



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Campus Acatlán

**EL DATA WAREHOUSE
UNA HERRAMIENTA
EN LA ADMINISTRACION INFORMATICA
PARA LA TOMA DE DECISIONES**

TRABAJO DE INVESTIGACION
QUE PRESENTA
ANGELICA GONZALEZ NAVARRO

Para obtener el Título de
Licenciada en Matemáticas Aplicadas y Computación

Bajo la Opción de

Seminario de Taller Extracurricular
"Análisis de Planeación"

Asesor: Lic. Gerardo Roldán Ceballos

Naucalpan, Estado de México, 1998



**TESIS CON
FALLA DE ORIGEN**

267335



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.

CONTENIDO	I
Introducción	V
Capítulo I. Fundamento Estructural: El <i>Data Warehouse</i>	1
1.1 Antecedentes	3
1.2 Tipos de Sistemas de Información	3
1.2.1 Sistemas de Procesamiento Transaccional en Línea (OLTP)	4
1.2.2 Sistemas de Información Ejecutiva(EIS)	5
1.2.3 Sistemas para Soporte en la Toma de Decisiones(DSS)	7
1.2.4 Sistemas de Consultas de Datos	7
1.3 Qué es <i>Data Warehouse</i>	8
1.3.1 Características de un <i>Data Warehouse</i>	11
1.3.2 Componentes	11
1.3.3 Requerimientos	18
1.3.4 La Administración Informática y su relación con el <i>Data Warehouse</i>	20
1.4 El Almacenamiento de Información en un <i>Data Warehouse</i>	21
1.4.1 Datamart	22
1.4.2 Almacén de Datos Operacionales	23
1.4.3 Repositorio	23
1.4.4 Metadatos	24
Conclusiones Capítulo I	27
Notas Bibliográficas	29
Capítulo II. Definición del Modelo de <i>Data Warehouse</i>, Diseño y Construcción	31
2.1 Ciclo del Desarrollo del <i>Data Warehouse</i>	33
2.1.1 Etapas del <i>Data Warehouse</i>	33
2.1.2 Fase Inicial del <i>Data Warehouse</i>	48
2.2 La Solución Integral de <i>Data Warehouse</i> : Modelo Propuesto	50
2.3 La Minería de Datos	51
2.4 El <i>Data Warehouse</i> un Proceso Iterativo	54
Conclusiones Capítulo II	55
Notas Bibliográficas	57

Capítulo III. El Valor Agregado del <i>Data Warehouse</i> para la Toma de Decisiones Estratégicas	59
3.1 La utilidad del <i>Data Warehouse</i> en la Planeación Estratégica	61
3.1.1 Análisis de Rentabilidad	66
3.1.2 Valor Competitivo	74
3.1.3 Productividad	75
3.2 Beneficios del <i>Data Warehouse</i>	76
3.3 Participantes de Mercado para la Solución del <i>Data Warehouse</i>	78
3.4 Perspectivas del <i>Data Warehouse</i>	79
3.4.1 La Administración Informática y el <i>Data Warehouse</i> en el proceso de Toma de Decisiones	81
3.5 Familia de Productos Informix para <i>Data Warehouse</i>	81
3.5.1 Caso Genérico Aplicado al Área de Ventas	84
Conclusiones Capítulo III	87
Notas Bibliográficas	89
Conclusiones	91
Glosario	99
Bibliografía	109

Este trabajo es la culminación de un periodo profesional en el cual he aprendido que cada objetivo debe ser alcanzado con constancia y compromiso continuo; así como con la plena convicción de verlo concluido.

De antemano doy gracias a todas y cada una de las personas que han estado apoyándome cada instante de mi vida.

A mis padres por su soporte y cariño incondicional, a mi pareja por compartir cada momento conmigo y a mis profesores por todas sus enseñanzas.

Gracias Siempre

Introducción

*“La posesión de hechos es conocimiento,
el uso de ellos es sabiduría.”*

- Tomas Jefferson

En el pasado, los mercados eran poco dinámicos, la competencia escasa y menos agresiva, los márgenes de utilidad altos y los recursos de apoyo amplio. Los negocios, se administraban mediante un estrecho contacto entre clientes y proveedores. Los administradores, confiaban en su intuición y sentido común para tomar decisiones.

La intuición, *la prueba y el error* ya no son métodos efectivos para administrar una empresa. La competencia se ha vuelto agresiva, los márgenes de utilidad se reducen y los recursos para análisis son menores. Dada esta volatilidad, los empresarios se encuentran cada vez más alerta sobre oportunidades que proporcionen un valor agregado a sus organizaciones. Para auxiliar dichas tareas, se ha acumulado una cantidad enorme de datos a través de sistemas operativos tales como aquellos que administran pedidos, inventarios, etcétera; dentro de estos datos, existe información con el potencial de incrementar participación de mercado, mejorar productividad, hacer crecer el rendimiento de la inversión y mejorar el servicio al cliente. Sin embargo, a pesar de lo anterior, las organizaciones se encuentran limitadas en su habilidad de hacer uso de la información, en beneficio de la empresa.

A lo largo de las últimas décadas, el reto para las áreas de *Tecnología de Información (TI)*, ha sido desde el almacenamiento de sus datos hasta el procesamiento masivo de la misma. Sin embargo, una vez cruzado el anterior umbral, la premisa de las áreas de TI es proporcionar una solución cuyo reto es obtener información de los datos que apoyen el proceso de toma de decisiones.

De esta manera, el presente trabajo: “El *Data Warehouse* una herramienta en la administración informática para la toma de decisiones”, tiene como objetivo el definir un modelo de *Data Warehouse* para la planeación estratégica por medio de la integración de datos y análisis de información para las empresas que requieren un valor agregado competitivo en su mercado. Como resultado de lo anterior, a lo largo del trabajo de investigación se encontrará información por medio de la cual se identificarán las diferentes etapas de los sistemas de información y su evolución hasta la propuesta por medio de la cual el *Data Warehouse* se reconoce como una herramienta con la cual los ejecutivos podrán trasladar el valor de los datos a información y por ende al CONOCIMIENTO de su negocio.

Planteamiento del Problema:

La información en las empresas privadas e instituciones del sector público, es uno de los activos más importantes en estas entidades, sin embargo, dicha información hoy en día no tiene un valor agregado por medio del cual se genere un proceso de toma de decisiones estratégico adecuado. El principal obstáculo que enfrentan dichas entidades, es que cuentan con datos dispersos que describen su negocio, pero, éstos no se pueden traducir a un modelo por medio del cual se puedan hacer estudios que den como resultado indicadores de control para la toma de decisiones ejecutiva.

Hipótesis:

- La planeación estratégica y su consiguiente toma de decisiones acertada, serán exitosas diseñando un modelo a través del cual se integren y transformen las diversas fuentes de datos que existen en los negocios para así poder interrelacionar las áreas pertenecientes a una empresa y verificar con esto sus indicadores internos y con ello estar en capacidad de tomar decisiones en pro de la productividad, competitividad y efectividad de la misma.
- La toma de decisiones, es clave para la correcta operación de una empresa en todos sus ámbitos; esta debe ser sustentada por un modelo mediante el cual sea posible la integración, transformación y análisis de los datos operativos e históricos de la empresa utilizando herramientas informáticas por medio de las cuales será factible el análisis de los indicadores internos y de mercado a través de los cuales se tomarán decisiones fundamentadas en información completa, veraz y oportuna, complementándolas a su vez con el criterio empresarial en pro del éxito de su empresa dentro del marco globalizado de mercado.

El presente trabajo consta de tres capítulos: I . Fundamento Estructural: El *Data Warehouse* -, en el cual se describen los antecedentes de organización de datos, tipos de aplicaciones, definición de un *Data Warehouse* y componentes; II. Definición del Modelo de *Data Warehouse*, Diseño y Construcción -, este capítulo describe el ciclo de desarrollo del *Data Warehouse* y propone un modelo para la construcción de un *Data Warehouse* empresarial, así mismo da guías para la implementación de descubrimiento de más información valiosa para las organizaciones por medio de herramientas de software para ejecutivos en la organización; III. El Valor Agregado del *Data Warehouse* para la Toma de Decisiones Estratégicas, este último apartado habla de los beneficios del uso e implementación de un *Data Warehouse*.

Fundamento Estructural:
El Data Warehouse

*“El talento para el éxito no es
más que hacer lo que puedes
hacer bien y hacer bien
cualquier cosa que hagas.”*

- Henry Wadsworth Longfellow

Objetivo del Capítulo:

Entender qué es un Data Warehouse, su fundamento, componentes, y arquitectura; así como comprender cuál es la interrelación que existe del Data Warehouse para con los sistemas operacionales de negocios que están alrededor del mismo.

1.1 Antecedentes

La historia en el manejo y explotación de la información ha evolucionando dramáticamente; día con día su uso y administración adecuada se vuelven imperantes. Las empresas modernas utilizan esta información como un arma competitiva, ya que mediante su oportuna explotación pueden tomar decisiones estratégicas que les permiten colocarse en una posición dominante en su nicho de mercado. La información empresarial es el activo más importante que estas poseen, bajo esta premisa, se ha buscado el almacenar dichos datos y a su vez manipularlos en el negocio.

A principios de la década de los sesenta comenzaron a aparecer los primeros sistemas de base de datos, en los cuales la información quedaba almacenada en un medio magnético y no sólo en papel; por medio de estos sistemas de base de datos se logró almacenar la información operativa de un negocio y salvaguardarla para su uso posterior. En las siguientes décadas han habido grandes innovaciones en el área de manipulación de datos, abarcando desde el nivel conceptual hasta el tecnológico. Sin embargo todos estos cambios tienen un objetivo común: proporcionar una solución integral para el manejo óptimo de información.

El principal componente para el almacenamiento de información es el manejador de base de datos, el cual consiste de un conjunto de datos relacionados entre sí y un grupo de programas para tener acceso a esos datos⁽¹⁾. El conjunto de información se conoce como base de datos y ésta contiene la información organizada del negocio, más estructuras internas que permiten la administración de este caudal de datos. El objetivo del manejador de base de datos es el almacenamiento de información, así como su recuperación para consulta, actualización o modificación de una forma adecuada y eficiente. Su función es el proporcionar una visión abstracta de la operación del negocio expresada en números, letras, imágenes, videos, etcétera; con ello, el manejador de base de datos es el responsable de almacenar los detalles relativos a la estructura de los datos, el lugar en donde se almacenan y como son almacenados.

1.2 Tipos de Sistemas de Información

Alrededor de las bases de datos se encuentran los sistemas que interactúan y se interrelacionan con ellas, estos sistemas pueden alimentar, transformar o bien explotar la información sobre la estructura de la base de datos, la acción que realiza cada sistema sobre los datos depende de la naturaleza del mismo. Cada uno de estos sistemas de información ha solucionado diferentes necesidades de negocio; algunas de ellas permiten mantener la operación diaria de los negocios y otros dan información condensada de los mismos. A continuación se describirán las categorías de sistemas de información y sus implicaciones en los ambientes empresariales.

(1) Henry F. Korth & Abraham Silberschatz, Fundamentos de Bases de Datos, McGraw Hill, México. 1988.

1.2.1 Sistemas de Procesamiento Transaccional en Línea (*OLTP - OnLine Transaction Processing*)

Los sistemas de procesamiento transaccional en línea, son una clase de programa que facilita y maneja aplicaciones orientadas a transacciones⁽²⁾, las operaciones que realizan son entradas de datos y extracción de los mismos, estos sistemas se encuentran en múltiples industrias incluyendo bancos, líneas aéreas, supermercados y empresas de manufactura entre otras⁽³⁾. Los sistemas *OLTP* controlan la operación diaria de las empresas, representando un componente crítico corporativo, ya que por medio de ellos se solucionan partes medulares de los sistemas operativos de una empresa; por ejemplo, los sistemas de cómputo en una tienda departamental tienen que ser capaces de mantener siempre en línea la operación de las transacciones realizadas por cada cliente en términos de sus compras realizadas, si esto no ocurriese de esta manera por un corto instante, esto causaría fuertes pérdidas en la operación de ventas de la empresa; este fenómeno se podría replicar en los ambientes bancarios y sus redes de cajeros para disposición de efectivo.

Un sistema de procesamiento transaccional en línea es aquel por medio del cual se mantiene la información del negocio al momento en que está siendo generada, los datos operacionales se almacenan a nivel de detalle, el cual cumple con los requerimientos de procesamiento inmediato para las funciones de negocio de la empresa.

Los sistemas *OLTP* coleccionan información y distribuyen datos operacionales tanto dentro las organizaciones como entre ellas. El uso de redes y computadoras personales ha permitido el correcto funcionamiento de estos sistemas de información.

Las características principales de los sistemas de procesamiento transaccional en línea son:

- Son sistemas que conforman la parte medular de la operación diaria del negocio.
- Deben ser operativos en ambientes 24*7 los 365 días del año.
- Deben contar con mecanismos de contingencia y alta disponibilidad.
- Deben ser sistemas sumamente eficientes en términos de respuesta al usuario, así como en términos de la integridad de los datos que almacenan.
- La información refleja el estado actual del sistema.
- Manejan un número considerable de usuarios.
- Los datos son volátiles en todo momento, es decir, siempre se están realizando movimientos sobre los mismos. Son dinámicos.

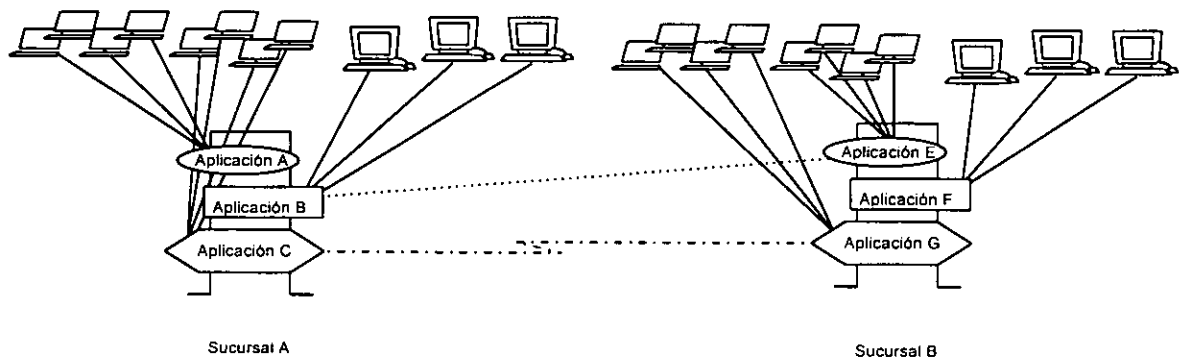
(2) Una transacción se define como una secuencia de intercambio de información y trabajo relativo que es tratado como una unidad con la finalidad de satisfacer requerimientos e integridad en una base de datos.

(3) Cita en Internet: <http://www.whatis.com/oltp.htm>

- La información se encuentra organizada de forma tal que resuelve el negocio operacional de la empresa.
- La historia que se maneja en términos de datos es típicamente hasta de un año.
- Los datos que los componen provienen de diferentes aplicaciones (véase Figura 1.1) del negocio; por ejemplo: nómina, compras, etcétera.

Este tipo de sistemas son la columna vertebral de los negocios, ya que por medio de estos se garantiza la operación correcta de las empresas. Los sistemas *OLTP* se encuentran afinados para satisfacer las demandas de sus usuarios, cuya tarea es mantener al día la información del negocio; esto implica el poseer un sistema integral con un manejador de base de datos que proporcione un buen desempeño para llevar a cabo las operaciones básicas de datos en la cual se realizan inserciones, actualizaciones y borrados de datos individuales y/o masivos. Cada operación que se realiza para modificar en cualquier sentido a los datos se conoce como transacción.

Los sistemas de procesamiento transaccional en línea también se conocen como sistemas operacionales del negocio.



Los Sistemas OLTP y su interacción con Aplicaciones

Figura 1.1

1.2.2 Sistemas de Información Ejecutiva (EIS - Executive Information Systems)

Definir un EIS, por sus siglas en inglés, no es una tarea fácil esto debido a las múltiples corrientes de estudio sobre los mismos. Así, el punto de partida es mencionar la primera definición acuñada en 1982 por John Rockard y Michael Treacy en un escrito de la Academia Americana de Consultores Gerenciales: *“Los sistemas de información ejecutiva tienen un propósito común, datos comunes y los métodos de utilización se definen como:*

-
- 1) Acceso del estado actual y tendencias proyectadas del negocio, y
 - 2) Análisis personalizado de los datos disponibles con el objetivo final de soportar la organización”.

Un EIS es un sistema de información basado en computadora, por medio del cual la información puede ser accedida, creada y liberada para altos ejecutivos que tienen poca experiencia en el área de computación. Esta información está disponible de acuerdo a la demanda y al nivel de detalle que es necesario para cada usuario, el poder explorar a múltiples detalles es conocido como *drill-down*. El usuario puede revisar los datos en forma incremental de detalle y con ello identificar problemas potenciales presentes por área. Los sistemas EIS, hablan más acerca del entendimiento que sólo de entrega de información; este entendimiento se aumenta agregando valor a los diferentes caminos en que la información es presentada, es decir, presentando la información clave de cada área del negocio.

El principal motivador para implementar un EIS, de acuerdo a la firma de consultoría Business Intelligence⁽⁴⁾, en orden prioritario es el siguiente:

- a) Mejorar el control estratégico.
- b) Mejorar el control financiero.
- c) Mejorar en su mercado y su información económica.
- d) Producir mejores análisis competitivos.

Los ocho componentes de un EIS ⁽⁵⁾ son:

1. Interfaz natural e intuitiva
2. Base de datos ejecutiva
3. Facilidades para conjuntar y consolidar datos
4. Técnicas de presentación
5. Técnicas de investigación
6. Técnicas de planeación
7. Técnicas de comunicación
8. Herramientas de desarrollo

Los sistemas ejecutivos de información se definen de acuerdo al modo de trabajo del ejecutivo, su principal objetivo es extraer la información al momento presente del negocio, con esta información el ejecutivo puede verificar el estado actual del mismo y con ello determinar en que punto se encuentra la

⁽⁴⁾ EIS's Report Management Overview, Business Intelligence, 1990.

⁽⁵⁾ Executive Information Systems, Pal Rolph & Peter Bartman.

compañía de acuerdo a los objetivos que tiene definidos en su plan estratégico de negocio. Estos sistemas permiten el monitoreo del progreso en metas críticas tales como: calidad, servicio al cliente, innovación y ventas entre otros.

1.2.3 Sistemas para Soporte en la Toma de Decisiones (DSS - Decision Support Systems)

Los DSS son sistemas de cómputo diseñados con el propósito de mejorar los procesos en la toma de decisiones, estos sistemas permiten hacer un análisis altamente estructurado y sumamente detallado, de igual manera, permiten realizar consultas ad-hoc para a sus usuarios; un ejemplo tipo de un DSS puede ser una hoja de cálculo financiera donde el análisis es hecho por un staff especializado y no por un ejecutivo. Estos sistemas se desarrollan para soportar a un sólo individuo en la toma de decisiones (un ejecutivo, gerente o ingeniero) y en algunos otros casos se utiliza para varias personas que toman de decisiones en conjunto con el objeto de llegar a una decisión en común, estos se conocen como sistemas para toma de decisiones grupales.

Los sistemas para soporte a la toma de decisiones son una herramienta por medio de la cual el sistema puede hacer un análisis más profundo de los datos contenidos en los medios ambientes operativos del negocio, típicamente estos sistemas arrojan información analítica de los negocios que pueden llevar a tomar decisiones de acuerdo a los indicadores que muestren como resultado para los ejecutivos.

Este tipo de sistemas esta diseñando para contestar preguntas y respaldar la toma de decisiones, el procesamiento analítico de esta información incluye análisis de tendencias, comparaciones e investigación de mercado. Para realizar esta tarea con un alto grado de confiabilidad es necesario el manejo grandes volúmenes de información.

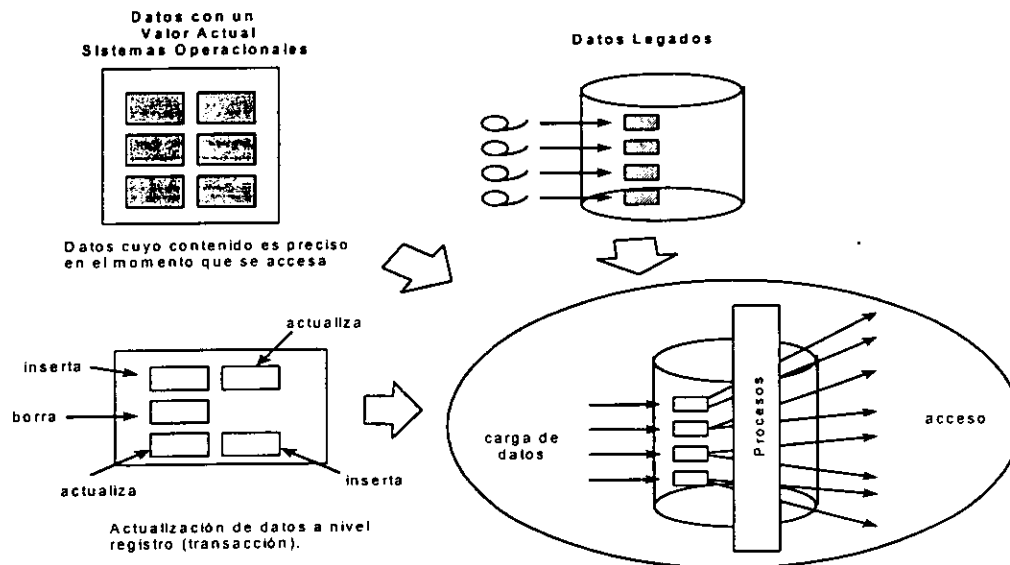
1.2.4 Sistemas de Consultas de Datos

Los sistemas de consultas de datos se definen como aquellos que únicamente obtienen información de los sistemas transaccionales en línea, o bien de los fuentes de datos provenientes de los sistemas DSS y EIS; su característica principal, es que estos sistemas no implican ningún tipo de análisis, es decir toman los datos planos del día a día del negocio. Por medio de esta información se obtienen datos genéricos de lo que esta ocurriendo al momento de su consulta. Por ejemplo, si se realiza una consulta al catálogo de clientes este sólo mostrará el número de clientes al día, más sin embargo no dará mas información específica del mismo.

1.3 Qué es *Data Warehouse*

La infraestructura de un “Almacén de Datos” (*Data Warehouse*) se compone por un conjunto de datos integrados orientados a una materia (Figura 1.2), que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones en la administración; esta es la definición histórica de *Data Warehouse* acuñada por el padre del mismo: William H. Inmon⁽⁶⁾.

El propósito del *Data Warehouse* es soportar el proceso analítico, este se compone de cualquier dato y/o repositorio de información para tomar una decisión o resolver problemas de negocio.



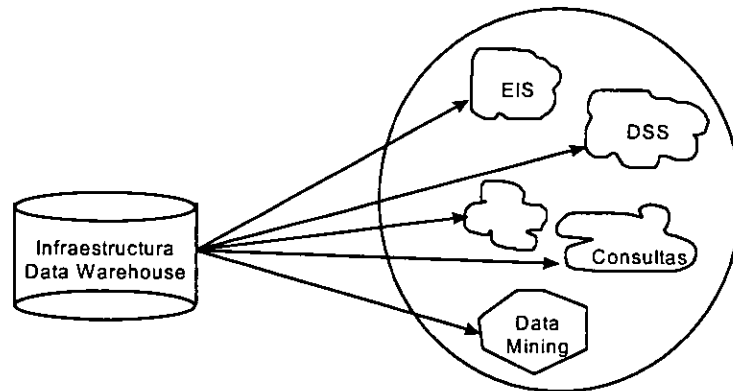
El *Data Warehouse* como Datos

Figura 1.2

El *Data Warehouse* es una arquitectura, es un depósito consistente de datos separados y que no interfieren con los sistemas operativos existentes, este depósito una vez modelado, permite satisfacer los requerimientos de acceso y reporte de datos. El *Data Warehouse* es un proceso continuo que mezcla los datos de varias fuentes heterogéneas, incluyendo datos históricos y adquiridos para soportar la constante necesidad de consultas estructuradas, reportes analíticos y soporte de decisiones.

El *Data Warehouse* es una infraestructura compuesta por múltiples piezas que llevan a la integración y transformación de datos para la toma de decisiones, los sistemas que están en capacidad de explotar esta información son: EIS, DSS y reportes con consultas específicas como se ilustra en la Figura 1.3.

(6) Building the Data Warehouse, W. H. Inmon



El *Data Warehouse* y su relación con EIS, DSS y de Consulta de datos

Figura 1.3

El *Data Warehouse* se define de la siguiente manera:

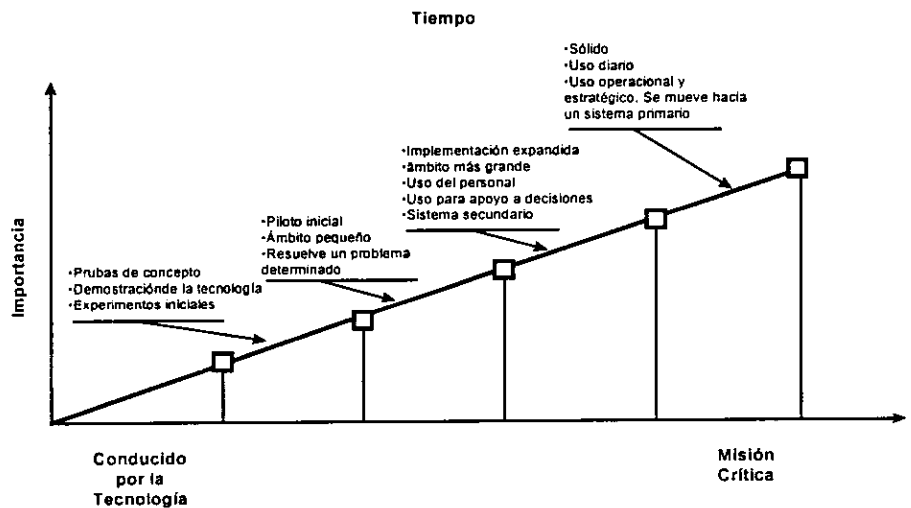
- Esta orientado a la materia: El *Data Warehouse* orienta y organiza los datos desde la perspectiva del usuario final o ejecutivo.
- Administra volúmenes masivos de información: La mayoría de los sistemas de *Data Warehouse* contienen información histórica del negocio, adicionalmente a estos datos se encuentra la información presente del mismo. El *Data Warehouse* debe proporcionar la capacidad del manejo de volúmenes masivos de información, ya que este acumula el universo de datos del negocio; de igual manera debe ofrecer opciones para condensar y adicionar datos clasificados, así como la posibilidad de manejar información a diferentes niveles de granularidad.
- Guarda información en diferentes medios de almacenamiento: Por los volúmenes que maneja el *Data Warehouse* la información generalmente se encuentra almacenada en diversos medios: cintas, discos y sistemas legados (sistemas de cómputo obsoletos o de arquitectura propietaria, pero funcionando).
- Comprende múltiples versiones de un esquema de base de datos: Debido a que el *Data Warehouse* tiene que guardar información histórica y administrarla, y como dicha información histórica ha sido manejada en distintos momentos por diferentes versiones de esquemas de base de datos, en ocasiones el *Data Warehouse* tiene que controlar información originada en organizaciones de bases de datos diferentes.
- Condensa y agrega información: Las bases de datos operacionales frecuentemente almacenan información muy detallada para la toma de decisiones adecuada. Un *Data Warehouse* condensa y

agrega información para presentarla en forma comprensible a los ejecutivos. La condensación y adición es esencial para retroceder y entender la imagen global del negocio.

- Integra y asocia datos de muchas fuentes de información: Debido a que las organizaciones han administrado históricamente sus operaciones utilizando numerosas aplicaciones de software y múltiples bases de datos, se requiere del *Data Warehouse* para recompilar y organizar en un sólo lugar la información que estas aplicaciones han acumulado al paso de los años. Este proceso es complejo por la diversidad de tecnologías de almacenamiento, de técnicas para la administración de bases de datos y por la semántica de los datos.

El propósito del *Data Warehouse* es ayudar a la administración a comprender el pasado y planear el futuro, incluso, si el *Data Warehouse* contiene información estratégica crucial, derivar el valor empresarial es una tarea complicada, para ello se requerirá una combinación de técnicas empresariales, habilidades técnicas, intuición y experiencia.

El *Data Warehouse* debe ser un sistema de misión crítica (Figura 1.4), un sistema se define de esta manera cuando se tiene confianza en sus operaciones y se usa diariamente, el sistema se torna fundamental para la misión del negocio que apoya y su falla puede proporcionar una falla en el negocio mismo.



Ciclo de Madurez de la Tecnología

Figura 1.4

Los requisitos para un sistema de misión crítica son:

- Disponibilidad.
- Consistencia y precisión.
- Estándares.
- Actividades basadas en requerimientos.
- Compatibilidad con la tecnología existente.
- Uso diario.

-
- Uso sencillo.
 - Alto desempeño.
 - Veraz.
 - Seguridad.

A medida que una compañía tenga la disponibilidad de un *Data Warehouse* y que el uso de la información procedente de él sea más solicitada, este sistema deberá cumplir con todas las condiciones de un sistema de misión crítica, con el objetivo de proporcionar información oportuna y veraz a sus usuarios.

1.3.1 Características de un *Data Warehouse*

Las características que definen a un *Data Warehouse* son:

- Contiene información para la toma de decisiones estratégicas.
- Desde el punto de vista de datos, proporciona el nivel más alto de condensación (agregación) de los mismos.
- Posee una estructura que permite hacer análisis de información desde lo más genérico a lo más particular.
- Permite hacer análisis de tendencias de acuerdo a los datos históricos.
- Integra datos para análisis corporativo.
- Sus datos son estáticos hasta que se hace una actualización de los mismos.
- Los datos que lo componen provienen de toda la empresa.

1.3.2 Componentes

La solución para implementar un *Data Warehouse* es un tanto compleja ya que se deberán combinar sus componentes o arquitectura a la par con los requerimientos provenientes de la empresa. Hoy en día en el mercado de soluciones de *Data Warehouse* se encuentran múltiples tecnologías, distribuidores y términos; por esta razón y para poder proporcionar un modelo de solución, el primer paso es determinar la arquitectura genérica y sus componentes para así garantizar la correcta integración de piezas de software para la generación de un modelo de solución integral.

La arquitectura de referencia se puede describir desde un punto de vista abstracto y simplificado de la siguiente manera⁽⁷⁾:

- Un conjunto de datos extraídos de bases de datos operacionales.

(7) HARDJINDER S. Gil, *Data Warehousing*, Prentice Hall, México, 1996.

- Un software que prepara los datos para que los accedan los usuarios.
- Un conjunto de aplicaciones y herramientas que ejecutan un conjunto de consultas y análisis complejos.

La arquitectura de referencia(Figura 1.5) del *Data Warehouse* divide los componentes del mismo de la siguiente manera:

- * *Bloque de fuente de datos*
 - * *Bloque del Data Warehouse*
 - * *Bloque del Datamart (8)*
 - * *Bloque de Acceso y Uso*
- *Capa de administración de datos*
 - *Capa de administración de metadatos*
 - *Capa de transporte e infraestructura*

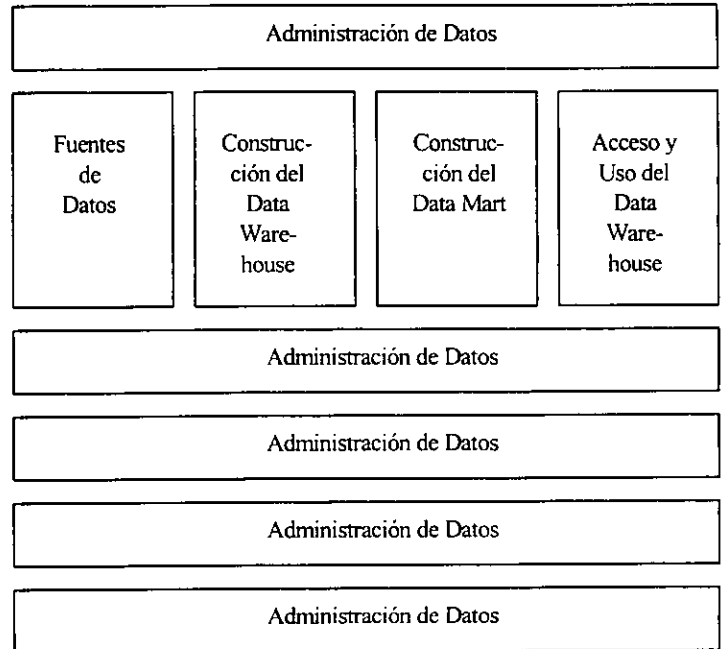


Figura 1.5

Los bloques se relacionan con la funcionalidad específica de *Data Warehouse* y las capas representan el ambiente necesario para implementar los bloques.

Las capas horizontales de Administración de datos y de metadatos corresponden a las actividades que se relacionan con la extracción, carga y actualización que se requieren para mantener el abasto de información para el *Data Warehouse* y el *Datamart*. Las otras capas horizontales corresponden a los servicios comunes que son esenciales para los bloques del *Datamart* y del *Data Warehouse*, estos servicios son la infraestructura, transporte, herramientas, y tecnologías.

Bloque Fuente de Datos

El bloque de fuente de datos se ilustra en la Figura 1.6 quedando de la siguiente manera:

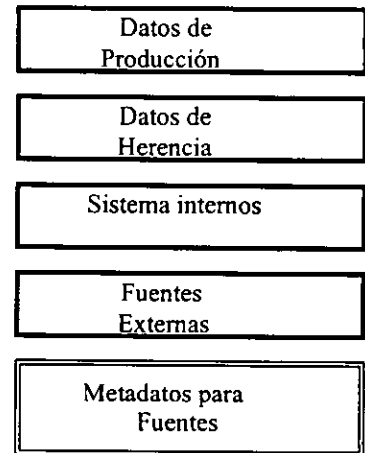


Figura 1.6

- Datos de Producción:** Se refieren a las bases de datos operacionales que contienen la información recopilada de las aplicaciones operacionales. Esta información puede provenir de diversas tecnologías.
- Datos de Herencia:** Estos datos son los que se encuentran fuera de línea, en archivos perdidos, debido a que ya no son necesarios para el apoyo de la operación actual. Sin embargo el valor de esta información es su contenido histórico que nos permite analizar tendencias de comportamiento de negocio.
- Sistemas Internos:** Existen fuentes de datos que no se almacenan en una base de datos operacional ni tampoco los utiliza una aplicación operacional de negocio. En general, estos datos son los siguientes:
- No estructurados - tal como formas electrónicas.
 - Estructurados - tal como reportes, hojas de cálculo y los documentos de procesamiento de palabras.
 - Semiestructurados - tal como en los reportes anuales.
- Sistemas Externos:** Estas fuentes de información no las controla, ni las posee la empresa. Pueden ser electrónicas, de análisis competitivo del mercado, artículos de revistas, periódicos u otros.
- Metadatos para fuentes:** Los metadatos son la información de la definición acerca de los datos de los fuentes. Los metadatos incluyen el nombre del dato capturado y extraído en las fuentes de información, la definición del contenido de los datos, la fecha en que fueron creados y la fuente u origen de los mismos.

(8) Datamart: Subconjunto de un Data Warehouse a nivel departamental, ver sección 1.4.1

Bloque del Data Warehouse

El bloque de construcción del *Data Warehouse* consiste de los siguientes componentes:

Componente de refinamiento es responsable de estandarizar, limpiar, filtrar y confrontar, así como imprimir la fecha de origen de la información extraída de las fuentes de datos seleccionadas.

- En el componente de refinamiento, se ubican los metadatos para los nombre y definiciones de datos estándar. Se crean y capturan metadatos adicionales para imprimir la fecha en los datos extraídos, para la fuente de los datos extraídos y para los campos faltantes que se hayan agregado.
- El componente de reingeniería es responsable de preparar los datos para que sean congruentes con las necesidades de análisis del usuario empresarial. La reingeniería, en el contexto del *Data Warehouse*, es distinta de la reingeniería de las aplicaciones o de los procesos empresariales. La reingeniería implica las siguientes actividades:
 - La integración de diferentes tipos de datos para la creación datos nuevos.
 - Los datos parcialmente divididos en series de fases de tiempo para el análisis.
 - El cálculo previo de la información resumida y la derivación de otra información que requiera el usuario empresarial.
 - La traducción y formato de datos de diferentes fuentes para poder combinarlos de manera uniforme y consistente.
 - La transformación y reubicación de los datos almacenados en las fuentes de información.

La actividad de modelado de la información deriva del modelo de datos del *Data Warehouse* a partir del modelo de datos de la empresa, si existe uno o crea uno nuevo. Entre los componentes del diseño de la base de datos se incluye la fragmentación de datos para de esta forma garantizar la consulta paralela sobre toda esta información. La administración de datos reduce los grandes volúmenes de datos que llegan y los datos previamente almacenados, utilizando técnicas de adición, conocidos también como sumarios y de resumen, o acumulados.

Bloque de construcción del Datamart

Este bloque es el segundo en términos de importancia dentro del *Data Warehouse*, este bloque se utiliza para crear un subconjunto especializado de datos para su consulta. El Datamart tiene componentes similares a los del *Data Warehouse*, la principal diferencia entre estos es el enfoque del usuario final. El Datamart es un subconjunto de todo el *Data Warehouse* y este puede ser opcional.

Bloque de acceso y uso del Data Warehouse

El tercer gran bloque de la arquitectura del *Data Warehouse* tiene dos componentes: Acceso y Recuperación, y Análisis y Reporte.

El componente de Acceso y Recuperación proporciona un acceso directo al *Data Warehouse*, evitando el Datamart, es decir, puede ser independiente de este último. Este componente también es responsable de transformar los datos recuperados en vistas multidimensionales⁽⁹⁾ o para almacenar en una base de datos también multidimensional. Los datos del *Data Warehouse* y del Datamart pueden dejarse en un depósito local con o sin traducciones y transformaciones, para su análisis local, minería y elaboración de reportes; de igual manera el acceso y examen de los metadatos también se incluyen en este componente.

El componente de análisis y reporte es responsable de herramientas y aplicaciones necesarias para aprovechar el *Data Warehouse* y el Datamart. Estas se clasifican en herramientas de reporte, de análisis y soporte de decisiones, de modelado empresarial y de minería de datos. Las herramientas para navegar por los metadatos para entender y reportar el contenido del *Data Warehouse* / Datamart son también responsabilidad de este componente.

Capa de Administración de Datos

Las tareas de extraer, cargar, actualizar, reforzar seguridad, archivar y restaurar el *Data Warehouse* a partir de archivos son sustentadas por la capa de administración de datos(Figura 1.7). Esta capa incorpora las políticas estándar de administración de datos, los procedimientos, los programas y operaciones de seguridad, las autorizaciones de acceso, archivar y restaurar y purgar datos.

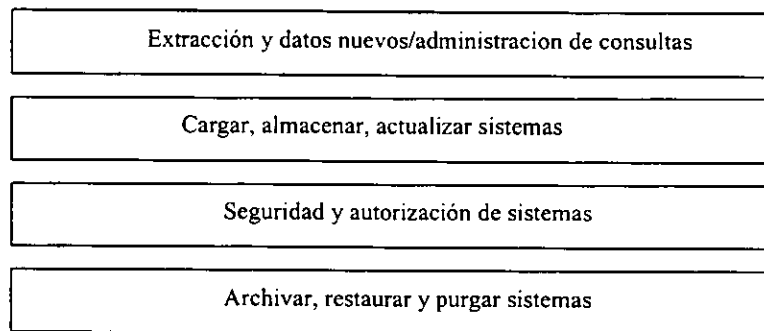


Figura 1.7

⁽⁹⁾ Multidimensional: Se define como la capacidad de representar la información de una base de datos en múltiples capas lógicas.

Capa de Administración de los Metadatos

La arquitectura del *Data Warehouse* se basa en el concepto de definiciones de datos, o metadatos; de esta manera la administración de los mismos nos dará un análisis detallado de dichas definiciones, los metadatos proporcionan guías y apuntadores hacia la información del *Data Warehouse*.

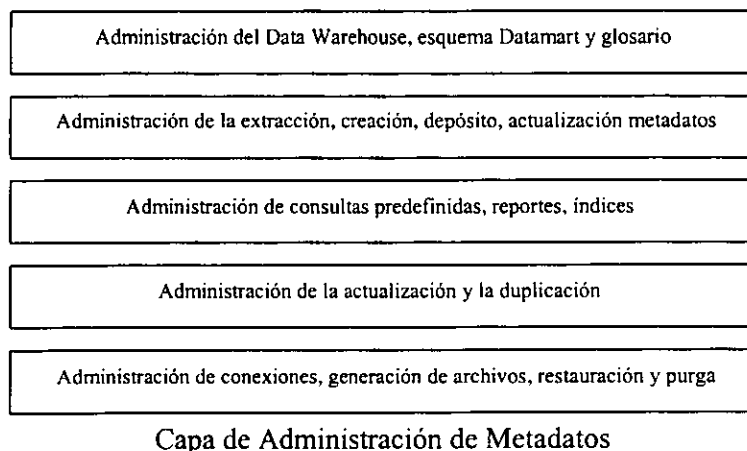


Figura 1.8

El *Data Warehouse* y los modelos y esquemas de datos lógicos y físicos del Datamart junto con el glosario técnico y empresarial, también se almacenan y administran en esta capa.

La capa de administración de metadatos es responsable de lo siguiente⁽¹⁰⁾:

- Definiciones estándar de los datos (incluyendo las definiciones técnicas y empresariales) de los datos depositados en el *Data Warehouse*.
- Los metadatos capturados y creados en los bloques de Refinamiento y Reingeniería.
- Los metadatos de granularidad, segmentación, áreas tema, adición y condensación.
- Los metadatos que describen las consultas y reportes predefinidos y diseñados.
- Los metadatos que describen los índices y perfiles que mejoran el rendimiento y acceso a la recuperación de datos.
- Los metadatos describen las reglas para preparar y programar el ciclo de actualización y duplicación.

Capa de Transporte

La tarea de transportar datos entre los diferentes bloques de la arquitectura la cumple la capa de transporte. Esta utiliza tecnología para transferencia y entrega de datos, y componentes de middleware⁽¹¹⁾. También aporta seguridad y la autenticación de solicitudes de transporte.

(10) HARJINDER, S. Gill. *Cliente/Servidor Data Warehousing*. Prentice Hall. 1996.

Gráficamente esta capa queda como sigue:

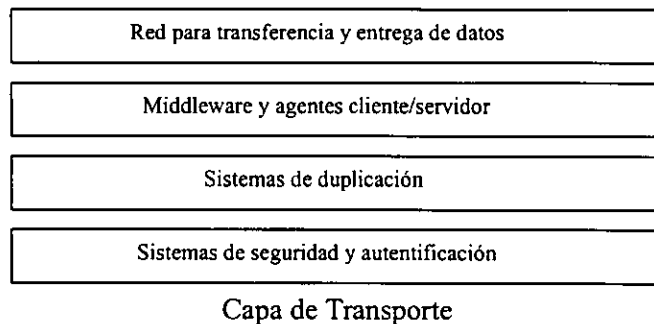


Figura 1.9

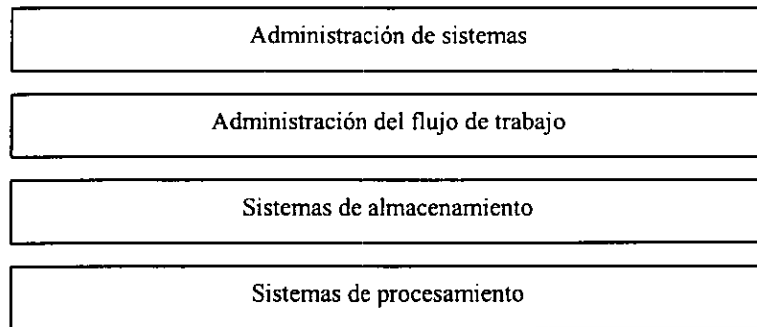
En el componente de redes para transferencia y entrega de datos, podemos tener múltiples protocolos de red, por ejemplo: TCP/IP, IPX, etcétera; múltiples tipos de red como Ethernet, Token Ring entre otros y entre ellos múltiples sistemas operativos.

Capa de Infraestructura

La capa de infraestructura arquitectónica es el que sustenta la operación del *Data Warehouse*, esta cuenta con 4 componentes:

- a) Administración de sistemas que ofrece la capacidad al constructor de sistemas y al empresarial el invocar, manejar y terminar las herramientas o aplicaciones.
- b) Administración de flujo de trabajo que sustenta el proceso de integración y administración para coordinar la ejecución ordenada y específica de las herramientas y aplicaciones y actividades para extraer, duplicar, agregar y resumir el *Data Warehouse* y el Datamart. Automatiza procesos para mantenimiento y actualización de reportes y consultas predefinidas.
- c) Sistemas de almacenamiento aportan servicios de administración de la base de datos para la fuente de datos, el *Data Warehouse* y los catálogos base de datos del Datamart y depósito multidimensional.
- d) Sistemas de procesamiento son los ambientes operativos implícitos en los bloques principales: Fuente de Datos, *Data Warehouse* y Datamart; es decir, herramientas, middleware y todos aquellos que se estén usando en el medio ambiente.

(11) Middleware se define como una capa de software intermediaria cuya función es comunicar a los componentes de software que están entre el mismo.



Capa de Infraestructura

Figura 1.10

1.3.3 Requerimientos

Los requerimientos de un *Data Warehouse* son iguales de variados y diversos como las clases de usuarios que lo utilizan para obtener beneficios empresariales, los requerimientos dependerán de la perspectiva que tiene cada persona con respecto al *Data Warehouse*. Por lo anterior, un punto crucial es la determinación adecuada de requerimientos para el *Data Warehouse*, de esta manera su construcción e implantación serán exitosas.

Así, el primer paso es definir las perspectivas del *Data Warehouse* en la empresa. Para definir la perspectiva son necesarios los diferentes puntos de vista asociados, es decir, conocer desde arriba hacia abajo, o sea la perspectiva conducida por el negocio, o el dueño; al igual que el punto de vista de abajo hacia arriba, correspondiente a la del constructor o implementador del sistema de *Data Warehouse*.

Una vez satisfechos estos puntos de vista, como primer paso, se deberán cubrir los requerimientos de acuerdo a las perspectivas ya sea del negocio o dueño, o bien del constructor. Esto se definirá bajo un consenso, proveniente de prioridades entre las áreas involucradas, de forma que se genere un valor a los ejecutivos, así como para los constructores.

Las perspectivas de los requerimientos (Figura 1.11) pueden clasificarse de la siguiente forma:

- El dueño/inversionista en el sistema de información intenta resolver un problema empresarial utilizando la tecnología de la información; el punto de vista del empresario se mide en dinero y la perspectiva que tiene del negocio es periodos y procesos empresariales. Por lo general, el interés del empresario en la tecnología se limita al análisis de inversiones y pagos, y la evaluación de riesgos y programas. Para él, el *Data Warehouse* es sólo un componente de soporte para la toma de decisiones dentro de la perspectiva empresarial total en su sistema(Figura 1.3). Más sin embargo, el empresario

deberá confiar en que un *Data Warehouse* bien construido resolverá sus problemas de análisis de datos. La toma de decisiones fundamentada en la información procedente del *Data Warehouse*, permitirá sustentar dichas acciones vislumbrando al negocio empresarial como un ente integral, en donde se conjugaran áreas tales como: ventas, mercadotecnia, finanzas, administración estratégica, planeación, desarrollo de productos, atención a clientes, recursos humanos, etcétera.

- El inversionista/dueño por lo regular participa en la generación del concepto del *Data Warehouse* por su capacidad de aprobación y revisión. El inversionista realiza el análisis de costo/beneficio de esta solución, incluye una análisis de recuperación de inversión y de recuperación de activos.

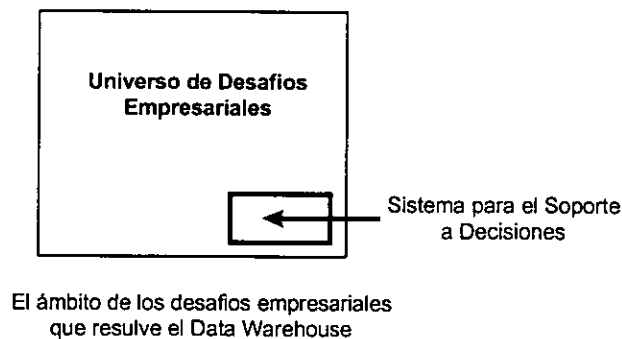


Figura 1.11

- El usuario empresarial utilizará el *Data Warehouse* para funciones corporativas, su perspectiva esta definida para realizar un conjunto de procesos y pasos que debe seguir para llevar a cabo el análisis empresarial; este usuario no necesita saber el cómo se forma ni cómo se encuentra estructurado el *Data Warehouse*, sólo requiere que éste realice las funciones señaladas correcta y eficientemente. Su principal interés de este usuario es utilizar la información del *Data Warehouse* para conocer los objetivos del negocio y mejorar la productividad y eficiencia tanto de la organización del personal (interna) como de su organización en el mercado (externa). Esta información en complemento con su conocimiento con relación a clientes, productos, servicios y la organización empresarial interna para así ofrecer un mejor nivel de servicio, o bien generar estrategias que ganen mercado, operando con mayores ganancias y menos costos. Estos usuarios empresariales necesitan convertir la información en hechos, transformar estos hechos en conocimientos y usar el conocimiento para llegar a decisiones y recomendaciones; buscan el presentar resultados de su análisis en forma coherente y lógica, para así compartirlos con los miembros ejecutivos que toman decisiones.
- Los diseñadores del *Data Warehouse* son personas que deberán entender tanto las necesidades empresariales de éste como la tecnología de implementación que se necesitan para solventarlas. El diseñador es responsable de entender las necesidades empresariales a corto y a largo plazo, las

inversiones en tecnología actual y planeada sobre plataformas de cómputo y tecnología de bases de datos, entender y asimilar la disponibilidad real de los datos que proporcionan los sistemas operacionales, así como entender los elementos necesarios de seguridad para la operación correcta del *Data Warehouse*. El diseñador es un usuario que observa los modos en los que la información deberá organizarse para satisfacer las necesidades del ejecutivo o usuario empresarial, de igual manera es su responsabilidad el mantener actualizados los datos del *Data Warehouse* contemplando cada uno de los sistemas que lo alimentan e interactúan con él.

- El constructor es el responsable de la instalación e integración física de los diversos componentes del *Data Warehouse*, este incorpora los elementos operacionales y de despliegue del *Data Warehouse*, incluyendo la disponibilidad del sistema, mecanismos de recuperación y de seguridad.

Tanto los diseñadores como constructores se responsabilizan en conjunto del desarrollo, instalación, implementación y mantenimiento del *Data Warehouse*.

Proveniente de la descripción anterior y basada en las perspectivas y responsabilidades de cada área asociada al *Data Warehouse*, se debe hacer un análisis en el cual se describa de forma matricial los requerimientos a cumplir en cada área, para con esto garantizar el éxito del *Data Warehouse* cubriendo las necesidades a lo largo de toda la empresa. Es importante señalar que en ocasiones será necesario tomar decisiones que se orienten al negocio y que pueden no ser las más adecuadas para todas las áreas que se involucran en la definición del *Data Warehouse*.

	Qué	Cómo	Dónde	Quién	Cuándo	Por qué
Dueño/ Inversionista	· Sistema para análisis de ingresos por cliente	· Analizar datos históricos	· Área de Mercadotecnia	· Gerente de Área	· Mensualmente · Trimestralment e · Anualmente	· Generar decisiones de mercadotecnia promocional.
Usuario Empresarial	· Áreas tema	· Reportar · Consultar	· Escritorio	· Analistas	· Información fluida	· Objetivos operacionales · Soporte de decisiones
Implementador de la tecnología de la información	· Modelado de la <i>Data Warehouse</i>	· <i>Data Warehouse</i> e información por área	· Trabajo corporativo a nivel empresarial en red	· Programador de aplicaciones · Administradore s de Base de Datos	· <i>Data Warehouse</i> y ciclo de actualización de carga	· Objetivos técnicos.

1.3.4 La Administración Informática y su relación con el *Data Warehouse*

La administración es un proceso el cual se definen generalmente cuatro etapas (fuente: SICOFIN Consultores, Control de Gestión, 1993):

Planeación: Consiste en anticipar esencialmente el futuro, tomando decisiones sobre los eventos que se espera ocurran en el tiempo por venir.

Integración: Implica obtener, estructurada y ordenadamente, los recursos humanos, materiales, financieros y tecnológicos que permitan cumplir con los planes establecidos, logrando su máximo aprovechamiento.

Dirección: Es la etapa del proceso administrativo donde se realiza lo planeado, delegando la autoridad correspondiente, para la obtención de los objetivos señalados, a través de la comunicación, motivación, supervisión y coordinación de una persona o grupo.

Control: Es un mecanismo informativo continuo, constituye un impulso correctivo para asegurar la concurrencia de todo el proceso.

El *Data Warehouse* se complementa desde el punto de vista informático con el proceso administrativo, como una herramienta que permitirá obtener indicadores para la planeación estratégica por medio de la integración de datos y el análisis de información para las empresas que requieren un valor agregado competitivo en su mercado. De esta manera, la meta de la administración en los negocios consiste en desarrollar estrategias y planes exitosos para aumentar las utilidades, mantener el nivel de satisfacción del cliente e incrementar la participación en el mercado.

La finalidad de un *Data Warehouse* consiste en auxiliar a la administración a comprender el pasado y planear el futuro. Todo esto se apoyará con el uso adecuado de componentes de tipo informático, entre los cuales se encuentran piezas de software, hardware y programas; los cuales trabajarán de una forma coordinada para transformar datos en información. Así, la administración desde el punto de vista informático, será apoyada por medio del *Data Warehouse*, proveyendo de información que permitirá desarrollar nuevas estrategias y planes de mercado para mejorar la posición global de una empresa mundial de negocios.

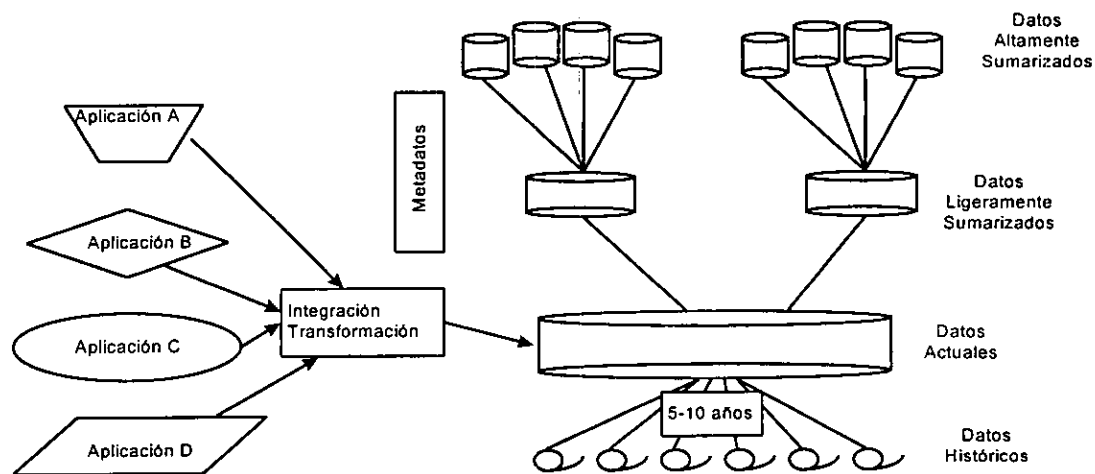
1.4 El almacenamiento de Información en un *Data Warehouse*

Un *Data Warehouse*, como se ha mencionado anteriormente, es un conjunto de datos inter-relacionados entre sí para tomar decisiones ejecutivas de negocio, esta información puede constituir al *Data Warehouse* de diferentes maneras.

El *Data Warehouse* es un almacén de información con propósitos a lo largo de la empresa, es decir, incorporará información de todas las áreas de la misma, provocando con ello la factibilidad de cruzar

información para hacer inferencia entre las áreas de forma interna, o bien, para generar decisiones que atañen a la organización por completo. Pero, el *Data Warehouse* no sólo proporciona esta estructura, también nos permite desmembrar el sistema vía diversos componentes a nivel estructura de datos y granularidad de información.

La fotografía del *Data Warehouse* corporativo, en términos de su estructura de datos se representa de la siguiente manera:



Esquematización del *Data Warehouse*

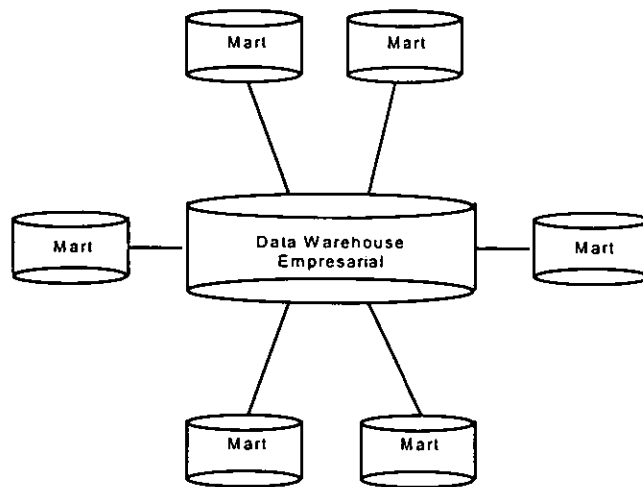
Figura 1.12

Existen estructuras adicionales que pueden sustituir y/o interrelacionarse con el *Data Warehouse*, estas se describen en los siguientes incisos.

1.4.1 *Datamart*

Un *Datamart* es un subconjunto de un *Data Warehouse* existente (Figura 1.13), este se encuentra orientado hacia un aspecto más fino dentro de una función de negocios. Los datos que se contienen en él generalmente guardan un estado altamente resumido, a nivel mes, trimestre o mayor. Los *Datamarts* contienen un número menor de datos, ya que estos se especializan por línea de negocio, por ejemplo: área de ventas, área de compras, área de recursos humanos, etc. Su información no es actualizada frecuentemente, pero si se prescriben intervalos de tiempo en los cuales se realiza esta actividad.

Los *Datamarts* son utilizados para la toma de decisiones a nivel departamental, o bien para analizar alguna característica específica o función de negocios; por lo tanto, el *Data Warehouse* proporciona información para la planeación estratégica a lo largo de la empresa y el *Datamart* proporciona información a un nivel más específico en las organizaciones.



El *Datamart*: El Rostro del *Data Warehouse*

Figura 1.13

1.4.2 Almacén de Datos Operacionales

El almacén de datos operacionales es una forma híbrida de un *Data Warehouse* que sirve a la comunidad operacional, es decir, permite a las aplicaciones transaccionales en línea utilizarlo como albergue para guardar la información. La granularidad de los datos que contiene no es a nivel transacción, pero si a un siguiente nivel. Los registros sumariados son sujetos a cambio cada vez que los detalles inherentes a estos sumariados cambian, de esta manera puede ser difícil el reconstruir la historia.

Algunos de estos sistemas son actualizados dentro de segundos después de la actualización del sistema fuente; algunos otros esperan lapsos máximos de 24 horas. Los datos almacenados en este tipo de almacén son actuales, generalmente no mayores a algunos días. Los datos reflejan el estado actual del negocio en lugar de representar múltiples estados en diferentes tiempos del negocio.

Un almacén de datos operacional puede ser utilizado para la toma de decisiones del día con día, y también puede servir como una fuente común de datos para el *Data Warehouse*.

1.4.3 Repositorio

Un repositorio de datos combina todos los datos fuente de un medio ambiente operacional aplicativo en una sola base de datos normalizada. Los registros de este repositorio son actualizados frecuentemente. Los datos contenidos en el mismo son completamente operacionales y no históricos. El repositorio puede

o no ser utilizado para consultas ejecutivas, todo ello dependerá de los requerimientos específicos del sistema.

Un repositorio cumple con los requisitos necesarios para un ambiente de procesamiento operacional en línea a lo largo de toda la empresa, integrando en él todos los datos fuente de la misma.

Las características de un repositorio son:

- Base de datos de alto volumen normalizada.
- El sistema relacionado a éste se encuentra afinado para ambientes de procesamiento transaccional en línea, soportando operaciones del día a día.
- La operación de datos es a nivel transaccional.
- El detalle de los datos se define por transacción individual.
- Puede servir de alimentador para otros sistemas.
- Representa un ambiente integrado de datos fuente.
- Contiene datos actuales.

1.4.4 Metadatos

Los metadatos contienen datos acerca de los datos, es decir, mantienen las relaciones entre los datos contenidos en el *Data Warehouse* y los datos operacionales. Los metadatos invaden todas las actividades del *Data Warehouse*, estos representan guías y apuntadores hacia la información del mismo.

Los metadatos constan de los siguientes tipos de elementos:

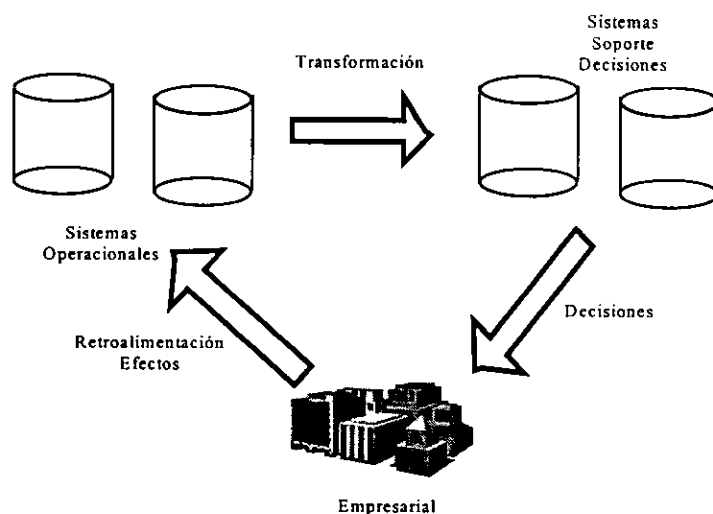
- Ubicación y descripción de servidores, bases de datos, tablas, nombres y resúmenes del *Data Warehouse*.
- Reglas de profundización al detalle, organizados por jerarquías de dimensión empresarial.
- Nombres o alias de encabezados y hechos de datos con nombres más técnicos.
- Reglas de cálculos personalizados.
- Seguridad a nivel personal, de grupo y de empresa.
- Descripción de fuentes originales y transformaciones.
- Definiciones lógicas y atributos del *Data Warehouse*.
- Definiciones físicas de tablas y de columnas, así como de sus características.
- Ubicación integrada del *Data Warehouse*.
- Criterios de envejecimiento de datos.
- Seguridad
- Unidades de medida.

La administración de los metadatos debe contemplar las siguientes condiciones para mantener el control de su operación:

- Definiciones estándar de los datos, incluyendo las definiciones técnicas y empresariales de los datos depositados en el *Data Warehouse*.
- Captura y creación de metadatos.
- Metadatos que describen las consultas y reportes predefinidos y diseñados.
- Metadatos que describen índices y perfiles que mejoran el rendimiento en el acceso y recuperación de datos.
- Los metadatos describen las reglas para preparar y programar el ciclo de actualización y duplicación.

La planeación es una fase importante de la implementación del *Data Warehouse*. Las decisiones tomadas durante la fase de planeación tienen un impacto significativo en el ámbito de implementación y en la magnitud del esfuerzo. Las decisiones clave de implantación incluyen la selección de lo general a lo particular o viceversa, o bien combinado; la selección de la información, fuentes de datos y tamaño del metamodelo; y la estimación de planes de programa, proyecto y justificaciones de presupuesto.

El nacimiento del *Data Warehouse* ha sido un proceso bastante complejo, en el cual se integran múltiples piezas arquitectónicas y de diseño, aunado a esto, día con día se incorporan más conceptos y nueva tecnología de punta que resuelven puntos específicos del *Data Warehouse*. Derivado de lo anterior, el siguiente capítulo definirá un modelo para el desarrollo del *Data Warehouse* con el objetivo y foco de entrega de información por medio de su estructura y no sólo en el uso de tecnología específica o el modelado de datos provenientes de una organización.



El Proceso de Toma de Decisiones

Figura 1.14

El *Data Warehouse* es una herramienta por medio de la cual se puede hacer más confiable el proceso de toma de decisiones, con lo cual será posible generar un ciclo de información (Figura 1.14) a través del cual se entregará un valor agregado a las empresas del hoy y del mañana.

Conclusiones del Capítulo

El manejo de datos de una forma adecuada y eficiente se ha convertido en un deber empresarial que debe ser ejecutado para la correcta operación de las instituciones; el manejo de los datos y su correcta explotación han evolucionado presurosamente hasta colocarnos en un medio informático y administrativo en el cual podemos utilizar nuevas soluciones de negocios para generar una cadena de valor por medio del conocimiento de información para la toma de decisiones. Este conocimiento puede ser obtenido por una solución integral de *Data Warehouse*, donde dicha solución contiene la fotografía empresarial histórica y presente de los datos de una organización.

La certidumbre en la toma de decisiones depende de la cantidad y la calidad de las evidencias con que se cuente en el momento que se analizan las alternativas; durante este proceso se considerarán las evidencias que se tengan al alcance, así el *Data Warehouse* proveerá la información necesaria con el objetivo de apoyar a la selección adecuada de las diversas alternativas empresariales en el proceso de toma de decisiones, lo cual se traducirá finalmente en tres beneficios principales para las empresas:

- Desarrollo
- Crecimiento
- Supervivencia

La solución para la toma de decisiones apoyada por un *Data Warehouse*, o bien un *Datamart* debe ser considerada por cualquier empresa seria de negocios cuyo objetivo sea un crecimiento y desarrollo en su nicho empresarial.

Notas Bibliográficas

- HENRY F. KORTH & ABRAHAM SILBERSCHATZ. Fundamentos de Bases de Datos. McGraw Hill. México. 1988.
- W. H. INMON. Building the Data Warehouse, Mc Graw Hill. 1989.
- HARJINDER GIL. Cliente/Servidor Data Warehousing. Prentice Hall. México. 1996.
- PAL ROLPH & PETER BARTMAN. Executive Information Systems. 1990.
- EIS's Report Management Overview. Business Intelligence. 1990.
- Internet: <http://www.whatis.com/oltp.htm>

II

Definición del Modelo de *Data Warehouse*, Diseño y Construcción

*“La educación no es una cosa,
que uno adquiere,
sino un proceso de toda
la vida.”*

- Gloria Steinen

Objetivo del Capítulo:

Entender y comprender el ciclo de desarrollo de un Data Warehouse, así como proponer un modelo genérico para el mismo, con la finalidad de dar una guía simplificada para su la solución integral de Data Warehouse.

2.1 Ciclo del Desarrollo del *Data Warehouse*

El *Data Warehouse* sigue el mismo ciclo de perfeccionamiento que todos los desarrollos de software. Las fases del ciclo son las mismas, lo mismo que su secuencia⁽¹²⁾. Gráficamente tenemos lo siguiente:

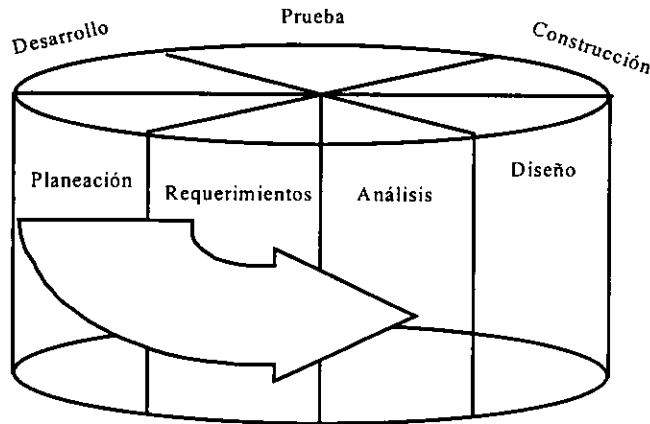


Figura 2.1

2.1.1 Etapas del *Data Warehouse*

I. Planeación

La planeación está constituida de múltiples pasos, algunos de ellos se pueden llevar en paralelo lo cual acorta el tiempo de duración de esta fase.

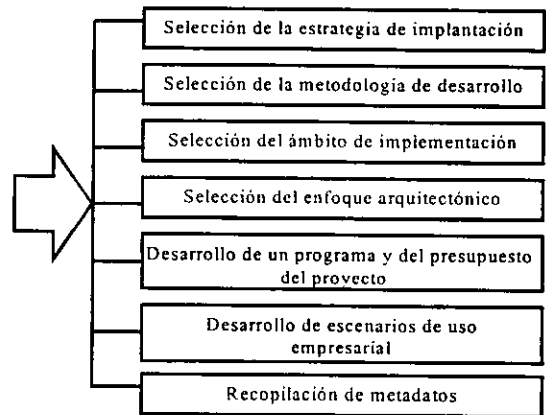


Figura 2.2

⁽¹²⁾ HARJINDER, S. Gill. *Cliente/Servidor Data Warehousing*. Prentice Hall Hispanoamericana. México. 1996.

⊙ Selección de la estrategia de implementación

Uno de los pasos más importantes consiste en decidir la estrategia general de implementación. La decisión se define dependiendo de la cultura de la organización y se basa como se llevan a cabo las tareas dentro de la organización. Las dos estrategias que son aplicables a este caso son:

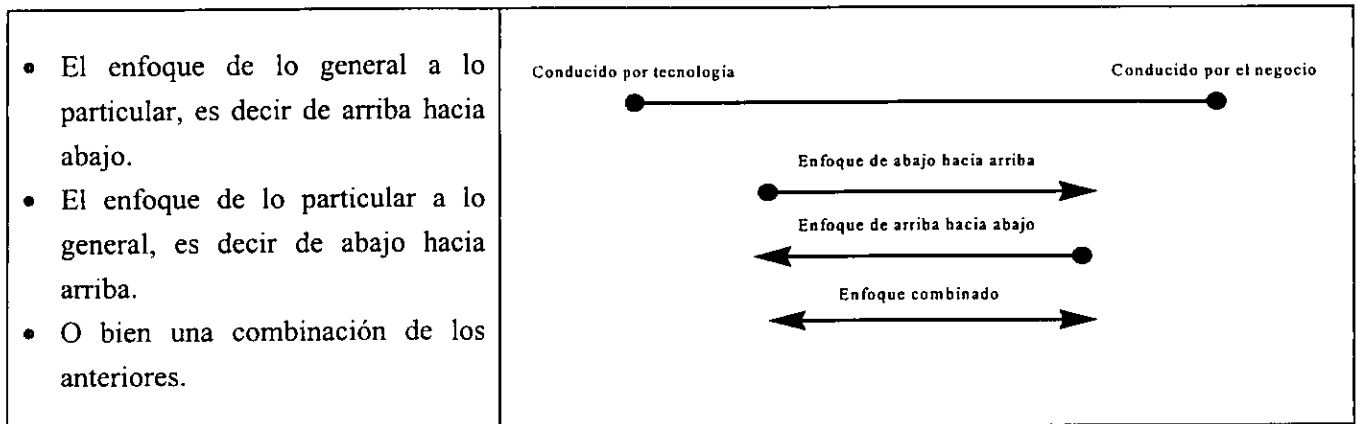


Figura 2.3

Enfoque de arriba hacia abajo: En este enfoque, se identifican primero los requerimientos empresariales que debe cubrir el *Data Warehouse* propuesto, estos son los principales conductores de la implementación del mismo; las ventajas(*) y desventajas(**) de este enfoque son:

- * Los requerimientos empresariales delimitan con claridad las fronteras de la implementación del *Data Warehouse* y puede ser, por lo tanto, de costo accesible y una forma enfocada a la implementación de la solución del *Data Warehouse*.
- * La tecnología es conducida por el negocio y no a la inversa.
- * Resulta fácil comunicar los beneficios del *Data Warehouse* a quienes toman decisiones y a la inversa.
- ** En ocasiones, pueden quedar oportunidades fuera del horizonte empresarial. Estas oportunidades perdidas son el resultado de demasiada concentración.
- ** La tecnología puede impulsar el negocio y ofrecer una ventaja competitiva que, en principio, no sea muy obvia para las actividades.
- ** Las expectativas iniciales pueden restringir la persecución de objetivos con recompensas potencialmente mayores, una vez implementado el *Data Warehouse*.

Un enfoque de arriba hacia abajo se recomienda cuando la tecnología está madura y se comprende bien, y cuando están claros y bien entendidos los problemas empresariales que deben resolverse. Aplicar este enfoque proporciona un excelente ajuste entre la tecnología y los objetivos empresariales y, cuando se hace de la manera adecuada, produce el máximo efecto para la inversión.

El enfoque de abajo hacia arriba comienza con experimentos y prototipos basados en la tecnología. Se selecciona un subconjunto de la problemática empresarial y se formula la solución para este subconjunto. Este enfoque permite que una organización avance con un gasto considerablemente menor y que evalúe los beneficios de la tecnología antes de establecer compromisos significativos. En el área del *Data Warehouse*, se toma por lo regular un enfoque de abajo hacia arriba para implementar un *datamart*, un pequeño sistema de información ejecutiva, o un *Data Warehouse* de departamental que esta claramente orientado a responder algunas consultas de un dominio determinado, como un área de contabilidad, análisis de mercado y administración de productos.

Un enfoque de abajo hacia arriba es útil para realizar evaluaciones de tecnología y es una buena técnica para organizaciones que no implementan tecnología de punta, también sirve para que una empresa aproveche las tecnologías que se encuentran en su ciclo temprano de madurez sin comprometerse a riesgos mayores.

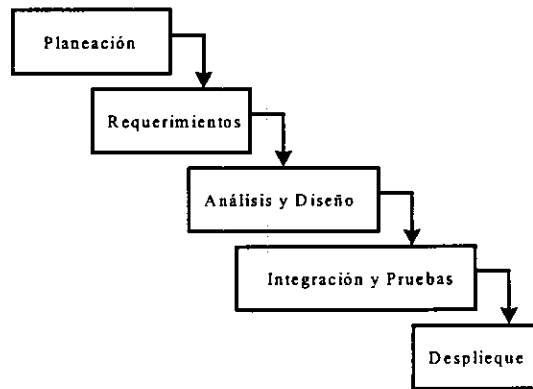
El enfoque combinado permite a una organización explotar la naturaleza planeada y estratégica del enfoque de arriba hacia abajo, al mismo tiempo que conserva la rápida implementación y aplicación oportunista del enfoque de abajo hacia arriba. Este enfoque depende de dos componentes:

- * Una arquitectura de arriba hacia abajo, estándares y un equipo de diseño que aplique la experiencia en el proyecto y que además pueda retroceder y convertir las decisiones tácticas en decisiones estratégicas.
- * Un equipo de proyecto de abajo hacia arriba que se concentre en implementar una solución empresarial muy enfocada, estrecha, pero de largo alcance, en un periodo de tiempo reducido. Este enfoque combinado se adapta mejor al despliegue (puesta en producción) rápido y oportunista de la tecnología del *Data Warehouse*, al tiempo que se reserva el derecho de construir una solución estratégica que tenga un valor a largo plazo.

© Selección de la metodología de desarrollo

Método de análisis y diseño en **cascada**. En este método, se reúne primero una serie de requerimientos. Estos requerimientos se analizan y se subdividen de manera progresiva. Después se construye un diseño mediante los resultados del análisis.

El diseño comienza como un nivel abstracto y se subdivide sucesivamente en niveles más concretos hasta que surge el código del sistema.

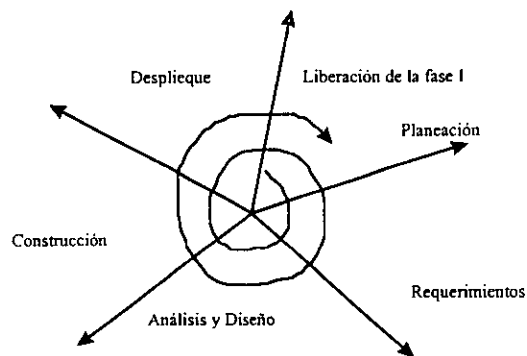


Diseño en Cascada

Figura 2.4

Método de desarrollo en **espiral**. En este método el énfasis está en la velocidad y la culminación, con un reconocimiento de que los requerimientos no se pueden identificar con claridad o especificar al inicio. El enfoque se basa en la observación de que es más fácil redirigir un sistema desplegado con nuevos requerimientos que construir una solución completa basada en requerimientos inadecuados o no disponibles.

De ahí, el método espiral es partidario de la rápida generación de sistemas cada vez más funcionales con intervalos cortos entre versiones sucesivas. El intervalo entre versiones se emplea para identificar e implementar funcionalidad adicional. El método espiral se presta principalmente al desarrollo de aplicaciones de base de datos, al desarrollo de un *Data Warehouse* y desarrollo de sistemas orientados a objetos.



Desarrollo en Espiral

Figura 2.5

En el dominio de desarrollo de productos, el método de desarrollo espiral se utiliza cuando surgen las siguientes situaciones:

- No se puede predecir con claridad ni anticipación la dirección de un mercado y sus requerimientos.
- El tiempo de colocación en el mercado es un factor importante en la implementación de un producto.
- Es necesaria una mejora iterativa para hacer correcciones de mercado.
- La ventaja competitiva proviene de mejoras súbitas en forma continua.
- A la organización le toma por lo menos seis meses absorber las versiones sucesivas de software.

El desarrollo del *Data Warehouse* presenta todas las características anteriores. Por lo tanto, el método espiral es la opción idónea para la metodología de desarrollo de un *Data Warehouse*.

Desarrollo de los objetivos empresariales. El primer paso para realizar cualquier tarea compleja consiste en desarrollar una lista de objetivos empresariales que el sistema debe de satisfacer, así el establecimiento de estos objetivos para el *Data Warehouse* es una misión difícil ya que existen diferentes perspectivas, terminología y definiciones de información. El considerar el *Data Warehouse* como un “producto” interno para la empresa puede ser una simplificación, es un sistema de software cuyos objetivos son apoyar decisiones para una audiencia puramente interna. Con esta visión es posible que el grupo de expertos en planeación de productos en una empresa se concentre en plantear los objetivos del *Data Warehouse*, lo que por lo regular puede ser sustentado usando preguntas como las siguientes:

¿Cuál es el mercado de destino (la audiencia potencial) del *Data Warehouse*?

¿Cuáles son las plataformas en uso actual o planeadas?

¿Cuáles son las capacidades en términos de características y funciones?

¿Cuáles son las diversas fuentes de datos y/o deben integrarse al *Data Warehouse*?

Respondiendo clara y concretamente estas interrogantes, el equipo de planeación podrá delimitar los objetivos empresariales a cubrir por el *Data Warehouse* y con ello tener un margen mínimo de error a tiempo de entrega de información que provendrá del *Data Warehouse*.

⊙ Selección de un ámbito inicial de implementación

En la mayoría de las organizaciones, la motivación principal del proyecto de *Data Warehouse* es una primera implementación que produzca beneficios inmediatos a un grupo de usuarios. Después de definir un rumbo general y un conjunto general de objetivos para el *Data Warehouse*, se hace necesario derivar con rapidez un ámbito limitado para una primera implementación. El ámbito del proyecto del *Data Warehouse* puede restringirse a múltiples dimensiones, como se ilustra en la Figura 2.6.

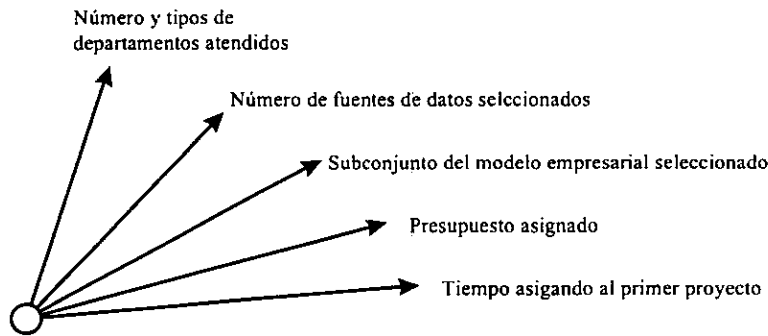


Figura 2.6

A grandes rasgos, las dimensiones se dividen en dos categorías principales:

- **Ámbito determinado a partir de la perspectiva del usuario empresarial del *Data Warehouse*.**
En este caso pueden formularse preguntas como: ¿cuáles son los departamentos que necesitan utilizar inicialmente el *Data Warehouse*?, ¿cuál es el rango de consultas empresariales a las que debe responder inicialmente el *Data Warehouse*?
- **Determinación del ámbito con base a consideraciones tecnológicas.**

Las preguntas asociadas a este rubro pueden ser: ¿cuál es el tamaño del metamodelo del *Data Warehouse*?, ¿cuál es el tamaño de los datos dentro del *Data Warehouse*?, ¿cuáles y cuántas son las fuentes de entrada de datos?, ¿qué tan utilizables son los datos de fuentes?, ¿qué tan bien están documentadas las fuentes de datos?, ¿cuál es la disponibilidad de modelos lógicos y herramientas de desarrollo?, ¿se cuenta con un modelo de datos empresarial?, ¿se pueden utilizar habilidades y recursos humanos existentes?, ¿se implementará el *Data Warehouse* sobre las plataformas existentes o sobre similares a las existentes?

⊙ **Enfoque de selección de una arquitectura**

En esta etapa dentro de la planeación se deberá seleccionar una opción de arquitectura para implementar, dentro de ellas se podrán considerar aquellas que se adapten con mejor con el negocio y con sus necesidades a cubrir:

Sólo Data Warehouse

Esta arquitectura reconoce que todas las aplicaciones del *Data Warehouse* requieren operaciones que se aplican sobre todas las fuentes de datos, esto implica que un sólo *Data Warehouse* lógico alimenta a todos los usuarios finales con información para apoyo a las decisiones.

Sólo <i>Datamart</i>	Esta arquitectura reconoce que cada departamento tiene sus propias necesidades y que un <i>Data Warehouse</i> corporativo no satisface estos requerimientos.
<i>Data Warehouse</i> y <i>Datamart</i>	En este caso se reconocen las necesidades específicas de un departamento junto con las de un <i>Data Warehouse</i> corporativo.
Separación de plataforma e infraestructura	Seleccionar que plataformas se utilizarán para el <i>Data Warehouse</i> , <i>datamart</i> , fuentes de datos y herramientas; se planteará en donde existirán o coexistirán dependiendo de su factibilidad entre todos estos componentes.
Arquitectura cliente/servidor de 2 capas	Se determinará si se utilizará una capa a nivel servidor (típicamente de base de datos) y clientes (herramientas gráficas de explotación/presentación).
Arquitectura cliente/servidor de 3 capas	Se usará una capa cliente(herramientas gráficas de explotación/presentación), un servidor de aplicaciones(incluye los <i>datamarts</i> y herramientas del <i>Data Warehouse</i>) y otro de base de datos(información del <i>Data Warehouse</i>).

© Desarrollo de un programa y del presupuesto del proyecto

Esta fase dentro de la planeación del *Data Warehouse* es de suma importancia, los pasos necesarios para su éxito son:

√ Articular un plan de programa como un conjunto de planes de proyecto

Un plan de programa es una visión general de la actividad del *Data Warehouse* y su función en la vida diaria y semanal de la organización. El *Data Warehouse* se emplea como auxiliar de la planeación para mejoras continuas de las operaciones, así como para proporcionar avances competitivos con base en el análisis de los datos. Por lo tanto, el plan de programa identifica los diversos departamentos y unidades empresariales que usarán el *Data Warehouse*. Así mismo establece prioridades sobre la implementación, para atender primero las necesidades empresariales críticas. Sin embargo, los planes de proyecto son los planes para implementaciones específicas del *Data Warehouse*. Si el plan de programa proporciona la estrategia, los planes de proyecto proporcionan la táctica. Adicionalmente los planes de proyecto incorporan las prioridades del plan de programa.

-
- √ Reservar un presupuesto adecuado para el programa al tiempo que se compromete el gasto para proyectos específicos.

La planeación del presupuesto se basa en dos enfoques:

- a) Estimación del costo, con base en el historial de la organización en el desarrollo del software. Se asignan porcentajes a las fases del ciclo de vida del software.
- b) Estimación del costo, con base a la arquitectura de referencia seleccionada. En este caso se asignará presupuesto para cada componente de la arquitectura de referencia: fuentes de datos, *Data Warehouse*, *datamart*, etc.

- √ Proporcionar medidas para la estimación de la retribución del *Data Warehouse*

La estimación de retribución del *Data Warehouse* es una tarea complicada por la prolongada demora entre el análisis y las acciones resultantes. Algunas de las medidas que pueden emplearse son las siguientes:

- Medidas de recuperación de costo o de ahorro de costo
- Medidas de creación de oportunidades
- Medida de creación de ingresos
- Medida de crecimiento de mercado
- Medida de ventaja competitiva
- Medidas de satisfacción al cliente

⊙ Desarrollo de escenarios de uso empresarial

Los escenarios empresariales son una importante herramienta del prototipo de requerimientos. Un escenario consta de los siguientes ingredientes:

- * Definir un usuario empresarial claramente identificado, como un gerente de producción de accesorios de automóvil.
- * Un área funcional que respalde el *Data Warehouse* o el *datamart* y que lo utilice cuando este terminado, por ejemplo: un área de análisis de mercado.
- * Una o más consultas empresariales de interés crucial para el área funcional que no satisfagan actualmente a los sistemas de información existentes (Por ejemplo: ¿cuáles son las ventas por trimestre de aire acondicionado para automóviles durante los dos últimos años en el mercado de Guadalajara?). Las consultas empresariales deben formularse en términos precisos empleando el lenguaje común de negocios que entiendan los usuarios finales y el área funcional.

Los escenarios empresariales ayudan a definir las expectativas del usuario final con respecto al *Data Warehouse*.

⊙ Recopilación de metadatos

Parte de la fase de planeación del *Data Warehouse* es la necesidad de recopilar varios elementos de diseño relacionados con los metadatos. Los metadatos es el término que se emplea para las definiciones de datos, estos representan ingredientes del anteproyecto que se utilizan para construir el *Data Warehouse*.

II. Requerimientos

La fase de requerimientos de la implementación del *Data Warehouse* es una especificación precisa de las funciones de las funciones que se obtendrán del *Data Warehouse*. Además de las características y funciones necesarias, los requerimientos describirán con claridad el ambiente operativo en el que se entregará el *Data Warehouse*.

Un elemento importante en la identificación de requerimientos será el ver estos desde distintos puntos de vista o expectativas. Así los requerimientos se pueden ver desde los siguientes ángulos:

a) Del propietario

Algunas de las cuestiones que proponen los propietarios o inversionistas del *Data Warehouse* son las siguientes:

- ¿Por qué construir un *Data Warehouse* o un *datamart*?
- ¿Qué problema empresarial abordaría?
- ¿Cuáles son los objetivos empresariales?
- ¿Quién es el cliente?
- ¿Cuál es impacto sobre la gente? ¿Sobre las habilidades?
- ¿Cómo afecta nuestras inversiones actuales de cómputo?
- ¿Tenemos capacidad para hacerlo?
- ¿Cuáles son los riesgos?
- ¿Cuánto costará? ¿cuándo estará listo?

Las áreas potenciales para acopio de los requerimientos son las siguientes:

-
- Objetivos empresariales.
 - Ámbito y objetivos del *Data Warehouse*, *datamart*, clientes.
 - Fuentes de datos.
 - Un plan, tal como presupuesto, calendario y recursos.
 - Impacto sobre las inversiones actuales, como en personal, tecnología y capacitación.
 - Requerimientos del cliente.

Las **áreas tema** son las que cumplen con diversas funciones empresariales. Una selección cuidadosa de estas áreas contiene el ámbito de la implementación del *Data Warehouse* al tiempo que maximiza su utilidad. Un ejemplo puede ser las áreas de comercialización y los temas pueden ser los siguientes:

- Investigación de los mercados
- Análisis de competitividad
- Comportamiento del comprador
- Producto de segmentación por mercado
- Decisiones sobre precios y presupuestos
- Decisiones sobre productos
- Decisiones sobre promociones
- Decisiones sobre canales
- Decisiones sobre tendencias
- Pruebas de calidad

El análisis de los temas seleccionados produce áreas específicas para el departamento de comercialización:

- Pedidos
- Promociones
- Mercados
- Ventas
- Ciclo de tiempo

La **granularidad** del *Data Warehouse* o *datamart* nos proporciona el nivel de detalle de la información requerida, esta tiene una relación directa con las actividades de resumen y adición que deben realizarse sobre los datos fuente. A menor granularidad, mayor cantidad de detalle. Un ejemplo de nivel de granularidad puede ser una pregunta como la siguiente: ¿Quién es el representante de ventas con más alto ingreso?, ¿cuál es el producto más vendido?.

Las dimensiones representan la categoría de los datos, esta información se puede asociar con respecto al tiempo. Así se puede obtener información que resuelva cuestiones en términos de intervalos de tiempo; por ejemplo: ¿cuál fue la venta de un producto el último trimestre del año?.

b) Del arquitecto

El arquitecto es la responsable de diseñar los diversos componentes del *Data Warehouse* para sustentar las necesidades actuales y futuras. La calidad del esfuerzo de arquitectura determinará lo siguiente:

- El rango de funciones y características a ofrecer.
- El rango de plataformas para la implementación.

- El uso de estándares e interfaces abiertas.
- La flexibilidad para incorporar mejoras.

Los arquitectos deben compilar una serie de requerimientos que coincidan con la visión del propietario, así como un conjunto de requerimientos que refleje la implementación de la tecnología.

c) Del desarrollador

La tarea del desarrollador es plasmar de una forma concreta lo que el arquitecto crea en abstracto. El desarrollador subdividirá el *Data Warehouse* en interfaces, computadoras, bases de datos, comunicaciones y aplicaciones; todas estas se utilizarán para llevar a la realidad los bloques de fuentes de datos, del *Data Warehouse*, del *datamart* y del bloque de acceso y herramientas. Otra de las actividades del desarrollador es verificar los requerimientos necesarios para poner en producción el *Data Warehouse*, en esta etapa se contemplan: consistencia de datos, confiabilidad de la información, manejo de los metadatos del *Data Warehouse*, mecanismos de transporte de fuentes, bases de datos, soporte técnico, políticas y procedimiento de acceso y seguridad, manejo de bases de datos extensas, tiempo de respuesta adecuado, contingencias y respaldos.

d) De los usuarios finales

El usuario final ve el *Data Warehouse* como una caja negra cuyo acceso principal es a través de las aplicaciones y herramientas de consulta y de reportes.



Las Preguntas de Usuarios en un *Data Warehouse*

Figura 2.7

III. Análisis

La fase de análisis del ciclo de desarrollo del *Data Warehouse* significa convertir los requerimientos acopiados en la fase anterior en un conjunto de especificaciones que puedan apoyar el diseño. En lo abstracto, hay tres especificaciones principales de entrada para el *Data Warehouse*. Estas son:

-
- ◇ Requerimientos de enfoque empresarial que delimitan las fronteras de la información que debe comprender el *Data Warehouse*. El enfoque empresarial determinará también la audiencia y sus requerimientos de información.
 - ◇ Especificaciones de requerimientos de fuentes de datos que delimitan las fronteras de información disponible en fuentes de datos actuales.
 - ◇ Especificaciones de requerimientos de usuario final y de acceso, las cuales definen cómo se utilizará la información del *Data Warehouse*. Junto con estas se encuentra la especificación de los tipos de herramientas y técnicas de exhibición que usa.

El proceso de análisis consiste en derivar modelos físicos y lógicos de datos para el *Data Warehouse* y los *datamarts* y definir los procesos necesarios para conectar las fuentes de datos, el *Data Warehouse*, los *datamarts* y las herramientas de acceso al usuario final.

Los modelos de datos deben ser analizados cuidadosamente, así como optimizados para las consultas empresariales de forma tal que el desarrollo nos deje un modelo de datos listo a ser utilizado para la consulta de datos ejecutiva.

La información que contiene el *Data Warehouse* se debe organizar de una forma diferente al que tiene la fuente de datos. A menudo, se almacenan totales precalculados y cuentas (resúmenes y adiciones) en el *Data Warehouse*. Es frecuente que se incorporen a la información que ingresa datos adicionales como la fuente, la fecha y la hora de origen, para proporcionar un contexto básico dentro del análisis conocido como *visión del Data Warehouse*⁽¹³⁾.

Visión de las fuentes de datos. El *Data Warehouse* carga y almacena datos de muy diversas fuentes. Estas fuentes se caracterizan por los siguientes aspectos:

- Tecnologías de almacenamiento múltiple.
- Definiciones múltiples de datos.
- Campos nulos.
- Diferencias de formato y codificación.
- Duplicidades.

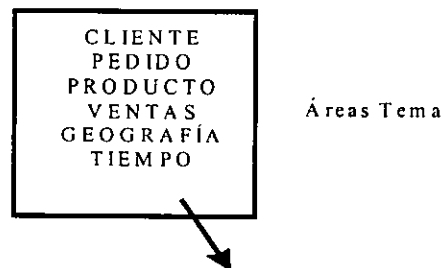
Visión del *Data Warehouse*. El *Data Warehouse* se organiza de manera diferente a los datos operacionales; el modelo de datos operacionales se encuentra construido para eliminar redundancias, coordinar transacciones y operar altos volúmenes de las mismas, este modelo contiene múltiples normas que definen su seguridad, integridad de datos referencial. En cambio el *Data*

⁽¹³⁾ HARJINDER, S. Gill. *Cliente/Servidor Data Warehousing*. Prentice Hall. México. 1996.

Warehouse se organiza de una forma diferente, este debe estar diseñado para responder a consultas de datos y no para satisfacer operaciones transaccionales masivas.

Debido a las diferencias fundamentales entre las bases de datos operacionales y las de *Data Warehouse* es necesario que cada una de ellas tenga un modelado de datos diferente. En el ámbito operacional el modelado de datos más conocido es el entidad-relación y en el ambiente de *Data Warehouse* el modelado estrella multidimensional.

Visión de consulta empresarial. La visión de consulta empresarial es la perspectiva de datos del *Data Warehouse* desde el punto de vista del usuario. El modelo de datos estrella multidimensional es un reflejo cercano a la forma en que un analista empresarial visualiza una consulta. De hecho una tabla multidimensional es una representación exacta de una consulta multidimensional. Por lo tanto, para el analista, proponer la consulta es lo mismo que consultar directamente la tabla multidimensional. De ahí que la consulta empresarial sea una solicitud de hechos, a veces llamados mediciones, a través de varias dimensiones. Por ejemplo, la tabla de hechos (véase Figura 2.8) contiene la información de venta, las dimensiones son el tiempo (últimos seis meses), el producto (aire acondicionado) y la ubicación (Guadalajara). Mediante la organización del *Data Warehouse* en estas dimensiones y tablas de hechos, se formula y se satisface con facilidad la consulta empresarial.

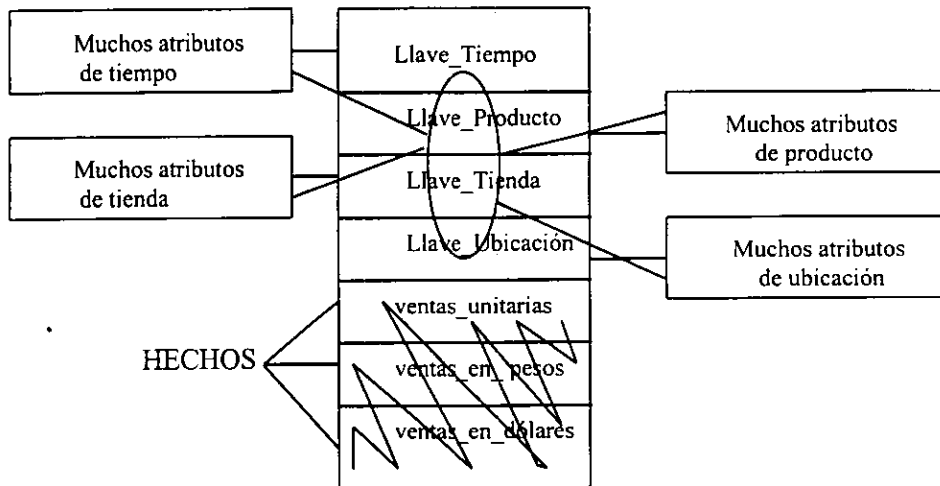


Derivados de reporte

Figura 2.8

El **modelo estrella multidimensional** representa los datos como un arreglo en el que cada dimensión es un tema alrededor del cual se hace un análisis. El tiempo es siempre una de las dimensiones. Las otras dimensiones dependen del problema empresarial que se analice. Este esquema consiste de un sólo objeto

en medio conectado con varios objetos de forma radial. El esquema refleja la visión del usuario final de una consulta empresarial: hechos tales como ventas, compensación, pago y facturas califican por una o más dimensiones (por mes, por producto, por región geográfica). El objeto en el centro de la estrella se denomina *tabla de hechos* y los objetos conectados a ella se denominan *tablas de dimensión*.



Esquema Estrella

Figura 2.9

Así el análisis debe incluir la visión desde todos los puntos de vista y un modelado eficiente de la información.

IV. Diseño

En la fase del diseño, los modelos lógicos desarrollados en la fase de análisis se convierten en modelos físicos. Los procesos identificados en la fase del análisis para conectar las fuentes de datos con el *Data Warehouse*, el *Data Warehouse* con los *datamarts* y el *Data Warehouse/datamart* con las herramientas de usuario final, se convierten en diseños para programas que realizan las tareas requeridas por los procesos.

En la fase de diseño se encuentran las siguientes dos actividades principales:

⇒ Diseño detallado de la arquitectura de datos.

→ Desarrollo de los modelos físicos de datos para la base de datos de almacenamiento del *Data Warehouse/datamart*.

→ Correspondencia de los modelos físicos de datos de las fuentes de datos con los modelos físicos del *Data Warehouse/datamart*.

⇒ Diseño detallado de la arquitectura de la aplicación.

→ Procesos que son internos a las fuentes de datos y se relacionan con depuraciones o extracciones parciales de información y procesos que conectan las fuentes de datos con el *Data Warehouse*.

→ Procesos que son internos al *Data Warehouse* y se usan para fines de manejo interno.

→ Procesos que conectan el *Data Warehouse* con el *datamart*.

→ Procesos que conectan el *Data Warehouse* con herramientas de usuario final.

→ Procesos que sustentan tareas de manejo, administración y prácticas internas para el *Data Warehouse* como sistema.

V. Construcción

La fase de la construcción es la responsable de implementar físicamente los diseños desarrollados durante la fase de diseño. Los programas que se deben crear en esta fase son:

- Programas que crean y modifican las bases de datos para el *Data Warehouse* y los *datamarts*.
- Programas que extraigan datos de fuentes relacionales y no relacionales.
- Programas que transformen datos, tales como integración, resumen y adición.
- Programas que actualicen las bases de datos relacionales.
- Programas que hagan búsquedas en bases de datos muy grandes. Varias herramientas de usuario final han optimizado utilerías de búsqueda para las consultas que generan.

VI. Despliegue

La fase de despliegue tiene que ver principalmente con los retos de instalación, puesta en servicio y uso de la solución de *Data Warehouse*. Las actividades que deben cubrirse son:

- Proporcionar la instalación inicial, incluyendo conexiones básicas de datos con fuentes, para actualización y sincronización de los datos.
- Planeación y entrega de una implementación por etapas.
- Proporcionar capacitación y orientación a todo tipo de usuarios.
- Planeación e implementación de la actualización de plataformas y mantenimiento necesario para el *Data Warehouse*.
- Proporcionar la administración de usuarios y sistemas.
- Proporcionar la capacidad de generar archivos permanentes y respaldos.

-
- Proporcionar la estrategia de recuperación en caso de contingencia.
 - Asegurar la integración dentro de la infraestructura existente.
 - Proporcionar los controles de acceso y seguridad.
 - Asegurar la completa disponibilidad y los procesos para manejar caídas de los sistemas y sus componentes de infraestructura.
 - Definición de los metadatos en términos de los usuarios finales y no técnicos.

El *Data Warehouse* es un método para ofrecer información para un grupo de usuarios finales que entienden el negocio pero no necesariamente la tecnología.

VII. Expansión/Desarrollo

El método de desarrollo espiral es un ingrediente clave para el despliegue rápido inicial de un *Data Warehouse*. Este método requiere también una rápida evolución de características y funciones basadas en las lecciones aprendidas en anteriores despliegues(iteraciones), con la información obtenida se podrán hacer las mejoras pertinentes al *Data Warehouse*.

2.1.2 Fase Inicial del *Data Warehouse*

Un despliegue inicial exitoso es esencial para el éxito de un proyecto de *Data Warehouse*. Los principales pasos que comprende el despliegue inicial son:

- Instalación inicial
- Planeación de la distribución
- Capacitación y orientación

Instalación inicial

En la mayoría de los casos, las organizaciones planean un despliegue inicial piloto de un *Data Warehouse* en pequeña escala, para apoyar a un número reducido de usuarios. El despliegue de cualquier tecnología nueva está lleno de riesgos, sin importar los méritos y beneficios finales de la tecnología. Administrar el despliegue inicial es fundamental para alcanzar un pronto éxito. El despliegue inicial de la tecnología de *Data Warehouse* representa una curva de aprendizaje para el usuario.

El procesamiento analítico en línea (OLAP, por sus siglas en inglés), la consulta y las herramientas de acceso de datos representan un reto importante de aprendizaje. La capacitación a usuarios debe incluir cómo resolver problemas por medio de las herramientas, además de operar ellos mismos las herramientas. Esta capacitación de aplicaciones debe proporcionarse utilizando datos reales para

solucionar problemas empresariales reales o simulados en el *Data Warehouse*, en vez de utilizar bases de datos estándar para capacitación. Solucionar ejemplos de retos empresariales ilustra el uso de herramientas de manera mucho más clara que muchos ejemplos inventados que se utilizan generalmente para una capacitación no dirigida.

Planeación de la distribución

Después de que un pequeño número de usuarios esta usando el *Data Warehouse* y de que los procesos de los aspectos relacionados están resueltos, la solución de sistema de información debe aumentar su escala a una mayor audiencia. La fase de distribución es aquella en la que se caracteriza una solución inicial y se distribuye a grandes audiencias de usuarios. Todos los componentes de la distribución están claramente identificados y los procesos para el manejo, instalación e integración de estos componentes están establecidos con claridad, verificados y probados. Por lo tanto la preparación para la distribución debe incluir los siguientes pasos:

- * Estandarizar los componentes de la solución de *Data Warehouse* que deben distribuirse. Estandarizar los componentes es el proceso de seleccionar un componente del mismo fabricante, adoptar los mismos procesos de instalación, utilizar los mismos parámetros de configuración y seleccionar las mismas plataformas operativas.
- * Estandarizar los procesos de instalación y despliegue para la distribución. Estandarizar un proceso comprende documentarlo de manera clara, paso a paso, describir una lista de requisitos previos(incluyendo habilidades del operador) y aplicar los pasos del proceso para verificar que funcionan.
- * Planear el método de distribución. Diferentes organizaciones utilizan una variedad de métodos de distribución. El más caro y probablemente el más eficaz consiste en enviar personal capacitado a instalar la solución en los distintos sitios de la distribución. El método menos caro es distribuir la solución a través de una red, con una lista de requisitos previos y un conjunto de manuales de procesos, esto se puede complementar con una ayuda de escritorio (help desk).

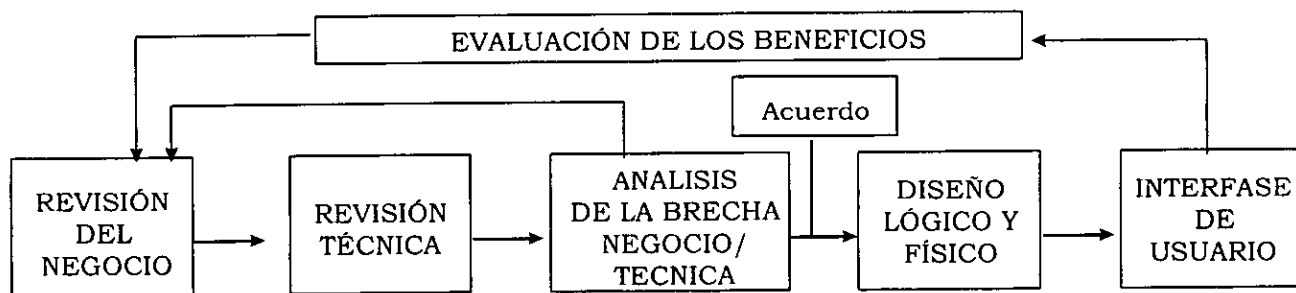
Capacitación y orientación

El *Data Warehouse* es una solución de sistemas conformada por varios componentes de tecnología. Se deben formular programas de capacitación específicos en el *Data Warehouse*, son necesarios programas por separado para los usuarios finales y al personal administrativo del *Data Warehouse*. Adicionalmente, puede ser necesaria la capacitación intensiva en el diseño físico de bases de datos, administración de sistemas de bases de datos y rendimiento de bases de datos.

2.2 La Solución Integral de *Data Warehouse*: Modelo Propuesto

En la sección anterior se describieron cada uno de los pasos a seguir para llevar a cabo la implementación del *Data Warehouse*, la metodología nos lleva desde la planeación, requerimientos, análisis, diseño, construcción, despliegue y la expansión de la solución, adicionalmente se hablo del enfoque que debemos tomar para planear el desarrollo del *Data Warehouse* contemplando los requerimientos desde un punto de vista de arriba hacia abajo o viceversa. De esta manera y apoyando todos estos pasos, a continuación se presenta un modelo genérico por medio del cual se simplifica esta tarea.

- I. Identificación de áreas susceptibles a creación de un *datamart* (sub-conjuntos del *Data Warehouse*).
- II. Establecer prioridades de estas áreas, de esta manera la primer área en esta lista será aquella sobre la cual se generará el primer *datamart*.
- III. Se realizará el análisis de negocio de esa área, identificando objetivos del mismo, problemas a resolver, recursos disponibles, es decir: el medio ambiente actual del negocio.
- IV. Análisis de los recursos técnicos en el área de estudio, incluyendo estado actual: recursos humanos, técnicos y de información.
- V. Hacer una relación entre el negocio y el área técnica homologando criterios con base a los recursos de estas áreas, en este paso se resolverán las diferencias entre estas 2 entidades y se entregará un documento que detalle la información a obtener del *datamart*.
- VI. Diseño y Construcción del *Datamart* usando alguna de las opciones tecnológicas presentadas en la figura 3.5 (Capítulo III), o bien apartado 3.5.
- VII. Generación de la interface para el usuario.
- VIII. Verificación de resultados de la iteración correspondiente.
- IX. Evaluación de beneficios.
- X. Comenzar una nueva iteración del *datamart* (I).



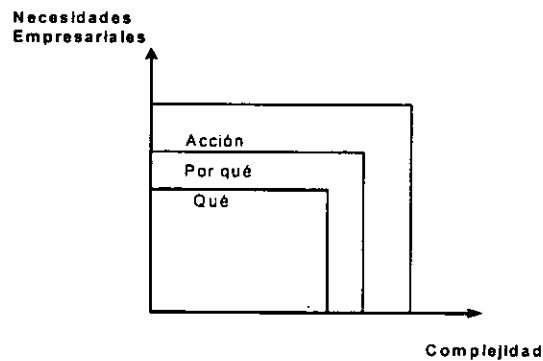
Modelo Propuesto para la Construcción de un *Data Warehouse*

Figura 2.10

2.3 La Minería de Datos

La minería de datos auxilia a los usuarios empresariales en el procesamiento de vastas reservas de datos para descubrir relaciones insospechadas, por ejemplo, entre productos y clientes o patrones de compra de clientes: la meta es descubrir “revelaciones estratégicas competitivas” para controlar la participación de mercado y las utilidades. Los seres humanos tienen agudeza para percibir excepciones y anomalías, pero no tienen la potencia y capacidad de inferir relaciones en grandes volúmenes de datos. Sin embargo, una vez extraídas las relaciones y presentadas a los analistas empresariales, estos las examinan y seleccionan las más útiles e interesantes.

Los analistas empresariales tienen un rango de necesidades. La primera necesidad es comprender qué está sucediendo en el negocio. La siguiente es comprender por qué está sucediendo. El valor de un análisis para los gerentes es más alto cuando genera una recomendación factible. Comprender el comportamiento y los pronósticos de los clientes y mercados, y lo que puede hacerse, son los retos para técnicas tradicionales de análisis. Las consultas, reportes y análisis multidimensional se concentran en lo que está sucediendo y, en menor medida, en el por qué. La minería de datos se concentra en llenar la necesidad de descubrir el por qué, para luego predecir y pronosticar acciones con cierto factor de confianza para cada predicción.



Las necesidades Empresariales y su Complejidad

Figura 2.11

Data Mining es un proceso llevado a cabo por un usuario o un programa de software en búsqueda de patrones de información en una base de datos.

Esta tarea puede ser hecha de una forma reactiva o a iniciativa del usuario, donde este define un conjunto de preguntas específicas y patrones de búsqueda, con lo cual la base de datos lleva a cabo la búsqueda de información.

Por el otro lado esta tarea también puede llevarse a cabo de una manera proactiva o a iniciativa del sistema, donde un software inteligente, el cual puede tener un motor usando redes neuronales, generando preguntas por sí mismo, analizando la base de datos, y busca por patrones sin instrucciones del usuario.

También existe una iniciativa donde el usuario y el sistema interactúan, el sistema proporciona cierta asistencia; por ejemplo: análisis multidimensional con la visión corporativa; donde el usuario puede ver a los datos desde múltiples puntos de vista, con preguntas realizadas al mismo. Las aplicaciones de *data mining* pueden realizar otras actividades:

Modelado predictivo: Los patrones descubiertos desde la base de datos se utilizan para predecir el futuro. El modelado predictivo permite al usuario enviar registros de uno más valores desconocidos mientras el sistema predice lo desconocido basado en patrones previos de la base de datos.

Análisis de pronóstico: Los patrones extraídos son usados para encontrar anomalías o elementos inusuales. Para descubrir lo inusual, la aplicación primero encuentra cuál es la norma, entonces detecta que campos se desvían del comportamiento típico de los datos.

El *data mining* y las bases de datos inteligentes ahora son más significativas en el mercado y las áreas de investigación. Algunas personas han sugerido que la tecnología está anticipada a las áreas de aplicación y que existe una necesidad para personalizar los sistemas actuales de los clientes para proporcionar soluciones dentro de industrias específicas tales como: marketing, banca, manufactura, etc.

Representación de Datos

La representación de datos es la abstracción del concepto a ser presentado. La elección de las estructuras de datos (al nivel de sistema) para la representación es muy importante ya que dicta como el nivel del sistema interactúa con la interface del usuario. Una buena representación puede tener las siguientes características:

Eficiencia: La representación debe ser seleccionada de forma que el acceso y las consultas puedan realizarse eficientemente.

Flexibilidad y portabilidad: Una abstracción debe ser suficientemente general para permitir el uso extensivo en diferentes partes de un modelo y en diferentes tipos de modelos.

Claridad: Una representación debe ser clara y bien definida mostrando adecuadamente la representación del objeto así como sus relaciones.

Transparencia: Tanto los diseñadores como los usuarios finales deben ser capaces de entender el modelo sin importar el nivel de abstracción de un objeto en particular.

Visualización de Datos

La visualización asistida por computadora puede dividirse en dos categorías: visualización científica y visualización intuitiva de datos.

La visualización científica estriba en convertir grandes cantidades de conjuntos de datos en fotografías comprensibles. Los conjuntos de datos se generan con instrumentos modernos y super computadoras, de forma que la visualización es esencial para su comprensión. Los resultados pueden requerir de un examen minucioso ya que su interpretación puede ser no obvia para la audiencia en general. La visualización científica usualmente envuelve un gran número de programación numérica: vectores aritméticos, inversión de matrices, etc.

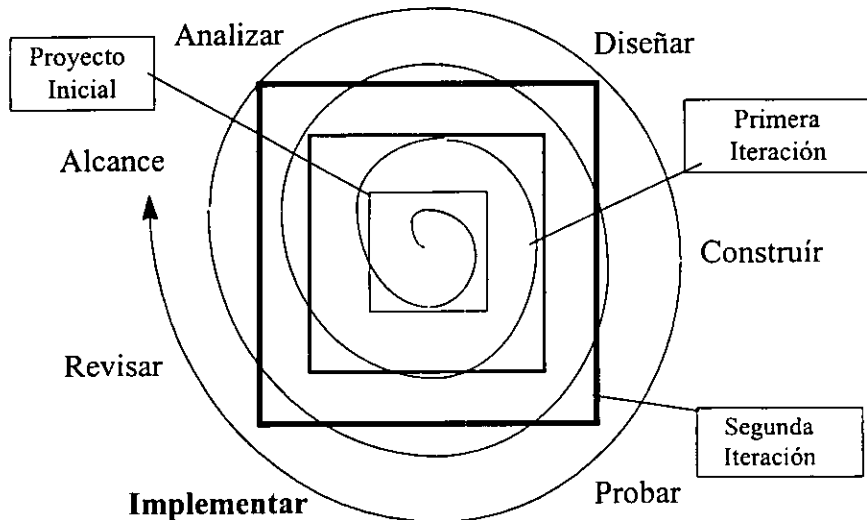
La visualización intuitiva permite el entendimiento de los datos por medio de una interpretación visual de estructuras de datos complicadas. Su meta es explicar o promover resultados para una amplia audiencia. La visualización intuitiva de datos es una tecnología que proporciona innovaciones tecnológicas que presentan los datos de una forma única y de una manera no tradicional. Los tipos de datos, formatos, y técnicas aplicadas incluyen descripciones multidimensionales y representaciones animadas. Por medio de la visualización se presentan cálculos computacionales, no geométricos y datos experimentales. Mientras las imágenes y datos interactivos son una herramienta presente, comienzan a surgir opciones como la animación en video para ilustrar estos conceptos.

Las propiedades de una interface de visualización de datos son:

- Manipulación de usuario y edición
- Soporte a multiusuarios
- Conectividad de red
- Interactividad

2.4 El Data Warehouse un Proceso Iterativo

El *Data Warehouse* es un proceso de mejora continua, cada ciclo completo dentro de un *Data Warehouse* o un *datamart* permite ir mejorando el modelo, o bien integrar nuevas necesidades de negocio dentro del mismo. De esta manera, el proceso de forma gráfica queda definido de la siguiente manera:



El *Data Warehouse* un Proceso Iterativo

Figura 2.12

Con todo lo descrito en este capítulo se da una guía de cómo comenzar un *Data Warehouse*, o bien un *datamart*; todo este proceso involucra múltiples piezas, desde tecnológicas, humanas, de negocios y de política; más sin embargo su exitosa construcción e implantación dejará un valor agregado a las organizaciones que lo lleven a la práctica.

Conclusiones del Capítulo

El *Data Warehouse* como todo proceso de sistemas involucra múltiples pasos a seguir para su correcta implantación. De aquí se concluye que las fases por las cuales debe pasar un *Data Warehouse* deben ser ejecutadas minuciosamente: *desde* la recopilación de requerimientos del área de negocios que utilizará esta solución como una herramienta para su proceso de toma de decisiones; el análisis adecuado del sistema, incluyendo las fortalezas y debilidades, así como, las amenazas y las oportunidades del mismo; el diseño en el que se deberán reflejar las necesidades cubiertas para generar los indicadores de negocios del *Data Warehouse* para la toma de decisiones; la construcción y desarrollo con una tecnología cuyo costo/beneficio sea satisfactorio para el negocio; *hasta* la implantación final y mejora continua del sistema para la toma de decisiones.

El uso de la información generada por un *Data Warehouse* más su integración con la minería de datos, permitirá a las empresas redescubrir su negocio, vislumbrando amenazas y generando oportunidades nunca antes vistas, las cuales servirán para planear en el presente para un futuro exitoso.

La correcta toma de decisiones y sus implicaciones inherentes llevan a concluir que sólo allegándose de los hechos y evidencias necesarias, se podrá crecer hacia el conocimiento completo en la dinámica empresarial actual, y este paso puede ser soportado por un proceso de planeación en el cual el *Data Warehouse* es el protagonista principal.

Notas Bibliográficas

- HARJINDER, S. Gill. Cliente/Servidor Data Warehousing. Prentice Hall. México. 1996.
- W. H. INMON. Building the Data Warehouse, Mc Graw Hill. 1989.
- Internet: <http://www.knowledgecenters.org/>

III

El Valor Agregado del *Data Warehouse* para la Toma de Decisiones Estratégicas

*“Todo acontecimiento en el mundo visible
es el resultado de una imagen, es decir,
de una idea en el mundo invisible.”*

- I Ching

Objetivo del Capítulo:

Entender cuáles son los beneficios de la construcción e implantación de un Data Warehouse dentro del ámbito empresarial, así como conocer los proveedores de productos y servicios para la integración de dicha solución.

3.1 La utilidad del *Data Warehouse* en la Planeación Estratégica

La planeación estratégica⁽¹⁴⁾ es una técnica que permite:

- Conocer en el presente el futuro de la empresa.
- Predecir y anticipar eventos futuros.
- Hacer que se presenten situaciones que de otra forma no se hubiesen presentado.
- Fijar objetivos corporativos y definir estrategias⁽¹⁵⁾ institucionales.

La empresa debe definir sus planteamientos estratégicos para eficientar su operación a través de principios básicos:

- Rentabilidad
Es obtener mejores resultados que el mercado.
- Productividad
Dar más rendimiento que antes o que otras empresas.
- Competitividad
Ofrecer ventajas sobre el mismo producto o servicio sobre el resto del mercado.

La producción de bienes o la prestación de servicios deben verse a través del triángulo virtuoso (Figura 3.1) que permite establecer un equilibrio entre los accionistas, los administradores y los clientes, de tal forma que la empresa permanezca siempre en el mercado, ofreciendo servicios de calidad que satisfagan las necesidades de la sociedad siendo competitivos, rentables y productivos, ya que si alguno de estos tres conceptos no se da en forma asociada y balanceada con los otros dos, es inminente que la empresa u organización morirá, a menos que se tomen medidas (toma de decisiones) que corrijan y equilibren la operación.

⁽¹⁴⁾ Control de Gestión basado en Planeación Estratégica. Consultores SICOFIN. Pág. 17.

⁽¹⁵⁾ Estrategia se define como cursos de acción general que muestran la dirección y empleo general de recursos y esfuerzos de toda la organización. Se deriva del griego "Strategas" que significa estrictamente un general. Consultores SICOFIN. Pág. 20.



Esquema de Análisis de Servicios

El Triángulo Virtuoso

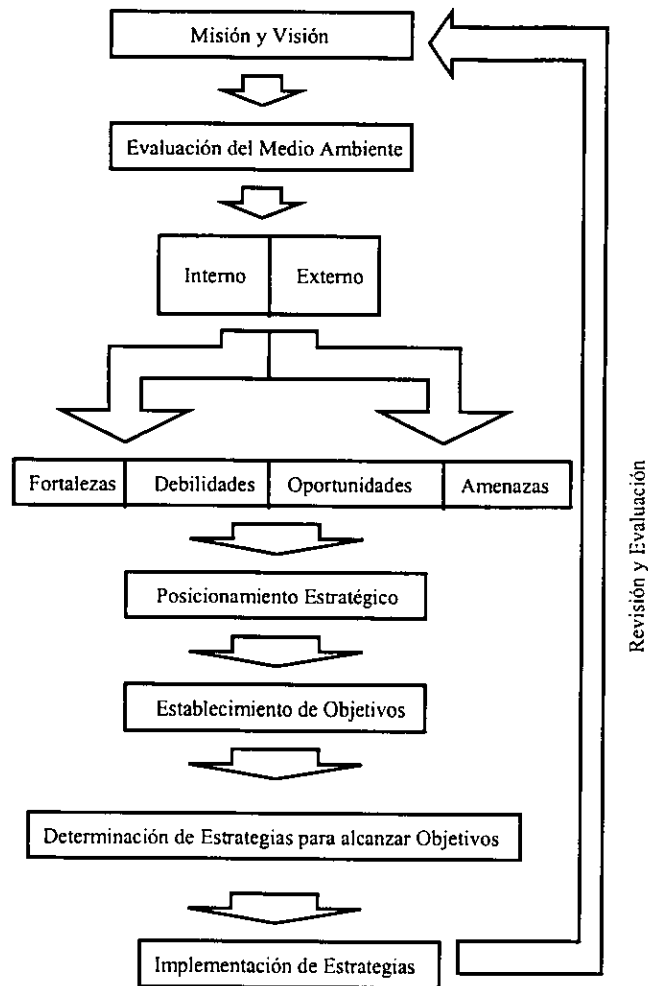
Figura 3.1

De esta manera la información que sea extraída por medio del *Data Warehouse* deberá ser utilizada para satisfacer este triángulo virtuoso, adicionalmente deberá considerar el marco estratégico de actuación, es decir, el medio ambiente que rodea e influencia el desempeño de las empresas y que deberá ser considerado para elaborar la planeación estratégica, en cuanto a las relaciones con proveedores, competidores, clientes, tecnología, personal de la organización, cambios en el mundo, política, mercados financieros, etcétera.

El *Data Warehouse* será usado como una herramienta que permitirá extraer información clave para los negocios, está misma deberá combinarse con una metodología de planeación estratégica en la cual el empresario analizará el ámbito interno de su empresa, su entorno y sus preferencias; con el uso de este método, el director analizará el ámbito interno de su empresa reflexionando sobre sus límites y alcances; también vislumbrará su entorno general, y finalmente, definirá sus preferencias en un marco de valores. Las metodologías de planeación son diversas⁽¹⁶⁾, pero para establecer un marco de trabajo se pueden considerar como procesos comunes a las mismas los siguientes puntos:

- Posicionamiento estratégico de la empresa, actual y futuro.
- Determinación de las estrategias para alcanzar los objetivos.
- Implantación de las estrategias.

⁽¹⁶⁾ Control de Gestión basado en Planeación Estratégica. Consultores SICOFIN. Págs. 24-25. México 1993.



El Proceso de Planeación Estratégica
SICOFIN CONSULTORES

Figura 3.2

La meta de la administración y de la planeación consiste en desarrollar estrategias y planes exitosos para aumentar las utilidades, mantener complacidos a los clientes e incrementar la participación de mercado. De acuerdo con el Business Week, “con la competencia instalada en cada industria, un *Data Warehouse* es la más reciente arma de comercialización que se debe tener (una mejor forma de conocer las necesidades de los clientes y cómo apearse a ellas)”⁽¹⁷⁾.

La finalidad del *Data Warehouse* consiste en auxiliar a la administración y planeación estratégica a comprender el pasado y planear el futuro. La planeación estratégica busca respuestas a preguntas como:

⁽¹⁷⁾ La integración de la Información para la toma de decisiones Data Warehousing. Prentice Hall. 1996.

-
- ¿ Cuáles son las fuerzas de mi empresa y cuáles sus debilidades ?
 - ¿ Qué puedo hacer ?
 - ¿ Qué podría hacer ?
 - ¿ Cuáles son las oportunidades y amenazas que enfrento ?
 - ¿ Qué quiero hacer ?
 - ¿ Cuáles son los valores éticos y sociales de mi organización ?
 - ¿ Qué están comprando nuestros clientes ? ¿ Qué no están comprando ? ¿ Cuáles incentivos han funcionado antes con los mismos clientes en esta época del año ?
 - ¿ Cuántos de nuestros vendedores visitan a un mismo cliente ? ¿ Qué piensan nuestros clientes de esto ?
 - ¿ Qué están haciendo nuestros competidores ? ¿ Cómo ha afectado esto a nuestros clientes ?
 - ¿ Cómo se comparan nuestros costos para cada línea de producto durante los últimos tres años ? ¿ Qué factores provocaron incrementos ? ¿ Afecto esto nuestros márgenes de utilidad ?

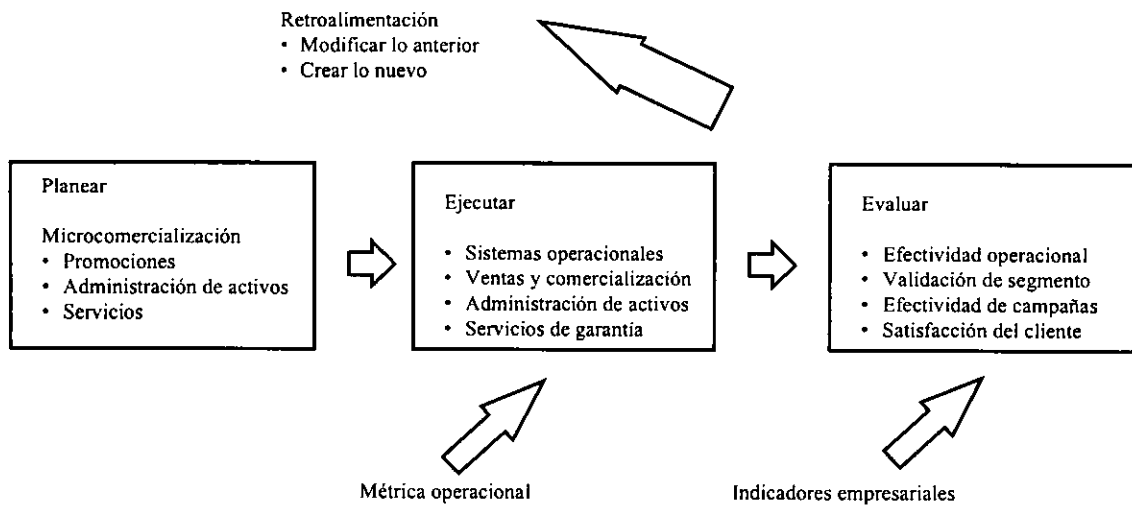
Cada pregunta conduce a otra, yendo desde lo más general y enfocándonos a cuestiones más particulares. La planeación desea obtener estas respuestas y tomar mejores decisiones. El reto es extraer los datos del *Data Warehouse* y convertirlos en información que ayude a la empresa a hacer mejores elecciones, lo cual redundará en mejores decisiones y en la creación de una ventaja empresarial sostenible.

El *Data Warehouse* es un depósito de conocimiento corporativo. Contiene información detallada, seleccionada y almacenada, por ejemplo:

- Cuáles clientes y proveedores se conocen?
- Qué se sabe de los clientes, proveedores y empresas?
- Cuál es el desempeño (rentabilidad y rendimiento)?

El extraer un valor empresarial es un esfuerzo complejo, aún cuando el *Data Warehouse* contenga información estratégica. Obtener un valor requiere de múltiples aptitudes empresariales combinadas, de habilidad técnica, intuición y experiencia. De esta manera, el llegar hasta la información requiere de una estrecha asociación entre los usuarios de la empresa y los profesionales de la tecnología de la información.

El *Data Warehouse* es una parte integral del sistema de “ciclo cerrado” en la retroalimentación de la administración de las empresas. Los segmentos básicos del sistema de “ciclo cerrado” de retroalimentación son planear, ejecutar y evaluar. Gráficamente tenemos lo siguiente:



Administración Cíclica

Figura 3.3

El *Data Warehouse* ayuda en la planeación de la empresa y en la evaluación de la eficiencia de la ejecución de un plan. Por ejemplo, es posible que una estrategia de comercialización defina los segmentos de clientes adecuados para la microcomercialización y las campañas promocionales, o una nueva estrategia de servicios para mejorar la satisfacción de clientes con un programa de garantías. Las actividades de evaluación – crecimiento y rentabilidad de un segmento de clientes del micromercado⁽¹⁸⁾ y la satisfacción de los clientes con el programa de garantía – ayudan a afinar las estrategias y planes de comercialización.

El *Data Warehouse* no es una herramienta mágica, es necesario el conocer, acceder y recuperar la información y convertirla en información y hechos. Estos hechos conformarán los cimientos de una base de conocimientos que sirve para determinar la salud de la empresa y la dirección futura del negocio.

Un *Data Warehouse* se usa para comprender, medir y administrar parámetros empresariales estratégicos, como el crecimiento del ingreso y la rentabilidad, la participación en el mercado y los segmentos de clientes, y la eficiencia a nivel de la administración de activos. Por ejemplo, para mejorar la administración de activos y el retorno sobre los activos, el análisis de compras a nivel empresa pudiera descubrir oportunidades de descuentos por volumen, menores costos de embarque y una reducción en el personal de apoyo sin afectar en forma significativa los niveles de servicio.

⁽¹⁸⁾ Micromercado se define como un sector de consumo en un ramo específico.

3.1.1 Análisis de Rentabilidad

En el medio competido de hoy en día, la lentitud no es una opción; y de acuerdo a la definición de rentabilidad, se deberá actuar con eficacia y velocidad para obtener mejores resultados en el mercado.

Con un *Data Warehouse*, los usuarios serán capaces de tomar decisiones adecuadas, ya que ellos contarán con la información que les dé una ventaja para tener mejores resultados en su ámbito empresarial. Los datos se localizarán en un contexto empresarial adecuado para producir información y hechos. Así, el usuario agregará a los hechos su habilidad empresarial, para destilar los conocimientos y después aplicarlos para hacer recomendaciones o decisiones. Un *Data Warehouse* produce la información correcta para los ejecutivos, gerentes y analistas de una empresa. Estos usuarios pueden aprovechar el *Data Warehouse* para construir un rango de estrategias y planes, desde administrar mejor las relaciones con el cliente hasta modificar las bases de la competencia.

El aprovechamiento del potencial del *Data Warehouse* sólo está limitado por el rendimiento, capacidad y creatividad de los usuarios de la empresa. Las experiencias anteriores señalan hacia un amplio rango de valores, en muchos sectores empresariales, derivados del *Data Warehouse*.

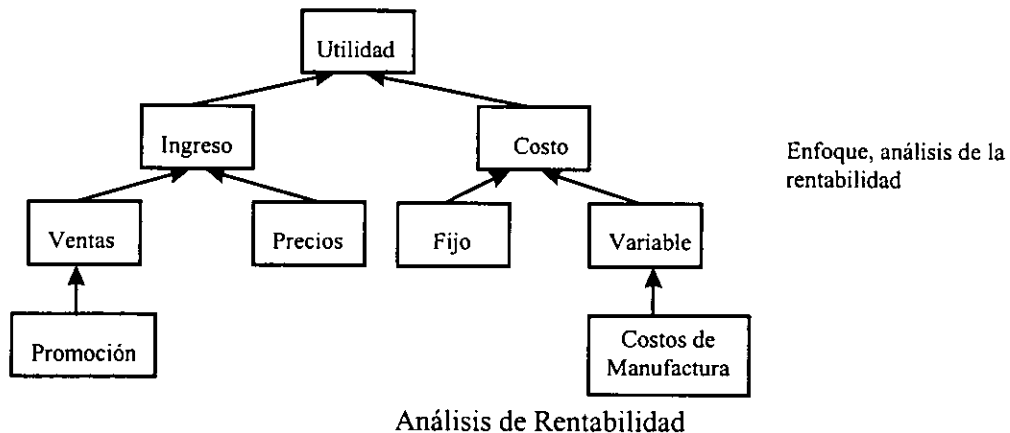
Un método para medir la rentabilidad para un *Data Warehouse* es el de construir el modelo económico para el mismo, definiendo y cuantificando los objetivos empresariales, estableciendo estrategias y tácticas a utilizar.

Seguido de este paso se deben desarrollar los *modelos proforma de ingresos y costos*, los cuales conducen al *modelo proforma*. Por último se deben medir las medidas de desempeño empresarial y de auditoría para el *Data Warehouse*. En esencia, el modelo económico consta de un modelo financiero de tres piezas, comprendidas entre los objetivos y las medidas de desempeño.

	Cuantificado Estrategias y tácticas Estrategias y tácticas
Modelo de Ingresos Proforma	Ingresos y Beneficios Tangibles Intangibles
Modelo de Costos Proforma	Costos del <i>Data Warehouse</i> Inicial Recurrente
Modelo Financiero Proforma	Análisis Financiero Análisis de Riesgos

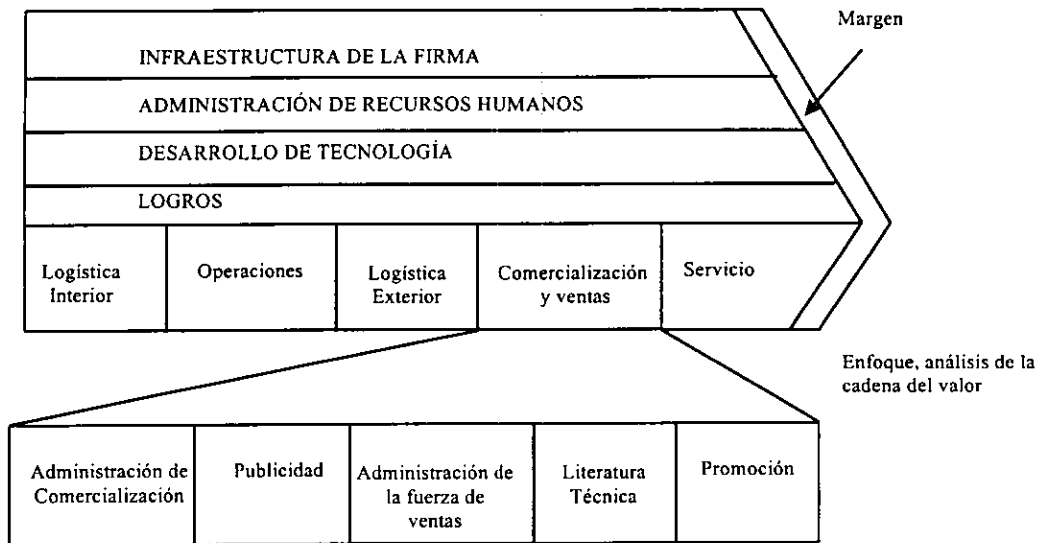
El paso inicial consiste en enfocar un aspecto o área empresarial, tal como la rentabilidad de clientes o comercialización dirigida. La elección del área o aspecto produce diferentes estrategias y objetivos empresariales. Existen diversos enfoques, los cuales se ajustan a distintos objetivos, para analizar el valor y beneficios de los objetivos establecidos y estrategias.

Para los objetivos de análisis de rentabilidad, de un producto, mercado o incluso un solo cliente, las actividades analíticas son determinar los costos e ingresos reales o prorrateados (ver Figura 3.4). Deben medirse y analizarse los factores que afectan los ingresos y los costos, lo que implica que, para mediciones de desempeño, el *Data Warehouse* debe proporcionar información relevante. Por ejemplo, para determinar la rentabilidad de un producto el *Data Warehouse* debe suministrar información sobre todos los ingresos y también sobre todos los costos (fijos, variables, de promoción, de ventas, etcétera) del producto. Sin este tipo de información, no se podrá seleccionar un objetivo de rentabilidad en el modelo económico de justificación de costos.



Análisis de Rentabilidad
Figura 3.4

Cuando el objetivo del *Data Warehouse* comprende diversas funciones de la organización, es necesario el análisis de la cadena de valor agregado para complementar el análisis de rentabilidad (ver Figura 3.5). El modelo de costos y de ingresos puede ser bastante complejo, pero si se analiza primero cada función como un centro de costos e ingresos, se puede aplicar un nuevo modelo.



Fuente: Competitive Advantage
 Michael Porter
 Free Press Mac Millan 1985

Cadena del Valor

Figura 3.5

Un objetivo de satisfacción del cliente implica equilibrar las necesidades del cliente con la capacidad de la empresa para ingresar una utilidad y agrega valor.

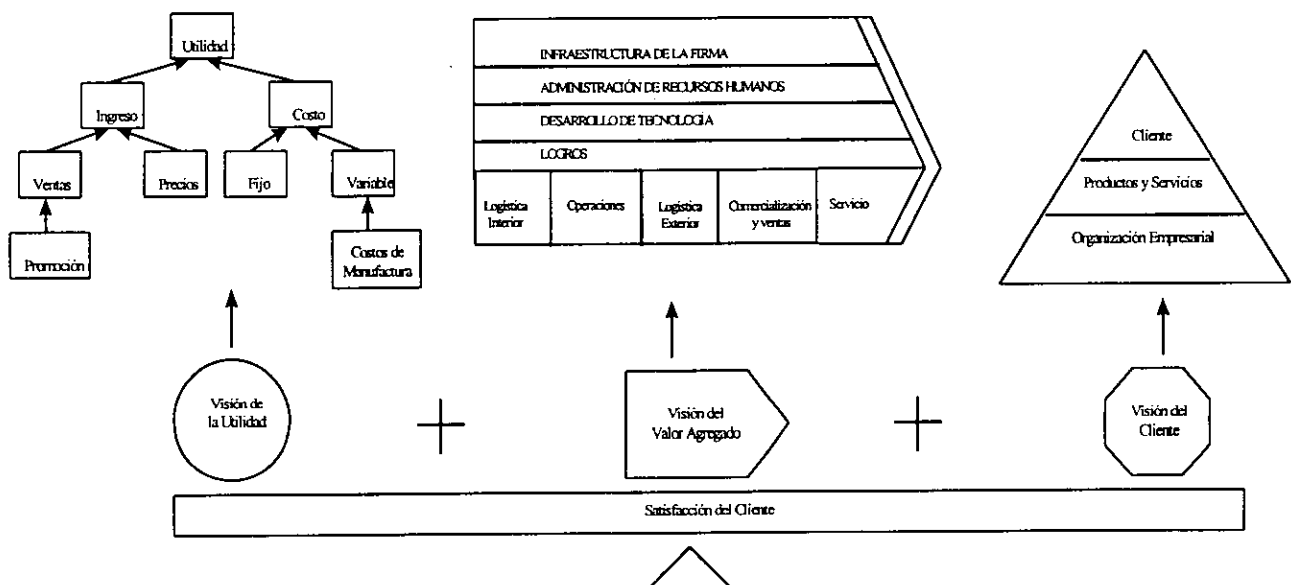


Figura 3.6

Por lo regular, los objetivos del *Data Warehouse* deben ser tales que cada objetivo se reduzca a una corriente de ingresos, un conjunto de costos y las metas de rentabilidad y beneficios establecidos como objetivos cuantificables. Los beneficios no cuantificables, tangibles o intangibles, deben establecerse en un lenguaje empresarial, de modo que puedan ser comunicados con facilidad y medidos cuantitativamente.

El *modelo de ingresos proforma* puede tener mediciones tanto tangibles como intangibles. Las mediciones tangibles pueden ser valores monetarios, o bien, declaraciones de objetivos. En el modelo de ingresos, el reto consiste en convertir todo a un valor monetario. Por ejemplo, si el objetivo del *Data Warehouse* es “la promoción del producto incrementará la cantidad de clientes en un 10%”, se necesitará traducirlo en un incremento de ingresos en X por ciento. Entonces ambas metas se establecen como “Incrementar el ingreso en X por ciento y los clientes en 10%”.

Los objetivos de ingresos se pueden clasificar en los cuatro tipos siguientes:

- Incrementar los ingresos, debido a un producto, mercado o canal.
- Ingresos nuevos, tal como un producto, mercado o canal nuevo.
- Mantener los ingresos, junto con la reducción de costos.
- Disminuir los ingresos, si se elimina un mercado, producto o canal.

Algunas mediciones potenciales de mejoramiento en los ingresos incluyen la rentabilidad del cliente, producto o canal y la productividad de áreas como el telemercado.

Para el desarrollo del *modelo de costos proforma* se requieren los dos siguientes submodelos:

De Costos Empresariales:	Del <i>Data Warehouse</i> :
<p>Hay dos tipos de costos empresariales: costos nuevos o incremento de costos actuales. Se incurre en costos para generar nuevos ingresos o incrementarlos. Para cuantificar los costos del incremento, es esencial primero establecer los costos actuales. Los costos del incremento deben incluir todos los costos de los profesionales de la tecnología informática para apoyar al usuario empresarial del <i>Data Warehouse</i>.</p>	<p>Existen dos costos del <i>Data Warehouse</i>: iniciales y recurrentes (continuos). Los costos iniciales son gastos por única vez (diseño, construcción y despliegue), o bien, costos de capital (hardware, software, infraestructura, etcétera). A su vez, los costos recurrentes se subdividen en dos clases: fijos y variables. Un ejemplo de un costo recurrente fijo es la actualización continua del <i>Data Warehouse</i> a partir de las fuentes de datos. Los costos variables continuos son por elementos como el crecimiento del <i>Data Warehouse</i> conforme aumenta el ámbito, cuando se incorporan más usuarios, etcétera.</p>

El modelo de costos proforma debe incluir todas las erogaciones – los costos en tecnología de la información y los costos de la organización de usuarios empresariales. Se deben clasificar, asignar e incluir en los costos totales de hardware y software incluyendo gastos de mantenimiento.

La filosofía de la tecnología de información del *Data Warehouse* debe ser: los costos iniciales más bajos, los menores costos recurrentes, el más rápido calendario de implementación (el tiempo es costo e inversión) y la máxima reutilización de las inversiones existentes.

El modelo financiero o de utilidades proforma es una medición de rentabilidad (beneficios tangibles) de la iniciativa del *Data Warehouse* y de los beneficios tangibles por alcanzar. Desde una perspectiva cuantitativa, la iniciativa del *Data Warehouse* es rentable en cualquiera de las cuatro formas que se muestran en la figura 3.7.

Los cuatro escenarios de rentabilidad para la iniciativa del *Data Warehouse* se describen a continuación:

- Escenario 1 Se incrementan los ingresos y los costos, pero el incremento en los ingresos es mucho mayor con relación al incremento de los costos. Un ejemplo son los ingresos aumentados, a raíz de una campaña promocional de comercialización.

- Escenario 2 Los ingresos aumentan, pero los costos se mantienen iguales. Un ejemplo es un proyecto de comercialización dirigida, sin incremento en el presupuesto actual de comercialización.

- Escenario 3 Los ingresos se mantienen, pero los costos disminuyen. Un ejemplo son los ingresos iguales para una familia de productos, pero se identifican y eliminan programas ineficientes de comercialización.

- Escenario 4 Los ingresos disminuyen, pero los costos disminuyen aún más con relación a la baja de ingresos. Un ejemplo es la eliminación de productos o servicios no rentables.

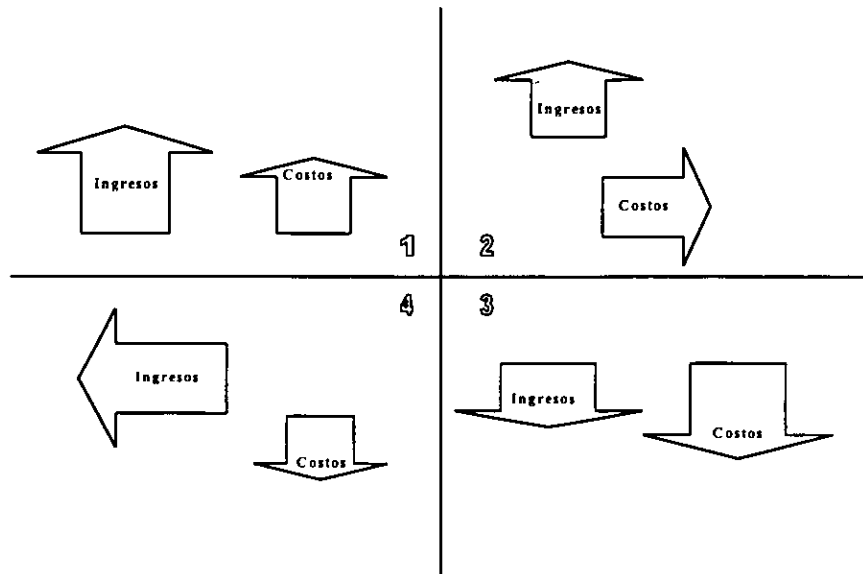


Figura 3.7

El quinto escenario es el ideal en el cual los ingresos aumentan de forma notable y los costos disminuyen en gran medida.

El modelo financiero proforma debe incluir proyecciones a no más de 3 años, debido a que el mundo de los negocios es demasiado dinámico y cambiante.

Finalmente, para afinar el *Data Warehouse* es esencial un plan de auditoría. Este plan debe incluir las mediciones tangibles y no tangibles del desempeño. Los planes de auditoría son necesarios para afinar los programas piloto antes que la empresa se comprometa a una estrategia mayor. Las mediciones del desempeño tangibles incluyen las monetarias, tales como el crecimiento de los ingresos y las utilidades y los objetivos de administración de costos, y las metas dictadas en lenguaje empresarial, tales como por ejemplo: incrementar en N la cantidad de clientes, aumentar las ventas en un M%, entre otras. Las mediciones intangibles incluyen el mejoramiento de la satisfacción del cliente, la percepción de la compañía por parte del mercado, etcétera.

La medición de Retorno del Ingreso (ROI, por sus siglas en inglés, Return of Investment) en el medio ambiente del *Data Warehouse* es complejo, ya que cuando se tienen que estimar valores de indicadores tales como la capacidad de tomar mejores decisiones, o bien, el tener un acceso más rápido a la información, dichos parámetros son un tanto subjetivos. Definitivamente existen ahorros tangibles debido a la automatización del proceso de distribución de información, pero estos típicamente son pequeños alrededor del todo en un medio ambiente empresarial.

De acuerdo a un estudio realizado por *Sentry Market Research*, los mayores beneficios del *Data Warehouse* son el acceso a la información y los beneficios no cuantificables tienen un valor incalculable, entre ellos se tienen: mejor acceso a la información, disponibilidad de los datos, confiabilidad de los datos, ventaja competitiva, mejor control de la información, consistencia de los datos, descentralización de los datos, ahorros en costos, menor dependencia de sistemas legados, mejor integración de los datos y flexibilidad. Para los usuarios activos de un *Data Warehouse* el retorno de la inversión es evidente por sí mismo y de acuerdo a *Bob Chin*, Director del área de informática de la empresa *Healthcare Inc.*, "el calcular el retorno de la inversión de un *Data Warehouse* es como preguntar cuál es el retorno de la inversión del equipo ejecutivo gerencial de una empresa". El cálculo del retorno de la inversión es difícil, y de acuerdo a un estudio realizado por IDC, en términos generales el ROI a tres años puede variar de un 3% hasta un 1838%.

El tipo de retorno de inversión que se puede esperar de un *Data Warehouse* depende de diferentes factores. Los más importantes y que cabe mencionar son: elección de tecnología, la mercadotecnia interna del *Data Warehouse* y la estrategia de negocios.

El primer caso, los *factores tecnológicos*, es una variable controlable y de costos manipulables. Por ejemplo, se considera que el 34% de los costos iniciales de un *Data Warehouse* son invertidos en hardware y otro tercio de los costos se utiliza para servicios de desarrollo interno (fuente *IDC Corp.*), los participantes del *Data Warehouse* que utilizan tecnología de hardware y almacenamiento viejas u obsoletas no gozan del beneficio del factor precio/rendimiento sobre las reducciones de los mismos que han traído consigo el retorno de la inversión.

El líder de más alto en el retorno de la inversión en el *Data Warehouse*, es la *estrategia de negocio*; las compañías que tienen el más alto ROI usan su *Data Warehouse* para cambiar su proceso de toma de decisiones, el *Data Warehouse* está embebido en su proceso de decisional. Calcular el ROI de un *Data Warehouse* debe ser contemplado como parte integral de la estrategia de negocios, de esta manera el ROI del *Data Warehouse* es una parte de la estrategia específica de negocios. Por ejemplo, si una compañía decide el atacar a ciertos clientes con un ofrecimiento dirigido específicamente a ellos, y si el *Data Warehouse* proporcionó dicha información, el ROI del *Data Warehouse* se deberá basar en el ROI de la estrategia de negocios, así, si se considera que dicha estrategia incrementará las ventas en un 25%, entonces el ROI del *Data Warehouse* también se deberá basar en ese incremento. De tal forma, el retorno del costo del *Data Warehouse* debe ser analizado como parte de la iniciativa de negocios.

Cuando se busca el ROI del *Data Warehouse* únicamente en los beneficios más tangibles de automatización del proceso de información puede ser posible que no se alcance un ROI muy positivo.

De acuerdo a lo anterior, las organizaciones que utilicen un *Data Warehouse* deberán basar la obtención del retorno de la inversión, separando dos rubros importantes: la parte tangible y lo intangible; se deberán modificar las técnicas de medición de una manera tal que se reflejen ambos factores.

El verdadero valor de un *Data Warehouse* debe reflejar tanto los beneficios cuantitativos como cualitativos, el cálculo tradicional del retorno de la inversión – que analiza los beneficios tangibles menos los costos – omite los beneficios no tangibles que dan el valor agregado al *Data Warehouse*. Como resultado de lo anterior, se están comenzando a evaluar técnicas alternas para reemplazar el uso del ROI para el *Data Warehouse*. Estas alternativas incluyen el Valor Presente Neto, el cual enfatiza el retorno de mañana en valor al día de hoy; Desplazamiento de Costo, el cual compara el costo del nuevo sistema contra el de reemplazar el anterior; y el de Valor Agregado de Negocios, el cual mide la tecnología no en unidades monetarias sino en términos de soporte a metas de negocio clave y métricas. Con ello, los beneficios intangibles pueden ser estimados.

Una vez que la organización ha medido los beneficios por anticipado y ha seleccionado el desarrollar un *Data Warehouse*, deberán ponerse en su lugar los mecanismos para garantizar un continuo retorno a lo largo de toda la vida del sistema. Un *Data Warehouse* deberá ser lo suficientemente flexible para adaptarse a cambios en el tamaño del negocio, metas, mercados, estructura y otras variables que pueden alterar las expectativas acerca de que debe proporcionar el *Data Warehouse*.

De acuerdo al reporte especial emitido por *International Data Corporation (IDC)* intitulado “Un Estudio del Impacto Financiero del *Data Warehouse*”, escrito por *Stephen Graham, Dirk Coburn y Carsten Olesen*, en 1996, encontró que cifras del Retorno de la Inversión del *Data Warehouse* en un período de tres años son los siguientes:

Promedio del Retorno de Inversión (ROI) Global	401%
Promedio del Retorno de Inversión (ROI) Proyectos Grandes	322%
Promedio del Retorno de Inversión (ROI) en Datamarts	533%
Periodo Promedio del Retorno de la Inversión	2.3 años
Costo Promedio de la Solución	2.2 Millones de Dólares

Los indicadores de rentabilidad tangibles que pueden ser incluidos en el *Data Warehouse* en términos de los ingresos y costos pueden ser los siguientes:

- Rendimiento sobre la inversión
- Rendimiento sobre los activos
- Rendimiento del capital social
- Margen de la utilidad

- Relación costo de ventas / ventas
- Relación de gastos de operación
- Factor de utilidad real
- Factor de utilidad presupuestada
- Eficiencia general de operación
- Rendimiento de la inversión total
- Porcentaje de cumplimiento en programa de producción y ventas

La rentabilidad se puede mejorar, por ejemplo, comprendiendo la correlación entre productos y servicios, entre márgenes y líneas de productos. El análisis puede ser una guía para decidir, por ejemplo, cuáles líneas de productos proporcionan oportunidades de servicio más prolongadas, en especial si no existe una relación obvia entre las ventas de un producto y el ingreso por servicio; sólo un análisis histórico mostraría dicha relación, y con ello y una decisión derivada de esta información se puede mejorar la rentabilidad de líneas productivas y servicios asociados.

Los costos y Beneficios del Data Warehouse

Costos	Beneficios
<p>Costos Preliminares</p> <ul style="list-style-type: none"> ◦ Planeación ◦ Diseño ◦ Ingeniería de la Información/Modelado 	<p>Beneficios Tácticos</p> <ul style="list-style-type: none"> ◦ Reducción de reportes escritos e impresos ◦ Reducción en la demanda de consultas específicas ◦ Entrega más eficiente de la información a los usuarios
<p>Costos Iniciales</p> <ul style="list-style-type: none"> ◦ Servidor de Hardware ◦ Software de Base de Datos ◦ Herramientas de Transformación y Transferencia 	<p>Beneficios Estratégicos</p> <ul style="list-style-type: none"> ◦ Herramientas de Acceso para Usuario Final y aplicaciones ◦ Decisiones más sustentadas ◦ Proceso Decisional más rápido
<p>Costos Constantes</p> <ul style="list-style-type: none"> ◦ Mantenimiento de los Datos ◦ Desarrollo de la Aplicación ◦ Capacitación y Soporte 	<ul style="list-style-type: none"> ◦ Habilidad para soportar y manejar información empresarial. ◦ Soporte para Estrategias de Negocios

Fuente: *Sentry Market Research*

3.1.2 Valor Competitivo

Un conocimiento verdadero y completo de un cliente, surgido de los datos históricos en el *Data Warehouse* y aumentado con un ciclo de retroalimentación más rápido de eficiencia operacional, pueden

hacer factibles formas nuevas y más rápidas de satisfacer con más precisión y amplitud los requerimientos del cliente. Por ejemplo la comercialización exacta de los productos adecuados, con el conjunto apropiado de características, en el momento oportuno, aumenta la satisfacción del cliente. Los clientes complacidos establecen relaciones cada vez más estrechas y exclusivas con los fabricantes. Tales relaciones conducen a mayores ingresos, mejores márgenes y esto es difícil romper por los competidores.

La administración de las relaciones con el cliente es la solución a la pregunta de siempre: conoce a tu cliente. La integración de organizaciones diferentes de comercialización conduce a una mejor comprensión del cliente. Por ejemplo, cuando un cliente que compra diferentes productos de distintas organizaciones de comercialización se le contempla como un cliente a nivel empresa, podría surgir nueva información sorprendente. El ingreso total generado por ese sólo cliente para todas las organizaciones bien pudiera colocar al cliente en los mejores 20 de la empresa. Una nueva estrategia de administración de las relaciones con este cliente podría ser darle un mejor servicio, promociones especiales, oportunidades de mayoreo, etcétera. Aprender mejor quién es en realidad el cliente y cuál es el espectro completo de sus necesidades, puede guiar la comercialización al envío preciso de la comunicación adecuada para este. Los clientes aprecian un trato especializado, lo cual los hace más fieles hacia la compañía.

3.1.3 Productividad

La productividad de la empresa tiene contemplados los siguientes factores (véase Figura 3.1 “Triángulo Virtuoso”):

Recursos	Humanos
	Tecnológicos
	Materiales
	Financieros

Un *Data Warehouse* puede apoyar a la productividad mediante la generación de información que nos permita analizar qué está pasando con activos fijos, con los costos de inventario, con equipos ociosos, con procesos internos en los cuáles pueden ser distribuidos de una mejor manera los recursos humanos, etcétera.

Los indicadores provenientes del *Data Warehouse* aplicados a este rubro permitirán tomar decisiones que mantengan un balance entre la rentabilidad, la competitividad y la productividad.

3.2 Beneficios del *Data Warehouse*

Los beneficios del *Data Warehouse* son múltiples, y estos dependerán de la empresa que se aplique y el área o áreas seleccionadas de la misma; más sin embargo siempre cubrirán los derivados de rentabilidad, competitividad y productividad.

La cosecha de beneficios se verá reflejada una vez que se definan los objetivos de la administración y con ellos el análisis empresarial proveniente del *Data Warehouse*:

Por ejemplo:

Objetivos de la Administración (Beneficios del *Data Warehouse* por medio de la extracción de indicadores correspondientes):

- I. Aumentar utilidades.
- II. Mejorar márgenes y reducir cargas.
- III. Incrementar participación de mercado.

Análisis empresarial y formulación de cuestiones (Búsqueda de Indicadores):

- Márgenes financieros a través del tiempo(anual, semestral, trimestral, mensual).
- Participación del mercado en el tiempo(anual, semestral, trimestral, mensual).
- Tendencias del mercado e indicadores demográficos
- Utilidades por área geográfica, organización, producto, periodo.
- Ingresos por área geográfica, organización, producto, periodo.
- Precios y costos de productos y servicios.
- Satisfacción al cliente por área geográfica, organización, producto y periodo.
- Capital y capacidad por área geográfica, organización, producto y periodo.
- Datos regulatorios y restricciones.

CONTENIDO DEL *DATA WAREHOUSE*

- Reportes financieros de balance.
- Reportes externos sobre competidores.
- Datos históricos sobre ingresos, utilidad, capacidad y capital.
- Área geográfica.
- Estructura organizacional.

La respuesta a estos indicadores llevará al cumplimiento de objetivos y por ende a los beneficios empresariales correspondientes.

Una fórmula sencilla por medio de la cual se puede llegar a la toma de decisiones aplicándola adecuadamente con el uso de información proveniente del *Data Warehouse*, puede ser llevada a cabo siguiendo los siguientes pasos⁽¹⁹⁾.

1. Definir el objetivo de la administración.
2. Formular cuestiones empresariales.
3. Verificar datos.
4. Formular indagación en el *Data Warehouse*.
5. Consultar el *Data Warehouse*.
6. Analizar los datos.
7. Recomendar y decidir.

El uso oportuno y adecuado del *Data Warehouse* serán la punta de lanza que permitirá que las empresas sobrevivan en el mercado globalizado mundial, adicionalmente, de acuerdo a la firma de consultoría Select-IDC, el uso del *Data Warehouse* en el mercado se da de acuerdo a la siguiente matriz denominada "El Origen del *Data Warehouse*".

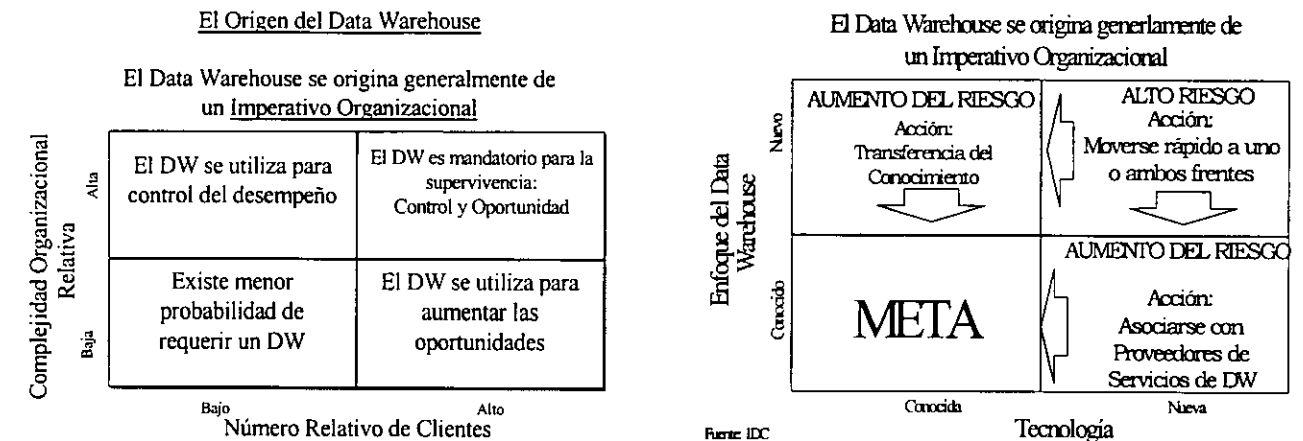


Figura 3.9

⁽¹⁹⁾ La integración de Información para la mejor toma de decisiones. Prentice Hall. Pag 184.

3.3 Participantes de Mercado para la Solución del *Data Warehouse*

El *Data Warehouse* es una solución completa, no es un producto que se pida en cantidades, es una metodología en la cual se involucran múltiples productos y servicios asociados a él.

El *Data Warehouse* es un área en la que las empresas invertirán recursos significativos. Por ello, en el mercado de los *Data Warehouses* presenta con gran rapidez un gran número de fabricantes – desde fabricantes con ingresos por varios millones de dólares hasta ingresos por sólo un millón; desde fabricantes con estructuras y múltiples productos hasta proveedores con un sólo producto; y desde compañías bien establecidas hasta recién creadas. Al mismo tiempo no hay sólo un fabricante que tenga la amplitud de productos para satisfacer por completo las necesidades empresariales.

En este ambiente, la tecnología de la información debe conducir la selección de fabricantes y productos comprendiendo tanto el mercado de la tecnología del *Data Warehouse* como la dinámica de los fabricantes. Esto ayuda a comprender y administrar los riesgos. Por lo tanto, el análisis de los fabricantes debe realizarse en términos del costo - beneficio para la empresa y el retorno de la inversión causada por la solución completa del *Data Warehouse*.

Como se vio en los dos capítulos anteriores, el *Data Warehouse* se compone por múltiples bloques y capas, de esta manera en cada uno de ellos se puede contar con soluciones de diferentes proveedores de tecnología y servicios. El cuadro sinóptico de los principales participantes en el mercado queda así:

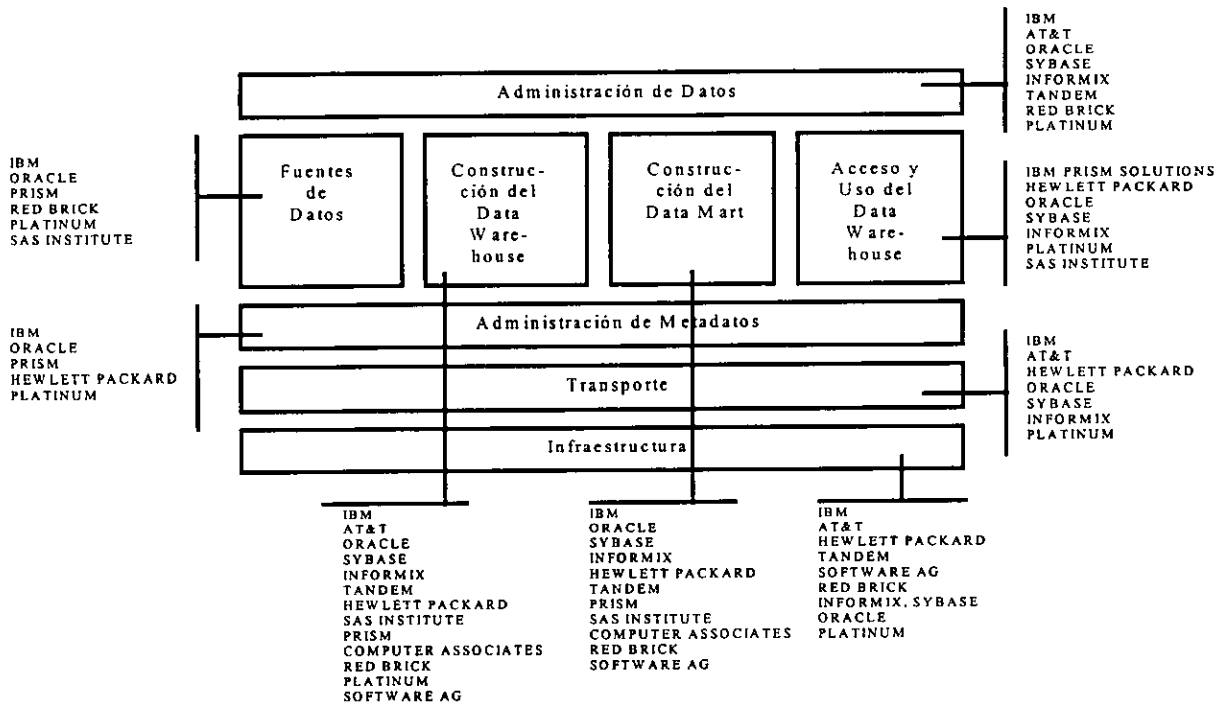


Figura 3.10

3.4 Perspectivas del *Data Warehouse*

El futuro del *Data Warehouse* es muy prometedor, nos estamos acercando a la era del conocimiento en la cual, los hechos y su uso adecuado nos van a dar la pauta para el crecimiento de la sociedad empresarial. El desarrollo se va alcanzando mediante la transición de datos, a información y finalmente a conocimiento; por el otro lado, una evolución de plataformas, a aplicaciones, llegando al contenido (fuente Select-IDC Marzo 1998).

El uso del *Data Warehouse* se masificará, usando como medio de acceso tecnologías al alcance de todos, con herramientas amigables a través de redes inteligentes donde se transmitirá la información para la toma de decisiones; al mismo tiempo esa información podrá ser vista con representaciones de datos simples hasta datos complejos, en los cuales se podrán incorporar sonidos, voz, imagen, video y otros, llegando a consultas vía Internet con el uso gráfico de interfaces. Todo ese cúmulo de información nos llevará hacia el conocimiento y comunidades virtuales en las cuales habrán múltiples individuos distribuidos en el globo terráqueo consultando la información para tomar decisiones individuales y de grupo, o bien para un ámbito personal o empresarial.

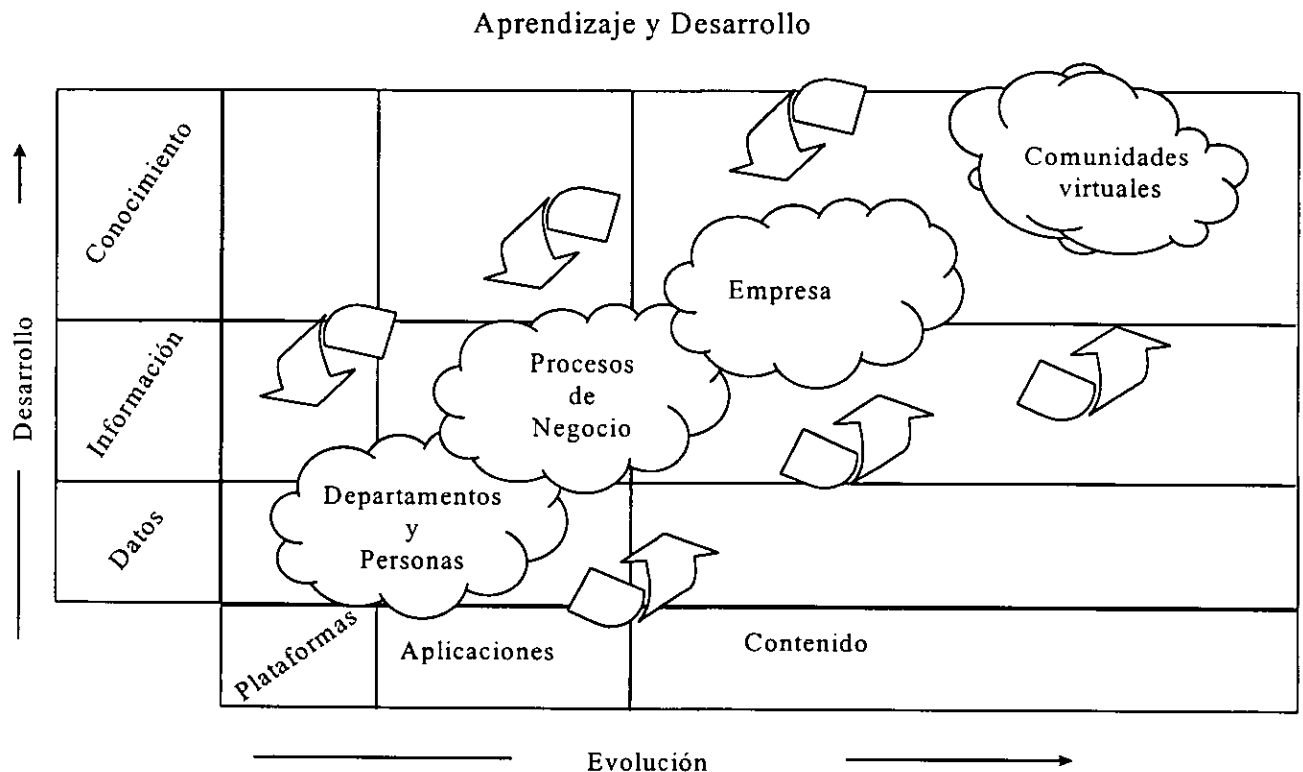


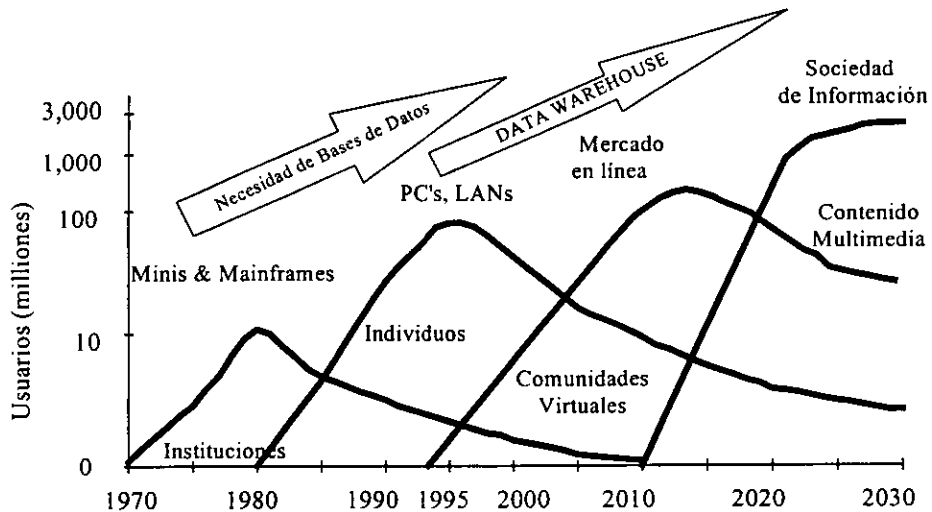
Figura 3.11

El mercado de la tecnología informática evolucionará hacia la sociedad de la información, en donde un

Data Warehouse en línea será crucial para dicha evolución.

El mercado de la tecnología informática, en donde se incluyen las inversiones para el rubro de *Data Warehouse*, está creciendo aceleradamente, y esto continuará por este camino hasta alcanzar la cadena del valor del conocimiento, en donde las empresas compartirán su información para así generar más dinamismo en el mercado global. Cada empresa tendrá la información estratégica que podrá compartir para su bienestar conjunto.

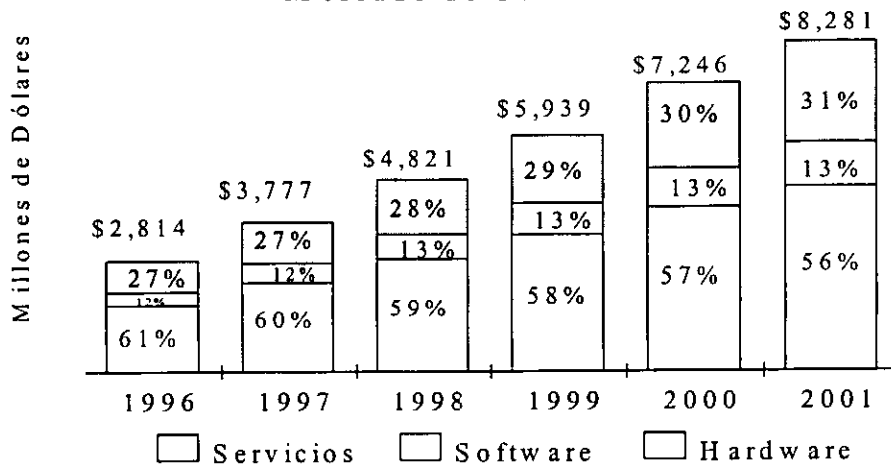
Evolución de la Industria de TI



Fuente: Select-IDC/Ing. Ricardo Zermeño, Marzo 1998.

Figura 3.12

Mercado de TI en México



Fuente: Select-IDC/Ing. Ricardo Zermeño, Marzo 1998.

Figura 3.13

3.4.1 La Administración Informática y el *Data Warehouse* en el proceso de Toma de Decisiones

La toma de decisiones es un proceso en el cual se deberá escoger entre dos o más alternativas. La administración informática, sustentada por medio del *Data Warehouse*, apoyará a los tomadores de decisiones en las fases de planeación, organización, dirección y control, a seleccionar de acuerdo a indicadores entregados por el *Data Warehouse*, alguna alternativa o alternativas para el diseño de estrategias de negocio para mejorar su rentabilidad, productividad y competitividad.

El *Data Warehouse* como una herramienta para la toma de decisiones apoyará al equipo ejecutivo y de gerencia a contar con indicadores que les permitirán tomar decisiones más racionales, es decir, el tomador de decisiones podrá ser totalmente objetivo y lógico. Tendrá clara la meta y todas las acciones en el proceso de toma de decisiones que llevarán de manera consistente a la selección de aquella alternativa que maximice la misma. La toma de decisiones racional se orientará a un objetivo, por ejemplo, incrementar las ventas en un 10% en el último trimestre del año; de ahí, con la información proveniente del *Data Warehouse* (indicadores de ventas), se identificarán todas las alternativas viables: cambio de campaña publicitaria o incremento en la producción del producto "A"; el siguiente paso será identificar las posibles consecuencias de cada alternativa, dándoles un valor de retorno al negocio: la campaña publicitaria incrementará las ventas en un 10% pero los costos se elevarán en un 8%, o bien, las ventas se incrementarán en un 10% con la producción del producto "A" pero esta última elevará los costos en un 15%; finalmente el tomador de decisiones escogerá la alternativa que califique más alto, es decir, la que le dé los beneficios máximos.

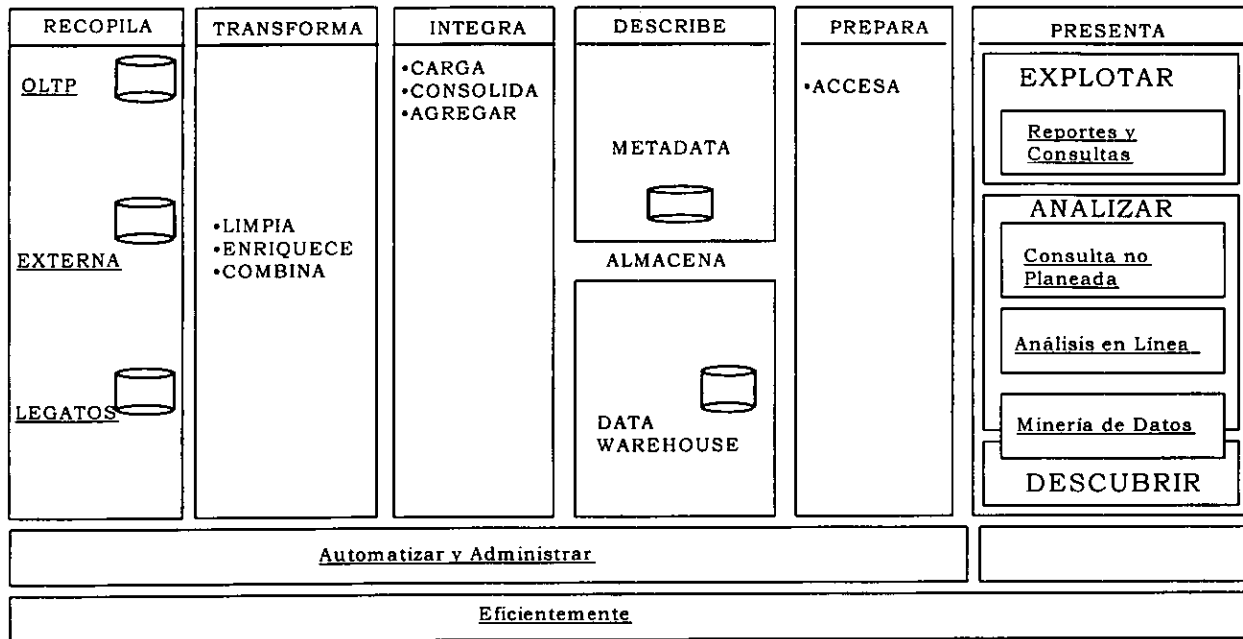
Así, concluyó que la toma de decisiones, sustentada por medio de una herramienta como el *Data Warehouse*, permitirá la integración, transformación y análisis de los datos operativos e históricos de la empresa por medio de las cuales será factible el análisis de los indicadores internos y de mercado a través de los cuales se tomarán decisiones fundamentadas en información completa, veraz y oportuna, complementándolas a su vez con el criterio empresarial en pro del éxito las empresas a nivel mundial.

3.5 Familia de Productos Informix para *Data Warehouse*

Informix Software Inc. es una compañía dedicada al desarrollo de soluciones de bases de datos cuyo objetivo es dar respuesta a las necesidades tecnológicas y de negocios a empresas en general. El foco principal de la compañía es proveer los productos y servicios necesarios para aplicaciones transaccionales en línea, ambientes distribuidos de bases de datos, ambientes en el Web y aplicaciones de *Data Warehouse*.

De esta manera y con la finalidad de proporcionar un camino tecnológico a seguir, a continuación se describe la solución para *Data Warehouse* de Informix Software Inc.

Los principales pasos a seguir de una forma esquemática (Figura 3.14) para la construcción de una solución completa de *Data Warehouse* son: Recopilación, Transformación, Integración, Descripción, Preparación y Acceso. En cada una de estas fases y etapas existen diferentes productos que cubren los requisitos para la correcta ejecución de cada una de ellas.



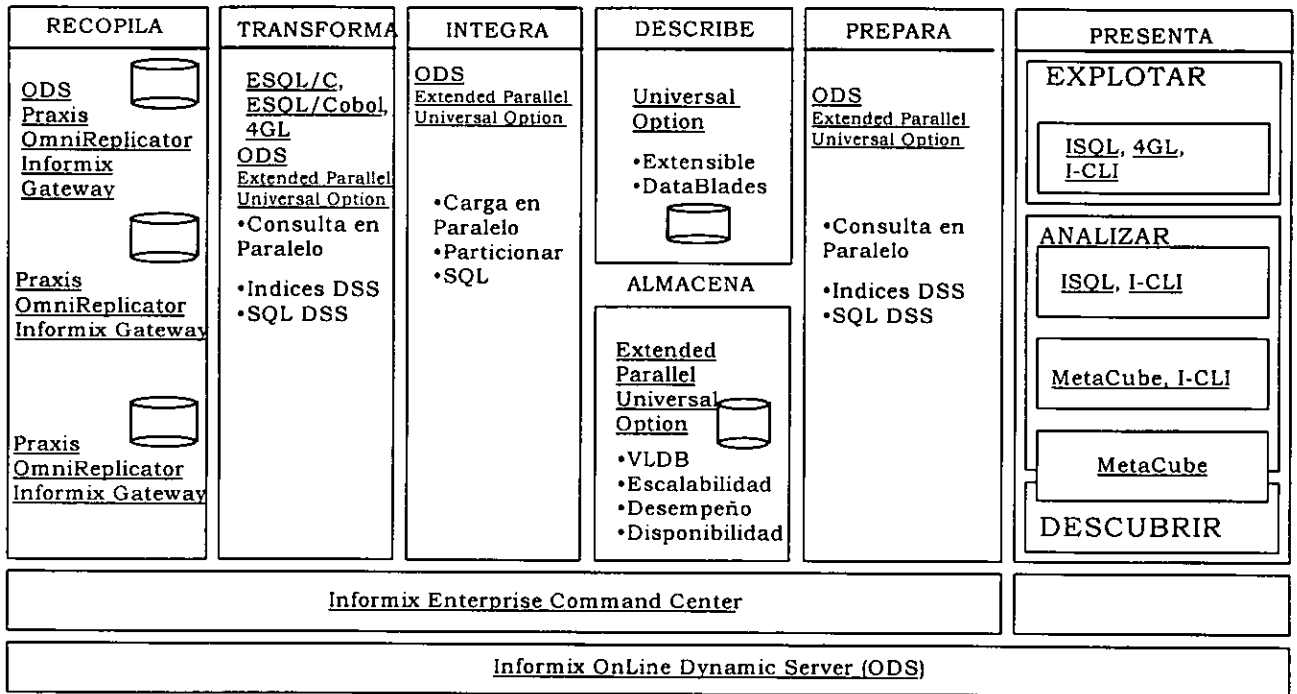
Solución Genérica de *Data Warehouse*

Figura 3.14

Descripción de Componentes para *Data Warehouse* de Informix

<ul style="list-style-type: none"> Informix Dynamic Server (ODS) 	Manejador de Base de Datos Relacional para ambientes transaccionales en línea, <i>Data Warehouse</i> y ambientes de Web. Su principal característica es su paralelismo interno utilizado en todas las actividades de base de datos.
Universal Data Option	Opción en el manejador de base de Datos ODS que permite el manejo de tipos de complejos (video, voz, imagen, texto, etcétera), en términos del contenido de información: por ejemplo, reconocimiento facial en consultas a la base de datos. Opción Objeto – Relacional.
Extended Parallel Option	Opción que proporciona al manejador ODS la capacidad de altos volúmenes de información, en el cual se pueden hacer consultas probabilísticas para la toma de decisiones.

<ul style="list-style-type: none"> • Informix SQL • Informix 4GL/ Informix ESQL/C 	Herramientas para desarrollo y de acceso a la base de datos IDS. Permite hacer programación de cualquier tipo para las aplicaciones.
<ul style="list-style-type: none"> • Informix Metacube 	Herramienta para procesamiento analítico en línea, que permite hacer consultas para la toma de decisiones, bajo un modelo multidimensional de información.
<ul style="list-style-type: none"> • Informix Gateway Manager • Informix Connectivity Class Library (I-CLI) • Praxis OmniReplicator 	Herramientas e interfaces de comunicación y conectividad para la manipulación de datos.



Solución de Informix para *Data Warehouse*

Figura 3.15

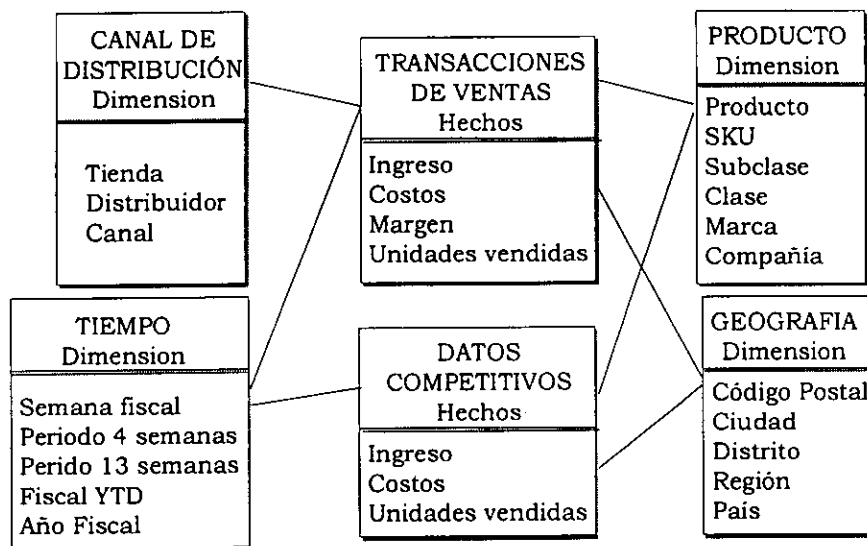
El mercado comercial de *Data Warehouse* ha producido con rapidez una gran cantidad de fabricantes: desde los que tienen ingresos de varios miles de millones de dólares hasta los de menos de un millón; desde fabricantes de muchos productos y estructuras, hasta los proveedores de un sólo nicho o de un sólo producto; desde las compañías bien establecidas hasta las de reciente aparición. Al mismo tiempo ningún fabricante por sí solo tiene la amplitud de productos para satisfacer totalmente las necesidades de las empresas. En este ambiente, en la idea de tecnología de la información se debe tener mucho cuidado al elegir los fabricantes y los productos, entendiendo tanto el mercado en tecnología de *Data Warehouse*

como de inversiones y compromisos de los fabricantes con las soluciones de *Data Warehouse*. Esto ayudará a las empresas a comprender y manejar los posibles riesgos en la construcción e implantación de un *Data Warehouse* corporativo.

3.5.1 Caso Genérico Aplicado al Área de Ventas

El proceso de desarrollo de *Data Warehouse* como ya se ha visto, es todo un proceso el cual va desde el análisis hasta la implantación del mismo; en cada una de estas etapas se deben considerar los puntos de negocio como técnicos de la empresa. A continuación se incluye un ejemplo genérico aplicado al área de ventas en el cual únicamente se muestra el modelado de datos, por medio del cual es posible la consulta final sobre la información contenida en el mismo.

El esquema estrella muestra la tabla de hechos (descrito en Capítulo II) y dimensiones con las cuales se harán las consultas ejecutivas de información.



Esquema Estrella

Figura 3.16

-
- I. Incrementar la madurez de la tecnología de base de datos relacional con la incorporación de ampliaciones de objetos y características del *Data Warehouse*.
 - II. Integración de herramientas de flujo de trabajo para automatizar los siempre crecientes requerimientos del *Data Warehouse*.
 - III. Determinar estándares de la industria para que los productos funcionen juntos.
 - IV. Considerar las lecciones aprendidas por los pioneros.

Finalmente, es importante considerar que hoy en día los negocios están experimentando una intensa competencia mundial y el cambio más acelerado jamás visto. Han aparecido competidores que ni siquiera figuraban en el radar empresarial. Esta situación afecta todas las decisiones y estrategias en el momento en que las empresas intentan dominar el mercado de su industria apenas para sobrevivir. Este estado en los negocios influye en todos sus aspectos - servicios a clientes, comercialización, ventas, ingeniería de productos, manufactura y todas las partes de la organización - desde el director general hasta el personal de apoyo, desde las oficinas corporativas hasta remotas oficinas de ventas. *Conforme las empresas desarrollan nuevas estrategias, el consenso de la comunidad en tecnología de información es que el Data Warehouse y la dirección hacia la que apunta, es uno de los factores críticos de supervivencia y competitividad efectiva de las mismas.*

"El Data Warehouse es una arquitectura, no un producto"... Consultores GARTNER GROUP.

Conclusiones del Capítulo

El *Data Warehouse* es una nueva concepción para el soporte a la toma de decisiones y de acuerdo a lo expuesto en el capítulo II y III se concluye que el *Data Warehouse* es una herramienta de apoyo la cual genera información sustantiva de un negocio; así mismo, para su correcta implementación es necesario pasar por un proceso el cual está soportado por diversas tecnologías y servicios. El *Data Warehouse* gestiona la porción del ambiente de datos global de una empresa y sirve como fuente de datos única e integrada para el proceso de toma de decisiones. Su desarrollo y crecimiento es hoy inminente, de acuerdo a estudios realizados en el ámbito de la tecnología informática, se ha vislumbrado que la adopción de esta solución año con año crece en un 20% aproximadamente (cifras de IDC Corporation); de igual manera las cifras del retorno de la inversión en estos proyectos deja ver que esta solución es considerablemente rentable en los primeros tres años. Por lo anterior considero que cada empresa deberá incluir en su plan estratégico la inclusión de un *Data Warehouse*.

De esta manera el *Data Warehouse* deberá ser adoptado por las organizaciones mundiales para su toma de decisiones adecuada y oportuna, más sin embargo hoy en día el mercado empresarial Mexicano aún esta en una posición conservadora, ya que sólo algunas de las empresas más grandes del país han comenzado a hacer este tipo de implementaciones por la complejidad y costo de la solución como tal. Sólo los pioneros están tomando el reto, pero definitivamente ellos llevarán la pauta para dirigir los medios ambientes empresariales, ya que su información los llevará al conocimiento pleno y por ende a su éxito en su hábitat. El *Data Warehouse* al día de hoy aún es una solución costosa, la razón principal es que se utiliza tecnología y servicios de vanguardia tecnológica; pero el potencial y beneficios que trae el uso de esta solución justifica su desarrollo, implantación y uso.

El *Data Warehouse* es un arma poderosa en el portafolio de la administración de negocios de la empresa y por ende de la planeación estrategia de las mismas, así esta solución completa proveerá de la información y conocimiento necesario para sustentar la cadena de valor en la sociedad, dicha cadena representa el conocimiento de cada persona y de cada grupo de trabajo así como el intercambio de información valiosa para el crecimiento conjunto de la sociedad.

En análisis de Retorno de la Inversión será un punto importante para evaluar la rentabilidad del *Data Warehouse*, sin embargo, deberán establecerse las medidas adecuadas para que dichos estudios y análisis incluyan de una forma adecuada tanto los beneficios tangibles, como los no intangibles; y con esto garantizar resultados confiables a este respecto.

Notas Bibliográficas

- BOAR BERNARD. Understanding Data Warehousing Strategically. Information Technology Consultants. 1998.
- HARJINDER, S. Gill. Cliente/Servidor Data Warehousing. Prentice Hall. México. 1996.
- STHROCECK MICHAEL. Data Warehousing Is Worth The Investment. InternetWeek Magazine. 1998.
- Control de Gestión basado en Planeación Estratégica. Consultores SICOFIN. México 1993.
- *Data Warehouse* Directions. SMR. EUA 1997. Internet <http://www.sentrytech.com/smrwp4.htm>
- Knowledge Through Data Warehousing: Measuring, Managing and Retaining ROI. DM Review. EUA 1998. Internet http://www.dmreview.com/issues/1998/jan/articles/jan98_16.htm
- Reaping the Rewards of Data Warehousing. Datamation. EUA 1997. Internet: http://www.datamation.com/PlugIn/whitepapers/data_warehouse/reaping_rewards.html
- Internet: <http://www.informix.com>
- Internet: <http://www.knowledgecenters.org/>
- Internet: <http://www.techweb.com/>

Conclusiones

*“La educación no es preparación para la vida,
la educación es la vida
misma.”*

- John Dewey

Las primeras décadas de la “Revolución de la Información”, se han usado para manipular datos. A lo largo de ese camino las tecnologías y herramientas, las organizaciones de sistemas de información, los métodos y la administración, nuestras mentes y todo alrededor sólo enfatizaban *datos*. A finales de los 90’s, se comenzó a vislumbrar que esto ya no era suficiente. El mundo empresarial y de negocios necesitaba una dieta contextual de información, sintetizada de múltiples fuentes, implicaciones e interpretaciones más que sólo hechos abstractos. Así, lo que el mundo global necesita en este momento es: “*conocimiento*”. De esta forma, el punto neurálgico será agregar valor a lo que hoy en día se tiene, transformando datos en información e información en conocimiento. Con todo lo anterior, la premisa de este trabajo ha sido el sustentar el camino para llevar a cabo la siguiente traducción, “*de los datos a la información*”.

Con base al objetivo general de este trabajo, *definir un modelo de Data Warehouse para la planeación estratégica por medio de la integración de datos y análisis de la información para las empresas que requieren un valor agregado competitivo en su mercado*, se concluye como parte medular lo siguiente: El *Data Warehouse* es una herramienta que permite dar sustento cuantitativo y cualitativo al proceso de toma de decisiones, esto se alcanza por medio de la integración, transformación, consolidación, construcción y explotación de datos históricos empresariales, en donde los datos se convierten en información, en indicadores de negocio que llevan al conocimiento de las organizaciones tanto de manera interna como externa. Todo este proceso en el cual se desglosa el *Data Warehouse*, quedo definido por medio del modelo propuesto en el capítulo II, en el cual se analiza el negocio y sus necesidades, llegando al diseño del *Data Warehouse* y finalmente a la presentación de información que requiere o demanda el usuario.

De esta forma, sí por definición una herramienta es un instrumento que facilita el desempeño de una tarea; se concluye que, el *Data Warehouse* (herramienta) facilita y sustenta la toma de decisiones (tarea), esto sucederá adecuadamente si y sólo si, se cuenta con dos elementos básicos: las necesidades presentes/futuras empresariales y los datos históricos que describen el comportamiento de dicha organización; con ambos elementos se estará en capacidad de generar el modelo de *Data Warehouse* con el cual se extraerá la información que proporcione el valor agregado a la organización para sustentar la planeación estratégica y satisfacer el triángulo virtuoso empresarial: rentabilidad, productividad y competitividad.

De acuerdo a lo visto en el primer capítulo de este trabajo, se puede vislumbrar que históricamente se ha cuidado de la información empresarial, esto se ha canalizado por medio de los diferentes tipos de sistemas de información (Sistemas Transaccionales en Línea), los cuales, por una parte describen el funcionamiento de la empresa (interna y/o externamente) y por otro, almacenan el desempeño empresarial de la misma. De lo anterior, concluyó que la verdadera diferenciación y ventaja competitiva se gana mediante el entendimiento global de uno de los activos más importantes de las organizaciones: su

información. El valor agregado que proporciona el *Data Warehouse* es muy importante ya que permite a la organizaciones entender *su pasado para planear su futuro*.

En cuanto a los beneficios cuantificables y cualificables del *Data Warehouse* se puede concluir que el retorno de la inversión de un proyecto de este tipo debe ser medurado considerando los beneficios tangibles e intangibles del mismo; no sólo se deberán contar los pesos y centavos que se están recuperando físicamente de la inversión tecnológica en términos de hardware, software o servicios de consultoría; también se deberán incluir medidas que puedan calificar factores como cuestiones que genera el *Data Warehouse*, como pueden ser directrices que provocan mayor satisfacción al cliente, o bien que atraen a más clientes. Por lo tanto, los beneficios deberán contemplar matrices que incluyan factores objetivos y subjetivos, cuantificables y no cuantificables; con los cuales sólo así se podrá medir en términos reales el impacto de la construcción e implementación de un *Data Warehouse*. Así, el *Data Warehouse* puede ser medido de acuerdo a la habilidad para atraer nuevos clientes, proporcionar un mejor servicio, reducir costos, mejorar resultados financieros, o similares. El retorno de la inversión en el *Data Warehouse* es muy prometedor de acuerdo a los estudios de mercado que se han realizado, sin embargo, nunca se deberá olvidar el incorporar la parte estratégica de los negocios y asociarla directamente con la información emitida por el *Data Warehouse*, de esta manera la parte subjetiva de beneficios se reflejará claramente en el retorno de la inversión de esta solución.

Otro factor clave que deberá ser considerado para el éxito de un *Data Warehouse* es la utilización de la tecnología adecuada, ya que esta es uno de los habilitadores principales de la solución aquí propuesta. Asimismo, otra sugerencia clave es que no se deberá olvidar que, el éxito en la inversión de un *Data Warehouse* será directamente proporcional a la habilidad del *Data Warehouse* de direccionar y resolver una necesidad de negocio. Si el *Data Warehouse* es percibido por la comunidad de negocios sólo como la más reciente novedad en el área de tecnología informática, y además, el equipo de desarrollo tiene poca o ninguna experiencia con el negocio, es muy probable que la iniciativa del *Data Warehouse* falle, ya que no se esta dejando ningún beneficio al negocio en sí. El *Data Warehouse* es un habilitador para generar conocimiento, este conocimiento requerirá que las múltiples funciones del negocio colaboren conjuntamente. Los grupos informáticos ayudarán con la tecnología, mientras que el resto de las áreas funcionales proporcionarán el conocimiento del dominio específico que les concierne a cada una de ellas.

De acuerdo a la información con la que se cuenta hasta este momento, las compañías a nivel mundial están implementando su *Data Warehouse* para alcanzar una ventaja estratégica; esto se logrará a través del adecuado acceso a la información en el momento que se necesita, dicha información les dará el conocimiento de su negocio, productos, clientes y competidores.

Los escenarios que vislumbro para las siguientes décadas que puede tener el *Data Warehouse* son muy prometedores; hoy en día esta solución se encuentra para ambientes de tipo gráfico y en arquitecturas

cliente/servidor; pero la apuesta tecnológica va mucho más allá, el escenario optimista que considero será implementado se encamina hacia ambientes de Web en donde por medio de un navegador podrá ser accesada la información para la toma de decisiones; esto abarcará desde los ambientes empresariales, en los que un ejecutivo podrá consultar desde cualquier lugar geográfico la información que él requiere para el crecimiento competitivo y productivo de su empresa a través del mundo; llegando hasta ambientes de *Data Warehouse* compartidos en el ámbito individual en los cuales todos podremos tomar decisiones con base a la información que se encuentre en el "World Wide Web". Aunado a esto, el *Data Warehouse*, incluirá la capacidad de manejar datos especiales en los cuales podremos manejar videos animados, mapas, fotografías, y similares los cuales podremos manipular en términos a su contenido.

Escenario Optimista a Nivel Mundial del Mercado de *Data Warehouse*

	1998	2000	2010	2020
Crecimiento de la inversión en el <i>Data Warehouse</i> Incluyéndose los siguientes rubros: <ul style="list-style-type: none"> • Hardware 34% • Software 32 % • Servicios Profesionales (Consultoría) 34% 	Inversión estimada de \$5,500 Millones de Dólares (Fuente International Data Corporation 1996)	Inversión estimada de \$15,000 Millones de Dólares (Fuente International Data Corporation 1996)	Inversión estimada de \$185,000 Millones de Dólares	Inversión en el sistema integral organizacional de \$1,000,000 Millones de Dólares
Medios de Distribución y Uso	Únicamente a nivel interno y centralizado en las empresas, arquitecturas cliente/servidor. Sólo datos numéricos y alfanuméricos.	Arquitecturas Cliente/Servidor, incluyendo distribución de los datos vía Intranet. Sólo datos numéricos y alfanuméricos. Comercio Electrónico.	Soluciones de <i>Data Warehouse</i> compartidas entre compañías que trabajan como socios de negocio. Inicio de manejo de contenido con datos especiales como video, huellas digitales, mapas, etcétera. Distribución con Internet, Intranet y extranets.	Era del conocimiento, conglomerados de información, células informáticas. Manejo de todo tipo de datos para la toma de decisiones, transferencia por medio de mecanismos electrónicos más sofisticados, veloces y seguros en internet.
Sistemas Operativos preferidos para <i>Data Warehouse</i>	Proprietarios 20% UNIX 70 % NT 10%	Proprietarios 10% UNIX 45 % NT 45%	Proprietarios 10% Abiertos 90%	Completamente Abiertos
Hardware preferido para <i>Data Warehouse</i>	Tandem, Teradata, HP, IBM, Sun	Teradata, IBM, HP	Teradata, IBM	IBM

Obstáculos para el desarrollo del <i>Data Warehouse</i>	Economicos, políticos y organizacionales. Costos altos, la información no es compartida entre la organización.	Políticos y organizacionales, los costos de la construcción de un <i>Data Warehouse</i> se habrán reducido de manera importante, el reto será vencer el monopolio de la posesión de la información entre las diversas áreas de la organización.	Sistemas de <i>Data Warehouse</i> en su etapa de madurez a nivel mundial.	Ninguno. Cambio de paradigma.
Escenario Global del <i>Data Warehouse</i>	La inversión global de productos crecerá constantemente, debido a que la adopción del <i>Data Warehouse</i> esta en pleno auge. Continuando así para la siguiente década.	Ya se habrá superado la crisis del 2000 en los sistemas de computo. Y el <i>Data Warehouse</i> será un requisito el cual deberán cubrir todas las compañías a nivel mundial durante esta década.	Se continuará en el crecimiento de la era de la Información a la era del Conocimiento, en la cual el <i>Data Warehouse</i> juega un papel primordial.	El <i>Data Warehouse</i> en esta década se complementará con los demás sistemas de la organización, existirá un sistema único en el que el proceso de toma de decisiones y el operacional serán un ente integral.

Escenario Pesimista/Conservador a Nivel Mundial del Mercado de *Data Warehouse*

	1998	2000	2010	2020
Crecimiento de la inversión en el <i>Data Warehouse</i>. Incluyéndose los siguientes rubros: <ul style="list-style-type: none"> • Hardware 34% • Software 32 % • Servicios Profesionales (Consultoria) 34% 	Inversión estimada de \$3,000 Millones de Dólares. Rescisión en soluciones informáticas por crisis financiera mundial.	Inversión estimada de \$10,000 Millones de Dólares.	Inversión estimada de \$40,000 Millones de Dólares.	Inversión enfocada a otro tipo de soluciones para el conocimiento de información y toma de decisiones.

Medios de Distribución y Uso	Únicamente a nivel interno y centralizado en las empresas, arquitecturas cliente/servidor. Sólo datos numéricos y alfanuméricos.	Arquitecturas Cliente/Servidor, incluyendo distribución de los datos vía Intranet. Sólo datos numéricos y alfanuméricos.	Uso limitado del <i>Data Warehouse</i> .	Sustitución del <i>Data Warehouse</i> por otra solución tecnológica.
Obstáculos para el desarrollo del <i>Data Warehouse</i>	Económicos, políticos y organizacionales. Costos altos, la información no es compartida entre la organización.	Económicos, políticos y organizacionales. Las crisis económicas no permiten crecer las soluciones informáticas para la toma de decisiones.	Las empresas que cuentan con un <i>Data Warehouse</i> , ya no encuentran rentable la solución, ya que la misma implica altos costos de mantenimiento.	Sustitución del <i>Data Warehouse</i> por otra solución tecnológica.

Los escenarios se plantean de manera prospectiva, el escenario optimista muestra datos de información recabada por firmas de analistas expertos en soluciones de *Data Warehouse*.

Para el *Data Warehouse* no considero que se dé un escenario pesimista, más bien sólo un conservador, en el cual deberemos aceptar que el *Data Warehouse* es una solución un tanto costosa, que sólo estará al alcance de empresas medianas a grandes, las cuales dominarán sus mercados por el propio valor competitivo que generará la información proveniente de su solución de *Data Warehouse*.

Las proyecciones de mercado para el ambiente de *Data Warehouse* indican que para el año 2000 éste equivaldrá a 15 mil millones de dólares, teniendo una tasa de crecimiento anual del 35% (fuente META Group 1998). En cuanto al momento actual por el cual está atravesando el *Data Warehouse* en términos de su penetración es la siguiente: un 24% del total de las empresas tiene en producción esta solución, un 4% esta en la implementación, un 3% se encuentra en la planeación, un 27% esta en la evaluación de dicha solución y el 42% aún no tiene planes (fuente IDC 1997). Por ende y de acuerdo a lo anterior reafirmo que esta solución tecnológica sigue creciendo y seguirá creciendo en número de adeptos, hasta que se llegue a la satisfacción total de conocimiento de las empresas.

Desde mi punto de vista, una solución de *Data Warehouse* es un imperativo competitivo, ya que por medio de su uso las empresas estarán en capacidad de centralizar y explotar a su máxima expresión su información para el beneficio de su negocio.

Las grandes industrias se están moviendo hacia el conocimiento intensivo, este conocimiento maximizará su valor agregado en la planeación estratégica con la finalidad de que las compañías planeen su futuro en el presente, tomando decisiones estratégicas sustentadas en el conocimiento.

El *Data Warehouse* es un requisito obligatorio para comprometerse en un área de mercado, esta solución permitirá y facilitará una forma dominante para el siguiente milenio para aquellas compañías que lo adopten como herramienta informática para su administración estratégica empresarial.

Finalmente se puede agregar que el valor del *Data Warehouse* se encuentra en tres categorías principales: soporte de decisiones, rearquitectura de aplicaciones existentes y reingeniería del proceso empresarial. El primer rubro cubre lo visto a lo largo del presente trabajo y se refiere a la capacidad de obtener indicadores para la toma de decisiones, el segundo se deriva como una implicación inherente a la reingeniería por medio de la cual se decidirá como las aplicaciones existentes alimentaran el *Data Warehouse* para el proceso de transformación de datos a información y por último, la reingeniería empresarial, en la que la empresa pasará de un uso del *Data Warehouse* para el soporte a las decisiones a la reingeniería del proceso empresarial, debido a que se elevará el desempeño del personal, los activos y la capacidad competitiva de la misma.

Como conclusión global y de acuerdo a lo descrito a lo largo del trabajo se tiene que “el *Data Warehouse* es una herramienta informática que permitirá tomar decisiones con un fundamento sustentado en la información empresarial, esto traerá consigo crecimiento y un valor agregado competitivo a las empresas que lo adopten para soportar sus decisiones estratégicas, todo esto gracias a los datos históricos empresariales”.

“La estrategia es el camino de la muerte y la vida, es la brecha hacia la sobrevivencia o destrucción, así es imperativa su examinación. Existe un camino para sobrevivir que ayudará y fortalecerá, existe un camino a la destrucción que empujará al olvido”. The Art of War, Sun Tzu, Shambhala Publications, 1988.

Así, el *Data Warehouse* es el camino hacia la sobrevivencia que ayudará y fortalecerá los negocios del futuro.

Actualización de datos	La actividad de actualizar continuamente el contenido de los datos del <i>Data Warehouse</i> a partir de las fuentes de datos después de terminar la carga inicial. Este es un proceso continuo donde las operaciones de datos de hoy se convierten en los datos históricos del mañana.
Adición	La actividad de combinar datos de diferentes tablas para formar una unidad de información más compleja que se necesita con frecuencia para responder a consultas al <i>Data Warehouse</i> con mayor facilidad y rapidez. Casi todas las bases de datos operacionales guardan datos en la unidad más sencilla en que es posible (normalización) para prevenir anomalías de actualización y eliminación.
Agregar	Incorporar múltiples fuentes de datos o dimensiones para crear una dimensión nueva.
Análisis dinámico de datos	Comparar valores de datos de diferentes fuentes de datos y de dimensiones distintas.
Análisis multidimensional de datos	Análisis de múltiples dimensiones de datos.
Base de datos multidimensional	Una base de datos diseñada alrededor de un conjunto de dimensiones; se usa en el análisis multidimensional.
Base de datos operacional	Una base de datos que apoya sistemas de software que están soportando las operaciones empresariales de una organización. Conocida como base de datos OLTP, por sus siglas en inglés OnLine Transaction Processing (Procesamiento Transaccional en Línea).
Bloque de acceso y uso	El componente del <i>Data Warehouse</i> que maneja las opciones para acceso y uso de la información que guarda el <i>Data Warehouse</i> . Estas opciones incluyen herramientas para consulta, generación de reportes y examen.
Bloque del <i>Data Warehouse</i>	El componente de la arquitectura del <i>Data Warehouse</i> que representa actividades directamente con el <i>Data Warehouse</i> .

Bloque de fuentes de datos	El componente de la arquitectura que representa las actividades que tratan con las bases de datos operacionales y las fuentes externas que aportan datos para cargarlos al <i>Data Warehouse</i> .
Capa de administración de datos	El componente del <i>Data Warehouse</i> que maneja las opciones de administración de datos.
Capa de administración de metadatos	Componente relacionado con los aspectos de la definición de los datos que se capturan y se guardan en el <i>Data Warehouse</i> . Todas las herramientas y procesos que deben interactuar con el <i>Data Warehouse</i> usan los aspectos de la definición.
Capa de infraestructura	Componente del <i>Data Warehouse</i> relacionado con las plataformas y el soporte de cómputo en términos de hardware y software del ambiente del cómputo que apoyan el <i>Data Warehouse</i> .
Capa de transporte	El componente del <i>Data Warehouse</i> relacionado con el transporte de datos entre componentes de la solución de <i>Data Warehouse</i> . El transporte maneja las opciones de transporte en todos los niveles para proporcionar el traslado de datos.
Ciclo de vida de desarrollo	El flujo de actividades que representa un proyecto de implantación y despliegue de <i>Data Warehouse</i> desde la concepción hasta su despliegue y uso.
Consolidación de datos	Proceso de sintetizar porciones de información en bloques únicos de conocimiento esencial.
Consulta empresarial inteligente	Consulta, análisis y generación de reporte de un <i>Data Warehouse</i> con más de dos dimensiones.
Consulta	Una petición formal y claramente especificada de información del <i>Data Warehouse</i> o del <i>Datamart</i> planteada por un usuario o una herramienta operada por un usuario.

Curva de madurez de la tecnología	La curva que representa la inserción de una tecnología, desde su concepción hasta su integración final en los sistemas de misión crítica.
Datos de herencia	Fuentes de datos históricos para el <i>Data Warehouse</i> , los cuales se guardan en medios fuera de línea.
Diccionario de datos	Un compendio de definiciones y especificaciones para categorías de datos y sus relaciones.
Dimensión de datos	Un calificador conceptual que provee el contexto o significado para una métrica; por ejemplo: tiempo del producto, geografía. Es también el nivel más alto de análisis.
Dimensión	Los <i>Data Warehouse</i> organizan un depósito grande de datos históricos y operacionales usando, múltiples dimensiones para establecer categorías.
Esquema de base de datos	La especificación de como esta organizada la base de datos.
Esquema de estrella	Como el nombre sugiere, el esquema estrella es un paradigma de modelado que tiene en medio un objeto único conectado a varios objetos en forma radial. El esquema estrella refleja la visión del usuario en una consulta empresarial: hechos tales como ventas, compensaciones, pagos, etcétera, calificadas por una o más dimensiones. El objeto en el centro de la estrella se llama tabla de hechos y los que se conectan desde la periferia se llaman tablas de dimensión.
Estructura	Una estructura es una manera de ver algo complejo y hacerlo entendiére usando un conjunto de analogías simplificadas para ayudar a convertir una solución compleja en componentes pequeños.
Extracción de datos	Modalidad de descubrimiento del análisis de datos, o analizar datos de detalle para revelar relaciones, patrones y asociaciones insospechados o desconocidos.

Extracción	Actividad de transferir datos de bases de datos operacionales la <i>Data Warehouse</i> .
Filtrar y cotejar	Eliminar datos sólo operacionales y seleccionar datos operacionales que pide el modelo de diseño de datos del <i>Data Warehouse</i> ; por ejemplo, cotejar con el área o segmento temáticos deseados.
Generalizar	Ver a niveles más altos de resumen o adición.
Granularidad	Término que se usa en los <i>Data Warehouse</i> para expresar en nivel de detalle. A más alto nivel de granularidad, más bajo nivel de detalle (más alto nivel de abstracción).
Hardware	Hardware se define como el aspecto físico de las computadoras, telecomunicaciones y otros dispositivos de información. El término se usa comúnmente para describir una "caja", circuitos electrónicos y componentes de una computadora.
Información histórica	Información que se recopiló en el pasado y que por lo general se archiva y retiró de las bases de datos operacionales. Debido a los volúmenes de datos que recopilan diario los sistemas operacionales, la información histórica se traslada a medios más lentos y menos usados, tales como dispositivos magnéticos.
Integración de datos	La actividad de combinar datos de diversas fuentes en el <i>Data Warehouse</i> a fin de presentar una colección única de datos para el usuario del <i>Data Warehouse</i> .

Internet	<p>Internet es un sistema compuesto por computadoras en red y la gente que las utiliza. El predecesor de este tipo de servicio fue conocido como ARPAnet concebida en la Agencia de Investigación de Proyectos Avanzados (<i>Advanced Research Projects Agency - 1969 - USA</i>). Su propósito original fueron causas bélicas (específicamente guerras nucleares). Internet es un servicio público, cooperativo y de mantenimiento masivo. Físicamente Internet usa una porción de los recursos totales de las redes de telecomunicaciones públicas.</p> <p>Dos términos adicionales se han agregado a este primero: intranet y extranet; el primer caso describe comunicación en dominios privados y el segundo comunicación entre intranets.</p>
Limpiar y pulir	<p>Añadir campos faltantes y reemplazar codificación secreta con reglas y estructuras de datos empresariales.</p>
Limpieza y refinamiento de datos	<p>La actividad de eliminar errores en las fuentes de datos antes de intentar cargarlos en el <i>Data Warehouse</i> o <i>Data Mart</i>.</p>
Mercado de datos (<i>Datamart</i>)	<p>Una implementación del <i>Data Warehouse</i> con un ámbito de funciones y datos más restringido de <i>Data Warehouse</i> corporativo, sirve a un departamento único o a una parte de la organización. Una compañía generalmente tiene varios <i>Datamarts</i>.</p>
Middleware	<p>El término común que se aplica al software que intercambia información de manera transparente entre aplicaciones y bases de datos. El middleware ofrece un mecanismo de conexión abstracta entre el software de aplicación y la base de datos, y oculta al programador de aplicaciones los elementos específicos que dependen de la implementación.</p>
Modelado de datos	<p>Actividad de representar las categorías de datos y relaciones entre ellas como una abstracción en un diagrama.</p>
Modelar	<p>Desarrollar el modelo de <i>Data Warehouse</i> y definir la base de datos física del <i>Data Warehouse</i>.</p>

Modelo lógico de datos	Una representación formal y abstracta de las categorías de datos y las relaciones entre ellos, en forma de diagrama. La representación es independiente del modo en que las clases y sus relaciones se implementan físicamente en la base de datos.
Navegador	Un navegador (Browser en inglés) es un programa que proporciona una forma de ver, leer y escuchar información en el <i>World Wide Web</i> . Técnicamente un navegador es un programa cliente que utiliza el protocolo de transferencia de hipertexto (HTTP Hypertext Transfer Protocol, siglas en inglés).
OLTP	Siglas de OnLine Transaction Processing (Procesamiento Transaccional en Línea). El término se usa para definir cualquier sistema de software que reúne datos usando transacciones (en el momento que ocurren) entre la fuente de datos y la base de datos.
Procesamiento analítico en línea	OnLine Analytical Processing (OLAP) es una tecnología de análisis multidimensional de datos y capacidad de generación de reportes.
Refinado de datos	Las actividades colectivas aplicadas a los datos de las fuentes a fin de prepararlos para que se carguen en el <i>Data Warehouse</i> .
Reingeniería empresarial	Término usado para las actividades relacionadas con procesos existentes de modelado empresarial para identificar áreas de mejora.
Reporte	Documento que se produce en respuesta a una consulta.
Resumen	Actividad de incrementar granularidad de la información en una base de datos. Un resumen reduce el nivel de detalle y es muy útil para presentar datos para soporte de decisión.
Sincronización de datos	La actividad de mantener los datos del <i>Data Warehouse</i> actualizados con los datos que se guardan en las bases de datos operacionales que se cargaron primero.
Sistema Cliente-Servidor	Un sistema de aplicación de software donde el procesamiento de la aplicación lo proporcionan en conjunto dos componentes arquitectónicos claramente separados: el cliente y el servidor.

Sistema de misión crítica	Un término de uso común para describir una aplicación de software que es esencial para la operación continua de una empresa. La falla de este sistema afecta la viabilidad de la empresa.
Sistema de soporte de decisiones (DSS Decision Support Systems)	Sistemas automatizados de aplicación que ayudan a la organización a tomar decisiones relacionadas con el negocio.
Sistemas de información ejecutiva	Un término común que usan los sistemas de generación de consulta y generación de reportes que ejecutan adiciones y resúmenes directamente en los datos operacionales, algunas veces guardando datos resumidos y agregados en forma privada, y proporcionando la capacidad de consultas y generación de reportes a quienes toman decisiones (ejecutivos).
Software	Software es un término genérico utilizado para describir múltiples programas utilizados para operar computadoras y dispositivos relativos a ellas. El software se divide típicamente en dos partes: software de aplicación (programas que hacen trabajo específico para un área usuaria) y software de sistema (que incluye sistemas operativos y cualquier programa que soporta una aplicación de software. El término middleware es usado para describir programas que son intermediarios entre software de aplicación y de sistema.
Tabla de dimensión	Un componente del esquema estrella (ver también <i>Esquema de estrella</i>).
Tabla de hechos	Un componente del esquema estrella (ver también <i>Esquema de estrella</i>).
Verificar	Comprobaciones de calidad de los datos definidos por las reglas empresariales, el modelo de <i>Data Warehouse</i> , etcétera; la congruencia y la integridad dominan la verificación.
Visualización de datos	Presentación de los datos de una forma gráfica para facilitar el análisis de datos complejos y voluminosos.

World Wide Web

La definición técnica del World Wide Web (red Ampla del Mundo) representa todos los recursos y usuarios de Internet (ver definición Internet) que utilizan el protocolo de transporte de hipertexto. El *World Wide Web* es el universo del acceso de información en red para el conocimiento humano. El *World Wide Web* (WWW) se compone de hardware, software y telecomunicaciones.

- BOAR BERNARD. Understanding Data Warehousing Strategically. Information Technology Consultants. 1998.
- HARJINDER, S. Gill. Cliente/Servidor Data Warehousing. Prentice Hall. 1996.
- HENRY F. KORTH & ABRAHAM SILBERSCHATZ. Fundamentos de Bases de Datos. McGraw Hill. México. 1988.
- INMON, W. Building the Data Warehouse. QED Technical Publishing Group. 1992.
- INMON, W. Information Systems Architecture for the 90's. QED Technical Publishing Group. 1992.
- INMON, W. Using the Data Warehouse. John Wiley & Sons. 1994.
- INMON, W.H. and Welch, J.D.; and Glassey; Katherine. Managing the Data Warehouse. John Wiley. 1997.
- KELMAN, Glenn. Using INFORMIX-MetaCube with Data Warehouses. Training Material. Informix Software Inc. Part ID502-55011-999999-3. 1996.
- KING, Mike. An Introduction to Data Warehousing. Training Material. Informix Software Inc. Part ID502-54591-999999-3. 1997.
- PAL ROLPH & PETER BARTRMAN. Executive Information Systems. 1990.
- STHROCECK MICHAEL. Data Warehousing Is Worth The Investment. InternerWeek Magazine. 1998.
- Control de Gestión basado en Planeación Estratégica. Consultores SICOFIN. México 1993.
- EIS's Report Management Overview. Business Intelligence. 1990.
- White Paper. Data Warehousing for Enterprise Decision Making. Informix Software Inc. Part ID. 000-20716-70. 1997.
- Info Sheet. Informix in Data Warehousing. Informix Software Inc. Part ID 000-20716-70
- Informix Corporation. Informix and Data Warehousing. Informix Times. Issue 3. 1995.
- Internet: <http://www.informix.com>
- Internet: <http://www.knowledgecenters.org/>
- Internet: <http://www.knowledgecenters.org/dwcenter.asp>
- Internet: <http://www.whatis.com/oltp.htm>
- Internet: <http://www.techweb.com>