *TI (tecnología de información)*, puesto que ha permitido pasar de un simple mecanismo de generación de informes a disponer de una plataforma tecnológica que proporciona información integrada. Definitivamente, el objetivo es aprovechar *Business Intelligence* en todas las áreas de la empresa.

Cabe aclarar que *Business Intelligence*, en realidad, no es parte de la tecnología de base de datos por sí misma, en su lugar, es un uso muy importante de esa tecnología. Los usos en cuestión reciben el nombre de *Datawarehouse*, *Datamart*, *OLAP*, entre otros.

**1.4 El ciclo de Business Intelligence**

Es importante entender que *Business Intelligence* es mucho mas que una actitud empresarial o una tecnología a disposición de la organizaciones; de hecho, es un marco de referencia para la gestión del rendimiento empresarial, un ciclo continuo por el cual las compañías definen sus objetivos, analizan sus progresos, adquieren conocimientos, toman decisiones, miden sus éxitos y comienzan el ciclo nuevamente.

A partir de este ciclo *Business Intelligence* ayuda a los gerentes a tomar mejores decisiones más rápidamente en los niveles estratégicos y operativos. Normalmente se analizan datos provenientes de muchas fuentes. El análisis conduce a ideas, muchas de ellas pequeñas, y se espera que unas cuantas grandes. Estas ideas sugieren maneras de mejorar el negocio cuando se actúa sobre ellas; asimismo, estas ideas pueden ser medidas para ver si funcionan. Estas mediciones también proveen más datos para el análisis, y el ciclo comienza de nuevo. A esta progresión (del análisis de las ideas a la acción a la medición) la llamamos ciclo de *Business Intelligence*.

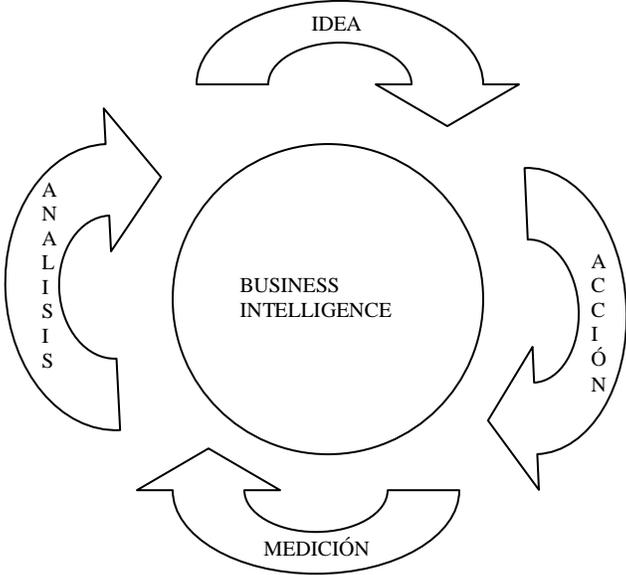


Figura 3. El ciclo de Business Intelligence

### **1.4.1 Análisis**

Cuando emprendemos el análisis de nuestro negocio, ¿cómo determinamos cuáles datos debemos recopilar y qué análisis deberemos desarrollar? Hacemos esto por un filtro mental de selección consciente y subconsciente que creemos que es importante. Este proceso de filtrado esta basado en nuestra comprensión básica y por las suposiciones de cómo nuestro negocio opera, incluyendo, por ejemplo, lo que es importante para nuestros clientes, proveedores y empleados. Este conjunto de todo lo que pensamos acerca de cómo funcionen las cosas (en este caso de nuestro negocio) es lo que se denomina como *modelo mental*.

Los modelos mentales son esenciales para los gerentes que deben tomar muchas decisiones a un paso creciente. Ellas son la base sobre la cual nosotros informalmente decidimos que lo que pensamos es una buena idea. Nuestro modelo mental, sin embargo, pueden causarnos daño porque este puede bloquearnos para distinguir lo que puede ser evidente para otros.

A veces las ideas vienen de diferentes tamaños. Hay ideas operacionales, tales como descubrir la causa de la varianza de precio en una compra de mercancía especializada. Hay ideas estratégicas. Por ejemplo, la mejor manera de ganar suscriptores de teléfonos celulares es regalar el teléfono y cobrar solamente por el servicio. A veces, las ideas han creado industrias completas. Los visionarios de computadoras se dieron cuenta de que los chips y componentes de computadores se estaban haciendo tan pequeños, potentes y baratos que al final todos podrían poseer un computador; de ahí se desarrollo el computador personal.

La idea es el producto del análisis amplio, sin restricciones nacido de preguntas que solo el ser humano se puede hacer –el descubrimiento de patrones que solamente los humanos pueden reconocer como útiles.

Hasta el siglo XVI, la sabiduría convencional mantuvo que la tierra ese el centro del universo y que el sol, la luna y los planetas giraban alrededor de ella. Este fue el modelo ptolemaico desde el cual científicos desarrollaron sofisticadas formulas matemáticas que realmente predecían los eventos celestiales tal como los eclipses. Luego, una generación diferente de pensadores y matemáticos tuvieron una idea revolucionaria: la tierra y los planetas giran alrededor del sol. Esta idea heliocéntrica fue una herejía, por supuesto, y esto cuestionó al basamento político y religioso, así como también la ciencia convencional. En pocas palabras, una idea por si sola no siempre es aceptada nada más porque ésta sea brillante o incluso correcta.

¿Cuál es el significado de esto para *Business Intelligence*? Si una persona tiene una idea importante, ésta generalmente tiene que ser compartida por otros para que sea de utilidad. Exactamente como el modelo heliocéntrico, los cuestionamientos a la sabiduría convencional son frecuentemente mal recibidos y propensos a encontrar resistencia. La inteligencia de negocios bien organizada nos conduce a la idea, pero también nos provee de datos claros, patrones, lógica, reportes, graficas, algoritmos de cálculo y otros análisis y herramientas de presentación para ayudarnos a vender la idea. Propagar ideas es como ayudar a la gente a ver el mundo de una forma nueva y comprende los beneficios de esto como información fundamental.

### **1.4.2 Acción**

La acción se conecta con el ciclo de BI a través del proceso de toma de decisiones. La acción es lo que sigue a una mejor y más rápida toma de decisiones proporcionada por el BI. Las decisiones bien fundamentadas (decisiones apoyadas en un buen análisis y sus conclusiones) proporcionan una motivación extra al que emprende la acción. Al contrario de tomar decisiones por el método de ensayo y error, que a menudo resulta en planes de acción limitados y en un menor financiamiento para los proyectos, la acción respaldada por un análisis sólido y por el *Business Intelligence* se caracteriza por un objetivo más claro y en un mayor apoyo por parte de la organización para su implementación.

Por último, un despliegue de las conclusiones obtenidas a través del *Business Intelligence* mejora los tiempos de la fase de acción. Hoy en día, las organizaciones tienen el imperativo de reaccionar más rápidamente, probar nuevos métodos, experimentar más y obtener prototipos más rápidamente y con mayor frecuencia. La toma de decisiones basadas en el *Business Intelligence*, que genera respuestas cada vez más rápidas y un ciclo óptimo de realimentación, proporciona mayores oportunidades de realizar las mencionadas pruebas orientadas a la acción.

### **1.4.3 Medición**

A través de una mejor recopilación de la información y una frecuencia en la generación de informes que se proporcionan, resulta posible medir los resultados de la compañía en comparación con los estándares cuantitativos, con la finalidad de desembocar en un nuevo ciclo de análisis, idea y acción correctivas.

Esto puede sonar al viejo ciclo de planificación y control que los antiguos contadores convirtieron en los cimientos de la gestión financiera durante décadas. Sin embargo, con el *Business Intelligence* existe una enorme diferencia: no se trata tan solo de la gestión financiera y el control de gastos, sino de la gestión completa de la compañía. *Business Intelligence* permite fijar los estándares para hacer un seguimiento de los resultados y proporcionar retroalimentación en cada área funcional del negocio, utilizando métricas más amplias que las medidas financieras tradicionales (medimos lo que creemos que es importante). El término empleado por *Business Intelligence* para designar las métricas importantes es *indicadores de gestión* (KPI- key performance indicators).

Los sistemas de BI están específicamente diseñados para asimilar grandes cantidades de datos complejos de diferentes fuentes y combinar los datos utilizando algoritmos complejos con el fin de asignar, agregar y, en definitiva, jugar con la información. El resultado es la obtención de informes con las métricas e indicadores del negocio (esto es, los auténticos KPI) que los gerentes necesitan identificar, analizar y utilizar para tomar decisiones de forma frecuente. En un sistema de BI realmente completo, todas las áreas funcionales del negocio deben ser incluidas en el ámbito de los KPI. El objetivo es que los gerentes gestionen aquello que es posible gestionar, y esto suelen ser los KPI.

Para detallar los conceptos expuestos en este y los de los siguientes capítulos, en el capítulo 4 se hace la explicación de un caso práctico utilizando la herramienta de Cognos power play.

## CAPITULO 2. Datawarehouse

Hoy en día las empresas cuentan en su mayoría con la automatización de sus procesos, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas constantemente actualizados, ya que el nivel de competitividad que existe entre ellas les ha exigido desarrollar nuevas estrategias de gestión. Esta necesidad de obtener información para una amplia variedad de individuos es la principal razón de negocios que conduce al concepto de *Datawarehouse*.

El énfasis no está sólo en llevar la información hacia los más altos niveles gerenciales sino que sea accesible a través de toda la organización, para que todos los empleados que la necesiten la tengan a su disposición. Reunir los elementos de datos apropiados desde diversas fuentes de aplicación en un ambiente integral centralizado, simplifica el problema de acceso a la información y en consecuencia, acelera el proceso de análisis, consultas y el menor tiempo de uso de la información.

### **2.1 Definición de Datawarehouse**

De acuerdo con William H. Inmon, quien es considerado como el padre del *Datawarehouse*: “Un *Datawarehouse* es un conjunto de datos integrados orientados a un material que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración.”

Los *Datawarehouse* surgieron por dos razones: primero, la necesidad de proporcionar una fuente única y consistente de datos para el propósito de apoyo para la toma de decisiones; segunda, la necesidad de que el sistema para el apoyo de toma de decisiones no afecte a los sistemas operacionales.

Normalmente el usuario de la información provista por un *Datawarehouse* es una persona especializada en una línea concreta del negocio de un departamento, el cual normalmente trata con una gran cantidad de transacciones y clientes. Estas actividades se reflejan en la infraestructura como bases de datos que llegan a cientos de terabytes y continúa creciendo.

El *Datawarehouse* convierte entonces los datos operacionales de una organización en una herramienta competitiva, por hacerlos disponibles a los empleados que lo necesiten para el análisis y toma de decisiones. Las aplicaciones para soporte de decisiones basadas en un *Datawarehouse*, pueden hacer más practica y fácil la explotación de datos para una mayor eficacia del negocio, que no se logra cuando se

usan sólo los datos que provienen de las aplicaciones operacionales (que ayudan a la operación de la empresa en sus procesos cotidianos), en los que la información se obtiene realizando procesos independientes y muchas veces complejos.

## **2.2 Características de un Datawarehouse**

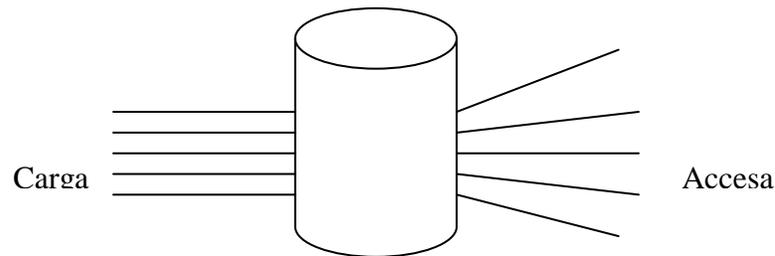
Un sistema *Datawarehouse* es una herramienta eficaz de organización y análisis de los grandes volúmenes de información que las compañías generan y que posteriormente permite el desarrollo de estrategias más efectivas y rentables. Pero la definición del nuevo modelo de datos y el método de carga y mantenimiento de la información, requiere un personal especializado que atienda las necesidades de cada empresa.

Las principales características de un *Datawarehouse* en comparación con una base de datos operacional (OLTP) se pueden resumir en las siguientes:

- **Variación temporal**
- **Manipulación de los datos**
- **Orientado al sujeto**
- **Integración**

**Variación temporal:** El almacén contiene un lugar para guardar datos con una antigüedad de 5 a 10 años o incluso más antiguos, para que puedan ser usados en comparaciones, tendencias y previsiones. De hecho, la principal función de un *Datawarehouse* es la consulta de datos históricos que ayude a tener una perspectiva general de cómo la empresa ha evolucionado con el tiempo y las estrategias a seguir para el futuro a partir de estos datos.

**Manipulación de los datos:** Los datos no serán modificados o cambiados de ninguna manera, una vez que han sido introducidos en el almacén de datos, solamente podrán ser cargados, leídos y/o accedidos. Esto es muy importante, ya que este tipo de almacenamiento es sólo para consulta, es imposible tratar de manipular los datos del *Datawarehouse* como si estuviéramos trabajando con una base de datos relacional (Operacional). Como se muestra en la figura 5.



### DATAWAREHOUSE

Figura 5. Transacciones que realiza un Datawarehouse

**Orientado al sujeto:** Los datos se organizan de acuerdo al sujeto o al tema en vez de la aplicación, por ejemplo: Una compañía de seguros, usando un almacén de datos podría organizar sus datos por clientes, premios y reclamaciones, en lugar de por diferentes productos (automóviles, vida, etc.). Los datos organizados por sujetos y/o temas contienen sólo la información necesaria para los procesos de la toma de decisiones.

**Integración:** Cuando los datos residen en muchas aplicaciones separados por los distintos entornos operacionales, la decodificación de los datos es, a menudo, inconsistente. Por ejemplo, en una aplicación la palabra SEXO podría codificarse como “m” y “f”, en otra como “0” y “1”, cuando los datos fluyen de un entorno operacional a un entorno de almacén de datos o de *Datawarehouse*, ellos asumen una codificación consistente, por ejemplo SEXO siempre se transformaría a “m” y “f”.

Desde un sistema *Datawarehouse*, la información se puede mostrar y representar de muchas maneras. La forma más común de analizarla, es utilizando un sistema de proceso de análisis en línea (*OLAP*). Los productos *OLAP* ofrecen un rango muy variado de capacidades de análisis avanzado, como el multidimensional y el estadístico.

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

### **2.3 Definición de OLTP**

A continuación se definirá brevemente lo que es la tecnología Transaccional *OLTP* con el fin de hacer una comparación de esta tecnología con la del *Datawarehouse*.

El sistema de gestión transaccional *OLTP* es un modo de proceso de información en tiempo real, en virtud del cual tiene lugar una actualización fiable de la base de datos con cada transacción, y garantiza un alto grado de la integridad de los datos, la eficiencia de cada transacción y la fiabilidad del sistema. Puede considerarse que el modelo relacional en el cual se basa *OLTP*, tiene como objetivo guardar la integridad de la información necesaria para operar un negocio de la manera más eficiente. Sin embargo, este modelo no corresponde a la forma en que el usuario percibe la operación de un negocio.

Mientras su tecnología subyacente ha cambiado dramáticamente con el tiempo, las bases de datos operacionales mantienen su misma funcionalidad básica: capturar, actualizar, almacenar y recuperar archivos de datos. Al final, un gran almacén corporativo con billones de registros, con terabytes de almacenamiento de una base de datos se conoce con el nombre de RDBMS, (relational database management system) que es el acrónimo de sistema de administración de bases de datos relacionales. Conjuntamente, los sistemas operacionales son frecuentemente llamados sistemas *OLTP* (online transaction processing).

Para comprender como funciona un sistema *OLTP*, utilizaremos un ejemplo común en el lenguaje *OLTP*, retirar dinero desde un cajero automático. Generalmente, introducimos la tarjeta en el cajero y el cajero lee el perfil del cliente. Introducimos el código secreto de cliente y el cajero lo verifica. Introducimos el importe que deseamos retirar, el cajero nos entrega el dinero y el recibo donde se muestra el saldo de la cuenta. En estas operaciones se esta interactuando con un sistema *OLTP*, y este proceso tiene tres características que fundamentalmente definen el sistema.

- Se procesa una transacción: podemos retirar de acuerdo a las reglas del juego, las cuales son que la cuenta bancaria sea correcta, el código secreto correcto, tener saldo en la cuenta y así en adelante. Esta es la actividad más básica para un banco.
- Se ejecutan todos los elementos de la transacción en tiempo real: verificar quien hace la transacción, entregar el dinero e inmediatamente actualizar el saldo de la cuenta. Todas las transacciones suceden más o menos simultáneamente.
- Se procesan muchas transacciones sobre una base interrumpida: los depósitos, retiros y otras operaciones que tengan que ver con las cuentas de clientes, son procesadas continuamente en paralelo todo el día, cada día. No existen cortes, paradas o inicios (al menos ninguno aparente), para los clientes que utilicen los cajeros automáticos.

*OLTP* esta diseñado para administrar los datos originales o datos crudos, los cuales requieren eficiencia y procesamiento de transacciones al minuto y al más bajo nivel de detalle.

Ahora, cambia la pregunta: ¿Por qué no utilizamos los sistemas *OLTP* para desarrollar análisis del negocio? La respuesta sería errónea en la pregunta que queremos responder eficientemente. Por ejemplo, ¿retiró dinero John Doe? Y si es así, ¿Cuánto? , es una pregunta de fácil respuesta por el sistema *OLTP*. A pesar de que esta información puede ser ofrecida rápidamente, esta información es útil para desarrollar el análisis de todo el negocio.

Para identificar alguna tendencia de los seis meses pasados ¿Cuál ha sido el promedio de transacciones de cajero automático diariamente en la ciudad de Naucalpan del estado de México, fuera del horario normal trabajo? Esta es una idea de una pregunta de potencial interés que podría afectar el número y ubicación de cajeros automáticos, pero ¿esta pregunta podría ser respondida por un sistema *OLTP*? Hay muchos registros que buscar, ordenar y totalizar, cálculos matemáticos que son necesarios para resultado (promedio de todos los cajeros automáticos). Imponer este tipo de consultas a un sistema *OLTP* y sobre una base regular causaría interferencia con las operaciones regulares del banco.

Mientras los sistemas *OLTP* son pésimos para el análisis, lo que ellos hacen es más bien recoger los datos originales o crudos que son la base para el análisis multidimensional. Y este es el desafío de cara al negocio. Es decir, aquellos millones de transacciones en los sistemas operacionales (típicamente múltiples sistema *OLTP* en grandes corporaciones) que son la base, rica en datos que necesitan ser convertidos en información útil para desarrollar el análisis del negocio. Conseguir esto no es sencillo, especialmente

cuando los datos residen en múltiples sitios, organizados de forma distinta y frecuentemente en sistemas de tecnología antigua.

## **2.4 Datawarehouse vs. OLTP**

Los sistemas tradicionales de transacciones y las aplicaciones de *Datawarehouse* son polos opuestos en cuanto a sus requerimientos de diseño y sus características de operación. Es de suma importancia comprender perfectamente estas diferencias para evitar caer en el diseño de un *Datawarehouse* como si fuera una aplicación de transacciones en línea *OLTP*.

Una diferencia importante radica en el número de usuarios: normalmente, el número de usuarios de un *Datawarehouse* es menor al de un *OLTP*. Es común encontrar que los sistemas transaccionales son accedidos por cientos de usuarios simultáneamente, mientras que los *Datawarehouse* sólo por decenas. Los sistemas de *OLTP* realizan cientos de transacciones por segundo mientras que una sola consulta de un *Datawarehouse* puede tomar minutos. Otro factor es que, frecuentemente, los sistemas transaccionales son menores en tamaño a los *Datawarehouse*, esto es debido a que un *Datawarehouse* puede estar formado por información de varios *OLTP*'s.

Existen también diferencias en el diseño, mientras que el de un *OLTP* es extremadamente normalizado, el de un *Datawarehouse* tiende a ser des-normalizado. El *OLTP* normalmente está formado por un número mayor de tablas, cada una con pocas columnas, mientras que en un *Datawarehouse* el número de tablas es menor, pero cada una de éstas tiende a ser mayor en número de columnas.

Los *OLTP* son continuamente actualizados por los sistemas operacionales del día con día, mientras que los *Datawarehouse* son actualizados en *batch* de manera periódica.

Las estructuras de los *OLTP* son muy estables, rara vez cambian, mientras que las de los *Datawarehouse* sufren cambios constantes derivados de su evolución. Esto se debe a que los tipos de consultas a los cuales están sujetos son muy variados y es imposible preverlos todos de antemano.

En la siguiente tabla se concretizan las principales características de ambas herramientas.

<i>Característica</i>	<i>Sistema OLTP</i>	<i>Datawarehouse</i>
Función	Soporta las transacciones de la organización	Soporta el análisis para la toma de decisiones.
Operación predominante	Altas, Bajas y Cambios	Consultas
Integración de información	Datos en general desagregados	Datos en distintos niveles de detalle y <i>agregación</i>
Tiempo	Estado actual de la organización	Histórico, Actual y Proyectado
Estructura de almacenamiento	Estructura relacional. Alto niveles de normalización	Visión multidimensional. Modelos estrella, tablas con cierto nivel de des-normalización
Perfil de usuarios	Usuarios de niveles operativos, medios mandos	Usuarios a niveles ejecutivos, gerenciales.
Alcance de la información	Explotación de la información limitada al segmento del negocio que soporta el sistema <i>OLTP</i>	Explotación de toda la información relacionada con el negocio.
Respuesta	Trabaja en tiempo real. Respuesta de la transacción instantánea. Devuelve datos puntuales	Procesamiento masivo de datos. Tiempo de respuesta desde unos cuantos segundos hasta varias horas.

### **2.5 Proceso de Preparación de los datos (ETL)**

Muchas de las cuestiones que rodean a los sistemas *Business Intelligence* para la toma de decisiones, se refieren en primer lugar a las tareas de obtener y preparar los datos ya que cada una de estas operaciones involucra sus propias consideraciones especiales.

El proceso de Extracción, Transformación y Carga (**E**xtraction, **T**ransformación and **L**oad) consiste en obtener datos de una o más bases de datos transaccionales, archivos o de la mezcla de fuentes. Transformar estos datos para eliminar inconsistencias, agregar información resumida y finalmente cargar estos datos en el *Datawarehouse*. El resultado final es contar con una base de datos dedicada que contenga información estable, integrada e histórica. A su vez esta información es presentada con la dimensión de tiempo, de manera tal que se pueda tener la información clasificada por fecha.

A continuación examinaremos cada uno de estos procesos:

### **1.- Extracción:**

La extracción de datos es el primero de tres pasos que se llevan a cabo para el poblado del *Datawarehouse*. A pesar de la simpleza del enunciado, es uno de los procesos más complejos.

Este proceso consiste en capturar los datos de la base de datos operacional u otras fuentes. Hay muchas herramientas disponibles para ayudar en esta tarea, incluyendo herramientas poco comerciales. El proceso de extracción tiende a ser intensivo en *E/S* y por lo tanto puede interrumpir las operaciones de la BD operacional.

Los datos de sistemas transaccionales o de archivos es la principal fuente de datos para el *Datawarehouse*. Las bases de datos transaccionales pueden ser archivos indexados o sistemas de bases de datos relacionales, etc.

En ocasiones los procesos de extracción no están bajo el control del equipo de desarrollo del *Datawarehouse*, puede ser que estén administrados por personal a cargo de los sistemas *OLTP* o administradores de base de datos (*DBA*). En tal caso el equipo de *Datawarehouse* debe proporcionar especificaciones de extracción necesarias para delegar esta tarea a la gente indicada. Es necesario enfatizar que cuando los diseñadores del *Datawarehouse* empiezan a trabajar con una base de datos desconocida por ser nueva para ellos, siempre es necesario el apoyo total de las personas que conocen exactamente la estructura de la base de datos, ya que ellos proporcionarán datos como: nombre de la base de datos, usuario, password, nombre de las tablas, campos de donde podemos extraer la información necesaria para poblar el *Datawarehouse*; además de otras características técnicas de las bases de datos que se utilizarán para este fin.

Las especificaciones de extracción deben:

- Determinar dónde se localizan los datos en los sistemas operacionales y en los legados. Puede ser difícil obtener parte de la información. Si esto es así se debe determinar la razón y contar con planes de contingencia para obtener la información en un periodo de tiempo razonable y previamente establecido.

- Definir los estándares que se van a usar para nombrar los archivos, el separador de campos dentro de los archivos, el formato para fechas, etc.
- Determinar la frecuencia de los procesos de extracción. Este punto debe acordarse entre las áreas que administran los sistemas transaccionales, de usuarios y la de desarrollo de *Datawarehouse*, ya que generalmente los usuarios desean tener la información tan actual como sea posible, en tanto que los administradores de los sistemas operacionales prefieren extracciones semanales o nocturnas. El área de desarrollo del *Datawarehouse* puede hacer las observaciones pertinentes buscando un balance entre la operación y la extracción de datos, entre los recursos materiales y el tiempo requerido y la complejidad y necesidad de los procesos de extracción.

## **2.- Transformación:**

Es particularmente importante cuando se necesitan mezclar varias fuentes de datos, un proceso al que le llamamos *consolidación*. En estos casos todos los datos que son de diferente fuente y a menudo con diferente estructura o formato, se deben de homogenizar con el fin de contar con una base de datos independiente de las fuentes. Aquí es importante recordar que una de las metas principales del *Datawarehouse* es tomar datos de varias fuentes dispersas y brindar una visión completa de la organización.

Algunos de los problemas que se presentan en este punto son:

- Datos en diferentes tablas con los mismos nombres, pero cuyo contenido representa aspectos diferentes de la organización.
- Datos en tablas con diferente nombre, pero que representan lo mismo.
- Tablas en el *Datawarehouse* que requieren datos de varias fuentes.
- Datos que son requeridos en las tablas del *Datawarehouse* que provienen de la operación de archivos diferentes, con operaciones tan complejas como se necesiten.

## **3.- Carga:**

La carga de datos es el proceso de *poblar* el *Datawarehouse*. Es una tarea fundamental, que generalmente involucra grandes volúmenes de datos y procesos repetitivos, típicamente deben hacerse en periodos breves de tiempo, por ejemplo diariamente durante momentos de baja actividad, semanalmente, en días no laborables, etc.

Este es el paso más sencillo, desde mi punto de vista, ya que sólo se tiene que direccionar, toda la información extraída y transformarla a una o varias de las tablas del *Datawarehouse*.

Como hemos visto una parte importante de los esfuerzos de implementación se enfocan en la extracción, transformación, y carga de los datos provenientes de los sistemas fuentes dentro del *Datawarehouse*. Para apoyar al equipo técnico a la implementación de procesos *ETL* (**E**xtraction, **T**ransformación and **L**oad), existe una categoría de herramientas de software, llamadas herramientas *ETL* que se utilizan con la finalidad de expandir la funcionalidad de estos procesos.

Normalmente los usuarios no tienen interacción con estas herramientas. En cambio, ellos verán sólo el producto final, datos depurados, consistentes, e integrados que les son entregados en el momento en que los necesitan. A continuación presentamos algunas categorías que el grupo de implementación debe utilizar normalmente en las herramientas *ETL* existentes:

\* **Lectura de cualquier fuente de datos:** Las herramientas *ETL* se enorgullecen de sí mismas de tener la capacidad de leer datos desde cualquier sistema fuente, tales como bases de datos relacionales, hojas de cálculo y archivos de datos como se muestra en la figura 6.

\* **Eficiencia:** Las herramientas *ETL* rutinariamente procesan grandes volúmenes de datos de una manera rápida y eficiente. Dependiendo de su aplicación particular, esto podría significar millones o billones de registros de transacción.

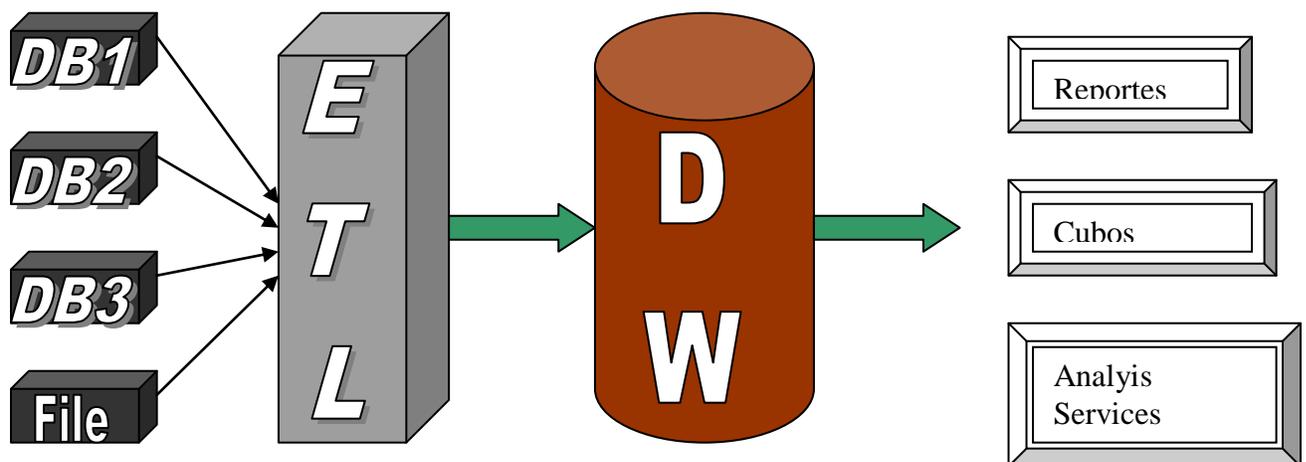


Figura 6. Flexibilidad del ETL para cargar la información de diferentes orígenes a nuestro Datawarehouse, para ser tomados por nuestras herramientas OLAP

\* **Multi-Plataforma:** Las herramientas *ETL* permiten escribir programas una vez y luego implementarlos para una gran variedad de plataformas de *hardware* y sistemas operativos.

## 2.6 Modelo Estrella

En general, el modelo multidimensional también se conoce con el nombre de esquema estrella, pues su estructura base es similar: una tabla central y un conjunto de tablas que la atienden radialmente. El esquema estrella deriva su nombre del hecho que su diagrama forma una estrella como se muestra en la figura 7. El centro de la estrella consta de una o más tablas de hechos, y las puntas de la estrella son las tablas estructurales (dimensiones). En algunos casos especiales sólo es necesaria la tabla de hechos y las dimensiones pueden ser tomadas desde una columna de la misma tabla de hechos.

Este modelo entonces, resulta ser asimétrico, pues hay una tabla dominante en el centro con varias conexiones a las otras tablas. Las tablas estructurales tienen sólo la conexión a la tabla de hechos y ninguna más.

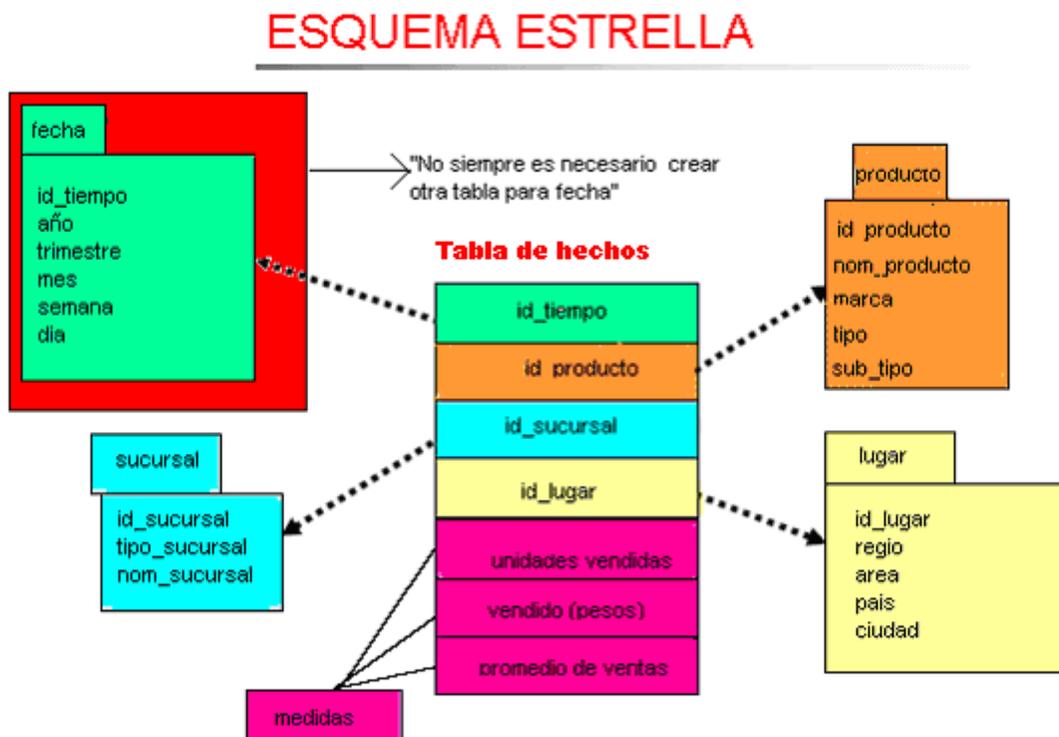


Figura 7. Modelo estrella, con una tabla de hechos central rodeada por tablas estructurales.

A continuación se explican de manera específica las características de las tablas que componen el modelo estrella.

### **2.6.1 Tabla Hechos o Transaccional**

Es en la(s) tabla(s) central en un esquema dimensional donde se almacenan las mediciones numéricas del negocio.

La *granularidad* de la información que se puede explorar queda determinada por el nivel de detalle que se almacenará en la tabla. Por ejemplo, para el caso de producto, mercado y tiempo, el grano puede ser la cantidad de madera vendida ‘mensualmente’.

Cada medida es tomada de la intersección de las dimensiones que la definen. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. La razón de estas características es que así se facilita que los miles de registros que involucran una consulta sean comprimidos en unas pocas líneas en un conjunto de respuesta.

Así entonces, se distinguen dos tipos de columnas en una tabla de Hechos: columnas Hechos y columnas key. Donde la columna Hechos es la que almacena alguna medida de negocio y una columna key forma parte de la clave compuesta de la tabla. Por ejemplo, la clave de un producto es la que se almacena en la tabla Hechos y ésta se liga a la tabla de dimensiones para producto, para de esta manera, poder traer toda la estructura de ese producto (Ejemplo: Línea, Familia, Tipo, Sub-tipo). Esta jerarquía se encuentra en la tabla de dimensión y sólo ligándola a la tabla Hechos se pueden obtener todos estos datos. También es importante señalar que es el corazón de un *Datawarehouse* ya que contiene datos de las transacciones de la empresa. Y que puede no necesitar de las tablas de dimensiones para la creación del *Datawarehouse*.

### **2.6.2 Tablas Estructurales o Dimensionales**

Estas tablas son las que se conectan a la tabla de Hechos y son las que la alimentan. Una tabla estructural almacena un conjunto de valores que están relacionados a una dimensión particular. Estas tablas no contienen cifras, sus valores son los elementos que determinan la estructura de las dimensiones, en ellas existe el detalle de los valores de la dimensión respectiva.

Una tabla estructural está compuesta de una *primary key* que identifica unívocamente una fila en la tabla junto con un conjunto de atributos, y dependiendo del diseño del modelo multidimensional puede existir una *foreign key* que determina su relación con otra tabla estructural.

Los atributos dimensionales toman un papel determinante en un *Datawarehouse*. Ellos son la fuente de todas las necesidades que debieran cubrirse. Esto significa que la base de datos será tan buena como lo sean los atributos dimensionales, mientras más descriptivos, manejables y de buena calidad, mejor será el *Datawarehouse*.

Ya que las tablas de dimensiones tienen cambios muy escasos se consideran como catálogos. Por ejemplo, para los clientes sólo tienen modificaciones cuando se cambia el nombre del representante o cliente, o cuando se da de baja. Lo que hace a este tipo de tablas muy estáticas, lo cual ayuda para que puedan ser usadas por otros *Datawarehouse* sin necesidad de crear otra.

Como ya se mencionó, tienen la función de describir las columnas de dimensión de la tabla de Hechos. Y como es un catálogo no se permite tener productos repetidos o diferentes productos con la misma descripción, esto es muy importante si no queremos tener problemas a la hora de generar el *Cubo* (base de datos multidimensional).

## **2.7 Datawarehouse y OLAP**

Es importante distinguir las capacidades de un *Datawarehouse* y las de un *OLAP*:

*OLAP* utiliza una vista multidimensional de datos globales para proporcionar un acceso rápido a información estratégica para el posterior análisis mediante la creación de una base de datos multidimensional (*Cubo*). *OLAP* le permite a los analistas, gerentes y ejecutivos acceder a los datos a través de un acceso interactivo, coherente y rápido hacia una amplia variedad de posibles vistas de información. Además, transforma los datos brutos a fin de que reflejen la dimensión real de la empresa tal como la entiende el usuario.

*OLAP* y los *Datawarehouse* se complementan. Un *Datawarehouse* almacena de forma estructurada los datos (tablas Hechos, tablas dimensionales) para que el *OLAP* los tome. *OLAP* transforma los datos del *Datawarehouse* en información estratégica e incluye la navegación y la búsqueda básica.

Entonces podemos decir que el *Datawarehouse* es el administrador de la información que el *cubo* mostrara, ya que cuando el *cubo* se genera se vuelve independiente del *Datawarehouse*, sin embargo si necesitamos hacer modificaciones al *cubo* es necesario modificar la información del *Datawarehouse* ya que la única forma de actualizar el *cubo* es regenerándolo, entonces, es necesario que la herramienta *OLAP* se vuelva a conectar al *Datawarehouse* para traer toda la información actualizada nuevamente para crear el *cubo*.

En el siguiente capítulo veremos con mas detalle lo qué es el *OLAP* (procesamiento analítico en línea)

## CAPITULO 3. El Procesamiento Analítico en Línea (OLAP )

”La tecnología *OLAP* se refiere a las herramientas de reporte y graficado que permiten desplegar información consolidada”

Al procesamiento analítico o análisis multidimensional se le conoce también como Procesamiento Analítico en Línea, *OLAP*. Este procesamiento se apoya en una visión multidimensional de los datos empresariales en el *Datawarehouse* y puede tener un motor de depósito de base de datos multidimensional.

*OLAP* es una tecnología de procesamiento analítica que crea nueva información empresarial a partir de los datos existentes, por medio de un rico conjunto de transformaciones empresariales y cálculos numéricos.

En un *Datawarehouse*, se depositan datos para consulta, análisis y divulgación, a diferencia del procesamiento de transacciones en línea *OLTP*, en donde los datos se reúnen y almacenan para operación y control.

Se puede resumir la definición *OLAP* en sólo 5 palabras “Fast Analysis of Shared Multidimensional Information” (FASMI), “Análisis rápido de información multidimensional compartida”

### **3.1 Definición**

*OLAP* debe su nombre a que contrasta adecuadamente con *OLTP*; el cual era un término que estaba ampliamente extendido cuando el termino *OLAP* fue creado. E. F. Codd (uno de los gurús de la tecnología de base de datos relacionales), fue quien acuñó el termino *OLAP*; evidentemente quiso resaltar las diferencias fundamentales entre el procesamiento de transacciones y el procesamiento analítico. Posteriormente muchos vendedores han saltado al tren de *OLAP*, proporcionando sistemas con un amplio rango de características y funcionalidades pero no todas aplicaron los criterios de Codd. Pero, ¿Por qué *OLAP* para *Business Intelligence*? esta pregunta puede ser respondida considerando tres capacidades críticamente importantes que todos los sistemas *OLAP* deben incorporar con alto nivel de eficiencia:

- *OLAP* primero proporciona un modelo de datos intuitivo y conceptual, para que los usuarios que no tengan experiencia como analistas puedan comprender y rápidamente relacionar. Este modelo es de hecho llamado análisis multidimensional, siendo habilitado para ver los datos a través de múltiples filtros o dimensiones. Los mejores gerentes y analistas fueron ya pensadores multidimensionales
- *OLAP* es rápido para el usuario, es la respuesta para conseguir la experiencia de información a la velocidad del pensamiento que mencionamos anteriormente. Rápidos tiempos de respuesta permiten que los gerentes y analistas puedan preguntar y resolver más situaciones en un corto periodo de tiempo, esto significa que podríamos alcanzar el síndrome de pregunta infinita, donde cada patrón de interés y cada pieza útil de información en una base de datos puede ser explorada teóricamente.
- Finalmente, los sistemas *OLAP* tiene un motor de cálculo bastante robusto para manejar las necesidades de cálculo especializado que una estructura multidimensional impone. Pensar que podamos escribir una hoja de cálculo con fórmulas que se conecten como mucho con diez dimensiones, como por ejemplo, tiempo, zona, producto, cliente, canal, estructura organizativa, etc., no es sencillo para el ser humano. El motor de cálculo de *OLAP* organiza los datos en una forma que permite a los analistas escribir fórmulas sencillas y directas que se ejecutan a través de múltiples dimensiones con solo unas pocas líneas de código.

El Procesamiento Analítico en Línea es una tecnología de análisis de datos que hace lo siguiente:

- Presenta una visión multidimensional lógica de los datos en el *Datawarehouse*. La visión es independiente de cómo se almacenan los datos.
- Comprende siempre la consulta interactiva y el análisis de los datos. Por lo regular la interacción es de varias pasadas, lo cual incluye la profundización en niveles cada vez más detallados o el ascenso a niveles superiores de resumen.
- Ofrece opciones de modelado analítico, incluyendo un motor de cálculo para obtener proporciones, desviaciones, etc. que comprende mediciones de datos numéricos a través de muchas dimensiones.
- Crea resúmenes, adiciones (conocidas como consolidaciones), jerarquías, y cuestiona todos los niveles de adición y resumen en cada intersección de las dimensiones.
- Maneja modelos funcionales de pronóstico, análisis de tendencias y análisis estadísticos.

- Responde con rapidez a las consultas, de modo que el proceso de análisis no se interrumpe y la información no se des-actualiza.
- Tiene un motor de depósito de datos multidimensional que almacena los datos en arreglos. Estos arreglos son una representación lógica de las dimensiones empresariales.
- Es aplicable en muchas áreas funcionales de una empresa, tales como producción, ventas y análisis de rentabilidad de la comercialización; mezcla de manufacturas y análisis de logística; consolidaciones financieras, presupuestos y pronósticos, planeación de impuestos y contabilidad de costos.

*OLAP* es una opción de análisis y de reporte, es un componente importante del bloque de acceso y uso de la arquitectura de referencia del *Datawarehouse*. Los componentes de la tecnología *OLAP* se capturan en sub-bloques de los bloques de acceso y uso. El componente *OLAP* del sub-bloque de análisis y reporte representa las opciones de análisis y reporte de los servicios *OLAP* requeridos, mientras que la transformación a la estructura multidimensional, así como el acceso a los componentes del *Datawarehouse* y de los *Datamarts*, son partes del sub-bloque de acceso y recuperación.

Estas opciones ilustran los componentes potenciales lógicos y físicos de la arquitectura *OLAP*.

Por otro lado, existen dos arquitecturas diferentes para los sistemas *OLAP*: *OLAP* multidimensional (*MOLAP*) y *OLAP* relacional (*ROLAP*) las cuales se presentaran a continuación.

### **3.2 Sistemas MOLAP**

La arquitectura *MOLAP* usa bases de datos multidimensionales (*cubo*), independiente de la base de datos operacional o del *Datawarehouse*, para proporcionar el análisis, su principal premisa es que el *OLAP* está mejor implantado almacenando los datos multidimensionalmente, para ser visualizada de una manera distinta a una base de datos relacional.

Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados o agregaciones son la base de las ganancias de desempeño de este sistema. Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores precalculados.

El sistema *MOLAP* utiliza una arquitectura de dos niveles: Las bases de datos multidimensionales y el motor analítico.

- La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato del *Datawarehouse*.
- El nivel del motor analítico es el responsable de la ejecución de los requerimientos *OLAP*. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través del cual los usuarios finales visualizan los análisis *OLAP*.

### **3.3 Sistemas *ROLAP***

La arquitectura *ROLAP*, accede a los datos de la base de datos operacional o al *Datawarehouse* para proporcionar los análisis *OLAP*. La premisa de los sistemas *ROLAP* es que las capacidades *OLAP* se soportan mejor contra las bases de datos relacionales.

El sistema *ROLAP* utiliza una arquitectura de 2 niveles:

- El nivel de base de datos: usa bases de datos relacionales para el manejo de los requerimientos de almacenamiento, acceso y obtención del dato.
- El motor analítico: proporciona la funcionalidad analítica de la aplicación y es el que ejecuta las consultas multidimensionales de los usuarios.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor *ROLAP*, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios. La arquitectura *ROLAP* es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Además soporta técnicas de optimización de accesos para acelerar las consultas.

Por otro lado, es importante mencionar que existe un tercer sistema, *HOLAP* (Hybrid Online Analytical Processing), el cual no es realmente un sistema diferente, más bien es la habilidad para diseminar los datos a través de bases de datos relacionales y multidimensionales con la finalidad de obtener lo mejor

de ambos sistemas, esto es tener la capacidad de consultar a una base de datos relacional o multidimensional en un mismo sistema.

### **3.4 Cubo Multidimensional**

Es una estructura de almacenamiento que nos permite realizar las diferentes y posibles combinaciones para visualizar los resultados de una organización hasta un determinado grado de detalle, esta estructura es independiente al sistema transaccional de la compañía así como del *Datawarehouse* que se utilizo para su creación, una vez creado el *Cubo* éste adquiere su propio conjunto de datos, y facilita consultar información histórica de manera rápida y eficiente; ofreciendo la posibilidad de navegar y analizar los datos requeridos.

En la figura 8 se muestra un ejemplo de un cubo en donde se aprecian 3 dimensiones la de productos, de lugar y tiempo (más adelante se detallara la definición de dimensión). En esta figura se ejemplifica una consulta a un cubo que responde a la siguiente pregunta, ¿Cuántos libros fueron vendidos en Montana durante Julio?.



Figura 8. Gráfica de un cubo y la forma en la que hace una búsqueda

A continuación se describirán los componentes más importantes de un *cubo*

### **3.4.1 Medidas o Métricas**

Son características cualitativas o cuantitativas, de objetos que se desean analizar en las empresas. Las medidas cuantitativas están dadas por valores o cifras porcentuales. “Lo que se puede medir se puede controlar y mejorar”. Por ejemplo, se tienen las ventas en dólares, el número de unidades de inventario, las horas trabajadas, el promedio de piezas producidas, el porcentaje de aceptación de un producto, el consumo de combustible de un vehículo y etc.

### **3.4.2 Dimensión**

Son los objetos del negocio con los cuales se puede analizar la tendencia y el comportamiento del negocio. La definición de estas dimensiones se basa en políticas de la compañía o del mercado, es decir, cómo interpretan o clasifican su información para segmentar el análisis en sectores que por sus características comunes facilitan la observación y el análisis. Para poder definir las dimensiones requeridas dentro de los *Cubos* multidimensionales, se debe responder a las siguientes preguntas:

- ✓ ¿Cuándo? → Permite hacer un análisis a través del tiempo y visualizar de manera comparativa el desempeño del negocio, en este caso nos permite seleccionar épocas de la historia para determinar el comportamiento en un momento dado.
- ✓ ¿Dónde? → Nos ubica en un área física o imaginaria donde se están llevando a cabo los movimientos que se desean analizar, estos lugares pueden ser zonas geográficas, bodegas de almacenamiento de mercancía, divisiones hacia el interior de la organización, centros de costo, clasificación de las cuentas contables, etc.
- ✓ ¿Qué? → Es el objeto del negocio o el objeto de interés para determinada área de la compañía, para estos casos se tienen los productos y/o servicios, la materia prima como elemento de interés para la división de abastecimientos, los vehículos para la sección de transportes, las maquinas para el área de producción, etc.

- ✓ ¿Quién? → En esta dimensión se plantea una estructura de los elementos que inciden directamente sobre el objeto de interés, en estos casos se hace referencia al área comercial o de ventas, a los empleados de la organización cuando se está realizando un análisis a nivel del talento humano, etc.
- ✓ ¿Cuál? → Es hacia donde está enfocado el esfuerzo de la organización o una determinada área del negocio, para hacer llegar los productos o servicios. En esta dimensión se trabaja con los clientes que pueden ser internos o externos a la organización. Por ejemplo, para el caso de un cliente interno, se puede tener un análisis de la cantidad de servicios que ofrece el departamento técnico a las áreas de negocio de la compañía y poder determinar qué áreas solicitan mayor soporte y quienes están aprovechando el recurso con que se cuenta.

La dimensión de tiempo es la más importante ya que virtualmente se garantiza que cada *Datawarehouse* tendrá una tabla dimensional de tiempo, debido a la perspectiva de almacenamiento histórica de la información. Usualmente es la primera dimensión en definirse, con el objeto de establecer un orden, ya que la inserción de datos en la base de datos multidimensional se hace por intervalos de tiempo, lo cual asegura un orden implícito.

En la siguiente tabla podemos ver un ejemplo del tipo de información que podemos obtener de cada dimensión:

CUANDO	DONDE	QUE	QUIEN	CUAL	MEDIDA	
Tiempo	Zona	Producto	Vendedor	Cliente	Ventas	Dimensiones
Año	Región	Tipo de Producto	Fuerza de Ventas	División de Clientes	Unidades	Nivel 1
Semestre	País	Grupo de Producto	Categoría	Tipo de Cliente	Horas	Nivel 2
Trimestre	Ciudad	Sub Grupo	Vendedor	Cliente	Litros	Nivel 3
Mes	Punto de Venta	Línea del Producto			Cajas	Nivel 4
Semana		Producto				Nivel 5

### 3.4.3 Niveles

Un nivel representa un grado particular de *agregación* dentro de una dimensión; cada nivel sobre el nivel base representa la sumarización total de los datos desde el nivel inferior. Para un mejor entendimiento, veamos el siguiente ejemplo: consideremos una dimensión Tiempo con tres niveles: Mes, Semestre, Año. El nivel Mes representa el nivel base, el nivel Semestre representa la sumarización de los totales por Mes y el nivel Año representa la sumarización de los totales para los Semestres.

Agregar niveles de sumariazi3n otorga flexibilidad adicional a usuarios finales de aplicaciones para analizar los datos. Ya que entre m1s niveles contenga un *Cubo* mayor es el detalle que se puede saber de algo en espec1fico y as1 ayudar al pleno entendimiento de las personas responsables de tomar las decisiones bas1ndose en el *Cubo*.

### 3.4.4 Categor1as

Las categor1as son datos individuales que describen o clasifican detalles de su organizaci3n. Las categor1as pueden ser diferentes niveles de informaci3n dentro de una dimensi3n o pueden ser agrupadas en m1s categor1as generales.

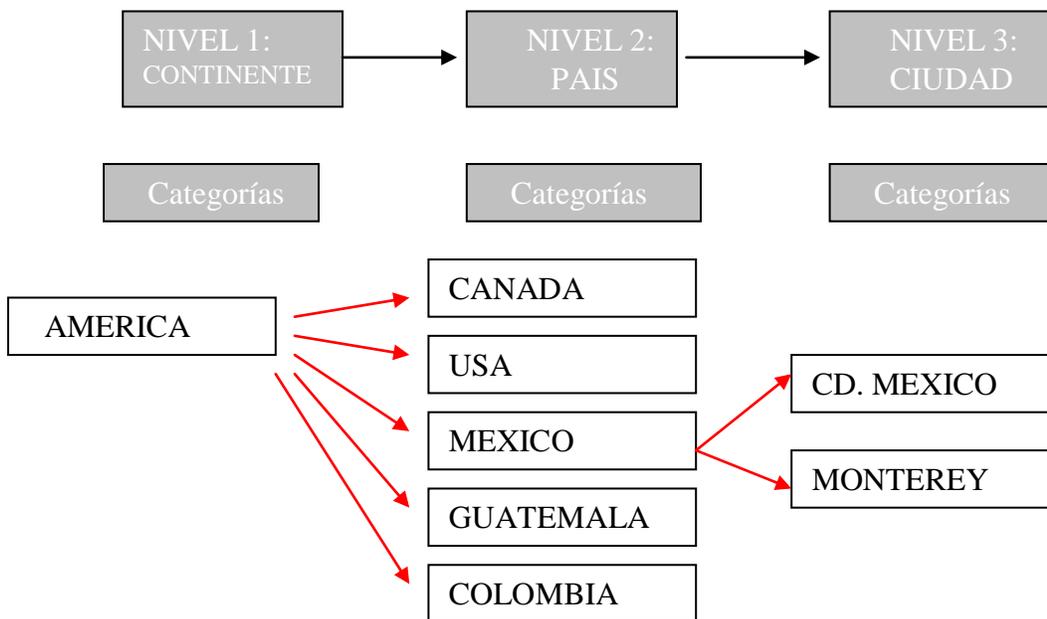


Figura 9. Relaci3n entre las categor1as y los niveles

Cada nivel en una dimensi3n tiene su propio conjunto de categor1as. Las categor1as son esenciales en los sistemas *Business Intelligence* porque ellas son los datos que se muestran en las columnas y renglones. El grado de detalle en las categor1as por cada nivel se profundiza con el siguiente nivel. Y se encuentra el m1ximo detalle con el nivel mas bajo. De lo anterior se puede ver que existen dos t1rminos que describen la relaci3n entre los tipos de categor1as:

Categoría padre: Una categoría en un nivel superior a otra categoría dentro de la misma dimensión

Categoría hija: Una categoría en el nivel inferior a otra categoría en la misma dimensión.

### 3.4.5 Análisis Multidimensional

Es la acción de combinar las diferentes áreas de la organización, de tal manera que permita ubicar cierto tipo de información que revele el comportamiento del negocio.

CUANDO	DONDE	QUE	QUIEN	CUAL	MEDIDA
2001	América	Farmacéutica	Gerente 1	Institucional	Ventas
1er Semestre	América	Urológico	Supervisor 2	Supermercados	Unidades
2do Trimestre	México	Viagra	Juan Pérez	Comercial 2	Cajas
Junio	DF				Litros

El análisis multidimensional que se puede obtener del ejemplo anterior es: el “numero de cajas (unidades)” de “medicamentos urológicos” vendidos específicamente a los “clientes institucionales” ubicados en una zona geográfica denominada “América” para el “primer semestre del año 2001”, por el grupo de vendedores clasificados como “Supervisor 2”.

En general, el análisis multidimensional se puede hacer combinando todas las posibles opciones presentadas en los *Cubos* multidimensionales, en busca de obtener tendencias de la información y cambios que permita visualizar oportunidades de negocio o situaciones a cambiar o mejorar.

Una vez que hemos expuesto los conceptos relacionados con el *Business Intelligence*, en el siguiente capítulo se presentará el caso práctico del desarrollo de un *Cubo* para lo cual se emplearan las herramientas de COGNOS, con el fin de generar reportes y aplicar los conceptos que se mencionaron anteriormente.

## CAPITULO 4.- Caso Práctico

Como se mencionó en la introducción, en este capítulo desarrollaremos un caso práctico con la herramienta de COGNOS (Powerplay), la cual es una de las herramientas mencionadas en el Capítulo 1. Empezaremos describiendo las características de la herramienta que utilizaremos para la explotación del *Cubo*, así como una breve explicación de la extracción, transformación y carga de los datos. También crearemos un modelo multidimensional y por último haremos algunos reportes de prueba.

### **4.1 Características de Powerplay**

Powerplay es una solución *OLAP* de COGNOS para *Business Intelligence* que ofrece un ambiente de análisis y reportes para la medición del rendimiento de los negocios, de tal forma que las personas involucradas en la empresa, tanto dentro como fuera de ella, puedan mejorar su productividad. Al tener acceso a información corporativa crítica, los usuarios explotan y analizan los datos desde cualquier ángulo y en cualquier combinación, con la posibilidad de elaborar reportes fácilmente y compartirlos con otros al publicarlos en Internet.

PowerPlay le permite a cualquier persona, más allá de los creadores de reportes tradicionales y analistas de negocios, hacer su propio análisis multidimensional y crear reportes *OLAP*, en ambientes Windows, Excel y de la Web.

La principal ventaja de esta herramienta es la fácil exploración de datos, ya que permite:

- Explorar los datos de la misma forma en la que se piensa, se puede analizar las tendencias de negocios organizando los datos según clasificaciones lógicas, tales como, períodos fiscales, regiones de ventas o grupos de productos, lo que tenga más sentido dependiendo de cada ambiente particular.
- Descubrir todos los hechos, explorando información multidimensional con el simple uso del ratón, profundizando con crecientes niveles de detalle, para visualizar los datos en diferentes dimensiones, tales como, ventas por región, por producto, etc. Observar y analizar gráficamente las relaciones entre los datos y moverse de una pantalla a otra fácilmente.
- Profundizar a través de los detalles: adentrándose de un nivel de detalle al otro, del resumen al detalle del segundo nivel de transacción mediante el *drill down* y el *drill up*, o de un *Cubo* (**PowerCube**) a otro, hasta encontrar la información deseada.

- Crear reportes dinámicos, convertir los análisis de precisos e impactantes a reportes de medición del rendimiento del negocio. Elaborar reportes de indicadores clave del rendimiento (Key Performance Indicators o KPI) y reportes tipo hoja de resultados, que ayuden a medir, monitorear y mejorar el negocio.
- Acceder a toda la empresa y mucho más, facilitar a los usuarios acceso a reportes en formato PDF a través de la red mundial. Los usuarios pueden convertir los PDF en reportes dinámicos para exploración y análisis posteriores, cortando, profundizando, ordenando, clasificando, filtrando, etc.

Por otro lado Cognos PowerPlay cuenta con una herramienta de modelado y creación de *Cubo* llamada ***PowerPlay Transformer*** con la que podemos hacer, de una manera muy sencilla, un modelo de algún problema y convertirlo a un *Cubo* para el análisis multidimensional. A continuación conoceremos un poco más de PowerPlay Transformer.

Esta herramienta de modelado no requiere programación, el *Cubo* que crea se conoce como PowerCube el cual es usado por los usuarios para el análisis multidimensional de una forma rápida y eficiente. Las bases de datos tradicionales presentan los datos como una serie de tablas bi-dimensionales que muestran la relación entre variables mientras que un PowerCube expande el número de dimensiones de dos, a tres o más.

PowerCube tiene las siguientes características:

- Son estáticos, significa que deben ser modificados periódicamente mediante la recreación del *Cubo*.
- Almacenan los datos en niveles sumarios. Esto es para almacenar mas detalles de la información en el *Cubo*, sin embargo, esto no es recomendado porque incrementa el tamaño del *Cubo* considerablemente.
- Fácil de personalizar, comienza con la valoración de los datos y la decisión de cómo se organizaran para soportar el modelo y tipo de análisis del negocio. Después se crea un modelo que especifique el lugar de los datos y la forma en la que se desea reestructurarlos. Finalmente, se crea un PowerCube para que los usuarios de PowerPlay puedan mirar el *Cubo*.

## Requerimientos del sistema

Para instalar PowerPlay, es necesario que la computadora cumpla con los siguientes requerimientos

Requerimiento	Especificación
Sistema operativo	Windows 2000, Windows Server 2003, or Windows XP
RAM	Como mínimo 512 MB
Espacio en disco	* Cognos products 260 MB (plus 150 MB free space on the drive that contains the temporary directory used by the installation program) * 105 MB for ObjectStore
CD-ROM	Un CD-ROM drive

### 4.2 Metodología

A continuación presentamos de manera concisa el proceso para la elaboración del *Cubo*.

1. Sistemas Transaccionales o fuentes de datos:

Permite acceder datos desde todos los sistemas operativos, tales como OS/400, Unix, Linux, Windows, etc. sobre las más comunes bases de datos, tales como: Microsoft SQL Server 2000, Oracle, Informix, DB2.

2. ETL:

Este proceso de Extracción, Transformación y Carga se hace utilizando el Data Transformation Services (DTS) de Microsoft SQL Server 2000, DecisionStream.

3. *Datawarehouse* o Repositorio corporativo:

El repositorio de datos es diseñado, poblado, validado y puesto en producción sobre las bases de datos más comunes, tales como: Microsoft SQL Server 2000, Oracle, Informix, DB2, etc.. Para Windows o Unix

4. Motor *OLAP*

La construcción de los *Cubos* multidimensionales se hace usando el Transformer de Cognos, Analysis Manager, etc. el cual con su tecnología *MOLAP*, permite manejar grandes volúmenes de datos y entrega *Cubos* altamente comprimidos, transportables y muy eficientes

5. Visualización de la Información:

Mediante la gran versatilidad y facilidad de uso del Cognos Powerplay, se hace el despliegue y análisis de información. Con la misma herramienta se efectúan la elaboración de los diferentes informes e indicadores de gestión.

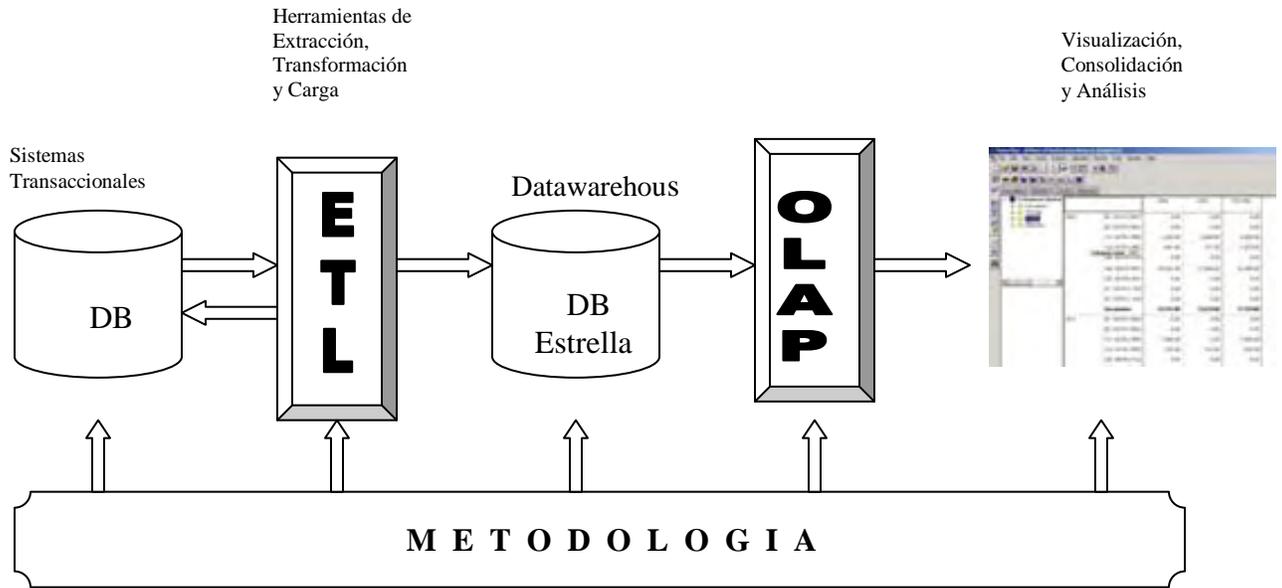


Figura 10. Forma grafica la metodología del Business Intelligence

### 4.3 Caso Práctico

Cuando buscamos oportunidades para *Business Intelligence* en una organización, identificar cual es la información de mas valor es un requerimiento indiscutible y fundamental, de forma mas simple, se necesita identificar las medidas sobre la cuales se quiere enfocar.

En nuestro caso práctico ocuparemos información de una empresa la cual por cuestiones de privacidad llamaremos GEM, esta empresa se dedica a la importación y distribución de productos alimenticios en todo el país, ellos tienen una gran variedad de productos los cuales están clasificados de dos diferentes formas. La primera clasificación de los productos es la siguiente: línea, tipo, familia, sabor. Y la segunda que solo los divide por “tipo de productos” las cuales solo tiene dos categorías (Dry y Frozen). Para el gerente tener el control de ventas por cada producto, es una tarea relativamente fácil ya que el sistema transaccional le puede dar esta información. Pero cuando trata de generar un reporte en el cual se hagan comparaciones entre las diferentes clasificaciones se convierte en una tarea que puede tomar varios días de trabajo. Esto nos da la pauta para interpretar estas necesidades como nuestras primeras dimensiones identificadas “PRODUCTO” y “TIPO PRODUCTO”.

De igual forma al seguir platicando del negocio no encontramos con que tienen una agrupación de clientes y estructura de ventas, las cuales se encuentran en la misma situación cuando tratan de generar los reportes relacionados a estas clasificaciones y mas aun cuando trata de generar reportes que involucren las agrupaciones anteriores conestas, entonces nos damos cuenta de que estas dos entidades también se convertirán en dimensiones.

Los clientes tienen la siguiente clasificación: canal mayor, canal menor, cadena y tienda. Y la estructura de ventas cuenta con la siguiente clasificación: Gerente, Supervisor y Vendedor. Todos los reportes son requeridos con las siguientes medidas:

- “EU” que es una medida interna, de la cual no se tiene mucha información porque el corporativo le asigna un valor a cada producto, pero no se sabe exactamente como se extrae.
- “Precio” que es el valor en pesos mexicanos de cada producto vendido.
- “Unidades” que es el numero de unidades vendidas.

Toda esta información fue proporcionada por cada una de las áreas involucradas en el ciclo de ventas de la empresa, con lo cual se podrá dar respuesta a las siguientes preguntas que ellos hacen comúnmente y que el *Cubo* deberá poder responderles. ¿Cómo se estaban vendiendo los productos?, ¿Qué Zonas del país eran las mas rentables? , ¿En qué porcentaje estaban creciendo las ventas por producto anualmente? , ¿Qué tipo de productos se consumían más? , tomando en cuenta que manejas dos tipos de productos “DRY y FROZEN”, etc. El cubo deberá poder responder a todas las preguntas relacionadas con las dimensiones que tendrá nuestro cubo así como las preguntas que involucren a mas de una dimensión por ejemplo ¿Cuántos productos de la línea x se venden en la zona y por el gerente z?

En base a la información recabada y con la que contaban en su sistema transaccional se hizo el análisis correspondiente para llegar a determinar el siguiente modelo estrella que les permitiera tener las respuestas a sus cuestionamientos.

## ESQUEMA ESTRELLA

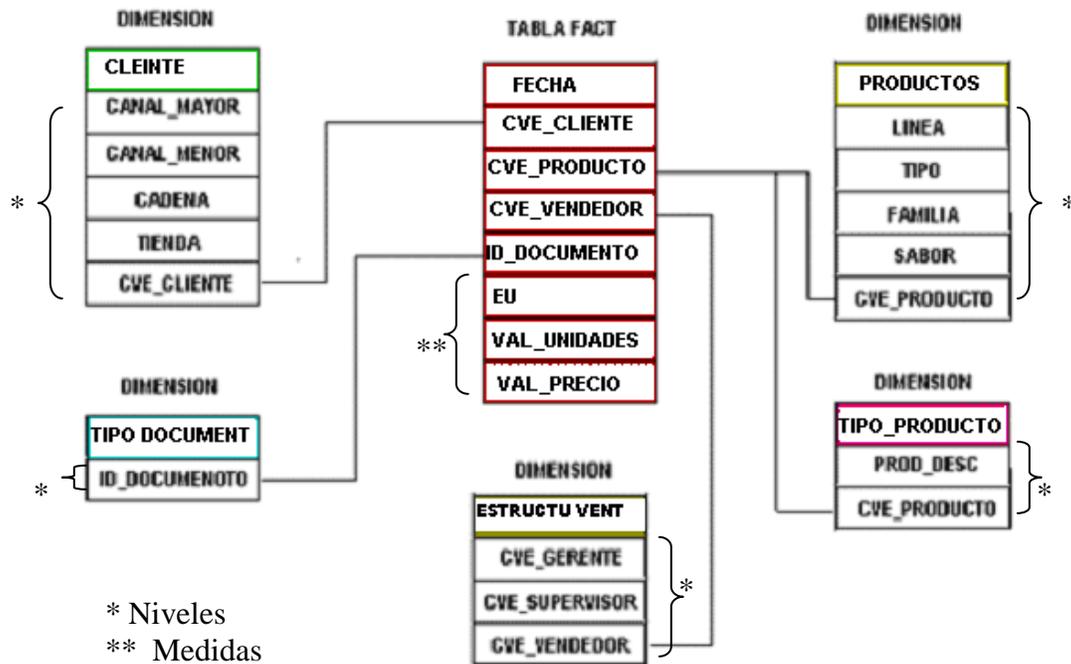


Figura 11. Modelo estrella creado para nuestro cubo

Una vez teniendo el modelo estrella procedemos con la Extracción, Transformación y Carga (ETL por sus siglas en inglés) de la información de la base de datos transaccional a nuestro *Datawarehouse*, la base de datos transaccional tiene el siguiente nombre JDE\_PRODUCTION y a nuestro *Datawarehouse* le llamaremos DSS, de esta manera ya tenemos definido el origen (JDE\_PRODUCTION) de la información y el destino (DSS) o *Datawarehouse* ambas bases de datos están en SQL SERVER. Para la extracción de la información utilizaremos (SQL Server Enterprise Manager).

### 4.3.1 ETL (Extraction, Transformation and Load)

Este proceso se realizara con la herramienta ENTERPRISE MANAGER (SQL Server) la cual se muestra en la figura 12. Esta herramienta se empleara para extraer la información y cargarla al *Datawarehouse*, aquí es importante recordar que este proceso es regenerativo, ya que cada vez

que se corre, toda la información del *Datawarehouse* se borra y se vuelve a cargar. Este proceso se compone por transformaciones que mas adelante describiremos.

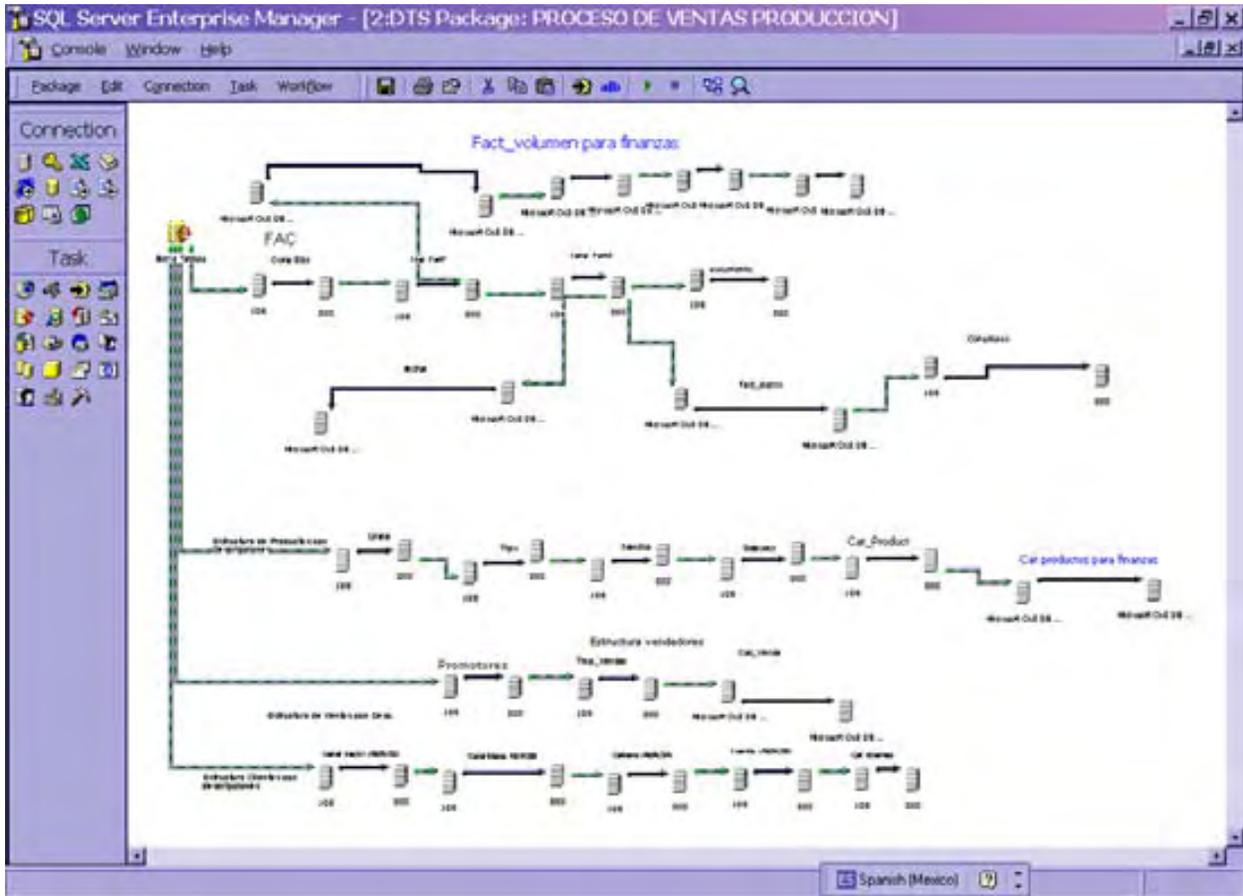
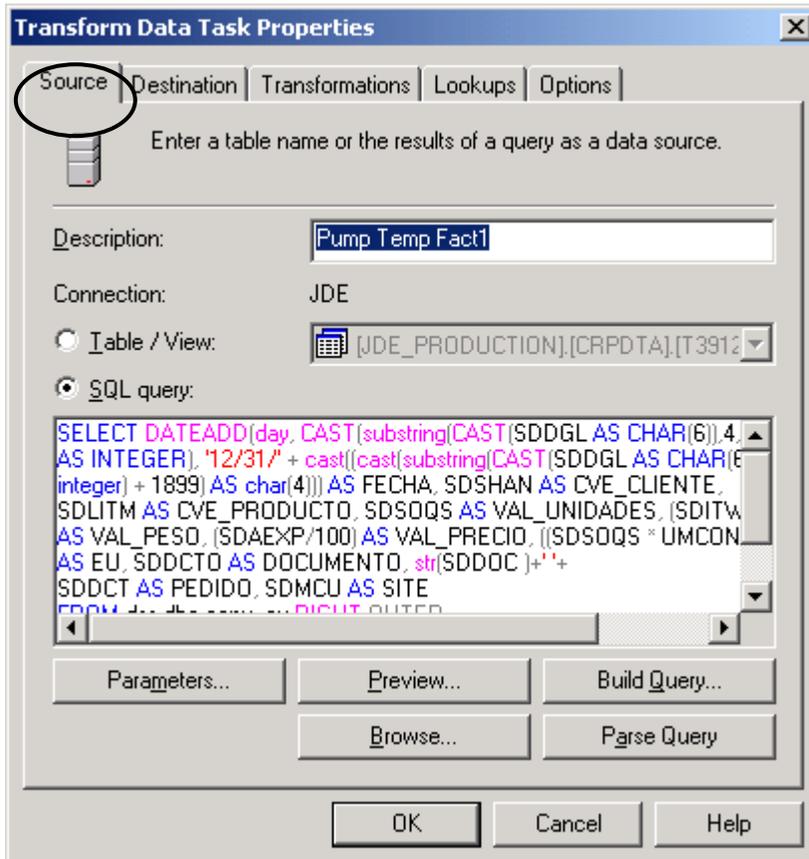


Figura 12. Package(Paquete) en ENTERPRISE MANAGER(SQLServer)

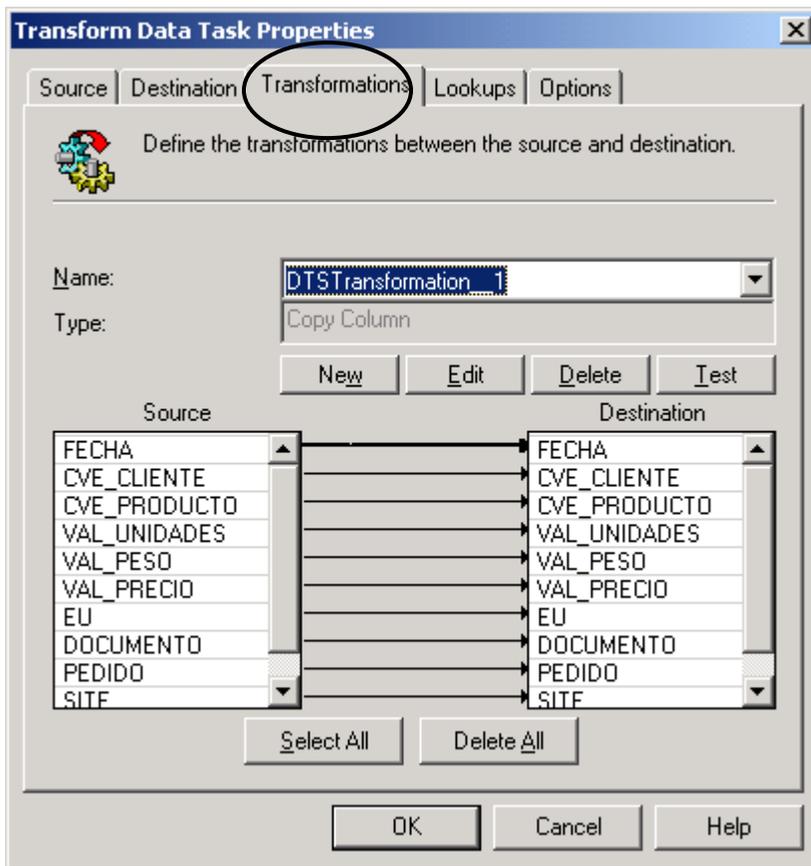
Cada transformación cuenta principalmente con tres partes fundamentales: SOURCE, DESTINATION, TRANSFORMATIONS, las cuales explicaremos brevemente con las siguientes figuras:



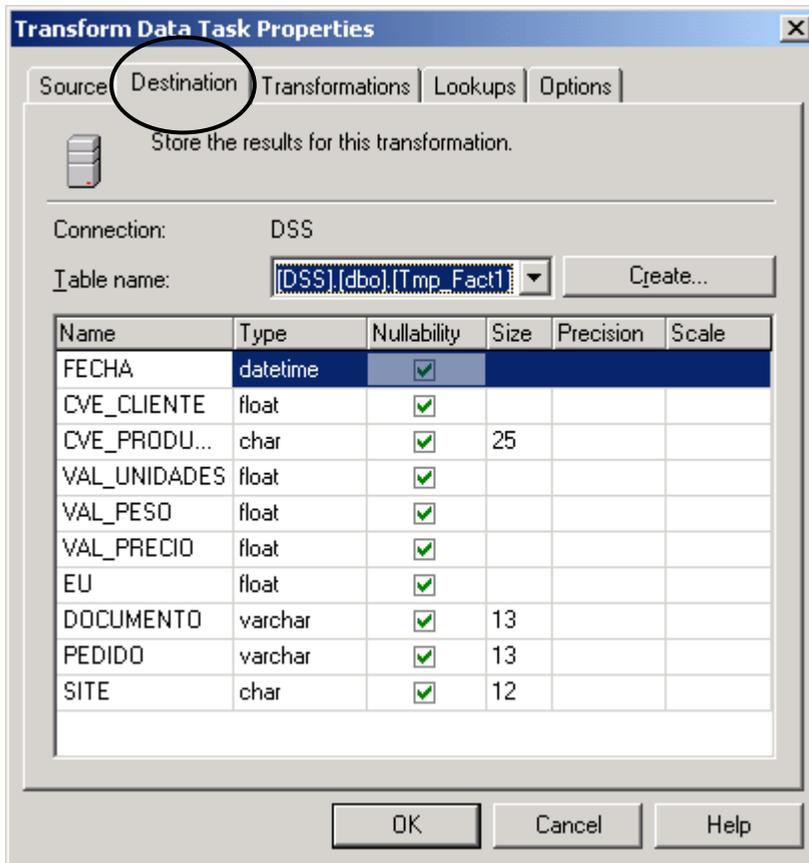
1.- SOURCE en esta parte se define el origen de los datos que puede ser una tabla completa o un Query de cualquier complejidad (Soportado por la base de datos),archivos de diferente formato, en este caso la figura muestra que la información viene de un Query.



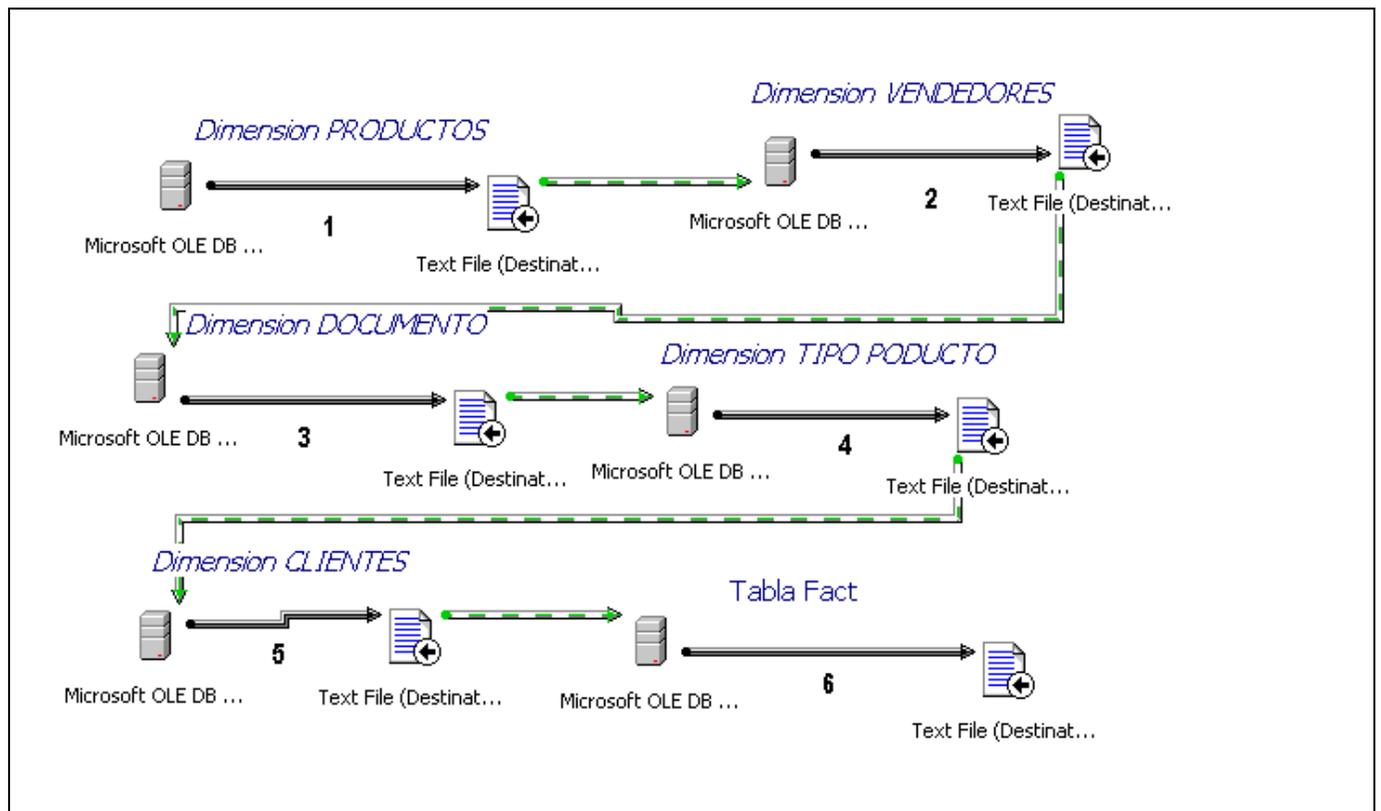
2.- TRANSFORMATIONS, aquí se define cómo se almacenará la información en el destino y en qué campos; también se puede agregar información que no traemos en el Origen pero que se puede derivar de la misma como cálculos, marcas, referencias. Para lo cual podemos utilizar los Activex Script en los que podemos agregar código de programación y de esta manera incluir información transformada, si es necesario. En la figura de abajo se puede ver que sólo se copia la información del Origen al Destino. Como lo comentamos anteriormente en esta parte consolidaremos la información ya que esta puede venir de diferentes orígenes y en nuestro Datawarehouse no puede entrar información que no este homogenizada. Ya que será el único orígenes para la creación de nuestros cubos.



3.- DESTINATION, en esta parte le indicamos en qué lugar ingresara la información, necesariamente estaríamos hablando de un Datawarehouse, es importante resaltar la flexibilidad de las herramientas de ETL ya que hay empresas que no tienen una base de datos relacional como Datawarehouse por el costo de estas sin en cambio nos permiten entregar la informaciones archivos planos que posteriormente utilizaremos para generar el cubo. En esta parte es cuando apuntamos hacia el *Datawarehouse* y es de esta manera como cargamos la información.



Una vez que ya tenemos la información en el *Datawarehouse* utilizaremos nuevamente la herramienta ETL para cargar la información de las tablas del *Datawarehouse* a archivos txt que es uno de los formatos reconocidos por nuestra herramienta OLAP (POWER PLAY TRANSFORMER). En la siguiente figura se puede ver el proceso que realiza el Packages para cargar la información en archivos txt.



#### 4.3.2 Modelo y Creación del Cubo (en Power Play Transformer)

Los principales elementos que componen el modelo son: Dimensión Map, Data Sources, Measures y PowerCubes. Las cuales se detallaran a continuación.

##### Dimension Map

Es en esta zona en donde se definen el número de dimensiones y la cantidad de niveles por cada dimensión, esta es la parte en donde se hace el mapeo de nuestro diagrama estrella.

##### PowerCubes

También se define el lugar en dónde se almacenaran el Cubos (.mdc)

**Ubicación física del Cubo:** c:\program files\cognos\inbi\tesina\tesina\practica\t\_ventas.mdc

##### Measures

Aquí se definen las características de las medidas con las que contará el *Cubo*, las cuales vienen de la tabla de Hechos.

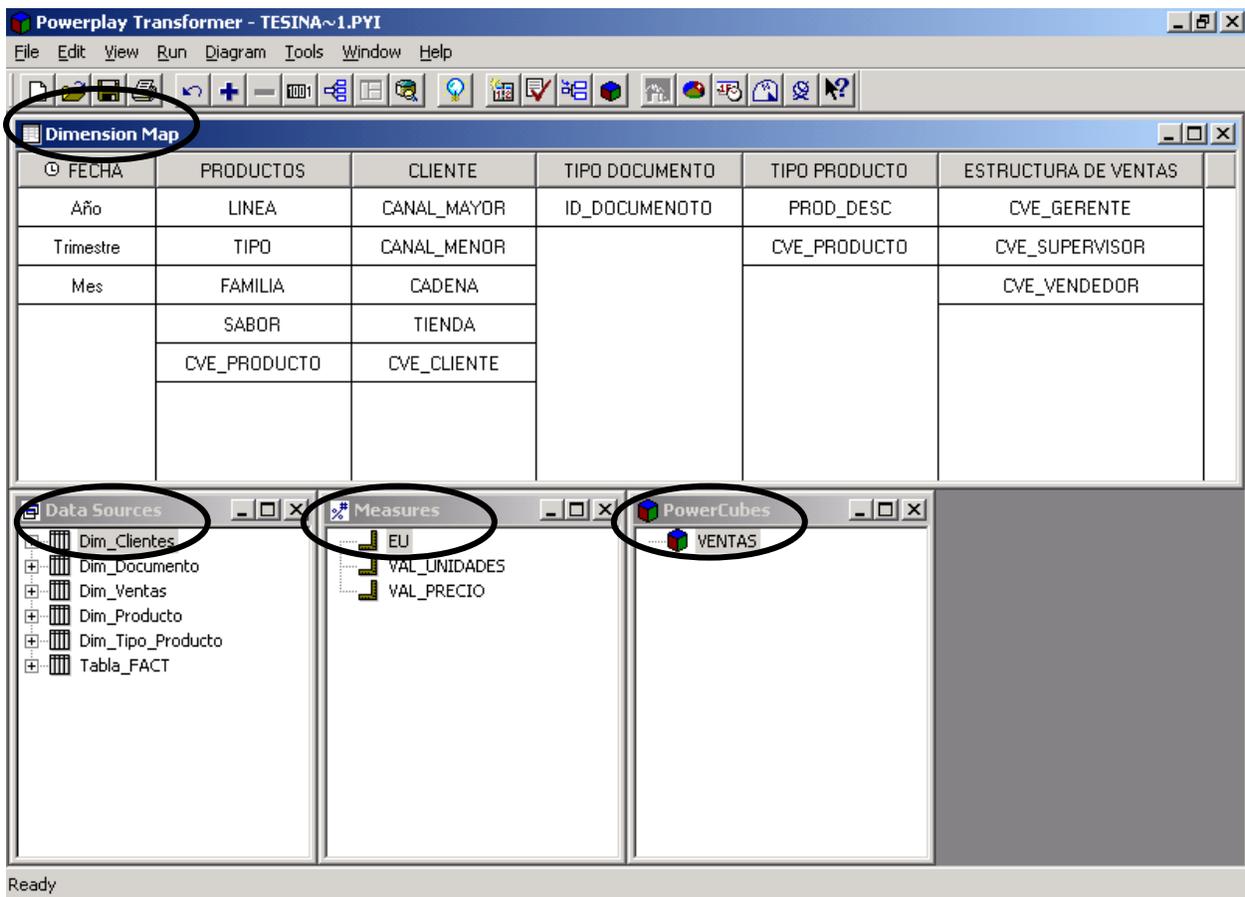
<b>Medidas</b>	<b>Descripción</b>
EU	Medida utilizada por la empresa, cada producto tiene diferente valor en EU
VAL_UNIDADES	Unidades vendidas
VAL_PRECIO	Valor en \$\$

### **Data Sources**

En esta parte se define de dónde se extraerá la información que el modelo utilizará para la generación del *Cubo*. Aquí es en donde utilizaremos los archivos txt que creamos a partir de nuestro *Datawarehouse*.

<b>Nombre Dimension</b>	<b>Ubicación Data Source</b>
<b>Clientes</b>	c:\program files\cognos\inbi\tesina\tesina\practica\tpractica2\dim_clientes.txt
<b>Tipo Documento</b>	c:\program files\cognos\inbi\tesina\tesina\practica\tpractica2\dim_documento.txt
<b>Estructura de Ventas</b>	c:\program files\cognos\inbi\tesina\tesina\practica\tpractica2\dim_ventas.txt
<b>Productos</b>	c:\program files\cognos\inbi\tesina\tesina\practica\tpractica2\dim_productos.txt
<b>Tipo Producto</b>	c:\program files\cognos\inbi\tesina\tesina\practica\tpractica2\dim_congelado.txt
<b>Hechos</b>	c:\program files\cognos\inbi\tesina\tesina\practica\fact.txt (De la tabla hechos)

Basándonos en el modelo estrella antes creado, se definió el siguiente modelo en Transformer PowerPlay. En la figura se muestra la herramienta (TRANSFORMER) y los elementos antes mencionados.



Como podemos ver el modelo cuenta con todas las dimensiones que se muestran en el modelo estrella.

### **4.3.3 Herramientas Front- End y Análisis de la información (Reportes)**

Para llevar a cabo los procesos de análisis y generación de reportes, las herramientas de *front-end* recogen los datos del *Datawarehouse* o del *Cubo* multidimensional y presentan éstos a los usuarios en forma de informes y vistas interactivas, las herramientas pueden estar clasificadas en dos categorías;

- Herramientas que responden a un paradigma de *reporting*: Estas herramientas tienen la inclinación de producir informes tabulares, muchas de las cuales proporcionan interfaces *web browser* para diseñar y visualizar informes adaptables para implementaciones de amplia escala. Estas herramientas frecuentemente tienen una gran capacidad de impresión y programación.
- Herramientas que obedecen a un paradigma de exploración de datos *ad hoc* multidimensionales: Estas herramientas consultan directamente cubos *OLAP*, dando soporte a análisis interactivo y vistas complejas con gráficos, aunque este beneficio no siempre resulta de importancia para los usuarios que normalmente necesitan capacidades robustas de formateo y distribución

De esta manera, cuando se diseñan sistemas de *Business Intelligence*, las compañías deben invertir tiempo en comprender los amplios objetivos de *Business Intelligence* y las necesidades específicas de los usuarios.

Una vez creado el *Cubo* se procede al análisis de la información para lo cual utilizaremos la herramienta de COGNOS llamada POWERPLAY que cae dentro de la segunda definición de herramientas *Front-End*, ya que se conecta directamente, lo que nos permite hacer un análisis *ad hoc*.

La siguiente figura muestra como es que PowerPlay presenta la información del *Cubo*.

PowerPlay - [PPlay1 of ventas (Explorer)]

File Edit View Insert Explore Calculate Format Tools Window Help

FECHA PRODUCTOS CLIENTE TIPO DOCUMENTO TIPO PRODUCTO ESTRUCTURA DE VENTAS EU

C:\Program Files\Cognos\ir

- FECHA
- PRODUCTOS
- CLIENTE
- TIPO DOCUMENTO
- TIPO PRODUCTO
- ESTRUCTURA DE VENTAS
- MEASURES

	SNACKS (new)	MEALS (new)	GREEN GIANT (new)	DOUGH (new)	PRODUCTOS
2004	98,219.00	0.00	10,050.00	17,125.00	316,259.00
2005	71,083.00	2,513.00	6,814.00	9,650.00	252,061.00
<b>FECHA</b>	<b>169,302.00</b>	<b>2,513.00</b>	<b>16,864.00</b>	<b>26,775.00</b>	<b>568,320.00</b>

Se muestran las Dimensiones así como las medidas.

SNACKS (new)

Con lo que hemos construido hasta el momento se pueden elaborar una gran variedad de reportes y consultas de los cuales, a continuación, se presentan algunos de ellos. El criterio que se tomo para elegirlos no es ninguno en particular, ya que esto depende de las necesidades de quien este utilizando el *Cubo* y de el conocimiento que se tenga de la herramienta visualizadora (Powerplay)

**Reporte 1.** Reporte que muestra las ventas por sub-canal de cada familia de producto.

		 <b>REPORTES CLIENTES V8 PRODUCTOS</b> INFORMACION ACTUALIZADA HASTA: 08/09/2005					
		DESSERT & SPECIALTY MIX (new)	SNACKS (new)	MEALS (new)	GREEN GIANT (new)	DOUGH (new)	HÁAGEN DAZS (new)
Supermarkets	Supermarkets	57,215.00	23,768.00	0.00	15,202.00	13,187.00	60,045.00
	Supermarkets	57,215.00	23,768.00	0.00	15,202.00	13,187.00	60,045.00
Distributors & Wholesalers	Wholesalers	37,612.00	0.00	0.00	0.00	0.00	0.00
	Distributors	1,727.00	0.00	0.00	0.00	0.00	0.00
	Distributors & Wh	39,339.00	0.00	0.00	0.00	0.00	0.00
Alternate	Other Alternate	7,288.00	18,758.00	0.00	62.00	140.00	5,371.00
	Video stores	0.00	0.00	0.00	0.00	0.00	24,336.00
	C-Stores	0.00	0.00	0.00	0.00	0.00	728.00
	QSR	0.00	0.00	0.00	0.00	0.00	137.00
	Alternate	7,288.00	18,758.00	0.00	62.00	140.00	30,671.00
C-Stores	C-Stores	0.00	0.00	0.00	272.00	278.00	4,338.00
	Restaurants	0.00	0.00	0.00	0.00	0.00	18.00
	Shops	0.00	0.00	0.00	3.00	4.00	46.00
	C-Stores	0.00	0.00	0.00	275.00	282.00	4,401.00
Clubs	Clubs	0.00	126,776.00	2,513.00	720.00	5,315.00	58,048.00
	Clubs	0.00	126,776.00	2,513.00	720.00	5,315.00	58,048.00
Independent bakeries	Independent bakery	7,963.00	0.00	0.00	0.00	9.00	312.00
	Pastry shops	0.00	0.00	0.00	0.00	354.00	0.00
	Independent bakery	7,963.00	0.00	0.00	0.00	363.00	312.00
Food service	Restaurants	2.00	0.00	0.00	553.00	338.00	22,973.00
	Hotels	25.00	0.00	0.00	51.00	7,009.00	29,820.00
	Other FS	6.00	0.00	0.00	0.00	53.00	2,536.00
	Independent bakery	0.00	0.00	0.00	0.00	17.00	11.00
	Supermarkets	0.00	0.00	0.00	1.00	4.00	13.00
	Distributors	0.00	0.00	0.00	0.00	41.00	0.00
	Food service	33.00	0.00	0.00	606.00	7,462.00	55,353.00
Shops	Grupo Tiendazs	0.00	0.00	0.00	0.00	15.00	24,199.00
	Shops	0.00	0.00	0.00	0.00	11.00	7,873.00
	Shops	0.00	0.00	0.00	0.00	26.00	32,072.00
Franchise chains	Restaurants	0.00	0.00	0.00	0.00	0.00	226.00
	Franchise chains	0.00	0.00	0.00	0.00	0.00	226.00
<b>TOTAL CLIENTE</b>		111,838.00	169,302.00	2,513.00	16,864.00	26,775.00	241,028.00

**Reporte 2.** En este reporte se puede ver el porcentaje cumplido (de ventas) respecto al año anterior, de esta manera pueden ver qué medidas tomar si fuese muy bajo el porcentaje. En este reporte también podemos mencionar el poder que tiene la herramienta ya que el porcentaje cumplido no es una medida del cubo pero la podemos calcular desde el momento que estamos generando el reporte .

		PORCENTAJE CUMPLIDO		
		INFORMACION ACTUALIZADA HASTA: 08/09/2005		
		2004	2005	% CUMPLIDO
Supermarkets	Supermarkets	96,670.00	72,747.00	75.25%
Distributors & Wholesalers	Wholesalers	24,088.00	13,524.00	56.14%
	Distributors	1,727.00	0.00	0.00%
	Distributors & Wh	25,815.00	13,524.00	52.39%
Alternate	Other Alternate	11,380.00	20,239.00	177.85%
	Video stores	11,206.00	13,129.00	117.16%
	C-Stores	430.00	298.00	69.30%
	QSR	101.00	38.00	35.64%
	Alternate	23,117.00	33,702.00	145.79%
C-Stores	C-Stores	2,625.00	2,263.00	86.21%
	Restaurants	18.00	0.00	0.00%
	Shops	0.00	52.00	/0
	C-Stores	2,643.00	2,315.00	87.59%
Clubs	Clubs	111,196.00	82,176.00	73.90%
	Clubs	111,196.00	82,176.00	73.90%
Independent bakeries	Independent baki	6,017.00	2,287.00	37.68%
	Pastry shops	226.00	128.00	56.64%
	Independent baki	6,243.00	2,395.00	38.36%
Food service	Restaurants	14,036.00	9,830.00	70.03%
	Hotels	20,037.00	16,868.00	84.18%
	Other FS	1,481.00	1,114.00	75.22%
	Independent baki	11.00	17.00	154.55%
	Supermarkets	0.00	18.00	/0
	Distributors	0.00	41.00	/0
	Food service	35,565.00	27,888.00	78.41%
Shops	Grupo Tiendazs	12,531.00	11,683.00	93.23%
	Shops	2,338.00	5,546.00	237.21%
	Shops	14,869.00	17,229.00	115.87%
Franchise chains	Restaurants	141.00	85.00	60.28%
	Franchise chains	141.00	85.00	60.28%
<b>TOTAL CLIENTE</b>		<b>316,259.00</b>	<b>252,061.00</b>	<b>79.70%</b>

**Reporte 3.** Muestra un informe general por cada unidad de medida, se puede ver como se distribuyen las medidas por productos, clientes y estructura de ventas.

 <b>REPORTE GENERAL</b> INFORMACION ACTUALIZADA HASTA: 08/09/2005 05 <b>GM</b>			
	VAL_PRECIO	VAL_UNIDADES	EU
DESSERT & SPECIALTY MIX (new)	\$10,228,746.00	54302	41,252.00
SNACKS (new)	\$15,597,571.00	130657	71,083.00
MEALS (new)	\$435,255.00	3769	2,513.00
GREEN GIANT (new)	\$1,800,009.00	6806	6,814.00
DOUGH (new)	\$2,857,376.00	11593	9,650.00
HÄAGEN DAZS (new)	\$37,283,043.00	72915	120,749.00
<b>TOAL PRODUCTOS</b>	<b>\$68,202,000.00</b>	<b>280042</b>	<b>252,061.00</b>
Supermarkets	\$20,894,996.00	73738	72,747.00
Distributors & Wholesalers	\$3,107,955.00	18034	13,524.00
Alternate	\$9,685,218.00	32654	33,702.00
C-Stores	\$877,530.00	1945	2,315.00
Clubs	\$19,083,237.00	127592	82,176.00
Independent bakeries	\$1,131,689.00	2362	2,395.00
Food service	\$8,772,867.00	16322	27,888.00
Shops	\$4,611,677.00	7310	17,229.00
Franchise chains	\$36,831.00	85	85.00
<b>TOTAL CLIENTE</b>	<b>\$68,202,000.00</b>	<b>280042</b>	<b>252,061.00</b>
601 CENTRO SUR	\$18,551,656.00	52459	61,517.00
611 BFS	\$955,060.00	1980	1,980.00
605 INSTITUCIONAL	\$3,384.00	8	7.00
606 CUENTAS CLAVE	\$33,031,172.00	178510	132,516.00
609 VENTA INTERNA GM	\$69,833.00	308	334.00
603 PACIFICO	\$8,196,633.00	30886	29,279.00
602 NORTE	\$2,782,585.00	8581	9,199.00
608 SECOS	\$0.00	0	0.00
607 SHOPS	\$4,611,677.00	7310	17,229.00
<b>TOTAL ESTRUCTURA DE VENTAS</b>	<b>\$68,202,000.00</b>	<b>280042</b>	<b>252,061.00</b>

Es importante mencionar que se pueden generar una gran cantidad de reportes, pero sólo puse algunos para que tuviéramos una perspectiva clara del tipo de reportes se pueden hacer, también

vale la pena aclarar que estos reportes se generan muy rápido sólo se necesita tener claro que información tiene el *Cubo* y conocer la herramienta de PowerPlay.

Con esto queda concluida la breve exposición de lo que es el *Business Intelligence*, su relación con el *Datawarehouse* y todos los conceptos que involucra una metodología como esta, así como la manera en que se lleva a la práctica mediante herramientas existentes en el mercado, en este caso COGNOS.

## CONCLUSIÓN



Todos los sistemas transaccionales, incluyendo los tradicionales paquetes ERP (Oracle, SAP, PeopleSoft, jdedwards, etc.), están diseñados y optimizados para capturar información y soportar las cadenas de procesos. Esta función la logran de manera extraordinaria.

Sin embargo, la misma optimización que logra la eficiencia para transacciones, los hace totalmente inefectivos para extraer y analizar información por lo que es necesario contar con un sistema de *Business Intelligence* si lo que la empresa desea es analizar sus datos de una manera eficaz para que les sea de utilidad en la toma de decisiones.

Lo más importante de los datos es la interpretación que de ellos se obtiene mediante la consolidación y la estructuración de éstos, sólo de esta manera podremos decir que tenemos el control de la organización, que sabemos como se encuentran cada área que compone nuestra organización, los logros, las carencias, etc. Definitivamente el *Business Intelligence* llegó para dar una solución a estas necesidades, por lo que se ha convertido en parte indispensable de aquellas empresas que desean ser competitivas en el mercado y mantenerse a la vanguardia.

La inteligencia de negocios en los últimos años ha desarrollado tecnologías y herramientas que ayudan a las empresas a poder extraer y analizar la información que queda sepultada en los

bancos de datos. Es verdad que se pueden tener todos los datos del mundo pero también es verdad que por si solos pueden no representar algo importante que señale el estatus de la organización,

Las estrategias de Inteligencia de Negocios le permiten tener a la mano la información más relevante de la operación de una empresa de forma rápida, flexible y oportuna. Las herramientas que integran la Inteligencia de Negocios consolidan la información de las distintas áreas, mejorando el proceso de decisión.

Es muy importante dar a conocer esta herramienta a los estudiantes de la Licenciatura en Matemáticas Aplicadas y Computación, ya que, dado que el perfil del egresado está enfocado a la toma de decisiones, les proporciona conocimientos y muchas posibilidades en esta área. De esta forma trato de abrir un panorama más amplio para los egresados o estudiantes acerca de las necesidades existentes en el mundo laboral, por lo que trate que esta tesina, más que una pura investigación fuera una posibilidad para conocer esta área del manejo de la información, el área del *Business Intelligence* “El análisis de la información”.

La validez y la importancia del presente trabajo radican primordialmente en que el perfil del egresado de Matemáticas Aplicadas y Computación incluye su capacidad analítica para la toma de decisiones. Entonces, no sólo es importante que conozca las herramientas que para este fin existen, como lo considere en el momento de incluir la aplicación con COGNOS, sino que entiendan el concepto de un *Cubo*, cómo crear un modelo y cómo con esto se puede ayudar a una empresa a mejorar se eficiencia.

Además considero que actividades de este tipo requieren una elevada capacidad de análisis que permita entender un problema o una necesidad, plantear el modelo y solucionarlo, la cual es una de las principales cualidades de los egresados de la Licenciatura en Matemáticas Aplicadas y Computación.

## GLOSARIO

**AGREGACIÓN**, son resúmenes de datos precalculados que mejoran el tiempo de respuesta de consultas al tener preparadas las respuestas antes de que se planteen las preguntas.

**BUSINESS INTELLIGENCE**, conjunto de metodologías y tecnologías orientadas a potenciar la gestión inteligente de la empresa que permitan a los equipos directivos controlar los negocios.

**CUBO**, unidad de datos del análisis en línea de la información, Es una Base de Datos multidimensional

**DATAMART**, almacén de datos departamental, no son más que datos históricos pero tratados para evitar datos duplicados, atributos no existentes, etc.

**DATAWAREHOUSE**, es un conjunto de datos integrados orientados a un material que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración.

**DBA**, (Data Base Administrador) Administrador de bases de datos.

**DRILL DOWN**, operación que muestra información más detallada, baja al siguiente nivel.

**DRILL UP**, operación que muestra información menos detallada sube al siguiente nivel.

**DSS** (Decision Support Systems), sistemas de ayuda a la toma de decisiones.

**E/S** , Entrada y Salida

**EIS** (Executive Information Systems), sistemas de información para ejecutivos, independientes de aplicaciones convencionales, ergonomía de presentación y manipulación de datos y alta disponibilidad de información y análisis.

**ETL** (Extraction, Transformation and Loading), herramientas dedicadas a la extracción de los datos desde las fuentes donde estos se encuentren a los *Datawarehouse*.

**GRANULARIDAD**, nivel de detalle en la información del cubo, a mayor granularidad mayor detalle.

**MODELO ESTRELLA**, organización física de los *Datamarts* que facilita el acceso a los datos y al análisis. Se caracteriza por tener una tabla central de hechos rodeada por tablas de dimensiones que contienen información desnormalizada de los hechos.

**FOREIGN KEY**, representación de la primary key de una dimensión en la tabla de hechos.

**MDDB**, Base de datos multidimensional

**METADATOS**, datos acerca de datos que describen los contenidos del almacén de datos.

**MIT** , Instituto Tecnológico de Massachussets

**ODS**, (Store Data Operational) Almacén de datos operacionales

**POBLAR**, Carga de información, que se hace generalmente con las herramientas de extracción, transformación y carga.

**QUERY**: Las consultas son la forma principal de hacer solicitudes de información a la base de datos. Las consultas están compuestas de comandos que se envían a la base de datos en un formato predefinido, generalmente utilizando el formato SQL.

**TI**, Tecnología de información

## BIBLIOGRAFÍA

- Elizabeth Bitt, Micheñ Luckevich, Stacia Misner ; “BUSINESS INTELLIGENCE Técnicas de análisis para la toma de decisiones estratégicas”, Editorial Mc Graw Hill 2003
- Date, C. J., “Introducción a los sistemas de bases de datos”, Editorial Prentice Hall, Séptima Edición 2001.
- Richar Connelly , Robin McNell, Roland Mosimann ,”The Munltidimensional Manager “ COGNOS Canadá June 2001
- Orea Guzmán Arturo “El Doz como base de un *Datawarehouse* caso practico aduanas” Universidad Nacional Autónoma de México, Febrero 2004.
- **DATAWAREHOUSE** (Ultima fecha de revisión 25-abril-2005)  
<http://www.Datawarehouse.com>
- **BITAM** (Ultima fecha de revisión 25-abril-2005)  
<http://www.bitam.com.mx/TecAnalisis.htm>
- **FREEDATAWAREHOUSE** (Ultima fecha de revisión 25-abril-2005)  
<http://freeDatawarehouse.com/tutorials/dmtutorial/Dimensional%20Modeling%20Tutorial.aspx>
- **INTELIGENCIA APLICADA AL NEGOCIO** (Ultima fecha de revisión 25-abril-2005)  
[http://www.eldiarioexterior.com/conocimiento/docs/BI\\_Inteligencia\\_aplicada\\_al\\_negocio.pdf](http://www.eldiarioexterior.com/conocimiento/docs/BI_Inteligencia_aplicada_al_negocio.pdf)
- **IMPLEMENTATION OF *DESIGN TO PROFIT* IN A COMPLEX AND DYNAMIC BUSINESS CONTEXT** (Ultima fecha de revisión 25-arbil-2005)  
<http://herkules.oulu.fi/isbn9514264509/html/x1596.html>
- **DATA WAREHOUSING** (Ultima fecha de revisión 25-arbil-2005)  
<http://www.sqlmax.com/dataw1.asp>
- **DATAWAREHOUSING** (Ultima fecha de revisión 25-abril-2005)  
<http://apuntes.rincondelvago.com/datawarehousing.html>
- **INTELIGENCIA DE NEGOCIOS** (Ultima fecha de revisión 25-abril-2005)  
<http://www.microsoft.com/latam/business/bi/default.asp>