



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

FACULTAD DE CIENCIAS

LA INFORMACIÓN Y LA TOMA DE DECISIONES A  
PARTIR DE UN DATA WAREHOUSE

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

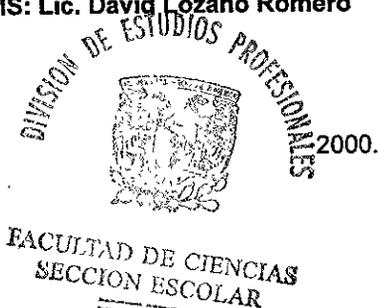
PRESENTA

202968

Fernando Rodríguez Ramírez

DIRECTOR DE TESIS: Lic. David Lozano Romero

MÉXICO, D. F.





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.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**MAT. MARGARITA ELVIRA CHÁVEZ CANO**  
Jefa de la División de Estudios Profesionales  
**Presente**

Comunicamos a usted que hemos revisado el trabajo de Tesis:

“La Información y la Toma de Decisiones a partir de un Data Warehouse”

realizado por: Fernando Rodríguez Ramírez

Con número de cuenta 9150821-0 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de tesis  
Propietario

Lic. David Lozano Romero

Propietario

Mat. Margarita Elvira Chávez Cano

Propietario

Lic. Fidel Eduardo Ventura Hernández

Suplente

Act. Ricardo Villegas Azcorra

Suplente

Act. Humberto Ramos Sánchez

~~FACULTAD DE CIENCIAS~~

Consejo Departamental de Matemáticas.

M. en C. José Antonio Flores Díaz.

DEPARTAMENTO DE MATEMÁTICAS

## **Agradecimientos.**

Agradezco a todas las personas que hicieron posible este trabajo: por su valiosa dirección, tiempo y acertados comentarios al Lic. David Lozano Romero. Asimismo a la Mat. Margarita Elvira Chávez Cano, por su acertada asesoría, constantes comentarios y revisiones así como por su entusiasta participación. También a Lic. Fidel Eduardo Ventura Hernández, Act. Ricardo Villegas Azcorra y Act. Humberto Ramos Sánchez, quienes revisaron el trabajo y participaron como Sinodales.

Agradezco a mis Padres por su comprensión y esfuerzo constante.

A mi familia por haber mantenido su apoyo y por su compañía inigualable.

A mis Amigos por su confianza.

A la Universidad Nacional Autónoma de México, con gratitud.

## **PREFACIO**

Desde el principio del tiempo el hombre como ser organizado y, bajo la influencia de su estructura social, ha emprendido el trabajo de predecir eventos de ocurrencia común o eventuales que alteran su medio ambiente ya sea en su entorno social, ambiental, económico y financiero por destacar algunos, para ello ha tenido que valerse de la experiencia (sucesos relativos al evento en cuestión, documentados), como única referencia y plataforma de entendimiento, encontrando valiosa ayuda a partir de múltiples variables asociadas las cuales ayudan a explicar o predecir eventos futuros, esta experiencia a desarrollado técnicas de manejo de información; hoy día en que se tiene plena conciencia de que el propósito fundamental del método científico es crear un modelo, partiendo de que un modelo es una manera conveniente de expresar el total de experiencias que se tienen para luego deducir de dicha experiencia, si se esta en presencia de leyes y patrones respectivos, y, de ser así, manipular dichas leyes y patrones para pronosticar, asignando una cierta probabilidad de ocurrencia dadas las tendencias observadas en el modelo, ello ha establecido métodos para auxiliar en diversos campos de interés.

Considerando lo anterior el propósito fundamental de este trabajo es mostrar como a partir de la información generada durante la operación de las empresas se tiene la posibilidad de establecer modelos que permitan obtener valiosas respuestas a interrogantes de clientes, productos, proveedores, costos, trayectoria de ventas, etc. Elementos todos ellos que sin duda ofrecen una ventaja competitiva en el entorno global que hoy se gesta, así mismo se consideran algunas técnicas de explotación de información a partir del concepto de Data Warehouse (repositorio central de datos, o almacén corporativo de datos), así como técnicas estadísticas de exploración y explotación de información.

El capítulo I atiende la importancia de la información y le atribuye el valor de activo, plantea su utilización en los niveles administrativos y considera la relación entre las metas de las organizaciones y la explotación de la información (gestión de la información). El capítulo II ofrece una revisión del proceso de toma de decisiones y considera la integración de la información para la mejor toma de decisiones, hasta este punto el material intenta situar al lector en un escenario donde la información producto de la operación de las empresas es el medio para generar respuestas y plantear estrategias de competencia que estén en función de la capacidad de respuesta histórica del

negocio, ya sea para producir un efecto de reposicionamiento en su forma de operar o reforzar los productos y servicios que ofrece con éxito, ello genera la necesidad de tener un almacén de datos corporativos (Data Warehouse) de tal forma que la información utilizada con fines de análisis haya sido validada y ofrezca certidumbre al analista y mantenga su carácter de información corporativa a través inclusive de reestructuraciones, ya que Data Warehouse considera el almacenamiento del detalle de la operación histórica, proporcionando una plataforma de análisis corporativo, eliminando la generación de datos heterogéneos (datos repetidos por distintas entidades de la organización), partiendo de esta necesidad el capítulo III provee una revisión de los conceptos de sistemas de información y muestra las ventajas que ofrece el Data Warehouse y algunas de sus variantes como Data Mart, y por último el capítulo IV se ocupa de algunas técnicas estadísticas de explotación de información en el Data Warehouse tales como Data Warehousing (organización de información dentro de un Data Warehouse), Data Mining (Proceso analítico de búsqueda de patrones), Redes Neuronales proceso que se refiere a modelos que aprenden de la información y pueden ser aplicados a diferentes conjuntos de información en busca de sub-conjuntos de la misma, que ofrezcan patrones de comportamiento.

Así el propósito final de este escrito es mostrar la tendencia actual a la construcción de este tipo de arquitectura dado el gran desarrollo de Software y Hardware de explotación y almacenamiento que se ha producido, considerando así mismo la sensible disminución en costos respecto hace algunos pocos años en la adquisición de estos insumos.

LA INFORMACIÓN Y LA TOMA DE DECISIONES A PARTIR DE UN DATA WAREHOUSE.

## **CONTENIDO**

Prefacio i

### **Capitulo I**

#### **El Impacto de la Información en las Organizaciones.**

1-1	La información considerada como recurso	1	
1-2	La toma de decisiones en los niveles de administración		5
1-3	La información en los niveles administrativos	7	
1-4	Los sistemas de información	8	
1-5	Los sistemas de información y los niveles de la administración	18	
1-6	Las metas de la organización y el papel de los sistemas de información	18	
1-7	Justificación para el desarrollo de un sistema de información	19	
1-8	Gestión de la información	21	

### **Capitulo II**

#### **La Toma de Decisiones.**

2-1	El Proceso de Toma de Decisiones	22	
2-2	Teoría de la Decisión	25	
2-3	Análisis de Decisiones	30	
2-4	Decisiones Estratégicas	32	
2-5	La Integración de la Información para la Mejor Toma de Decisiones	35	

## Capítulo III

### Sistemas de Soporte en la Toma de Decisiones.

3-1	Evolución de los Sistemas de Información	38
3-2	Situación Actual de los Sistemas de Información	42
3-3	Definición de Data Warehouse	50
3-4	Elementos Principales del Data Warehouse	52
3-5	Componentes del Data Warehouse	54
3-6	Características del Data Warehouse	59
3-7	Estructura del Data Warehouse	60
3-8	Valor del Data Warehouse	63
3-9	Terminología Relacionada con el Data Warehouse	64
3-10	Algunos Esquemas de Diseño para Bases de Datos	71
3-10	OLAP, On Line Analytical Processing	80

## Capítulo IV

### Análisis y explotación del Data Warehouse.

4-1	Concepto de Data Warehouse y Explotación	88
4-2	Data Mining	91
4-3	Análisis Estadístico	93
4-4	Descubrimiento de Conocimientos	98
4-5	Redes Neuronales	102
4-6	Otras Técnicas y Herramientas de Data Mining	103

**Conclusiones** 105

**Glosario** 107

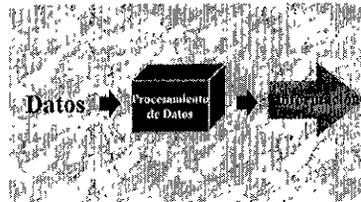
**Bibliografía** 111

**LA INFORMACIÓN Y LA TOMA DE DECISIONES A PARTIR DE UN DATA WAREHOUSE.****Capítulo I****El Impacto de la Información en las Organizaciones.**

1-1	La información considerada como recurso	1	
1-2	La toma de decisiones en los niveles de administración		5
1-3	La información en los niveles administrativos	7	
1-4	Los sistemas de información	8	
1-5	Los sistemas de información y los niveles de la administración	18	
1-6	Las metas de la organización y el papel de los sistemas de Información	18	
1-7	Justificación para el desarrollo de un sistema de información	19	
1-8	Gestión de la información	21	

**-CAPITULO PRIMERO****EL IMPACTO DE LA  
INFORMACIÓN EN LAS  
ORGANIZACIONES.****1-1 La Información Considerada como Recurso.**

La información es un conjunto de datos arreglados y ordenados en forma útil. La palabra **datos** (*clata*, en latín e inglés) es el plural de dato, que significa **hecho**. Los datos son hechos o material original de información. Son representados por símbolos, pero sólo son información en un sentido limitado, dado que, la información es resultado de un proceso de transformación. El procesamiento convierte los datos originales en información.



Toda información consiste de datos, pero no todos los datos producen información significativa y específica.

La información es el conocimiento relevante producido como resultado del procesamiento de datos y adquirido por la gente para promover el entendimiento y cumplir propósitos específicos. La información producida es de poco valor, a menos que aporte ideas para las decisiones y acciones humanas.

La acumulación de datos adicionales puede con frecuencia agregar nuevas dimensiones a la información existente, pero la interpretación de la información generalmente requiere un juicio humano y puede variar de una persona a otra.

Los datos proporcionan una base de la cual se deriva la información y que sirve para el análisis, la información

ensamblada por la recopilación y el manejo de datos, proporciona argumentos para tomar decisiones.

La información puede ser considerada un recurso ya que cumple con los atributos de un recurso físico y a través de un recurso, las empresas pueden llegar al cumplimiento de sus objetivos, de hecho, para algunas compañías la información es considerada un activo.

### **Atributos de la Información.**

- Tiene valor como el dinero, las materias primas o la fuerza laboral.
- Tiene características que permiten su medición, en términos de uso, duración y efecto sobre otros recursos:
  - Puede ser valorada en términos de recolección, almacenamiento y recuperación.
  - Puede ser presupuestada y controlada.
  - Puede evaluarse en términos de costo y valor de uso con fines de administración.

### **El Valor de la Información.**

Entre más sirva la información para reducir incertidumbre en las decisiones efectuadas por analistas en cualquier nivel administrativo, mayor será su valor. El costo de la información que obtenemos, debe compararse con los beneficios que ofrece a partir de su uso. La información exacta, oportuna, íntegra y concisa tiene un mayor valor que la que carece de una o varias de estas características, a pesar de que con frecuencia se sacrifique alguna de ellas por razones, las más de las veces, económicas.

A continuación nos enfocaremos con más detalle en las características que debe contener la información para considerarla de valor:

### **Exactitud.**

Es el porcentaje de información correcta con respecto al total de la información generada en un periodo determinado. La exactitud se refiere a que los datos se encuentren libres de errores con respecto a su manejo y manipulación.

**Oportunidad.**

O bien, puntualidad. El tiempo de respuesta debe ser lo suficientemente breve para que tenga poco volumen y bajo costo, todo esto con el fin de descubrir las tendencias importantes que señalen la necesidad de una acción.

**Integridad.**

Por lo general la información es incompleta, una mejor integración de los hechos disponibles que se encuentran diseminados en una empresa es el medio para proporcionar a los directivos información más completa.

**Concisión.**

El resumen de datos importantes que señalen las áreas de excepción de las actividades normales o planeadas es lo que necesitan los directivos actualmente.

---

**La Confiabilidad en la Información.**

La presencia de los elementos anteriormente mencionados en la información implica la existencia de confiabilidad en la información.

La confiabilidad de la información está en función de su grado de exactitud, se refiere a la dependencia de los datos y al grado en que todos los aspectos de una situación o problema son medidos. Además, debe existir un compromiso entre confiabilidad y oportunidad. La confiabilidad de la información está correlacionada con su grado de completamiento o integración.

La mayor parte de las interpretaciones gerenciales de las operaciones están basadas en información incompleta. La información parcial sobre ciclos de tiempo, tales como informes mensuales de ventas y producción, si son analizados y relacionados con el ciclo total, pueden ofrecer una base aceptable para el análisis y la acción. Por lo que, cuando se usan datos de sólo una parte del periodo que se reporta, hay también un considerable descenso en la confiabilidad de los datos relativos a las fases iniciales de una operación. Los datos reunidos al comienzo de una operación no son característicos de la operación

total o de sus últimas fases y son una fuente de información poco confiable.

Sin embargo, la información debe también permitir pronósticos de desviaciones futuras mientras éstas estén ocurriendo. Si bien es cierto que la información utilizada en esos pronósticos no es exacta al cien por ciento, el proceso de pronosticar y revisar obliga a desarrollar y depender de la información oportuna. El escoger entre lo oportuno y lo exacto de la información, varía dependiendo de la situación, responsabilidad no tan simple que recae en la persona que toma la decisión.

Actualmente, la información recopilada ha perdido confiabilidad debido principalmente a que presenta las siguientes inconsistencias:

- No se encuentran validados los datos, ya sea por los sistemas operacionales internos, o bien, por fuentes externas.
- Muchas veces la información representa periodos anteriores perdiendo veracidad respecto de la agrupación actual, debido al proceso de transformación que van sufriendo los datos y variables expresadas.
- No se encuentra una ubicación unificada para la consulta de información originada de varias fuentes de datos.
- Existen diferentes fuentes de información para un mismo dato
- La generación de información con características de explotación particulares, y después utilizada en otro sentido puede ocasionar desconocimiento en la interpretación de la misma.

A partir del uso de computadoras dentro del entorno empresarial, se han ido realizando esfuerzos para proveer información concisa y objetiva a los usuarios finales como un apoyo en la toma de decisiones. Paulatinamente, se ha reconocido que la información no es sólo un recurso dentro de una organización, sino que además es la base primordial para la toma de decisiones, pues es un elemento decisivo para determinar el éxito o fracaso de un negocio.

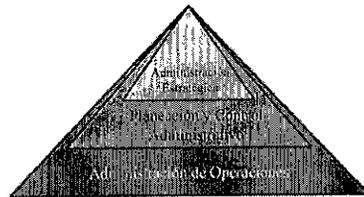
## 1-2 La Toma de Decisiones en los Niveles de Administración.

De manera general, se puede estructurar la administración en tres niveles básicos a saber: la operación, la planeación del control administrativo, y la dirección estratégica.

Cada nivel, colabora al logro de metas y objetivos en la organización.

- Sistema

Administrativo -



El tipo de decisiones que se realizan dentro de una organización, está en función de las actividades que realiza cada nivel dentro del sistema administrativo.

### **Administración de Operaciones.**

La operación ocupa el plano base de los tres planos administrativos. Los gerentes de operaciones apoyan sus decisiones en una serie de reglas preestablecidas, las cuales al implantarse en forma correcta otorgan resultados predecibles.

Se maneja una alta seguridad en la toma de decisiones. Las decisiones de los gerentes de operaciones afectan la implantación de los programas de trabajo, el control del inventario, de embarques, de recepción y el control de procesos como la producción.

Los gerentes de este nivel supervisan los detalles operativos de la organización, verificando que las tareas básicas se ejecuten oportunamente y sin rebasar las restricciones de la misma. Se trata generalmente de problemas cotidianos, siempre a corto plazo.

**Planeación y Control Administrativo.**

La administración media constituye el segundo plano de los niveles del sistema administrativo. Los gerentes de mandos intermedios toman decisiones sobre la planeación y el control a corto y mediano plazo y de la forma de asignar recursos para cumplir con las metas de la organización.

Los gerentes de mandos intermedios toman decisiones en ambientes controlados. Sus decisiones son de muy diversa naturaleza, desde el pronóstico de las necesidades futuras de recursos, hasta la solución de problemas laborales que puedan afectar la productividad. Por lo general, la toma de decisiones carece de un alto grado de estructuración. La toma de decisiones se puede caracterizar por tener cierto contenido operativo y estratégico.

**Administración Estratégica.**

La administración estratégica constituye el tercero de los tres niveles de la estructura administrativa. La toma de decisiones de los gerentes estratégicos, guiará a los gerentes de operaciones y de mandos intermedios en los meses y años próximos.

Los gerentes estratégicos, actúan bajo un clima de gran incertidumbre en la toma de decisiones, la cual es aclarada mediante el postulado de metas, establecimiento de estrategias y políticas para su logro, los gerentes estratégicos definen a la organización como un todo. Para ellos la amplia imagen es, donde la empresa decide el desarrollo de nuevas líneas de productos, la separación de negocios no redituables, la adquisición de otras compañías compatibles, la decisión de venta de la empresa, estrategias de ventas cruzadas de productos a partir del conocimiento del comportamiento de sus clientes, etcétera.

Los estrategias tienen objetivos de decisión múltiple, mientras que los gerentes de operaciones son de decisión sencilla. Con frecuencia, es difícil para un gerente de operaciones identificar aquellos problemas que serían fáciles a un gerente de mandos superiores ya que deben considerar las variables que conectarán la decisión con el resto del entorno operativo y estratégico de la empresa.

Los responsables de la planeación estratégica se enfrentan a problemas complejos o semi-estructurados, es decir su juicio está

basado en el conocimiento a priori de variables que dado el planteamiento de alguna situación otorgan algún valor o comportamiento en función del supuesto inicial. Por lo regular la amplia gama de opciones que contiene un problema resultan de difícil articulación. La naturaleza de las decisiones es diferente para cada uno de los distintos niveles; para los gerentes estratégicos, las decisiones se presentan con frecuencia para casos únicos; para los gerentes de operaciones, las decisiones tienden a ser repetitivas.

Las decisiones que se toman en la administración estratégica, sirven de guía a los niveles de operación, planeación y control administrativo, lo cual garantiza a estos niveles tomar decisiones en un ambiente de certidumbre, la cual está en función de las suposiciones que se hacen sobre el problema, lo que implica el cumplimiento de variables, las cuales van enfocadas al desarrollo de objetivos y a la asignación de recursos, para su cumplimiento se establecen estrategias y políticas.

### **1-3 La Información en los Niveles Administrativos.**

Los gerentes de operaciones requieren de información interna de naturaleza repetitiva y de bajo nivel. Son dependientes de información que manifiesta el desempeño actual y son grandes usuarios de información en línea, llamada de tiempo real. La necesidad de información que represente el desempeño pasado y de información periódica es bastante moderada por parte de los gerentes de operaciones. Para ellos, tiene poco valor la información externa que apoye proyecciones hacia el futuro o que permita la creación de escenarios de simulación.

En el siguiente nivel administrativo, los gerentes de mandos intermedios, que planean y controlan al mismo tiempo, necesitan tanto la información inmediata como la acumulada en periodos anteriores de operación del negocio. Debido a la naturaleza de sus tareas, los gerentes de este nivel tienen una gran necesidad de información de tiempo real, y con el fin de llevar un control adecuado, también necesitan la información actual del desempeño para compararla contra los estándares establecidos. Los gerentes de mandos intermedios dependen mucho de información interna. En contraste con los gerentes de operaciones, tienen un gran interés por la información histórica, además de toda la información que les permita pronosticar los acontecimientos futuros, así como la simulación en diversos escenarios.

Los gerentes estratégicos difieren por sus necesidades informativas de los gerentes de operaciones y de gerentes de mandos medios. Ellos dependen en mayor medida de información externa que provee las noticias de las tendencias del mercado y las estrategias de las corporaciones competidoras.

Puesto que las tareas de los gerentes estratégicos demandan proyecciones hacia un futuro incierto, necesitan aquella información de naturaleza productiva, así como la información que les permita crear diversos escenarios de simulación. Los gerentes estratégicos también hacen uso de los informes periódicos, ya que son corresponsables de la adaptación o cambios de gran dinamismo.

Los planificadores estratégicos requieren de un resumen de información general, más que contar con datos inconclusos o de mucho detalle, como lo requieren los gerentes de mandos operativos. La información para los planeadores estratégicos puede no ser reciente y valorada, mientras que los gerentes operativos necesitan la información precisa y actualizada.

Los planeadores estratégicos necesitan más información de tipo cualitativo, de origen externo, que de información cuantitativa e interna como la que utiliza un gerente de operaciones.

Las necesidades de información de los miembros de la organización también pueden visualizarse por áreas funcionales. Cada una de las divisiones, tales como compras, producción, distribución y mercadotecnia, se tratan como si las funciones del sub-sistema determinaran las necesidades de información. Con énfasis en los requerimientos funcionales de información, cada uno de los gerentes poseerá una visión muy distinta de la importancia de su propio sub-sistema funcional.

La perspectiva del gerente de operación ubica a su departamento como el elemento central del negocio, donde todas las demás áreas funcionales dependen de él.

#### **1-4 Los Sistemas de Información.**

Hasta ahora, el trabajo ha estado centrado en comentar la importancia, el valor y las características que tiene la información dentro de una organización, pues bien, el paso siguiente es el tema de los sistemas de información, por lo que es conveniente comentar

acerca de los sistemas; para más tarde continuar con los sistemas de información y adentrarnos más al tema que nos compete.

### **Definición de Sistema.**

Sistema es un grupo de partes integradas que tienen el propósito común de lograr algún o algunos objetivos.

Sistema es un conjunto de elementos organizados e integrados que se encuentran en interacción, que buscan alguna meta común, operando para ello sobre datos o información en una referencia temporal para producir como salida información con un valor agregado, lo cual le concede un estatus de importancia a esta nueva información la cual tiene una vigencia en su valor.

Sistema es un conjunto de partes interrelacionadas que trabajan juntas para alcanzar un objetivo en común.

### **Elementos de un Sistema.**

#### **Variables.**

Todo sistema es un procesador. Las entradas del sistema están normalmente pre-establecidas, es decir los datos entran en un formato único, aunque sus valores pueden ser distintos y por tanto, son sus variables. Las salidas están en función de la entrada en ese sentido varía su magnitud y significado.

#### **Parámetros.**

Un parámetro define un lineamiento de la información en función de directrices de negocio ó técnicas ya establecidas que actúan sobre la información procesada agrupándola, filtrándola ó ya sea incluso descartando algunos datos por su naturaleza, dando como resultado información dentro de ciertos rangos que permitirán el acceso de la misma en siguientes etapas del proceso .

#### **Componentes.**

Son las partes identificables del sistema. Por lo que los componentes pueden tener tareas particulares una vez realizadas entregan información a otros componentes en una secuencia de producción, a lo cual se suele llamar sub-sistema, partiendo de elementos individuales que en su interacción forman sub-sistemas y éstos a su vez sistemas considerando componentes tales como

objetos o personas, que poseen propiedades o características que influyen en la operación del sistema, en su velocidad, precisión, confiabilidad, capacidad, etcétera.

---

**Estructura.**

Es el conjunto de relaciones entre los objetos y atributos de los objetos del sistema. El grado en que los elementos funcionan juntos para alcanzar los objetivos totales sirve de igual forma para definir la estructura. Aunque se tiende a pensar en sistemas cuyos elementos trabajan totalmente para la consecución de metas comunes, hay pocos que verdaderamente funcionan de esa manera. Los elementos de un sistema trabajan juntos en grados variables.

**Definición de Sistema de Información.**

Considerando que la información es la base de operación de actividades estratégicas en todas las compañías, deben desarrollarse sistemas para producirla y administrarla. El objetivo de los sistemas de información es asegurar que la información exacta y confiable esté disponible cuando se necesite y sea presentada en forma práctica para su mayor aprovechamiento.

Un sistema de información es un grupo de personas, datos y procedimientos que funcionan en conjunto. Los componentes del sistema buscan un objetivo común para apoyar las actividades de la organización.

Un sistema de información es un método organizado de suministrar información pasada y presente y, proyectarla en relación con operaciones internas y de inteligencia externa. Respaldar la planeación, el control y las funciones operacionales de una organización mediante el suministro de información uniforme en el marco de tiempo apropiado para asistir al proceso de la toma de decisiones.

Un sistema de información es:

- Herramienta para planear y tomar decisiones.
- Modelo de la organización.
- Banco de información.

- Ayuda para encontrar y resolver problemas.

Un sistema de información ejecuta tres actividades principales:



1. Recibe datos de fuentes internas o externas de la organización como elementos de entrada.
2. Actúa sobre los datos para producir información (es un sistema generador de información).
3. El sistema produce la información o salida para el futuro usuario.

Los sistemas de información no necesitan estar basados en las computadoras, pero con frecuencia lo están, debido a que el volumen de información en ocasiones es tan grande que no es posible mantener un sistema de información en forma manual. El factor determinante es si un sistema puede ser mejorado incluyendo en él la capacidad de procesamiento por computadoras, cuando crece el volumen de información los procedimientos aumentan en complejidad o las actividades llegan a estar más interrelacionadas, obteniéndose grandes mejoras al introducir la ayuda de un sistema de cómputo.

Un sistema de información está constituido por una base de datos, procedimientos, programas de consulta e informe, así como de procedimientos manuales. Hay que tomar decisiones en el diseño de sistemas respecto al uso de personas o de máquinas y también entre varias clases de máquinas teniendo siempre presente los atributos y los costos.

En la búsqueda de apoyar las actividades de una empresa, se ha llegado al planteamiento de los sistemas de información, y precisamente sus objetivos básicos son la automatización de los procesos operativos, el proporcionar información que sirva de

apoyo al proceso de toma de decisiones y el logro de ventajas competitivas a través de su implantación y uso.

Ahora bien, existen diversos tipos de sistemas de información, entre los cuales se mencionan por el momento y de manera general los siguientes:

- Aquéllos que logran la automatización de procesos operativos son llamados con frecuencia Sistemas Transaccionales o Sistemas de procesamiento de datos operacionales.
- Sistemas de Información para la administración, los cuales se encargan de generar información que eventualmente se utiliza en la toma de decisiones, pues unifica funciones dentro de una empresa.
- Los Sistemas de Información que apoyan el proceso de toma de decisiones son los Sistemas de Soporte a la Toma de Decisiones que son adaptables a sus usuarios y soportan la toma de decisiones.
- Sistemas Expertos de Soporte a la Toma de Decisiones los cuales asimilan la experiencia de quienes toman las decisiones en la solución de problemas.

### ***Tipos de Sistemas de Información.***

Cada uno de los diferentes tipos de sistemas de información está destinado a procesar datos por alguna de las siguientes razones:

- Capturar los detalles de las transacciones.
- Permitir que se tomen decisiones.
- Comunicar información entre personas y localidades.

A partir de esto, se mencionan las características principales de los siguientes tipos de sistemas de información:

### ***Sistema de Procesamiento de Transacciones.***

Procesa datos referentes a las transacciones. Una transacción es un suceso que implica o afecta a una organización.

Los sistemas de procesamiento de datos son aquellos sistemas de información computarizados que se desarrollan para procesar grandes volúmenes de información generada en las funciones administrativas, tales como la nómina o el control de inventarios. Los sistemas de procesamiento de datos liberan del tedio y la rutina a las tareas que se realizan manualmente; sin embargo, el elemento humano sigue participando al llevar a cabo la captura de la información requerida.

Tales sistemas ejecutan periódicamente programas de manera automática. Una vez preparados, escasamente se requiere tomar decisiones. Los sistemas de información ejecutan las actividades de carácter rutinario de las empresas.

Los sistemas transaccionales permiten lograr ahorros significativos en mano de obra, pues automatizan tareas operativas. Estos sistemas requieren integrar datos para poder realizar sus operaciones y como resultado pueden generar grandes volúmenes de información. Por medio de estos sistemas se cargan grandes bases de información para una explotación en línea y posterior. Sus beneficios son visibles y palpables y en el corto plazo se pueden evaluar los resultados y las ventajas que deriva el uso de este tipo de sistemas, constituyendo así una de las partes más costosas dentro de la operación de una empresa.

Las razones del procesamiento son: *registro, clasificación, orden, cálculo, sintetización, almacenamiento y visualización* (o despliegue) de los resultados. Los sistemas de procesamiento de transacciones procesan datos referentes a las actividades de la empresa. La función en sí de procesamiento de transacciones se basa en la operación metódica de cualquier empresa u organización.

- La importancia de *registrar* la información generada en la operación de las empresas es que permite analizar el comportamiento y tendencia de los productos y clientes, lo cual promueve un mayor conocimiento del negocio y provoca el planteamiento de cursos de acción.
- La *clasificación* implica agrupar datos según las características comunes, a manera de estratos para una explotación posterior.
- La distribución y *ordenación* de datos consiste en disponerlos según una secuencia o sucesión. Facilita el procesamiento y a veces torna a los datos menos complicados de localizar, ya que se encuentran ordenados según las necesidades.

- Las operaciones de *cálculo* son, tal vez la razón más común para que se realice una acción durante el procesamiento de datos de transacciones, que consiste en efectuar alguna operación sobre los datos para generar resultados útiles.
- La *síntesis* o resumen reduce gran cantidad de datos de transacciones a una forma más breve y concisa.
- A través del *almacenamiento* las organizaciones mantienen registros de los sucesos o eventos que afectan sus operaciones.

### **Sistemas Informáticos para la Administración.**

#### (Management Information Systems - MIS - )

Es el sistema que examina y recupera los datos provenientes del ambiente que los captura a partir de las transacciones y operaciones efectuadas dentro de la empresa. Filtra, organiza y selecciona datos, los cuales son presentados en forma de información a los gerentes, proporcionándoles los medios para generar la información deseada.

Los sistemas de información para la administración no sustituyen a los sistemas de procesamiento de datos, más bien toman en cuenta a las funciones de procesamiento de datos. Son sistemas que se sustentan en la relación que surge entre las personas y las computadoras. Los sistemas de información para la administración requieren para su operación de personas, de software (programas de cómputo) y de hardware (equipo físico de cómputo). Estos sistemas de información para la administración soportan un amplio espectro de tareas de las organizaciones, más aún que los sistemas de procesamiento de datos.

### **Sistemas de Apoyo para la Toma de Decisiones.**

Son sistemas de información cuyo propósito es auxiliar a los administradores con decisiones únicas que no se repiten y que carecen relativamente de estructura. Los sistemas de apoyo para la toma de decisiones existen a fin de responder a condiciones inesperadas y propias de información. Estos sistemas son particularmente importantes para los altos niveles de dirección que deben tratar constantemente los problemas cambiantes y tomar decisiones en todo tipo de casos.

El sistema de apoyo para la toma de decisiones es similar a los sistemas de información tradicionales para la administración en el sentido de que ambos dependen de una base de datos como fuente de información, aunque la decisión será tomada en circunstancias de oportunidad, necesidad legal entre otras lo cual considera el elemento de juicio de la persona responsable de tomar la decisión. Los sistemas de apoyo para la toma de decisiones se diseñan con una orientación hacia la persona o el grupo que los utilizará, y no como los sistemas de información tradicionales para la administración.

---

**Características de un Sistema de Soporte para la Toma de Decisiones.**

1. Un sistema computacional tiene la posibilidad de interactuar en forma amigable y con respuestas en tiempo real con el analista.
2. Sirve de apoyo tanto en decisiones estructuradas y no estructuradas.
3. Tiene un uso frecuente por parte de la administración media y alta para el desempeño de sus funciones.
4. Puede emplearse por usuarios de diferentes áreas del negocio.
5. Permite acoplarse a una variedad determinada de estilos administrativos.
6. Permite que el usuario desarrolle de manera directa modelos de decisión sin la participación operativa de profesionales en Informática.
7. Permite la posibilidad de interactuar con información externa como parte de los modelos de decisión.
8. Facilita la comunicación de información relevante de los niveles altos a los niveles operativos y viceversa.
9. Tiene la capacidad de acceder a información de bases de datos corporativas.
10. Es simple, fácil de aprender y utilizar por el usuario final.

Estos sistemas necesitan una plataforma de información, la cual es proporcionada por los Sistemas Transaccionales, la información que generan sirve de apoyo a los mandos intermedios y a la alta administración en el proceso de toma de decisiones. Por lo general son intensivos en cálculos y en agrupación de información. Apoyan la toma de decisiones que, por su misma naturaleza son decisiones repetitivas.

### **Sistemas de Soporte para la Toma de Decisiones**

#### - Decision Support Systems - (DSS)

Apoyan la toma de decisiones mediante la generación y evaluación sistemática de diferentes alternativas o escenarios de decisión, con el uso de modelos y herramientas computacionales. No soluciona problemas, sólo apoya el proceso de toma de decisiones.

Soporta el proceso de decisión de los gerentes con acceso flexible a los modelos e información relevantes. Estos procesos se identifican en dos tipos, estructurados y no estructurados. En el caso de los no estructurados, la toma de decisiones juega un papel vital en proveer juicio y evaluar la capacidad como insumos dentro del problema, mientras que en los estructurados, los problemas son resueltos por procesos de decisión automáticos.

Los sistemas de soporte para la toma de decisiones explotan las bases de datos en un sentido definido, es decir por temas, estratos, clientes con valores extremos, etc. así mismo, los datos representan periodos diferentes de tiempo por lo que se tiene dos tipos de procesos:

- El primer tipo identifica un proceso repetitivo de cálculo de datos como reportes periódicos.
- El segundo tipo, es el no repetitivo, es un método de proceso más selectivo, manejado a través del resultado de cada una de las interacciones y finalmente, un sistema de soporte para la toma de decisiones debe distinguir cuando estadísticamente signifique un evento, es decir, deberá seleccionar la mejor variable como medida.

Una de las actividades del usuario final es aplicar la información para la toma de decisiones con el objetivo de maximizar los beneficios que se puedan rescatar del medio ambiente, minimizar costos (economía) y las transferencias entre sectores deberán

existir por parte de este criterio múltiple, económico, medio ambiente, político, regulatorio y tecnológico.

### **Sistemas de Información para Ejecutivos**

#### - Executive Information Systems - (EIS)

Dirigidos a apoyar también el proceso de toma de decisiones de los altos ejecutivos de una organización presentando información relevante y usando recursos visuales y de fácil interpretación, con el objetivo de mantenerlos informados.

### **Sistemas para la Toma de Decisiones de Grupo**

#### - Group Decision Support Systems-

Logran la participación de un grupo de personas durante la toma de decisiones en ambientes de confidencialidad y consenso, con el fin de apoyar decisiones simultáneas.

### **Sistemas Expertos de Soporte para la Toma de Decisiones**

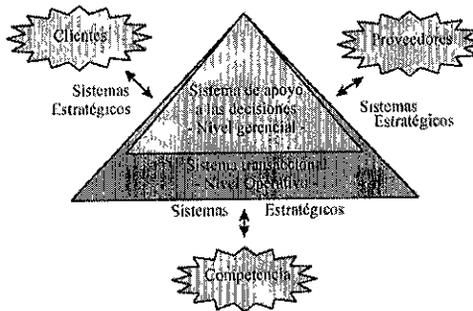
#### - Expert Decision Support Systems -

Cargan bases de datos que se integran con información ligada según el carácter de los productos y puntos de venta para que diferentes usuarios la exploten y apoyen la toma de decisiones. Es un área especial de los sistemas de apoyo para las decisiones, utilizan programas de computación que almacenan datos y reglas con el objeto de reproducir el proceso de operación de la empresa. Tales sistemas se utilizan para tratar situaciones caracterizadas por un alto grado de incertidumbre por tanto en los sistemas de apoyo para la toma de decisiones, se debe aplicar una combinación de experiencia y criterio para tomar una decisión. Usualmente los sistemas expertos se enfocan en un área limitada, estudiando hipótesis y hechos que permitan determinar variables extremas, así los sistemas de toma de decisiones modifican el entorno de la empresa al consolidar una decisión.

Los sistemas expertos son, en sí, un tipo especial de sistemas de información, que tienen un uso práctico en los negocios debido a la reciente y amplia disponibilidad del hardware y software. Un sistema experto (también conocido como sistema basado en el conocimiento) captura y en efecto utiliza el conocimiento de un experto, para la solución de un problema particular de la organización.

### Sistemas Estratégicos.

Su función primordial es proporcionar información para apoyar la toma de decisiones. Existen sistemas estratégicos pre-diseñados así como los desarrollos a la medida, su forma de desarrollo es con base en necesidades particulares logrando su perfeccionamiento a través de su evolución dentro de la empresa, tiene como objetivo lograr ventajas que los competidores no posean, ya sea en costos y servicios diferenciados con clientes y proveedores. Apoyan el proceso de innovación de productos y procesos dentro de una empresa, pues buscan ventajas respecto a los competidores.



### 1-5 Los Sistemas de Información y los Niveles de la Administración.

Cada uno de los tres niveles de administración repercute en diferente grado en el desarrollo de los sistemas de información. Es oportuno recalcar que mientras ciertos requerimientos de la gerencia se encuentran bien definidos, otros son ambiguos o se sobrepone.

Los sistemas de información diseñados para los gerentes de operaciones serán de valor si proveen información que se use en el sector operativo, que promuevan el acceso en línea a la información y de fácil visualización.

### 1-6 Las Metas de las Organizaciones y el Papel de los Sistemas de Información.

Un principio fundamental en el desarrollo de los sistemas de información para las empresas es que las aplicaciones son una herramienta y no un instrumento que debe tenerse para utilizar la tecnología de la información. Los sistemas deben desarrollarse

sobre la base de su propia capacidad con el fin de mejorar el desempeño de la organización.

La forma como se adaptan los sistemas de información en una organización depende de la naturaleza comercial de ésta. Por ejemplo, si los clientes son lo más importante, entonces es probable que los analistas de sistemas orienten la mayor parte de su trabajo hacia aplicaciones que se encarguen de mejorar el servicio proporcionado a los clientes. Si el control de los costos es el principal indicador del éxito en la compañía, entonces la mayor parte de las aplicaciones estarán dirigidas hacia la identificación y preservación de la ventaja en costos.

Se puede asegurar que es casi imposible que los sistemas de información resulten eficaces si son desarrollados en forma independiente de los objetivos, valores y metas de la organización, para la que fueron diseñados. Por tanto, las solicitudes de proyecto siempre se deben preparar y evaluar de acuerdo a estos principios.

### **1-7 Justificación para el Desarrollo de un Sistema de Información**

Los sistemas de información facilitan el aprovechamiento de la información y del personal que son dos elementos clave en una organización. La gerencia de una organización necesita los sistemas de información principalmente por las siguientes razones:

#### ***La Utilización de la Información.***

Los administradores están siempre manejando datos (detalles y hechos) en forma constante. Lo que más necesita un administrador es que la información esté a la altura de las tareas que se realizan o las decisiones que se toman.

#### ***El Rápido Ritmo del Cambio.***

Mantenerse al día es una preocupación constante de la gerencia. La información que recibe la administración debe actualizarse. Y los medios para lograrlo están en constante evolución. No obstante a eso, la necesidad de mejor información es más crítica cada vez.

### ***La Creciente Complejidad de la Administración.***

Debido en parte al ritmo de vida de una organización y por otra al alcance y dimensión de las tareas administrativas, el trabajo de la gerencia está creciendo en complejidad. Entre los factores que contribuyen a esto, se encuentran la evolución en la forma contratar personal especializado, la calidad de los productos terminados y la competencia real, así como la contracción de los límites del tiempo. La mayor diversidad en todo lo anterior añade una nueva dimensión a la toma de decisiones administrativas.

### ***La Interdependencia de las Unidades de la Organización.***

Las organizaciones no son agrupamientos incoherentes de trabajadores o estaciones de trabajo. Dado que todas las actividades están relacionadas, cada individuo interactúa con sus colegas. Los éxitos y problemas en un extremo de la empresa afectan a las actividades de la misma, aunque se encuentren físicamente separadas.

### ***El Mejoramiento de la Productividad.***

La productividad es la aptitud para incrementar la eficiencia de un proceso. Los sistemas de información computarizados, desarrollados y utilizados adecuadamente, pueden mejorar la productividad aumentando el volumen de trabajo realizado y la velocidad con la cual se ejecutan las transacciones. De igual manera la posibilidad de reducir errores o de aumentar la precisión.

### ***La Disponibilidad de las Computadoras en red para Usuarios Finales.***

Los sistemas de información computarizados son accesibles a una gran variedad de usuarios. Con una computadora personal en red se puede manejar información contable y administrativa para probar el impacto de estrategias alternativas, así como evaluar los resultados actuales de la empresa, resumir grandes volúmenes de datos, transmitir o recibir registros de información, elaborar informes, propuestas y correspondencia. Además, efectuar revisiones cuando sea necesario.

### **El Reconocimiento de la Información como Recurso.**

La información tiene un valor porque influye en la manera como opera la empresa. Carecer de información vital puede ocasionar que los administradores cometan errores, pierdan oportunidades y se enfrenten a grandes problemas de rendimiento. Los sistemas de información también son un recurso ya que incrementan la capacidad de los administradores y de los trabajadores. También hace posible lograr nuevos niveles de eficiencia.

### **1-8 Gestión de la Información.**

Existen tres importantes razones que exigen un enfoque de administración total de la información. La primera es eliminar el desperdicio, la segunda es asegurar la disponibilidad de la información y la tercera es detectar el peso de las cargas impuestas. La administración de los recursos de información es aplicable a cualquier organización que genera, adquiere y usa información. La administración de los recursos de información ajusta la tecnología a la necesidad, no el producto a la tecnología.

**LA INFORMACIÓN Y LA TOMA DE DECISIONES A PARTIR DE UN DATA WAREHOUSE.****Capítulo II****La Toma de Decisiones.**

2-1	El Proceso de Toma de Decisiones	22
2-2	Teoría de la Decisión	25
2-3	Análisis de Decisiones	30
2-4	Decisiones Estratégicas	32
2-5	La Integración de la Información para la Mejor Toma de Decisiones	35

**- CAPITULO SEGUNDO****LA TOMA  
DE  
DECISIONES.****2-1 EL PROCESO DE TOMA DE DECISIONES.**

El propósito fundamental de este capítulo es proveer una revisión de los conceptos y métodos asociados con la toma de decisiones. Los estudios de varios autores se han orientado principalmente al problema de la toma de decisiones en las organizaciones, dichos estudios no sólo muestran cómo las personas toman decisiones, sino que también han facilitado el desarrollo de una teoría de cómo las personas deberían de hacerlo; sobre todo en aquellas decisiones que afectan múltiples áreas.

Generalmente la visión de la alta dirección al tomar el rumbo del negocio plantea la necesidad de formular los indicadores más adecuados, que permitan percibir con certeza diversas situaciones a partir de la operatividad del negocio; así pues, podemos distinguir los elementos esenciales en la toma de decisiones y se mencionan a continuación.

**Elementos en la Toma de Decisiones.**

- |                     |  |
|---------------------|--|
| <b>El objetivo.</b> | Que constituye un resultado o situación a la cual se desea evolucionar.  |
| <b>El problema.</b> | Se refiere a la diferencia existente entre la realidad actual y la deseada.  |
| <b>La decisión.</b> | Que es la elección de un curso de acción determinado entre las variantes posibles. En efecto, es evidente que no se trata de un problema sencillo para el cual respuestas diferentes podrían ser apropiadas. |
| <b>La acción.</b>   | La decisión sin acción carece de valor, ello implica la necesidad de implantar la decisión.  |

Una vez que se ha trazado el *objetivo*, existen factores de importancia relevante como lo es el TIEMPO, de esta forma, el proyecto tiene un *ciclo de vida*, la *decisión* está necesariamente ubicada en un punto en el tiempo. Existen proyectos que dado su ciclo de vida, al presentarse una desviación en su ejecución caducan sin haber rendido ningún fruto. El factor tiempo contribuye a que cambie el contexto del problema.

Algunos autores incluyen la obtención de la información como un paso adicional generalmente ubicado dentro de la formulación del problema, en nuestra opinión la obtención de la información no debe presentarse como un paso separado. El problema es normalmente descrito dentro de un marco idealista, es decir, se supone:

- a) Que el analista tiene información suficiente en un gran número de variables posibles las cuales incluyen variables relevantes y funciones apropiadas para ellas, además posiblemente, algunas otras variables ajenas y variables de funciones externas y,
- b) que el analista tiene disponible información en la cual basa las conclusiones eventuales. En la práctica, la falta de satisfacción de estas suposiciones puede hacer de la toma de decisiones un ejercicio insignificante.

Se podría afirmar que la parte de definición más importante es la toma de la decisión, ya que en esta parte, necesariamente se debe de considerar una secuencia de etapas y pasos. Este modelo muestra cierto orden secuencial.

- ♦ La identificación clara de los objetivos pertinentes es condición para especificar debidamente un problema.
- ♦ A su vez, la especificación del problema es requisito para encarar racionalmente los posibles cursos de acción.
- ♦ Es necesario el planteamiento de cursos de acción antes de su evaluación a fin de tomar la decisión. Cuantos más cursos de acción se consideren, mayor será la probabilidad de elegir el más adecuado.
- ♦ Finalmente, para que sea efectiva la decisión, debe ir acompañada de una adecuada planificación para su correcta ejecución.

Sin embargo, el orden secuencial indicado anteriormente, podría no seguirse en términos absolutos, la secuencia definida es tentativa.

**Por ejemplo:**

- 
- El proceso de especificación de un problema bien puede provocar la revisión de un objetivo que en un primer momento se había fijado, lo cual implica un cambio.
  - La elección de curso de acción a seguir se presenta como una etapa posterior a la identificación de objetivos. No obstante, la evaluación de los puntos de acción puede traer a colación nuevos elementos de juicio que ameriten una revisión de los objetivos fijados originalmente.
  - La planeación detallada inherente a la ejecución de una decisión puede indicar inconvenientes no evaluados debidamente al momento de tomar la decisión, en consecuencia será necesario replantear la situación.

---

Bajo las consideraciones hechas anteriormente se habrá de evaluar un curso de acción a partir de un problema y la concepción de posibles cursos de acción. A continuación se presenta el modelo:

**Modelo General para la Toma de Decisiones.**

- Formulación del problema.
- Identificación de criterios para la evaluación de cursos de acción.
- Concepción de posibles cursos de acción.
- Evaluación de cursos de acción.
- Restricciones u otras pautas en cuanto al empleo de recursos humanos, financieros, físicos, etc.
- Principios, políticas o reglas que deban considerarse en función del negocio.

- ♦ Atributos o estándares que tiene que satisfacer el producto objeto de la decisión.

## 2-2 TEORÍA DE LA DECISIÓN.

### **Comentarios Introdutorios.**

En cada organización debe efectuarse la toma de decisiones. Frecuentemente, la decisión es la respuesta a una pregunta.

#### **Por ejemplo:**

---

¿Qué producto debe ofrecerse en venta cruzada en función de la demanda de algún otro y la relación entre ambos?

¿Cómo evaluar la rentabilidad de los clientes bajo el concepto de cliente integral?

¿De qué manera utilizar la mercadotecnia para minimizar los costos de investigación de mercado?

¿Cómo obtener información útil acerca de los costos y comisiones que se derivan a partir de la captación de clientes?

---

En cada una de estas situaciones son posibles de dos a más cursos de acción. Una decisión se toma cuando se elige alguno de estos.

En las últimas décadas se ha desarrollado una teoría de las decisiones que permite evaluar la eficacia de una decisión al medir el grado en el que sus resultados satisfacen el objetivo u objetivos especificados de antemano por la persona o grupo de personas que tomaron la decisión.

El contenido de la teoría de decisiones se originó de tres corrientes principales:

- ♦ La teoría de las preferencias y de la utilidad.
- ♦ La teoría de las probabilidades.
- ♦ La teoría de la inferencia estadística.

**Elementos de un Problema de Decisión.**

Los elementos de un problema de decisión son:

**Los actos.**

O cursos de acción. Al tomar una decisión, el decisor selecciona uno o algunos de ellos.

**Los eventos.**

O cosas que suceden. El decisor tiene un escenario previo, sujeto a cambio de acuerdo al entorno.

**Las probabilidades de ocurrencia de cada evento.**

Las cuales a menudo se conocen como probabilidades a priori.

**El pago de cada resultado.**

Donde un resultado es una combinación de un acto y un evento. El resultado se estimaba en función de su probabilidad a priori.

**Teoría de Decisiones.**

Un proceso de decisión es el que requiere un solo conjunto de decisiones o una secuencia de decisiones para su conclusión. Cada decisión tiene una ganancia o pérdida asociada, la cual se determina conjuntamente a partir de situaciones externas que influyen al proceso, el conjunto de posibles circunstancias, conocido como estados de la naturaleza, y una distribución probabilística que controle la ocurrencia de cada estado, se consideran como conocidos. Tanto el conjunto de decisiones permitidas como el conjunto de estados de la naturaleza se consideran finitos.

Las decisiones permitidas se denotan,  $D_1, D_2, \dots, D_m$ ; los estados de la naturaleza  $S_1, S_2, \dots, S_n$ ; el rendimiento asociado con la decisión  $D_i$ , y el estado  $S_j$ , se denotan por  $g_{ij}$ , ( $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ ). Un proceso que requiere tomar sólo una decisión se define en la tabla 3-2. Esta tabla se conoce como matriz de ganancias, siempre que las anotaciones  $g_{ij}$  estén en términos de ganancia para quien toma la decisión. A las pérdidas se las representa como ganancias negativas.

Tabla 3-1

	Estados de la naturaleza	
	S <sub>1</sub>	S <sub>2</sub>
D <sub>1</sub>	60	120
D <sub>2</sub>	-100	2000

Tabla 3-2

	Estados de la naturaleza			
	S <sub>1</sub>	S <sub>2</sub>	...	S <sub>n</sub>
D <sub>1</sub>	g <sub>11</sub>	g <sub>12</sub>	...	g <sub>1n</sub>
D <sub>2</sub>	g <sub>21</sub>	g <sub>22</sub>	...	g <sub>2n</sub>
...	.....	.....	.....	.....
D <sub>n</sub>	g <sub>n1</sub>	g <sub>n2</sub>	...	g <sub>nn</sub>

Ejemplo:

Una compañía de energéticos ofrece al dueño de un terreno \$60,000 por los derechos de exploración de gas natural en un sitio determinado y la opción para un desarrollo futuro. La opción, si se ejerce, vale 60,000 adicionales para el propietario, pero esto ocurrirá sólo si se descubre gas natural durante la etapa de exploración. El propietario, considerando que el interés de la compañía es una buena indicación de que existe gas, esta tentado a desarrollar por su propia cuenta. Para hacer esto, deberá contratar equipos locales con experiencia en exploración y desarrollo. El costo inicial es de \$100,000 los que se perderán si no se encuentra gas. Sin embargo, si descubre gas, el propietario estima una ganancia de \$2 millones de dólares.

Las decisiones para el propietario son D<sub>1</sub> (aceptar la oferta de la compañía de energéticos) y D<sub>2</sub> (explorar y desarrollar por su cuenta). Los estados de la naturaleza son S<sub>1</sub> (no hay gas en el sitio) S<sub>2</sub> (hay gas en el sitio). Las ganancias (en miles de dólares) para el propietario en cada combinación de eventos se dan en la tabla 3-1. Quedan por estimarse las probabilidades correspondientes a los dos estados de la naturaleza, P(S<sub>1</sub>) y P(S<sub>2</sub>).

En un proceso de decisión, sólo quien toma la decisión es capaz de realizar decisiones racionales; la naturaleza no es capaz de esto. El estado de la naturaleza existente en cualquier tiempo dado es un evento aleatorio, pero la distribución probabilística no se considera como una estrategia contra la decisión, diseñada para provocar pérdida para quien toma la toma la decisión.

**Criterios de decisión.**

Criterios "Ingenuos" (Naive), estos criterios permiten seleccionar la decisión que minimiza la pérdida máxima (minimax) para quien toma la decisión. En términos de una matriz de ganancia, es aquella decisión que maximiza la ganancia mínima posible, el

criterio optimista es seleccionar la que maximiza la ganancia posible, el criterio del punto medio, es seleccionar la decisión para la cual el promedio de las ganancias mínima y máxima es mayor. Como ninguno de estos tres criterios se basa e el estado probable de la naturaleza, se consideran inferiores a otros basados en ello. A continuación se presentan dos criterios probabilísticos.

Considere la tabla 3-1

La ganancia mínima para la decisión  $D_1$  es 60, mientras que para  $D_2$  es -100. ya que  $\max \{60, -100\} = 60$  que es la ganancia asociada a  $D_1$ , bajo el criterio minimax se recomienda  $D_1$ .

El valor mayor en la matriz es 2,000 ganancia asociada a  $D_2$ . por lo tanto,  $D_2$  es la decisión recomendada bajo el criterio optimista.

Los promedios de las ganancias mínima y máxima para  $D_1$  y  $D_2$  son, respectivamente,  $(120+60)/2 = 90$  y  $(2000+(-100))/2 = 950$  Ya que  $\max \{90, 950\} = 950$  está asociada con  $D_2$ , bajo el criterio punto medio se recomienda  $D_2$ .

Criterio a Priori (o de Bayes), este criterio consiste en seleccionar aquella decisión que maximiza la ganancia esperada. Necesariamente considera una probabilidad asociada a el o los estados de la naturaleza.

Considere que el propietario estima una probabilidad de encontrar gas de 0.6.

Con  $P(S_2) = 0.6$ , entonces  $P(S_1) = 1 - 0.6 = 0.4$ . Empleando los datos de la tabla 3-1, se calcula la ganancia esperada de  $D_1$  como

$$E(G_1) = (60)(.04) + (120)(0.6) = 96$$

Y la ganancia esperada de  $D_2$  como

$$E(G_2) = (-100)(.04) + (2000)(0.6) = 1160$$

Ya que el máximo es 1160, asociado con  $D_2$ ,  $D_2$  es la decisión recomendada bajo el criterio a priori.

Criterio a posteriori, si se puede realizar un experimento que proporcione información sobre el verdadero estado de la naturaleza, entonces los datos de este experimento pueden combinarse con las probabilidades iniciales de los diversos estados, para dar una distribución probabilística actualizada, denótese el resultado del experimento con  $\phi$  y considérese que la

confiabilidad del experimento esta dada por las probabilidades condicionales  $P(\phi | S_1), P(\phi | S_2), \dots, P(\phi | S_n)$ . Las probabilidades actualizadas (o a posteriori) de los estados:  $P(S_1 | \phi), P(S_2 | \phi), \dots, P(S_n | \phi)$  están determinados por el teorema de Bayes.

El propietario del terreno ha decidido realizar pruebas con sonido en el sitio donde se sospecha hay gas natural, con un costo de \$30,000 las pruebas indican que no hay gas, pero la prueba no es perfecta. La compañía que realizó las pruebas acepta que 30% de las veces la prueba indicara que no hay gas cuando en realidad éste existe. Cuando no existe gas la prueba es acertada 90% de las veces.

Inicialmente  $P(S_2) = 0.6$ , entonces  $P(S_1) = 0.4$ . Nótese con  $\phi_1$  al evento: las pruebas de sonido indican que no hay gas. Entonces la confiabilidad de la prueba esta dada por las probabilidades condicionales  $P(S_1 | \phi_1) = 0.9$  y  $P(\phi_1 | S_1) = 0.3$ . El teorema de Bayes da las probabilidades actualizadas como:

$$P(S_1 | \phi_1) = \frac{P(\phi_1 | S_1) P(S_1)}{P(\phi_1 | S_1) P(S_1) + P(\phi_1 | S_2) P(S_2)} = \frac{(0.9)(0.4)}{(0.9)(0.4) + (0.3)(0.6)} = \frac{2}{3}$$

$$P(S_2 | \phi_1) = 1 - P(S_1 | \phi_1) = \frac{1}{3}$$

La matriz de ganancia a posteriori se obtiene de la tabla 3-1 restando 30 (miles de dólares) de cada anotación para reflejar el costo de la prueba, las ganancias esperadas para las decisiones  $D_1$  y  $D_2$ , respectivamente, en términos de las probabilidades actualizadas son:

$$E(G_1 | \phi_1) = (60 - 30)(.66) + (120 - 30)(.33) = 230$$

$$E(G_2 | \phi_1) = (-100 - 30)(.66) + (2000 - 30)(.33) = 570$$

Ya que la ganancia máxima esperada está asociada con  $D_2$ ,  $D_2$  es la decisión recomendada bajo el criterio posteriori.

A fin de contar con una estructura adicional, se clasifica un problema de decisión dependiendo si ésta se toma en condiciones de certeza o incertidumbre la cual es proporcionada por la probabilidad a priori y reforzada por la probabilidad a posteriori.

### Decisiones con Certeza.

Cuando se toma una decisión con certidumbre, el decisor conoce cuál será el resultado para cada curso de acción que pudiera seleccionar. En primera instancia, pudieran parecer estas decisiones como triviales, ya que se conoce el resultado para

cualquier acto que pudiera escogerse. Sin embargo, existen factores ajenos que en principio no hubieran sido considerados tales como cambios legales, desastres naturales, guerra, es decir los estados de la naturaleza.

### **Decisiones con Incertidumbre.**

Una decisión se toma bajo incertidumbre si el decisor no conoce cual evento ocurrirá y por consiguiente no sabe el resultado para cada curso de acción. Considérese que:

- El decisor no tiene un conocimiento a priori de la probabilidades de los resultados para cada acto que pudiera elegirse.
- Generalmente Las decisiones bajo incertidumbre involucran a un competidor que piensa y genera una estrategia de competencia, este tópico es contemplado por la Teoría de Juegos.

Algunos ingredientes que faltan son el conocimiento, la experiencia, el juicio y el sentimiento de la persona ante determinada situación. Es importante incorporar estos insumos dentro de la evaluación.

## **2-3 ANÁLISIS DE DECISIONES.**

### **Comentarios Introdutorios.**

Uno de los aspectos más difíciles y comunes de decisiones directivas es hacerse una de estas dos preguntas: ¿Cuándo deberé decidir? y, ¿Cuánta información habré de tener antes de que yo decida?. Estas preguntas a menudo son similares una de la otra. Una razón para decidir posteriormente es, que la incertidumbre se resolverá en el tiempo, de tal manera que el decidir más tarde aumenta la cantidad de información disponible, pero ello va mermando la oportunidad de la decisión.

### **Concepto de Análisis de Decisiones.**

El análisis de decisiones es una disciplina que se ha estado constituyendo a partir de la experiencia de las grandes corporaciones que combinan la filosofía, la teoría, la metodología y la práctica profesional necesaria para formalizar el análisis de

decisiones. Específicamente, el análisis de decisión se ocupa de la distribución eficaz de recursos limitados (finitos), refleja los conceptos eficiente y limitado, ya que la óptima distribución de dinero, mano de obra, energía y muchos otros factores limitados es importante para quienes toman decisiones, la mezcla consciente de estos conceptos promueve la maximización de recursos en las corporaciones.

### ***Información Perfecta o Clarividencia.***

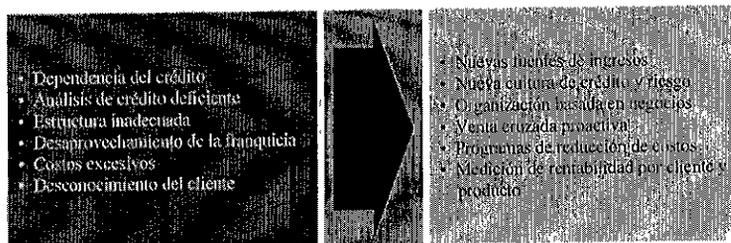
No obstante que los términos de información directa y clarividencia pueden parecer fuera de lugar en decisiones directivas, éste no es el caso. La información perfecta significa que el decisor es capaz de conocer qué evento ha acontecido o sucederá antes de que se haga la elección de un acto. Esto es lo mismo que clarividencia: Un clarividente es capaz de predecir el futuro, como si el evento hubiera sido observado antes de que se seleccionara el acto.

Para relacionar estos conceptos a las decisiones directivas, a menudo existen formas diferentes en que un directivo puede encontrar antes de tiempo lo que va a suceder. En algunos casos una llamada telefónica resolverá una incertidumbre permitiendo que la decisión se haga con información perfecta. En otros, una visita personal proveerá la información perfecta. En algunas situaciones la información se sepulta en un conjunto de archivos que se generan todos los días durante la operación del negocio. Para ello es necesario construir un prototipo de producto para resolver las incertidumbres.

Hoy en día, el volumen de información generada durante la operación carece de una explotación oportuna y se pierde información valiosa de los antecedentes de la evolución del negocio de una etapa a otra. Afortunadamente la tecnología permite con un alto grado de eficiencia realizar análisis de información útil para el negocio, a través de bases de datos relacionales o multidimensionales en donde se puede guardar el detalle de la operación para los distintos rubros de influencia en el negocio, como pueden ser la captación, colocación, costos operativos, comisiones y cartera; satisfaciendo así la expectativa de conocimiento de los distintos clientes en función de los productos ofrecidos por el negocio, tendiendo a poder brindar nuevos productos que satisfagan los perfiles de diferentes extractos de clientes, contribuyendo al conocimiento de las necesidades del mercado. Ello promueve el reposicionamiento del negocio a través del perfeccionamiento de sus productos y la venta cruzada de los

misimos, anticipando las necesidades de los clientes y constituyendo una ventaja competitiva en cuanto a servicio.

### Un ejemplo de reposicionamiento:



## 2-4 DECISIONES ESTRATÉGICAS.

Una estrategia es un plan previamente determinado que establece la secuencia de movimientos y contra-movimientos que se realizarán durante una operación o parte de la misma, usualmente se caracterizan por la novedad, complejidad y su amplia amplitud. Con frecuencia implican variables incontrolables como las acciones de los competidores o el estado de la economía. Estas decisiones inciden en toda la organización y tienden a influenciar los planes de largo plazo.

### Establecer Criterios de Decisión.

El establecimiento de estándares para juzgar una evaluación debe ser explícito, así como considerar el orden de prioridad antes que las estrategias de solución sean desarrolladas, ya que están en función de las directrices del negocio y en este punto deberán ofrecer soluciones que impliquen el óptimo desenvolvimiento de la corporación en convivencia con la evolución de la economía y mercado.

### Tipos de Decisiones.

#### Decisiones de Oportunidad.

Una decisión de oportunidad es una acción positiva que se aprovecha de un potencial de crecimiento, de mayores ganancias o

de adición al aceleramiento de algún otro objetivo ya trazado. Que el decisor aproveche las oportunidades será tarea de las herramientas de explotación.

### **Decisiones Operacionales.**

Las decisiones operacionales comprenden las operaciones de detalle de la empresa y van paralelas a los planes de corto plazo. El gerente medio pone en acción las políticas de la compañía al implantar decisiones operacionales que en principio se podrían pensar que son decisiones triviales, pero que en su conjunto constituyen la intermediación entre el cliente y la compañía.

### **Características de las Decisiones No Estructuradas.**

Las características más importantes de una situación no programable o no estructurada son las siguientes:

- *Una extensa base de datos*, es decir, más extensa de lo que los administradores pueden convenientemente tratar sobre una base manual.
- *Altos requerimientos para el manejo de los datos*, los datos tienen que procesarse a través de un conjunto de filtros para aportar señales que posean un significado para el decisor.
- Existe un *juicio* para determinar cuál es el problema fundamental implicado en la situación y para generar las alternativas adecuadas.
- Existen *interrelaciones complejas entre las variables de interés*, lo cual dificulta el detectar los eslabones de causa y efecto entre ellas.
- Existe la *multidimensionalidad*: El problema posee varias dimensiones a lo largo de las cuales el analista puede medir las ejecuciones.
- Existe un gran *beneficio económico*, el cual se deriva de una solución aceptable.
- Existe un ambiente o *contexto dinámico* del problema, el cual cambia rápidamente con el paso del tiempo.

Entonces, el decisor necesita contar con información pertinente, organizada y consistente para explorar las distintas dimensiones de la información a través sus variables de interés con el fin de

analizar adecuadamente el problema. También necesita apoyo computacional, para el análisis de datos. Finalmente, los canales de comunicación que conducen a su área administrativa deben estar libres de ruidos, y de prejuicios personales.

### **Situaciones que se Presentan en una Decisión No Estructurada.**

1. El analista no cuenta con reglas de aplicación frecuente al elegir su curso de acción.
2. Resulta complejo incorporar las especificaciones a priori de un escenario en la implantación de una solución ideal para su problema.
3. El número de criterios de decisión que él hubiese deseado aplicar en el análisis no eran operables.
4. Desconocimiento de alternativas cuando se empezó el análisis de la situación considerada.
5. La respuesta del entorno en la aceptación.

### **Decisiones Estructuradas.**

Se toman generalmente en niveles intermedios dentro de la organización y se caracterizan por ser predecibles en cierta medida, por lo que se pueden desarrollar modelos o programas que preparen el momento de la toma de decisión. El impacto de estas decisiones, que son a corto plazo, afecta las operaciones cotidianas de la empresa. En la mayoría de los casos existe la posibilidad de predecir los eventos, pues se presentan con cierta regularidad, que a pesar de que puede ser variable, las decisiones son similares o equivalentes.

Cuando una situación decisiva se repite con cierta frecuencia, por lo común puede hallarse una solución general, y usarse una y otra vez. Esto le evita al encargado de tomar decisiones el problema de estudiar todos los datos de nuevo para tomar la decisión con respecto a la situación particular conocida.

### **Decisiones de Crisis e Intuitivas**

Una decisión de crisis ya sea al nivel estratégico o al operacional se toma bajo presión. La situación de crisis se caracteriza por la tensión, la sorpresa, el tiempo limitado para responder y la amenaza a las metas de alta prioridad. El factor tiempo obstaculiza

una extensa recopilación y análisis de los datos, o requiere que sean hechos apresuradamente.

Debido a la planeación inadecuada, los gerentes operacionales trabajan usualmente bajo intensas presiones de tiempo y más aún, pueden considerar las decisiones de crisis como algo común más que como una excepción. Las decisiones estratégicas, en particular, sufren cuando se hacen bajo condiciones de crisis. Las decisiones de crisis con frecuencia son *intuitivas*; es decir, son basadas en la experiencia, el análisis instantáneo y un elemento emocional que colectivamente le dan al gerente un sentimiento de que el curso de acción elegido es el correcto. La mayoría de las decisiones operacionales rápidas que los gerentes hacen constantemente son de este tipo. Aunque estas decisiones a veces carecen de objetividad y son difíciles de justificar sobre una base lógica, pueden ser exactas y efectivas.

## **2-5 LA INTEGRACIÓN DE LA INFORMACIÓN PARA LA MEJOR TOMA DE DECISIONES.**

¿Cómo puede una empresa deslumbrar a sus clientes, acrecentar sus dividendos e incrementar sus acciones de mercado, todo al mismo tiempo? Esperamos que la tecnología de la información nos lleve a eso.

Una de las respuestas es el Data Warehouse (almacén de datos). El Data Warehousing (almacenamiento de datos) representa el proceso de reunir información histórica de una organización en un depósito central, y se ha convertido en una tecnología común y fundamental. Algunos analistas han dedicado a la literatura, planteamientos de administración de información en donde plantean que el Data Warehousing es la tendencia más grande dentro de la administración de información en nuestros días.

En el estudio anual de 1995 de los presidentes ejecutivos a fin de identificar los diez aspectos principales de administración para los ejecutivos de sistemas de información, CSE Parterns listó los siguientes, en orden de prioridad:

- 1. Igualar las metas de los sistemas de información y de las corporaciones.**
- 2. Instituir sistemas de información interrelacionados.**
- 3. Organizar y utilizar datos.**

4. Implantar una reingeniería de las actividades.
5. Mejorar la capacidad de los ejecutivos de alto nivel.
6. Habilitar el cambio y el dinamismo.
7. Establecer conexiones con clientes o proveedores.
8. Crear una arquitectura de información.
9. Actualizar los sistemas obsoletos.
10. Mejorar los procesos de desarrollo de sistemas.

**La tecnología del Data Warehouse resuelve los primeros tres elementos.**

Así como hay gran divergencia para establecer una definición precisa de Data Warehouse, existe un claro consenso de que la tecnología del Data Warehouse es un ingrediente esencial en el conjunto de soluciones para el soporte de decisiones en una empresa. Su meta es muy clara:

---

*Obtener información correcta para las personas adecuadas en el momento conveniente para que tomen decisiones.*

---

Son escasas las aplicaciones actuales para un almacenamiento más allá de los soportes de decisión rudimentarios (recursos de acceso de último usuario para acceder información, reportar y unirse a un sistema de soporte de decisiones, o bien, transferencias de información de bases operacionales de datos a bases de datos multidimensionales unidas a herramientas de análisis y explotación de las mismas).

Si bien es cierto, muchos de los problemas fundamentales implícitos todavía no se han solucionado en forma adecuada. Actualmente se presentan los siguientes retos:

- ♦ La integración de datos de varias fuentes.
- ♦ Calidad de la información (limpiar y refinar).
- ♦ Condensación y adición de datos.

- ♦ Sincronización de las fuentes con el Data Warehouse para asegurar una actualización constante conforme se generan nuevos datos de las fuentes.
- ♦ Aspectos del desempeño relacionados con compartir la misma computadora y plataformas.

Entonces, ¿Qué constituye un Data Warehouse completo?, ¿Cómo se anticipa el analista a los retos tecnológicos y al correcto diseño de soluciones dentro de la arquitectura del Data Warehouse? y ¿Cómo se construye un Data Warehouse en forma sistemática?. Todas estas preguntas son válidas y se tratarán en la siguiente parte de la investigación.

**LA INFORMACIÓN Y LA TOMA DE DECISIONES A PARTIR DE UN DATA WAREHOUSE.****Capítulo III****Sistemas de Soporte a la Toma de Decisiones.**

3-1	Evolución de los Sistemas de Información	38	
3-2	Situación Actual de los Sistemas de Información		42
3-3	Definición de Data Warehouse	50	
3-4	Elementos Principales del Data Warehouse		52
3-5	Componentes del Data Warehouse	54	
3-6	Características del Data Warehouse	59	
3-7	Estructura del Data Warehouse	60	
3-8	Valor del Data Warehouse	63	
3-9	Terminología Relacionada con el Data Warehouse		64
3-10	Algunos Esquemas de Diseño para Bases de Datos		71
3-11	OLAP, On Line Analytical Processing	80	

**-CAPITULO TERCERO****SISTEMAS DE SOPORTE  
A LA TOMA DE  
DECISIONES.****3-1 EVOLUCIÓN DE LOS SISTEMAS DE INFORMACIÓN.**

Muchas organizaciones han desarrollado modelos conceptuales, en lo que toca a la evolución de los sistemas de información. Estos planes suelen tomar en cuenta la integración gradual de la organización en los sistemas productores de información. Las alternativas estarán en función de las metas y tecnología que considera el desarrollo de la organización, así el diseño deberá permitir modificaciones que integren circunstancias cambiantes del entorno económico y legal. Lo que inicialmente se necesita para operar en un Sistema Transaccional, posteriormente, se introducen los Sistemas de Apoyo a las Decisiones. Por último, se desarrollan los Sistemas Estratégicos que dan forma a la estructura competitiva de la empresa, este tipo de sistema es alimentado por información generada en los sistemas operacionales. Sobre este aspecto, Richard Nolan, autor y profesor de la Escuela de Negocios de Harvard, desarrolló, en la década de los setenta la *Teoría de las Etapas*, según la cual, la función de la Informática en las organizaciones evoluciona a través de ciertas etapas de crecimiento, a continuación se mencionan las características principales de cada una:

**Primera Etapa - Inicio -**

- ♦ Se adquiere equipo con la justificación primordial del ahorro en mano de obra y exceso de gasto de papeles. Ayudando a automatizar la operación.
- ♦ Las aplicaciones típicas que se implantan son los Sistemas Transaccionales, como nómina, contabilidad, etcétera.
- ♦ El tipo de administración empleado sobre este aspecto es escaso y la función de los sistemas suele ser manejada por un administrador que hasta hace pocos años adquiriría la mayor parte de su conocimiento de la experiencia.
- ♦ El departamento de sistemas depende en la mayoría de los casos de otra área de la organización, ya que el negocio

utiliza a sistemas como una plataforma de apoyo para el propósito de negocio.

### **Segunda Etapa - Contagio o Expansión -**

- Se implantan el resto de los Sistemas Transaccionales no desarrollados en la etapa de inicio, tales como facturación, inventarios, control de pedidos de clientes y proveedores, cheques, etcétera.
- Existe una proliferación de aplicaciones en toda la organización, que debido a la falta de estándares e infraestructura adecuada, se realiza de manera desordenada y sin control.
- Se inicia la contratación de personal especializado (como analistas, programadores, etcétera.). Además, los analistas son asignados a las áreas funcionales de los usuarios, con el fin de agilizar el desarrollo de nuevas aplicaciones.
- Las aplicaciones desarrolladas carecen de interfases automáticas entre ellas, de tal forma que las salidas que produce un sistema se tienen que alimentar en forma manual a otro sistema.
- Las aplicaciones se desarrollan con escasos o nulos estándares de trabajo, lo cual trae como resultado sistemas de muy baja calidad, sistemas que en forma frecuente causan problemas a sus usuarios, el porcentaje de reprocesos de los sistemas crece, el mantenimiento a los sistemas ya desarrollados se dificulta de manera creciente, impactando negativamente en la productividad, no sólo del departamento, sino de la organización en sí.
- Los gastos por concepto de sistemas empiezan a crecer en forma importante, lo que marca la pauta para iniciar la racionalización en el uso de los recursos computacionales dentro de la empresa. Dando origen a la siguiente etapa de evolución.

### **Tercera Etapa - Control o Formalización -**

- Esta etapa se inicia debido a la necesidad de controlar el uso de los recursos computacionales a través de la técnica

presupuesto base cero, es decir, partiendo de que no se tiene nada y la implantación de sistemas con cargos a usuarios, por el servicio que se presta.

- Las aplicaciones están orientadas a facilitar el control de las operaciones del negocio para hacerlas más eficientes, tales como sistemas para control de flujo de fondos, control de órdenes de compra a proveedores, control de inventarios, control y manejo de proyectos, etc.
- El departamento de sistemas de la empresa suele ubicarse en una posición gerencial.
- El tipo de administración utilizado dentro del área de informática se orienta al control administrativo y a la justificación económica de las aplicaciones por desarrollar. Se establecen criterios en las prioridades para el desarrollo de nuevas aplicaciones, por lo que la cartera de aplicaciones pendientes empieza a crecer.
- Se inicia el desarrollo y la implantación de estándares de trabajo dentro del departamento, tales como estándares de documentación, control de proyectos, desarrollo y diseño de sistemas, auditoría de sistemas y programación.
- Se integra a la organización del departamento de sistemas, personal con habilidades administrativas y preparado técnicamente.
- Se inicia el desarrollo de interfaces automáticas entre los diferentes sistemas.
- Nace la función de la planeación de sistemas enfocada hacia el control presupuestal, que incluye la planeación de requerimientos de cómputo y la planeación de adquisición de equipo de computo.

#### **Cuarta Etapa - Integración -**

- La integración de datos adicionales a los sistemas surge como resultado directo de centralizar el departamento de sistemas bajo una sola estructura administrativa, ya que ahora tiene más clientes.

- Las nuevas tecnologías relacionadas con bases de datos, sistemas administradores de bases de datos y lenguajes de cuarta generación, hacen posible la integración.
- Los usuarios inician haciendo sus propias aplicaciones, lo cual propicia que realicen su propio trabajo sin tener que esperar a que sus propuestas de sistemas sean cumplidas.
- El costo del equipo y del software se ve disminuido, por lo que se encuentra al alcance de más usuarios.
- Se modifica el rol del usuario y del departamento de sistemas. El departamento de sistemas evoluciona hacia una estructura descentralizada, permitiendo al usuario utilizar herramientas de explotación más eficientes.
- Los usuarios y el departamento de sistemas inician el desarrollo de nuevos sistemas, reemplazando los sistemas antiguos, en beneficio de la organización.

#### **Quinta Etapa - Administración de Datos -**

- El departamento de sistemas reconoce que la información es un recurso muy valioso que debe estar accesible a usuarios (depende el perfil autorizado).
- Para poder cumplir con lo anterior, resulta necesario administrar los datos en forma apropiada, es decir, almacenarlos y mantenerlos en forma útil para que los usuarios puedan utilizar y compartir este recurso.
- El usuario de la información adquiere la responsabilidad de la integridad de la misma y debe manejar niveles de acceso diferentes.

#### **Sexta Etapa - Madurez -**

- Los sistemas que se desarrollan son Sistemas de Manufactura Integrados por Computadora, Sistemas Basados en el Conocimiento, Sistemas Expertos, Sistemas de Soporte a las Decisiones, Sistemas Estratégicos y, en general, aplicaciones que proporcionan información para las decisiones administrativas y estratégicas.

- ♦ Se tienen aplicaciones desarrolladas con tecnología de bases de datos y se logra la integración de redes de comunicaciones con terminales en lugares remotos, a través del uso de recursos de comunicaciones.
- ♦ Se mejoran los controles anteriormente utilizados, siendo menos rígidos en su aplicación.
- ♦ Se establecen en muchos casos, precios para los servicios de cómputo y en algunos otros se define al área de informática como Centro de Utilidades, en vez de Centro de Costos. Nace la idea de independizar al área de sistemas desde el punto de vista económico y organizacional ("Outsourcing").
- ♦ Existe una planeación más rigurosa de los recursos de cómputo y las aplicaciones con horizontes de planeación no menores a cinco años.

### **3-2 SITUACIÓN ACTUAL DE LOS SISTEMAS DE INFORMACIÓN.**

Tomando como punto de partida en la evolución de los sistemas de información la Teoría de las Etapas, se puede considerar como regla general que las organizaciones siguen a grandes rasgos los mismos patrones de crecimiento, aunque no uniformemente, pues en ocasiones se estancan en algunos aspectos, en otros se encuentran a la vanguardia; algunos problemas son superados y otros más se van rezagando. Con base en esto surge el cuestionamiento en cuanto a la situación que enfrentan en la actualidad las organizaciones en relación al nivel de evolución de sus sistemas de información. A continuación se presenta un panorama general al respecto.

#### **Crisis en la Información.**

La alta demanda de información actual, exacta, integrada y oportuna, no es completamente satisfecha debido a las grandes cantidades de *datos heterogéneos* que almacenan y producen las organizaciones de una manera tan veloz como nunca antes. Las empresas han caído en un ciclo que hace continua la creación de *datos heterogéneos* y las tendencias actuales en tecnología, están fomentando la creación de los mismos.

Existe una crisis de datos que la mayoría de las organizaciones enfrentan actualmente lo que dificultan la urgente necesidad de

integrar datos actuales y exactos para soportar las necesidades cambiantes de un negocio. Este rápido crecimiento de *datos heterogéneos* se debe a la orientación en la visión de las empresas hacia el rápido desarrollo de sistemas de información independientes capaces de atender las necesidades actuales y críticas de un segmento del negocio; sin considerar las necesidades futuras del mismo. Los constantes cambios en las necesidades de las empresas tienen como resultado del cambio constante en el ambiente organizacional.

### **La Demanda de Información.**

La situación empresarial en nuestros días está caracterizada por un ambiente dinámico. Dentro de ese ambiente dinámico, una empresa busca el cambio para ser más competitiva o simplemente por subsistencia. Por esta razón, las organizaciones presentan una alta demanda de información integrada para soportar las actividades corporativas. La única ventaja real competitiva que una organización tiene es el ser capaz de aprender y responder a los cambios más rápido que sus competidores.

En relación a la información, una organización debe ser capaz de conocer qué datos almacena, qué es lo que los datos representan en la realidad y cómo esos datos pueden ser integrados y aprovechados para poder plantear y establecer una verdadera ventaja competitiva. Esto no significa que la habilidad para conocer qué datos existen e integrarlos resuelva todos los problemas que enfrenta actualmente una empresa, pero hace estable una de las variables y permite que una empresa se concentre en otros problemas críticos.

### **Ambiente Dinámico.**

El ambiente organizacional hoy en día se basa en rápidos cambios. La proporción y la magnitud del cambio se incrementan constantemente. El cambio constante parece que permanece y las empresas deben aprender a manejar ese cambio, para volverse competitivas.

El problema es que existe muy poco tiempo para asimilar un cambio cuando llega el siguiente y no se cuenta con el tiempo suficiente para entender, evaluar e implantar correctamente un cambio, antes de que llegue el próximo. El ambiente organizacional obliga a muchas organizaciones a pronosticar posibles cambios,

pero se encuentran con que el tiempo para planear sobre cambios futuros se dificulta, cuando muchos recursos se han gastado ya en ejecutar los cambios actuales.

### **Cambios Organizacionales.**

Las organizaciones están evaluando alternativas y definiendo lineamientos para prepararse ellas mismas ante la situación del cambio, analizando las líneas de negocio que quieren o que son requeridas y determinando qué actividades son necesarias para soportar esas líneas de negocio de manera efectiva y eficiente. Se está cambiando la orientación de las actividades empresariales existentes para agregarles un valor a las líneas de negocio. Cualquier actividad empresarial que no proporcione un valor agregado es mejor cambiarla o eliminarla.

Para poder cambiar alguna actividad empresarial se requiere contar con la habilidad de entender e integrar *datos heterogéneos* internos y externos a la organización, el siguiente paso es su explotación, de tal forma que puedan brindar un soporte adecuado a las necesidades corporativas, Fortalecido en el ajuste a los sistemas de información para eliminar viejas asignaciones de datos y soportar las nuevas actividades organizacionales.

Ahora bien, un mayor problema surge cuando la organización es incapaz de ajustar sus sistemas de información porque se encuentran altamente ligados a bases de datos operacionales.

Cuando un sistema de información pierde eficiencia, las empresas en primera instancia buscan la manera de cambiar ese sistema de información, pero es necesario profundizar en el verdadero origen del problema, que bien puede ser la fuente que alimenta ese sistema de información: *los datos*.

### **Demanda de Información Corporativa.**

Las necesidades de información cambian también rápidamente. Más información es necesaria para soportar las actividades de una corporación y no sólo eso, sino que se necesita en un lapso de tiempo muy corto, además esta información debe ser actual y exacta, ya que, como se he comentado con anterioridad, la supervivencia de muchas empresas depende de contar con mejor información que sus competidores.

Esta demanda se basa en que los recursos de información deben satisfacer las necesidades críticas de la organización, lo cual sólo se puede lograr con información integrada, que pueda ser compartida y que contenga datos actuales y exactos. Una organización exitosa es aquella que es capaz de encontrar la información que necesita a través de la comprensión de los datos y usarlos como ventaja competitiva, así como aprender a utilizar los *datos heterogéneos* con el fin alcanzar una posición relevante en el desarrollo de su entorno.

### **Situación de los Datos.**

Los recursos destinados a proveer información no son capaces de satisfacer la expectativa, principalmente porque contienen grandes cantidades de *datos heterogéneos* y la proporción de los mismos se está incrementando rápidamente. Muchos sistemas de información están siendo designados y desarrollados para satisfacer las necesidades actuales, en vez de que puedan proveer flexibilidad para el futuro.

### **Datos Heterogéneos.**

Los *datos heterogéneos* se producen cuando existen diferentes fuentes que producen un mismo dato con significado distinto en cuanto a: tipo, calidad y significado. Son desiguales y no pueden ser fácilmente integrados para satisfacer adecuadamente las necesidades de información del negocio.

Los *datos heterogéneos* no se encuentran identificados, nombrados, definidos, estructurados y mantenidos por reglas uniformes. Se encuentran inconsistentemente nombrados, inadecuadamente estructurados y no están totalmente documentados. Su significado, contenido y formato es variable, por lo que su comprensión es difusa y su significado puede variar a través del tiempo.

Los *datos heterogéneos* se encuentran fragmentados en diferentes localidades, en diferentes bases de datos, con diferentes estructuras. Se encuentran almacenados de manera redundante.

### **Ciclo de Datos Heterogéneos.**

La disponibilidad de herramientas CASE, herramientas para desarrollo de aplicaciones, generadores de código, bases de datos y muchos otros productos, promueven la capacidad para crear datos

heterogéneos, es una tendencia que empezó con los archivos planos, procedió con los sistemas manejadores de bases de datos y continúa con la tecnología actual.

La producción de datos heterogéneos es el resultado de un ciclo, en el cual estos datos son continuamente producidos. Las personas no saben sobre los datos que almacenan o no quieren utilizar los datos existentes. A continuación se mencionan algunas razones que dan lugar a esta situación:

- No se conoce qué datos existen y son duplicados para su uso.
- Se conoce qué datos existen, pero son pobres en su forma o requieren mucho tiempo para su comprensión, por lo que no pueden ser fácilmente utilizados, así que nuevos datos son generados.
- Sólo se quiere controlar datos de manera individual y manipularlos, sin considerar el negocio como un todo.

Todas estas situaciones generan datos heterogéneos. Este ciclo avanza y por lo general las organizaciones no se percatan de que existen.

### **El Dilema de los Datos.**

Muy poco se está haciendo con respecto al problema de los datos heterogéneos en las organizaciones, ya que se encuentran concentrando esfuerzos en actividades como:

- Procesamiento de aplicaciones y desarrollo de comunicaciones para acceder y transportar datos.
- Cambiar el negocio y el desarrollo de aplicaciones: Redes de trabajo cliente/servidor, sistemas Data Warehouse, sistemas de información geográficos, sistemas de imágenes y una gran variedad de tecnologías están siendo implantadas.

Pero las corporaciones no se están concentrando en controlar sus datos heterogéneos en pro de nuevas técnicas y herramientas para la administración de datos.

El dilema de los datos se expresa en la necesidad de datos actuales, exactos y oportunos, para satisfacer la demanda de información corporativa, lo que propicia el desarrollo de datos heterogéneos. Esta situación parece requerir una elección entre satisfacer las demandas de información actual corporativa a expensas de la flexibilidad para el futuro o asegurar la flexibilidad para el futuro a expensas de satisfacer la demanda de información corporativa actual. Esta es una situación que parece desafiar cualquier solución simple.

---

**Dos tendencias están actualmente emergiendo:**

1. Una necesidad por entender e integrar los datos para soportar rápidamente las necesidades cambiantes del negocio: Las organizaciones se están poniendo alerta sobre su situación de datos heterogéneos y están empezando a planear sobre la integración de sus datos.
  2. Con el incremento en la demanda de información y la disminución en los recursos, más personas están creando datos locales más rápidamente para soportar las actividades actuales organizacionales. Como resultado, tenemos una rápida creación de datos heterogéneos.
- 

**Tendencias Tecnológicas.**

Cada una de ellas promete enormes beneficios, si se implanta adecuadamente, si no se instala apropiadamente o no se cuenta con una fuente de información integrada, es más probable que se genere un desastre mayor al problema que resulta de la existencia de datos heterogéneos.

**Arquitectura Cliente/Servidor.**

Esta arquitectura ofrece una estrategia para la distribución de datos para soportar las necesidades corporativas de información. Provee la oportunidad de compartir datos dentro y entre las organizaciones.

Existen dos problemas sobre los datos en la tecnología cliente/servidor:

1. Esta tecnología aparece para resolver el problema de datos heterogéneos brindando un acceso transparente a una variedad de diferentes bases de datos. Pero el hecho es que proveer acceso a diferentes bases de datos no resuelve el problema de los datos heterogéneos, porque no integra los datos en una sola fuente.
2. Su implantación generalmente crea mayor cantidad de datos heterogéneos y hace continuo este problema. Las bases de datos son diseñadas de manera independiente para satisfacer necesidades muy específicas y frecuentemente se van modificando, según las necesidades vayan variando.

Un modelo cliente/servidor implica el contar con una fuente de información integrada, donde los datos son integrados y compartidos a través de actividades y límites organizacionales. Implica forzar la integración de los datos para satisfacer la demanda informativa dentro de las empresas.

### **Sistemas Data Warehouse.**

Se están volviendo cada vez más populares para la administración de datos que varía a través del tiempo en vez de analizar tendencias de negocio y hacer proyecciones. Un Data Warehouse provee la oportunidad de brindar grandes cantidades de datos juntos, así mismo tiene la habilidad de resumir datos a altos niveles de explotación, así como búsqueda en niveles de mucho detalle. Provee un soporte primario para procesamiento analítico en línea (OLAP), sistemas de soporte a la toma de decisiones y sistemas de información ejecutiva.

Un Data Warehouse ofrece grandes oportunidades para una organización que quiera volverse competitiva, o seguir siéndolo. Los datos ubicados en el Data Warehouse deben ser consistentes y la organización debe utilizar técnicas formales para la administración y documentación del resumen de los datos. Una empresa que construya un Data Warehouse con datos heterogéneos está limitando su negocio basándose en el análisis de datos heterogéneos. Los sistemas de Data Warehouse constituyen una fuerte tecnología para ayudar a entender el negocio, ya que se convierte en una fuente única de datos, integrando al negocio.

**Otras Tendencias.**

1. En los últimos años hemos observado tendencias sobre manejo de datos ya sea a través de voz, texto e imágenes. Este tipo de información se está volviendo muy popular y el volumen se está incrementando de manera alarmante. Todo esto contribuye sustancialmente al problema de datos heterogéneos, a menos que sean administrados como parte de una fuente de datos integrada.
2. La programación orientada a objetos es una tecnología donde tanto los datos como los procesos son encapsulados. Esta tecnología promete ofrecer que el desarrollo de sistemas de información sea fácilmente ajustable para apoyar las necesidades de información del negocio. Su principal problemática es que puede dificultarse el determinar cuáles datos heterogéneos serán encapsulados con los procesos y qué es lo que los datos significan. El resultado es a menudo la definición de nuevos datos independientes de los datos existentes.
3. Sistemas de voz, imagen o texto, multimedia, sistemas de intercambio electrónico de datos, sistemas de información ejecutiva, sistemas expertos, redes neuronales y una gran variedad de otras tecnologías constituyen alternativas de explotación de información, pero deben de estar basadas en información bien definida, todas estas tecnologías, serán más exitosas si se basan en una fuente de información integrada. Las limitaciones de estas tecnologías y de los datos pueden ser compartidos y mezclados entre diversas tecnologías para satisfacer correctamente las necesidades de información del negocio.

---

Las organizaciones presentan actualmente el riesgo de que se desborde el nivel de datos respecto de su capacidad real de manejo, debido a esta acumulación constante, la cual esta siendo superior al ritmo usual de generación. Lo más crítico de esta situación es que muy pocas organizaciones están haciendo algo al respecto.

Estos datos crean un dilema real para las organizaciones, debido a que los datos integrados son necesarios para la competencia, aún cuando la continua producción de datos heterogéneos es a menudo ignorada en favor del desarrollo de sistemas de información y/o

adquisición de hardware. Los desarrollos de aplicaciones cliente/servidor, construcción de sistemas Data Warehouse y la distribución de datos a través de redes sin un control sobre los datos, da como resultado muchos problemas de magnitudes mucho mayores a los que se tenían con anterioridad. La fragmentación de los datos y la pérdida de datos históricos y clase de los datos pueden resultar desastrosos para una organización. La verdadera crisis sucede cuando las organizaciones se dan cuenta de que sus recursos informativos no son capaces de soportar las necesidades dinámicas del negocio.

Las empresas deben reconocer que existe y tienen un caos de datos, es necesario que se den cuenta de que están perdiendo el control de sus recursos de información y tomar acciones correctivas para detener el desarrollo de datos heterogéneos y construir un recurso de datos integrados junto con una arquitectura común de información. Es necesario administrar adecuadamente los datos, antes de que los datos los administren. Las corporaciones deben desarrollar un recurso común integrado de información que sea compatible con las necesidades de información tanto actuales como futuras, además de proveer un antecedente histórico del negocio.

Las necesidades empresariales de información van encaminadas hacia la creación de estructuras de información transparentes para el usuario, que funcionen automáticamente con cualquier aplicación que el usuario elija para analizar, distribuir y procesar la información a su disposición. Esta búsqueda de soluciones plantean un nuevo reto, pues la información ya no fluye sólo a través de los canales previstos, donde puede ser fácilmente indexada, recuperada o consultada. El nuevo modelo de empresa que permite que la información sea distribuida por encima de las divisiones tradicionales, exige la integración de distintas fuentes de información y aplicaciones.

### **3-3 DEFINICIÓN DE DATA WAREHOUSE.**

*“ Conjunto de datos integrados orientados, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración ”.*

W. H. Inmon.

Un Data Warehouse es un depósito consistente en datos – separados y que no interfieren con los sistemas operacionales existentes– que llenan por completo los diferentes requerimientos de acceso y reporte de datos.

Un Data Warehouse es una colección de datos críticos de información corporativa designados para proveer una fuente validada y única de información corporativa. Los datos que contiene el Data Warehouse son usualmente históricos, integrados y surgen tanto de fuentes internas como externas. Los datos están físicamente separados de los datos operacionales.

Un Data Warehouse es un proceso continuo que tiene la capacidad de proveer informes de datos de varias fuentes, incluyendo datos históricos y adquiridos para soportar la constante necesidad de consultas estructuradas, reportes para el análisis y soporte de toma de decisiones.

Haciendo una conjunción entre varias definiciones, tenemos que un Data Warehouse:

- Es un conjunto de datos integrados históricos y actuales generados en la operación del negocio, es decir contiene el detalle de la operación en distintos grados de granularidad.
- Estos datos no interfieren con los sistemas operacionales o transaccionales más bien se alimenta de ellos pero con la ventaja que los datos son validados por las áreas que los generan.
- Está designado para proveer información validada y de una fuente corporativa de datos para al soporte de la toma de decisiones, consultas estructuradas y reportes para análisis, con el fin de satisfacer las necesidades de información en una organización.
- Considera el detalle de las diversas unidades de negocio.
- Es un proceso continuo que consta básicamente de incorporación de datos y de los cambios que éstos sufran.

---

**Un Data Warehouse está compuesto por dos componentes principales:**

- a) Un almacén de información de eventos históricos y

- b) las herramientas que complementan el análisis estratégico de esa información (un sistema de soporte para la toma de decisiones).

---

Los fundamentos en la arquitectura de un Data Warehouse son mucho más importantes que el mencionar alguna herramienta en específico. Existen actualmente gran cantidad de herramientas en el mercado, las cuales tienen un ciclo de vida corto dentro del mismo. La característica central que debe que debe satisfacer un Data Warehouse es que tiene que ser lo suficientemente flexible, tanto para incorporar los datos y los cambios que éstos sufran, como para satisfacer las necesidades de análisis de información dentro de la organización a través del tiempo.

### **3-4 ELEMENTOS PRINCIPALES DEL DATA WAREHOUSE.**

#### **Administra Grandes Cantidades de Información.**

La mayoría de los Data Warehouses contiene información histórica que se retira con frecuencia de los sistemas transaccionales porque ya no es necesaria para las aplicaciones de la organización. Por el volumen de información, se debe contar con varias opciones para la adición y condensación que clasifican los datos. El Data Warehouse maneja información a diferentes niveles de detalle (llamado también, granulado).

#### **Almacenamiento de INFORMACIÓN.**

Por razones las más de la veces de volumen de información se utilizan diferentes medios de almacenamiento lo que permite, migrar información histórica y conservarla en medios más baratos como cintas en vez de mantenerla discos.

### **COMPRENDE MÚLTIPLES VERSIONES DE UN ESQUEMA DE BASE DE DATOS.**

Muchas veces la información a través del tiempo ha sido manejada en distintos momentos por diferentes versiones de esquemas de bases de datos. En estos casos, el Data Warehouse debe controlar la información que ha sido originada a lo largo de este periodo.

### **Condensa y Agrega Información.**

A diferencia de las bases de datos operacionales donde es muy alto el nivel de detalle, el Data Warehouse condensa y agrega la información para retroceder y entender la imagen global del negocio.

### **INTEGRA Y ASOCIA INFORMACIÓN DE MUCHAS FUENTES DE INFORMACIÓN.**

Esto es, la recopilación y organización en un solo lugar de la información que se ha ido acumulando con el paso de los años a través de las diversas aplicaciones utilizadas dentro de una organización.

---

Un Data Warehouse contiene la colección de información basada en computadora que es crítica para la ejecución exitosa de un negocio. Es mucho más que un archivo con datos corporativos y más que una nueva manera de acceder a la información corporativa. Provee herramientas para satisfacer las necesidades informativas de los administradores de empresas en todos los niveles organizacionales.

Un Data Warehouse debe ser diseñado de tal manera que sus usuarios puedan reconocer la información y este disponible mediante herramientas sencillas.

Una de las principales razones para desarrollar un Data Warehouse es poder integrar la información operativa a partir de varias fuentes en una arquitectura consistente capaz de soportar el análisis y la toma de decisiones. Los sistemas operacionales generan, actualizan y borran información en producción, el resultado final es lo que alimenta al Data Warehouse.

Entonces un Data Warehouse es generalmente una mezcla de tecnologías, la cual comprende bases de datos relacionales, multidimensionales, arquitectura cliente/servidor, programas de extracción y transformación, interfaces gráficas y más.

### 3-5 COMPONENTES DEL DATA WAREHOUSE.

#### **Resumen de los Datos.**

El resumen ligero de datos es el umbral de un Data Warehouse. Todos los elementos de una organización (áreas, departamentos, regiones, funciones, etc.) no tienen los mismos requerimientos de información. De esta manera, un diseño efectivo de un Data Warehouse provee, con el fin de estandarizar, información ligeramente resumida para cada elemento de la organización.

Los datos muy resumidos son primordialmente utilizados por los ejecutivos de la empresa. Estos datos pueden provenir de los resúmenes ligeros de datos utilizados por los elementos de la organización. El volumen de datos en este nivel es mucho menor que en otros niveles y representa una gran variedad de necesidades e intereses. Además para acceder los datos altamente resumidos, los ejecutivos también tienen la capacidad de ir accediendo diferentes niveles que vayan incrementando el nivel de detalle de la información.

#### **Detalle.**

El corazón del Data Warehouse es su detalle actual o detalle de la operatividad de la organización. El detalle proviene directamente de los sistemas operacionales y pueden ser almacenados como datos en bruto o como agregaciones de datos. El detalle operativo organizado por área, representa a toda la empresa.

El detalle de la operación es el nivel más bajo de granularidad en el Data Warehouse. El detalle operativo es generalmente de dos a cinco años de antigüedad y su actualización ocurre tan frecuentemente como sea necesario para soportar los requerimientos corporativos.

#### **Sistema de Registro.**

El sistema de registro es la fuente de la información que alimenta el Data Warehouse. La información en el Data Warehouse difiere de la información de los sistemas operacionales en que éstos únicamente pueden ser leídos, no modificados. Es por eso que el Data Warehouse debe asegurar mejor calidad en los datos disponibles. El Data Warehouse debe contar con información

oportuna, precisa y completa, además debe estar bien estructurada.

### **Programas de Integración y Transformación.**

Aún ante la más alta calidad en la información operacional, no puede ser completamente copiada como tal al Data Warehouse. Los datos operativos son muchas veces incomprensibles para la mayoría de los usuarios. Además, los datos operativos raramente están conformados por una lógica y estructura adecuadas para un Data Warehouse.

Mientras los datos operativos son trasladados de sistemas de registro al Data Warehouse, los programas de integración y transformación convierten los datos de una aplicación específica en datos corporativos. Estos programas ejecutan funciones tales como:

- a) Reformatear, volver a realizar cálculos o modificar estructuras.
- b) Adicionar el efecto del tiempo.
- c) Identificar valores de fallas.
- d) Suplir la lógica por la opción de escoger entre varias fuentes de datos.
- e) Resumir, cuadrar y cruzar información a partir de múltiples fuentes.

Conforme cambian tanto el ambiente operacional, como el Data Warehouse, los programas de integración y transformación deben ser modificados con el fin de que sean capaces de reflejar ese cambio.

### **Archivos.**

Los archivos del Data Warehouse contienen información histórica significativa que es de interés y valor para toda la organización – generalmente dos años de anterioridad-. Esto significa que usualmente los archivos del Data Warehouse almacenan una cantidad masiva de información. Estos archivos generalmente son utilizados para el planteamiento de proyecciones y el análisis de estrategias. Los archivos no sólo deben incluir información histórica (tanto en detalle como a niveles más generales –

información resumida-), sino que también deben incluir metadatos los cuales describen las características de los datos antiguos, de los cuales hablaremos a continuación.

### **Metadatos.**

Son un aspecto muy importante dentro del Data Warehouse, describen las características de los datos contenidos. Generalmente, los metadatos –o datos acerca de los datos– describen la estructura del Data Warehouse desde la perspectiva de los usuarios y programadores definiendo el sistema operacional o fuentes externas que suplen los datos, describen cualquier regla de transformación utilizada en los datos, son una historia de cuándo los datos fueron actualizados y la estructura de la base de datos del Data Warehouse.

---

Los metadatos son utilizados para responder los siguientes tipos de preguntas:

*¿Cuál es la información que estoy obteniendo? ¿De dónde viene?  
¿Cuándo fue actualizada? ¿Qué significa? ¿Cuál es su descripción corporativa?  
¿En dónde encuentro la información que necesito?*

---

La información que se encuentra en los metadatos permite a los usuarios determinar la utilidad de los datos contenidos en el Data Warehouse.

### **Por ejemplo**

---

Un usuario puede entender mejor el efecto de la compensación en la productividad de los empleados en los últimos cinco años. El analista puede también revisar las reglas de cómo la compensación y la productividad fueron determinadas para asegurar que los datos almacenados en el Data Warehouse son consistentes y exactos a través de los cinco años.

---

Conforme va incrementándose el volumen y la complejidad de los datos, la administración de los metadatos se va convirtiendo en un aspecto de mayor importancia. Los metadatos deben estar

disponibles y con la habilidad de localizar y entender los datos de acuerdo a las necesidades de información de una corporación.

El objetivo principal de los metadatos es que los analistas conozcan lo mejor posible todo lo referente a los datos reales que poseen para que puedan utilizarlos para satisfacer la demanda de información corporativa. La demanda de metadatos es también una necesidad de las corporaciones para completar datos exactos sobre los datos que almacenan.

Para satisfacer la demanda en los metadatos, es necesario que sean administrados como parte de una fuente de información integrada, tal y como se administran los datos financieros, datos de personal o datos de los clientes.

#### **Elementos de los Metadatos.**

- Descripción y ubicación de servidores, bases de datos, tablas, nombres y resúmenes del Data Warehouse.
- Reglas para las consultas al detalle o a resúmenes a través de jerarquías de dimensiones.
- Nombres elegidos o alias definidos por el usuario final para los encabezados y hechos de datos con nombres más técnicos.
- Reglas para cálculos personalizados definidos por el usuario final.
- Seguridad a nivel personal, de grupo de trabajo y de empresa, para visualizar, cambiar y distribuir resúmenes adaptados, cálculos y otros análisis de usuario final.
- Descripciones de fuentes originales y transformaciones.
- Definiciones lógicas de tablas y atributos del Data Warehouse.
- Definiciones físicas de tablas y columnas, así como de sus características.
- Ubicación integrada de las tablas del Data Warehouse.
- Antecedentes de extracción.
- Información de alias.
- Ubicación de área tema.
- Antecedentes de relaciones.

- Propiedad/Gerencia.
- Tablas de referencia y datos codificados.
- Criterios de envejecimiento y purga.
- Indicador de calidad de datos.
- Seguridad.
- Unidades de medida.

Los metadatos pueden también contener componentes de ubicación para auxiliar en las siguientes tareas:

- Identificación de fuentes operacionales.
- Ubicación sencilla de atributo a atributo.
- Conversiones de atributos.
- Conversiones de características físicas.
- Conversiones de codificación y tabla de referencia.
- Cambios de nombre.
- Cambios de llave.
- Valores predeterminados que se utilizan.
- Razón predeterminada.
- Lógica para elegir entre varias fuentes operacionales.
- Formulación algorítmica empleada.

Como ejemplo de un elemento de metadatos de atributos. Los atributos contienen información sobre las características físicas necesarias para conversiones.

Nombre de atributo:	Nombre del cliente.
Nombre alterno:	Nombre de cuenta, nombre de cliente.
Definición:	Un cliente es el nombre oficial que usa la persona o empresa cliente.
Fecha de creación:	30 marzo 1995.
Fecha de última actualización:	5 abril 1997.
Indicador de llave:	(sin llave).
Fuente de datos:	Sistema de ingreso de pedidos, el atributo archivo de Cliente es NuCta.
Información de transformación/conversión:	De 20 a 35 caracteres.
Algoritmo de resumen/derivación:	Ninguno.
Valor predeterminado que se usa:	Ninguno.
Fuentes operacionales múltiples:	No.

### **3-6 CARACTERÍSTICAS DEL DATA WAREHOUSE.**

Algunas características que distinguen a los Data Warehouses de otras fuentes de información incluyen las siguientes:

- ◆ Orientación a áreas o temas.
- ◆ Uso de información histórica para el soporte de la toma de decisiones.
- ◆ Información integrada.
- ◆ Función elemental como fuente de solo lectura con propósitos analíticos.

#### **Historia.**

Un Data Warehouse mantiene un registro completo de los eventos y condiciones del negocio a través del tiempo, a diferencia de las bases de datos operacionales, las cuales representan únicamente el estado actual del negocio. Un Data Warehouse contendrá, por ejemplo, todos los registros de ventas de un periodo de algunos años, no solo aquellas ventas que todavía no son completadas. Conservar la historia permite al analista desarrollar estrategias y análisis estadísticos a partir de una basta muestra de datos.

#### **Detalle.**

Información comprensible y al mínimo grado de generación por la operación está disponible en un Data Warehouse, aunque también es una propiedad de las bases de datos operacionales, la diferencia reside en la práctica histórica de ejecutar todo el análisis del negocio en datos agregados. La adición de datos juega un papel importante en el diseño de un Data Warehouse, el objetivo es guardar tanto detalle como lo permita la tecnología.

#### **Organización por Área o Tema, no por Proceso.**

Las bases de datos operacionales están organizadas para maximizar el desempeño de las transacciones que soportan un proceso particular del negocio. Esta separación organizacional genera dos efectos interesantes, uno es que proporciona cierta independencia al área en su operación y el otro es que fragmenta la

organización. El Data Warehouse esta diseñado para recoger esta información y hacerla disponible al resto de la organización.

**Ejemplo:**

Un asegurador típico tiene bases de datos operacionales clasificadas por vida, gastos médicos mayores, automotriz, etc. las cuales representan las clases de transacciones que son soportadas. Ahora bien, el asegurador con un Data Warehouse, podrá concentrar la información por primas, reclamaciones, clientes, productos y periodos de tiempo. Todos listo para ser analizado.

**La Mayoría de los Accesos son a la Medida (i.e. Ad hoc).**

En una base de datos tradicional, cada tabla tiene definido accesos a la información generalmente a través de campos únicos como número de cliente o Curp (clave única de registro poblacional), etc. Llamados índices, lo cual determina el acceso a la información, es decir toda consulta que se realice a la tabla será resuelta en basa a la información contenida en los índices (como modo de acceso) lo que los hace hasta cierto punto predecibles. Cada tipo de acceso utiliza un campo índice para consultar un registro en particular o un grupo de registros.

**3-7 ESTRUCTURA DEL DATA WAREHOUSE.****Data Warehouse Físico**

Base de datos física en donde es almacenada toda la información para el Data Warehouse, junto con los metadatos y el procesamiento lógico para tallar, organizar, empaclar y procesar el detalle de los datos.

**Data Warehouse Lógico**

También contiene metadatos, incluye reglas de la organización y procesamiento lógico para manipular los datos, así como la dirección lógica de almacenamiento para acceder a la información en la localidad de memoria en que resida.

## **Data Mart.**

### Mercado de Datos.

Un Data Mart es muy similar a un Data Warehouse en su estructura, es como una vista departamental o funcional de una empresa, cuyos contenidos provienen de información centralizada de la corporación a partir de un Data Warehouse. Un Data Mart puede ser caracterizado como un Data Warehouse por departamento.

Por lo general, el Data Mart es utilizado únicamente cuando el acceso de datos o requerimientos de uso para una función o departamento son considerablemente diferentes a aquellos que son provistos por el Data Warehouse corporativo.

Un Data Mart es como un subconjunto de un Data Warehouse el cual generalmente soporta solamente un elemento corporativo (un departamento, una región, una función, etc.).

Existen muchos factores que van dando pie al incremento en la popularidad de los mercados de datos. Mientras el Data Warehouse no contiene mucha carga de información, puede satisfacer las necesidades de diferentes departamentos como una base para el proceso de sistemas de soporte para la toma de decisiones, pero los Data Warehouses no permanecen tan pequeños mucho tiempo, lo que motiva a la construcción de mercado de datos.

### El Crecimiento Rápido de un Data Warehouse.

La competencia por adentrarse a un Data Warehouse crece con fuerza. Cada vez más procesos de soporte para la toma de decisiones departamentales se hacen dentro del Data Warehouse, hasta el punto en que el consumo de recursos se vuelve un verdadero problema.

El costo de hacer el procesamiento en un Data Warehouse se incrementa tanto como el volumen de los datos aumente.

Los analistas de sistemas de soporte a las decisiones departamentales encuentran a los mercados de datos más accesibles. Los Data Mart's se vuelven como una extensión natural de un Data Warehouse y son atractivos principalmente por las siguientes razones:

- Cuando un departamento tiene su propio mercado de datos, puede adecuar (personalizar) los datos tal y como fluyen dentro del mismo. No hay necesidad de tener que servir a toda la corporación. El departamento puede resumir, ordenar, seleccionar y estructurar sus propios datos departamentales sin tomar en cuenta a cualquier otro departamento o área.
- El monto de los datos históricos están referidos al departamento, no a la corporación. En casi todos los casos, el departamento puede seleccionar un grupo mucho más pequeño de datos históricos que aquél que es almacenado en el Data Warehouse.
- El departamento puede llevar a cabo cualquier proceso de soporte para la toma de decisiones que quiera y cuando quiera, sin tener consideraciones sobre el impacto para el uso de recursos sobre otros departamentos.
- El departamento puede seleccionar software según lo desee. Existe una riqueza de software de accesos y análisis en el nivel de procesador que cubre el Data Mart.
- El costo unitario de procesamiento y almacenamiento es apropiado y es significativamente menor que el costo del procesamiento y almacenamiento que se requiere para un Data Warehouse.

Existen muchas razones del porqué un Data Mart se vuelve más atractivo en la medida en que el Data Warehouse crece en volumen. También existen razones organizacionales, tecnológicas y económicas del porqué el mercado de datos es tan seductor y es considerado como una consecuencia natural del Data Warehouse.

**Contenido de un Data Mart.**

Un Data Mart contiene información departamental para soporte a la toma de decisiones. El Data Mart puede contener una gran diversidad de datos del departamento.

El Data Mart contiene información tanto en detalle como en resumen, además contiene datos preparados y datos *ad-hoc*.

**Data Warehouse Estratégico.**

Es el resultado de información corporativa basada en los requerimientos de negocio y en una arquitectura común de datos.

**3-8 VALOR DEL DATA WAREHOUSE.**

- ♦ **Convierte los datos en una herramienta competitiva:** Identifica dónde el dinero se gasta innecesariamente, busca oportunidades perdidas, identifica recursos gastados, ayuda a reducir el número de clientes insatisfechos.
- ♦ **Bases de datos:** Identifica nuevos mercados con base en la historia de ventas exitosas a través de mercados meta (no se limita a realizar búsquedas por código postal o estado).
- ♦ **Servicio al cliente.**
- ♦ **Producto de análisis.**
- ♦ **Orientación a una materia:** Permite crear un perfil de cada cliente, segmento de los mismos.

**Por ejemplo**

---

En la industria de seguros, la orientación a una materia significa revisar cada segmento del negocio como: automóviles, casa habitación, vida e información sobre salud, pero si se analizara por cliente se podría determinar que productos tiene contratados un cliente con la compañía, esto servirá de ayuda para promover ventas cruzadas e identificar

posibles áreas que no proveen el rendimiento adecuado a las organizaciones, productos que no se están ofreciendo etc.

---

### **3-9 TERMINOLOGÍA RELACIONADA CON EL DATA WAREHOUSE.**

#### **Data Warehousing.**

Data Warehousing es un proceso que reúne y organiza información para su uso en el proceso de toma de decisiones. El resultado de este proceso es un Data Warehouse que contiene información perteneciente a una o más áreas de negocio de interés para el usuario final. Es importante mencionar que varias metodologías, arquitecturas y tecnologías son utilizadas durante este proceso.

El propósito elemental del data Warehousing es el habilitar aplicaciones que soporten la toma de decisiones y fomentar el conocimiento de las actividades del negocio. El data Warehousing puede ser caracterizado más como una arquitectura que como una tecnología específica. Un rango amplio de tecnologías, tales como procesamiento en paralelo, bases de datos multidimensionales y sistemas manejadores de bases de datos relacionales son útiles como bloques para construcción durante la construcción de una arquitectura Data Warehousing.

#### **Almacén de Datos Operacionales.**

Su nombre en inglés, Operational Data Store (ODS). Es una colección de datos integrados de la empresa designados a soportar actividades de negocio operativas y tácticas, actividades de día a día. Un almacén de datos operacionales sirve como una confiable fuente actual de datos que pueden ser compartidos a través de aplicaciones y funciones corporativas.

Aunque un almacén de datos operacionales está muy relacionado con el Data Warehouse, existen ciertas diferencias. Su objetivo, es la principal diferencia, ya que, mientras el propósito del Data Warehouse es el soporte para la toma de decisiones, el propósito del almacén de datos operacionales es habilitar la integración de los sistemas operacionales. Para soportar los sistemas operacionales de una manera efectiva, un almacén de datos operacionales necesita cerca los datos actuales. Debido a que las tendencias y el análisis basado en el tiempo no son los intereses

del almacén de datos operacionales, no necesita tanta información histórica como el Data Warehouse. El almacén de datos operacionales y el Data Warehouse, son también orientados a una materia e integrados y puede apoyar a una aplicación de Data Warehouse.

### **Data Warehouse Virtual.**

Un Data Warehouse virtual accesa datos operacionales en su forma y ubicación originales. Los Data Warehouses virtuales son vendidos como una viable, rápida y poco costosa alternativa para implantar un Data Warehouse. De cualquier manera, existen varias limitaciones y desventajas por lo que los Data Warehouses virtuales no deberían ser considerados como una alternativa eficiente para la implantación de un Data Warehouse.

### **Procesamiento Informático.**

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

---

*¿Están las ventas de artículos deportivos sujetas a tendencias por temporada? ¿Cuáles elementos se afectan y cuando?*

*¿Cuántos automóviles se vendieron el mes pasado? ¿Cómo se compara esta cantidad con el mismo mes en los últimos cinco años?*

*¿Cuáles son los artículos de menor rentabilidad en el catálogo de ventas?*

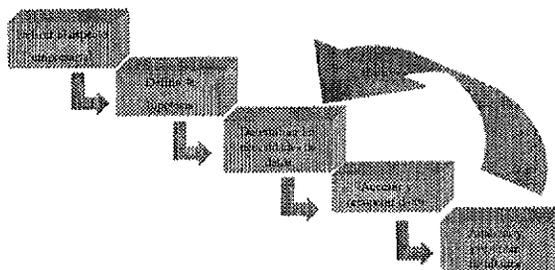
---

El procesamiento informático consta de tres distintos componentes:

1. Consultas para acceder y recuperar datos del Data Warehouse.
  2. Análisis de los datos.
  3. Presentación del análisis en forma de reportes, hojas tabulares y matrices, tablas y gráficas que van de sencillas a complejas.
- 

El procesamiento informático explota variables de la información en todas sus dimensiones.

**Pasos que Comprende el Procesamiento Informático.**

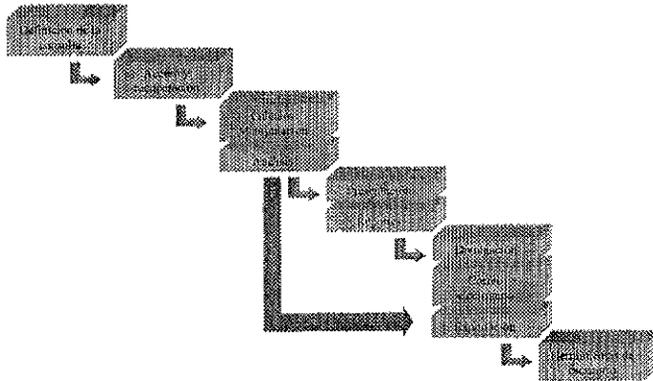


- 1) El usuario empresarial debe definir el problema a abordar.
- 2) Deben definirse las hipótesis. Es importante expresar el problema empresarial como una hipótesis, la respuesta será contestada en función de la información del Data Warehouse. Esta transformación requiere de habilidad empresarial, experiencia e ingenio en el análisis.
- 3) Para verificar esta hipótesis se debe acceder y analizar diversos datos. El usuario empresarial determina si se cuenta con los datos apropiados y si son suficientes para responder.
- 4) Se emplean herramientas de procesamiento informático para acceder al Data Warehouse y recuperar los datos.
- 5) Se analizan los datos y se presentan resultados como recomendaciones factibles.

En ocasiones se requiere de un análisis iterativo, en este caso, el usuario empresarial vuelve a definir la hipótesis original o accesa datos adicionales para verificar dicha hipótesis. Después, se realiza un ciclo de tres pasos de análisis de decisión:

- |    |                                 |
|----|---------------------------------|
| a) | necesidades de datos,           |
| b) | acceso y recuperación de datos, |
| c) | análisis de resultados,         |

y se repite las veces que sea necesario.  
Procesamiento informático:



#### Definición de la Consulta.

Una hipótesis empresarial se debe traducir a una consulta de cómputo para proponerla al Data Warehouse. Esta traducción la lleva a cabo un analista que combine las habilidades y conocimiento de sistemas y negocio. El reto consiste en traducir los términos empresariales en términos de lenguaje de consulta informática.

#### Acceso y Recuperación de los Datos.

Las herramientas de acceso proponen la consulta al Data Warehouse y recuperan los datos adecuados. El proceso de acceso y recuperación puede incluir la capacidad de realizar cálculos, como clasificar resultados o crear subtotaes por artículo.

#### Cálculo, Manipulación y Análisis de los Datos.

El usuario empresarial puede realizar mayores cálculos y manipulación sobre los datos reunidos por la consulta. El propósito del análisis adicional sería convertir los datos recuperados en información. Con el fin emitir recomendaciones, se

utilizan diversas técnicas de análisis, tales como cuadros y gráficas.

#### Presentación de la Información.

Los resultados del análisis se presentan como reportes, cuadros y gráficas o como datos preprocesados para un mayor análisis. Existen varias opciones, tales como reportes de corte, gráficas de pastel, de barras, histogramas, de dispersión y la elección depende de la naturaleza de la información analizada y de las necesidades particulares, estilo de comunicación y cultura de la organización.

#### Divulgación de la Información.

Las recomendaciones se podrían expresar como una copia en bruto o con formato, o enviarse por correo electrónico al usuario empresarial. Cada vez con mayor frecuencia, los resultados del análisis se exportan a herramientas de escritorio como hojas de cálculo o procesadores de texto para incorporarse a reportes y documentos.

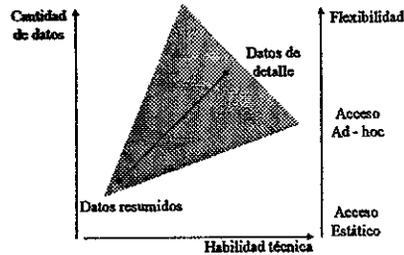
#### **Usuarios y Ambiente del Procesamiento Informático.**

Un Data Warehouse se construye para satisfacer las necesidades de soporte de decisiones de un amplio rango de usuarios empresariales, desde ejecutivos hasta analistas. Estos usuarios reciben apoyo de especialistas de tecnología de información para obtener el máximo valor del Data Warehouse.

#### Usuarios del Procesamiento Informático.

Las necesidades de los usuarios comprenden un amplio espectro: desde cuadros y gráficas hasta reportes, consultas predefinidas y análisis que requiere de consultas hechas a la medida (i.e. *ad-hoc*). La habilidad técnica de los usuarios es variable. Algunos accesan reportes y cuadros preparados, mientras que otros crean con facilidad complejas consultas *ad-hoc*. Para cubrir las necesidades de procesamiento informático de dichos usuarios, las capacidades de consulta, reporte y análisis del Data Warehouse deben ejecutar una gama que va desde el acceso de cuadros y resúmenes, hasta el acceso y consulta de datos de detalle.

## Complejidad de la capacidad de procesamiento informático.



Un Data Warehouse contiene datos con varios niveles de granularidad. En respuesta a los requerimientos de sus usuarios empresariales, el Data Warehouse podría tener los siguientes datos:

### **Cuadros y Reportes Preprocesados.**

Esta información está lista para su acceso y presentación y se puede crear como parte del proceso de carga o alimentación del Data Warehouse o la crean especialistas en tecnología de la información. La cantidad de datos es comparativamente pequeña. Los datos se consideran estáticos, ya que el usuario los recupera como están almacenados en el warehouse.

### **Resumen Pre-calculado e Información Agregada.**

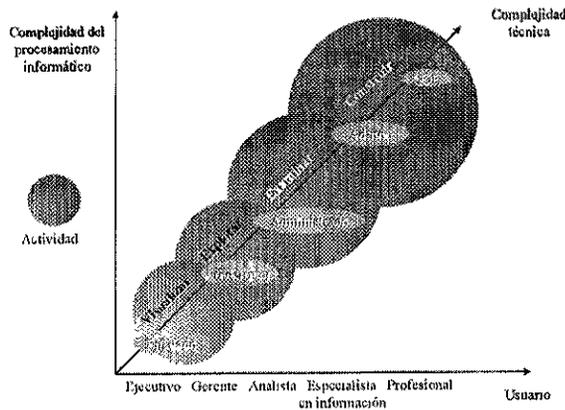
Esta información –tanto muy resumida como poco resumida– se crea como parte del proceso de carga del Data Warehouse o por aplicaciones de resumen programadas periódicamente. Este proceso de cálculo previo puede comprender la creación de resúmenes acumulativos con una actualización posterior del Data Warehouse. La finalidad principal del cálculo previo consiste en mejorar el tiempo de respuesta del acceso y de la consulta, ofreciendo al mismo tiempo flexibilidad en la información que se accesa y en cómo se analiza.

**Datos de Detalle.**

Estos datos están disponibles para una máxima flexibilidad de acceso y análisis *ad-hoc*. El acceso *ad-hoc* a los datos de detalle es esencial debido a que el usuario empresarial busca información que dará una ventaja competitiva a través de probar diversas hipótesis empresariales.

Al aumentar las necesidades de procesamiento informático por parte de los usuarios, aumenta la complejidad técnica del mismo. Por una parte, la cantidad de datos en el Data Warehouse incrementa la complejidad y, por otra, se requiere una enorme habilidad técnica para definir de manera inteligente la consulta. El enfoque de muchas herramientas de consulta y reporte consiste en abordar este reto. La meta del equipo de instrumentación del Data Warehouse es abordar este amplio rango de necesidades del usuario con un conjunto mínimo de diferentes herramientas de procesamiento informático.

El amplio espectro de usuarios empresariales tiene requerimientos ininterrumpidos de procesamiento informático, con una sofisticación creciente.



Los analistas de sistemas utilizan herramientas de procesamiento informático para extraer, resumir y convertir en cuadros y gráficas información del Data Warehouse, las cuales se ponen al alcance de los analistas de negocio quienes explotan estos resultados a través de "iconos" en las estaciones de trabajo del personal ejecutivo, en donde es posible observar los cuadros, gráficas y reportes. Como

regla general, la información se actualiza periódicamente. La capacidad y riqueza de las opciones de presentación de las herramientas de explotación son fundamentales para comprender mejor la información y ayudar a tomar mejores decisiones.

### 3-10 ALGUNOS ESQUEMAS DE DISEÑO PARA BASES DE DATOS

#### Modelo Multidimensional.

Considera tablas relacionales está basado en dos diferentes tipos de tablas: Una tabla central llamada de dimensión y al alrededor de esta tablas medida las cuales expresan la magnitud de las variables de dimensión.

Como ejemplos de dimensiones tenemos: tiempo, región geográfica, productos, etc. Ya que las dimensiones del tiempo pueden ser: año, semestre, trimestre, mes, quincena.

El modelo tradicional de datos de *entidad-relación* comprende la creación de tablas altamente normalizadas que son utilizadas para aplicaciones de proceso de transacciones en línea (OLTP– “*On Line Transaction Processing*”). Como resultado de los modelos de datos generalmente contienen muchas tablas que están altamente normalizadas y no bien situadas para consultas “*ad-hoc*” y soporte a la toma de decisiones por el número uniones de consultas (i.e. “*joins*”) requeridas. Un esquema de estrella simplifica el modelo de datos con la reducción del número de tablas y la alta desnormalización de datos, así que se minimiza el número de uniones de consultas (i.e. “*joins*”) requeridas. Aunque este esquema no es tan efectivo en aplicaciones “*OLTP*” por las anomalías en la actualización, es una buena herramienta para aplicaciones de soporte a la toma de decisiones y es utilizado para diseñar la estructura de tablas de un Data Warehouse.

Dentro de ambientes para soporte a la toma de decisiones es común encontrar requerimientos de consultas que pueden resolverse satisfactoriamente con modelos multidimensionales. Este tipo de consultas consiste en resumir algunos datos para obtener información agregada por algún o algunos criterios (o dimensiones).

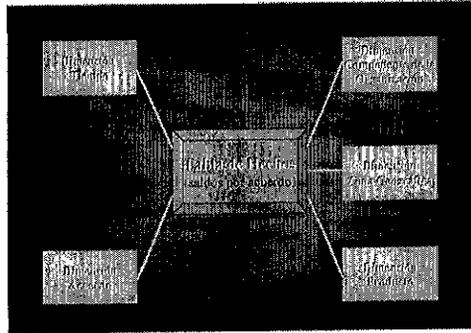
Por ejemplo, a partir de saldos promedio a nivel cuenta, se desea poder obtener el saldo promedio a nivel funcionario, sucursal, producto, zona geográfica.

Algunas consultas requerirán hacer la agregación por alguno de estos criterios, mientras que otras combinarán varios (e.g. sucursales en la zona norte del país, durante el primer trimestre de 2000) para obtener el resultado.

Para poder satisfacer estas consultas con un desempeño adecuado es muy importante que el diseño físico de las tablas involucradas sea cuidadosamente analizado, de tal forma que se cumplan las siguientes condiciones:

- ♦ La elaboración de la consulta debe ser lo más simple posible para el usuario final.
- ♦ Las estructuras de datos deben estar libres de anomalías que pudieran llevar al usuario a obtener resultados equivocados sin que él se percate de esto.
- ♦ El diseño debe ofrecer suficiente flexibilidad para permitir que el usuario haga análisis usando varias combinaciones de criterios y niveles de agregación.
- ♦ El desempeño de la consulta deberá ser óptimo considerando los volúmenes de datos que deben ser examinados y los recursos de cómputo disponibles.

El uso de un diseño físico basado en el modelo *entidad-relación*, por lo general, no podrá cumplir con estas premisas. Una alternativa es generar un diseño físico basado en un modelo multidimensional o esquema estrella. Este último cumplirá mejor con las condiciones anteriores.



Un modelo multidimensional es una alternativa de diseño físico que consiste en definir una tabla de hechos que está conectada con varias tablas de dimensiones.

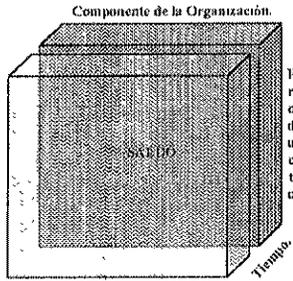
Siendo rigurosos, no deberíamos usar la palabra modelo, puesto que no se trata de tal, el esquema estrella o modelo multidimensional son técnicas de diseño físico.

#### **Tabla de Dimensión.**

Entidad sobre la que se almacenan atributos numéricos a los que se pueden aplicar operaciones aritméticas. Los atributos numéricos se almacenan por separado en la tabla de hechos. Por lo general, la entidad que forma una dimensión se forma a partir de una estructura jerárquica (año, semestre trimestre, Mes, día, etc.). Se les denomina dimensiones haciendo una analogía con un espacio geométrico. Las tablas de dimensión contienen atributos de negocios que se emplean como criterio de búsqueda "SQL" ("Solution Query Language")

La siguiente figura ilustra tres dimensiones:

- 1) Componente de la organización.
- 2) Producto.
- 3) Dimensión temporal.



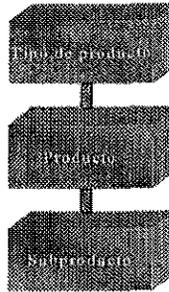
En la intersección de las tres dimensiones se encuentra un saldo.

La dimensión "Componente de la Organización" podría estar a nivel de sucursales y haber resultado de desnormalizar la jerarquía organizacional:



Es decir, la tabla "Componente de la Organización" tendría como atributos el Grupo Financiero Bancomer, la Banca, el Centro Regional y la Sucursal.

Por otra parte, la dimensión "Producto" pudo resultar de desnormalizar una jerarquía de productos:



Este diseño permite al usuario formular consultas como:

- El saldo de un producto para una sucursal en una fecha específica.
- Sumarizar los saldos de un producto para las sucursales que pertenecen a un centro regional a una fecha determinada.
- Obtener el saldo promedio en un período para un tipo de producto a nivel banca.

Algunas observaciones importantes para las tablas de dimensiones son:

- Todos los renglones de la dimensión "Componente de la Organización" describen una sucursal y su estructura jerárquica.
- Todos los renglones de la dimensión "Producto" describen a un subproducto con su correspondiente producto y tipo de producto.
- Una tabla de dimensión está desnormalizada para poder satisfacer completamente una consulta sin tener que recurrir a otras tablas (excepto otras tablas de dimensión y la tabla de hechos). Una consecuencia de esto es que almacenará muchos datos redundantes.
- El contenido de una tabla de dimensión es independiente del tiempo.

**Tabla de Hechos.**

Esta tabla asocia a todas las dimensiones a través de llaves foráneas y para cada combinación de dimensiones registra uno o varios atributos (hechos) numéricos a los que puedan aplicarse operaciones aritméticas. Una tabla de hechos contiene las mediciones básicas de los negocios.

---

Retomando el ejemplo anterior, la tabla de hechos tendría entre sus atributos:

- ♦ Producto.
- ♦ Componente de la organización.
- ♦ Tiempo.
- ♦ Saldo.

Es importante mencionar, que cada renglón almacena saldos a nivel de sucursal, subproducto y fecha.

---

**Granulidad de la Tabla de Hechos.**

Es el nivel de detalle en el que están expresados los atributos numéricos. La granulidad debe ser la misma para todos los renglones de una tabla de hechos.

**Granulidad de la Tabla de Dimensión.**

Es el nivel de detalle que se almacena en cada dimensión. La granulidad debe ser la misma para todos los renglones de una tabla de dimensión.

**Atributo Semi - Agregable.**

Es un atributo numérico almacenado en una tabla de hechos que no puede ser sumariado en todas las dimensiones. Puesto en otros términos, sólo son válidas las sumalizaciones de algunas dimensiones.

### Atributo No Agregable.

Es un atributo numérico almacenado en una tabla de hechos que por su naturaleza no puede ser sumariado. Por ejemplo, un porcentaje no puede ser sumariado.

Los esquemas estrella reflejan la visión del usuario final de una consulta empresarial. Hechos tales como ventas, compensación, pago y facturas se califican por una o más dimensiones (por mes, por producto, por región geográfica).

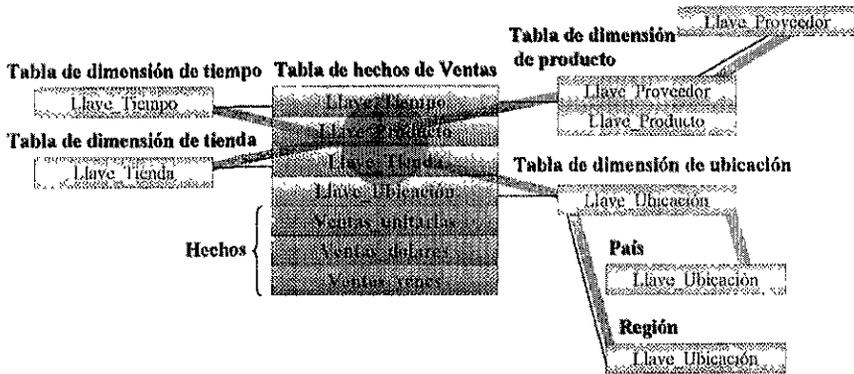


La figura, sólo exhibe las columnas privadas de la tabla de hechos y las llaves externas a las tablas de dimensión. Las tablas de dimensión contienen atributos no de llave que describen los valores de la dimensión. Por ejemplo la tabla de dimensión de Tienda podría tener atributos no de llave (área en pies cuadrados, tipo de arrendamiento, propiedad/franquicia de la compañía) y atributos de llave (Clave de la tienda). Las consultas en la tabla de hechos simplemente conectan listas de apuntadores a las tablas de dimensión. Cuando las consultas en la tabla de hechos se unen con consultas en la tabla de dimensión, se recupera información diversa o cuantiosa. Las tablas de dimensión permiten profundizar y hacer adiciones en la búsqueda por medio de uniones.

### Esquemas Copo de Nieve.

Es una extensión del esquema estrella en donde cada uno de los puntos de la estrella se divide en más puntos. En esta forma de esquema, las tablas de dimensión del esquema estrella contienen más normas. Las ventajas que proporciona el esquema copo de nieve son mejorar el desempeño de consultas debido a un mínimo almacenamiento en disco para los datos y mejorar el desempeño mediante la unión de tablas más pequeñas con normas, en vez de las grandes tablas sin normas. Asimismo, el esquema copo de

nieve incrementa la flexibilidad de las aplicaciones debido a la aplicación de normas y, por lo tanto, disminuye la granularidad de las dimensiones.



El esquema copo de nieve aumenta el número de tablas con las que el usuario debe tratar e incrementa la complejidad de algunas de las consultas que debe localizar. Algunas herramientas nuevas ocultan a los usuarios el esquema físico de la base de datos y le permiten trabajar a un nivel conceptual. Las herramientas ubican las consultas del usuario en el esquema físico. Necesitan de un administrador de base de datos para efectuar las ubicaciones, una vez instaladas las herramientas.

**Esquemas Mixtos.**

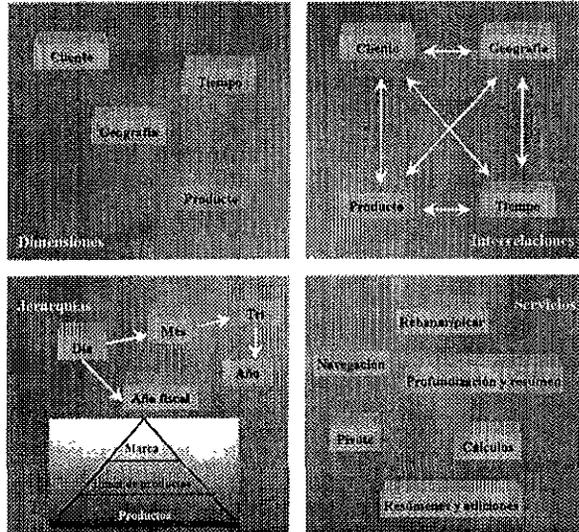
El esquema mixto es una mezcla entre el esquema estrella, basado en tablas de hechos y tablas de dimensión sin normas y el esquema copo de nieve, en donde todas las tablas de dimensión tienen normas. En el esquema mixto, sólo se aplican normas a las tablas de dimensión más grandes. Por lo general, estas tablas contienen volúmenes de columnas de datos sin normas (duplicados).

**Análisis Multidimensional.**

Tanto para la eficiencia operativa como para la planeación a futuro, se deben analizar muchos datos empresariales interrelacionados. Esta necesidad empresarial se aborda mediante el procesamiento

analítico. En éste, el enfoque está en el análisis de los datos, específicamente en el análisis multidimensional.

En el análisis multidimensional, los datos se representan mediante dimensiones como áreas geográficas, productos, clientes, etc. Como se puede observar en el siguiente gráfico las dimensiones muestran las variables susceptibles de contener detalle y la forma en que pueden ser relacionadas con otras para construir consultas mas elaboradas.



ESTA TESIS NO DEBE  
CALAR DE LA BIBLIOTECA

Por lo regular las dimensiones se relacionan en jerarquías, por ejemplo: ciudad, estado, región, país y continente, o estado, territorio y región. El tiempo es también una dimensión estándar con su propia jerarquía como día, semana, mes, trimestre y año, o día y año calendario.

Para facilitar un análisis complejo, el procesamiento analítico o análisis multidimensional, presenta una visión empresarial sencilla de los datos. Un usuario empresarial puede acceder a los ingresos por departamento y tienda para los últimos cuatro trimestres, para un conjunto dado de productos. Los resultados se pueden analizar desde la perspectiva de una variable (i.e. "pivotar") o girar para cambiar los ejes. Además, los usuarios empresariales pueden "navegar" por las dimensiones, profundizando y obteniendo

resúmenes a lo largo de los elementos de una dimensión o penetran a través de las dimensiones para ver otras perspectivas. Un equipo computarizado ofrece grandes capacidades como posiciones al principio o al final de la lista; promedios móviles; tasas de crecimiento; cálculos financieros de interés; tasas internas de recuperación y depreciación y conversiones monetarias y funciones estadísticas.

El procesamiento analítico se usa para entender lo que está sucediendo en la empresa y promete análisis del tipo: "¿Qué pasaría si ...?".

El procesamiento analítico se emplea para análisis históricos complejos, con amplia manipulación:—análisis de datos dinámicos—, así como para planeación a futuro y pronósticos, considerando al pasado como prólogo del futuro.

Al procesamiento analítico o análisis multidimensional se le conoce también como procesamiento analítico en línea (OLAP). Se apoya en una visión multidimensional de los datos empresariales en el Data Warehouse y puede tener un motor de depósito de base de datos multidimensional.

### 3-11 OLAP, ON LINE ANALITYCAL PROCESSING.

#### "On Line Analytical Processing."

Por lo general se refiere a procesos de consulta que involucran un diseño multidimensional. Algunos lo definen como un término que describe un enfoque dimensional en el soporte para la toma de decisiones.

"OLAP" se usa para referirse a procesos de consulta donde el usuario puede analizar sus datos agrupados y resumidos bajo distintos criterios y a diferentes niveles de detalle. Por lo regular incluye la capacidad de "drill up", "drill across" y "drill down", es decir, a partir de ciertos valores, empezar a agregarlos más o desglosar las cifras con mayor o menor detalle.

En un Data Warehouse se depositan datos para consulta, análisis y divulgación, a diferencia del procesamiento de transacciones en línea ("OLTP"), en donde los datos se reúnen y almacenan para operación y control. "OLTP" es una tecnología de procesamiento analítica que crea nueva información empresarial a partir de los

datos existentes, por medio de un rico conjunto de transformaciones empresariales y cálculos numéricos.

---

El procesamiento analítico en línea hace lo siguiente:

- ♦ Presenta una visión multidimensional lógica de los datos en el Data Warehouse. La visión es independiente de la manera en que se almacenan los datos.
  - ♦ Comprende siempre la consulta interactiva y el análisis de los datos. Por lo regular la interacción es de varias pasadas, lo cual incluye la profundización en niveles cada vez más detallados o el ascenso a niveles superiores de resumen y adición.
  - ♦ Ofrece opciones de modelado analítico, incluyendo un motor de cálculo para obtener proporciones, desviaciones, etcétera, que comprende mediciones de datos numéricos a través de muchas dimensiones.
  - ♦ Crea resúmenes y adiciones (también conocidas como consolidaciones), jerarquías y cuestiona todos los niveles de adición y resumen en cada intersección de las dimensiones.
  - ♦ Maneja modelos funcionales de pronóstico, análisis de tendencias y análisis estadísticos.
  - ♦ Recupera y exhibe datos tabulares en dos o tres dimensiones, cuadros y gráficas, con un “*pivoteo*” fácil de los ejes. El “*pivoteo*” es fundamental, ya que los usuarios empresariales necesitan analizar los datos desde perspectivas diferentes y el análisis desde una perspectiva conduce a otra cuestión empresarial que se va a examinar desde otra perspectiva.
  - ♦ Responde con rapidez a las consultas, de modo que el proceso de análisis no se interrumpe y la información no se desactualiza.
-

La tecnología "OLAP" se aplica en muchas áreas funcionales de una empresa, tales como producción, ventas y análisis de rentabilidad de la comercialización; mezcla de manufacturas y análisis de logística; consolidaciones financieras, presupuestos y pronósticos, planeación de impuestos y contabilidad de costos.

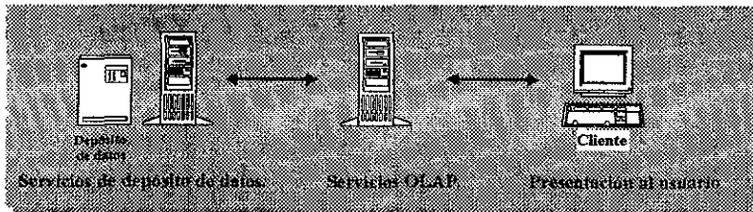
### Arquitectura OLAP.

#### Visión OLAP.

La presentación multidimensional y lógica de los datos del Data Warehouse o del mercado de datos para el usuario empresarial, sin importar cómo y dónde están almacenados los datos.

#### Tecnología de Depósito de Datos.

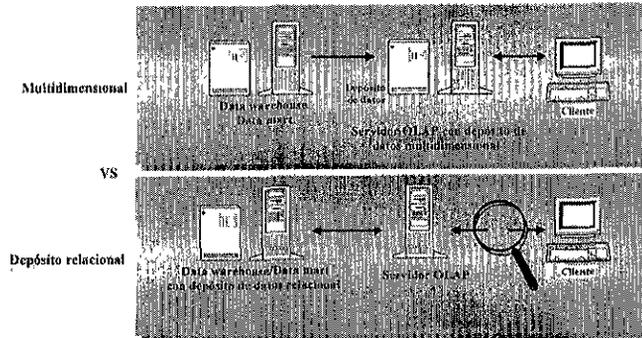
Las opciones de tecnología de cómo y dónde están almacenados realmente los datos. Las dos opciones populares son el depósito de datos multidimensional y el depósito de datos relacional.



Al usuario empresarial sólo le interesa la visión multidimensional de los datos y en un nivel de desempeño aceptable. Al personal de apoyo en tecnología de la información le interesa dónde y cómo se almacenan y acceden los datos, para asegurar un desempeño aceptable y una administración eficaz.

### Arquitectura Física.

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



### **ROLAP.**

#### Relational On Line Analytical Processing.

Es la implantación de "OLAP" (On Line Analytical Processing) usando un "DBMS" (Data Base Management System) relacional. Los datos se almacenan en tablas relacionales.

Aunque los datos se almacenan en forma relacional (columnas y registros), se presentan al usuario empresarial en forma de dimensiones empresariales. A fin de *ocultar* la forma de depósito, se crea una capa semántica de metadatos. Esta capa ubica las dimensiones para las tablas relacionales. También se crean metadatos adicionales para cualquier resumen o adición, con el fin de mejorar el tiempo de respuesta. Todos estos metadatos se almacenan en la base de datos relacional –creando otro depósito de metadatos en la solución general del Data Warehouse– para su mantenimiento y administración.

La actividad inicial de diseño y configuración se conduce mediante el diseño técnico de una base de datos. Se deben seguir los siguientes pasos básicos:

1. Construir el "diseño dimensional" utilizando técnicas como el retiro de normas, el esquema estrella, de copo de nieve y el esquema mixto.
2. Incorporar los datos adecuados de adición y resumen.
3. Dividir los grandes conjuntos de datos en segmentos más pequeños y manejables para mejorar el desempeño. Por ejemplo, se dividen las unidades de tiempo u organizacionales.
4. Agregar índices creativos o de *mapa de bits* para mejorar el desempeño (esto incrementa el tamaño de la base de datos y el tiempo para construir los índices).
5. Crear y almacenar los metadatos. Los metadatos incluyen las definiciones de las dimensiones, la ubicación de dimensiones para las tablas relacionales, las relaciones jerárquicas entre dimensiones, la información de segmentos, las definiciones y descripciones de resúmenes y adiciones, las fórmulas y cálculos, la vigilancia del uso y muchas otras.

Desde una perspectiva operacional, los pasos para ejecutar una consulta son los siguientes:

1. Construir la herramienta cliente utilizando una visión dimensional o empresarial de los datos.
2. Consultar el servidor "OLAP" desde la herramienta cliente y examinar los metadatos en tiempo real.
3. Crear declaraciones "SELECT" de pasos múltiples y/o consultas correlacionadas y someterlas a la base de datos relacional.
4. Sobre los resultados de la consulta a la base de datos, realizar las funciones multidimensionales, como cálculos y

fórmulas, traducción de bits para descripciones empresariales, etcétera.

5. Devolver los resultados a la herramienta cliente para un mayor procesamiento y exhibición o para su exhibición inmediata.

Las funciones principales que se ofrecen a los usuarios empresariales y administradores, incluyen las siguientes:

- ♦ Una visión empresarial de los datos relacionales.
- ♦ Soporte de la jerarquía dimensional.
- ♦ Funciones de cálculo, estadísticas y financieras, con ampliaciones del usuario.
- ♦ Profundización al nivel de detalle.
- ♦ Elección de herramientas de proceso frontal.
- ♦ Aprovechamiento de las inversiones existentes para respaldo y restauración de la administración de base de datos y subdivisión de la base de datos para análisis individuales.
- ♦ Navegación utilizando metadatos.
- ♦ Niveles múltiples de seguridad con privilegios de usuario.

La administración general y la administración de sistemas de este enfoque requieren de lo siguiente:

- ♦ Que no haya carga inicial o actualizaciones periódicas del servidor "OLAP".
- ♦ Uso de los procesos estándar existentes para respaldo, restauración y seguridad.
- ♦ Diseño del "modelo dimensional" para proporcionar la visión empresarial de los datos. Esto requiere de capacitación y pericia en el modelado con esquemas estrella, copo de nieve y mixtos; segmentación de datos; niveles de adición y resumen.

- ♦ Administración, sincronización y mantenimiento de todos los metadatos nuevos dentro de la solución general del Data Warehouse.
- ♦ Vigilancia de la utilización para afinar el desempeño. Dicha afinación puede afectar el modelo de base de datos, los segmentos o los niveles de adición y resumen. La afinación es complicada. El retiro de normas o la indexación para el desempeño pueden incrementar la escasez y tamaño de la base de datos, requiriendo así de búsquedas más largas y más operaciones de entrada y salida de disco.
- ♦ Algunos de los aspectos que se presentan por el uso de este enfoque son:
  - ♦ El uso de esquema estrella y copo de nieve, segmentos y retiro de normas. Esto mejora el desempeño, pero el intercambio afecta en forma adversa la flexibilidad y la capacidad de expansión de la base de datos relacional. Esto hace que actualizar la base de datos sea todo un reto y pueden requerirse actualizaciones en bloque.
  - ♦ El esquema de estrella, con sus variantes, adiciones y resúmenes, supone que los datos son estáticos, salvo cuando se cargan en bloque.
  - ♦ Los cálculos a nivel de fila, por ejemplo, cuando la utilidad es igual al ingreso menos el costo, requieren de traspasar columnas y filas. Existe un límite para estas acciones, incluso con declaraciones "*SELECT*" de pasos múltiples.
  - ♦ La administración y mantenimiento de metadatos representa un reto y un costo continuos.

Un buen ajuste para el enfoque de base de datos relacional y servicios "*OLAP*" es cuando los requerimientos solicitan lo siguiente:

- ♦ Aplicaciones intensivas de datos con necesidades significativas de examen de datos.
- ♦ Capacidad de sólo lectura con un mínimo de escritura.
- ♦ Un mínimo de cálculos cruzados entre dimensiones a nivel fila.

La exposición anterior se concentró en las funciones del usuario empresarial, el desempeño, el volumen y la viabilidad de cambio de escala, la administración y la conveniencia de los propósitos. Algunos otros factores de interés incluyen los siguientes:

- ♦ Interfaz de usuario deseada.
- ♦ Características y funciones de la herramienta cliente.
- ♦ Percepción de la arquitectura.
- ♦ Cuál enfoque aprovecha mejor las inversiones existentes en tecnologías "cliente/servidor", bases de datos y procesos formales de habilidades en la administración de datos.

Una vez establecidas la capacidad técnica y el alcance del data Warehouse el siguiente capítulo se enfocará a la explotación del Data Warehouse y algunas herramientas estadísticas para ello.

LA INFORMACIÓN Y LA TOMA DE DECISIONES A PARTIR DE UN DATA WAREHOUSE.

## Capítulo IV

### Análisis y Explotación del Data Warehouse.

4-1	Concepto de Data Warehouse y Explotación.	88
4-2	Data Mining.	91
4-3	Análisis Estadístico.	93
4-4	Descubrimiento de Conocimientos.	98
4-5	Redes Neuronales.	102
4-6	Otras Técnicas y Herramientas de Data Mining.	103

**- CAPITULO CUARTO****ANÁLISIS Y EXPLOTACIÓN  
DEL  
DATA WAREHOUSE.****4-1 CONCEPTO DATA WAREHOUSE Y EXPLOTACIÓN.**

El concepto "*Data Warehouse*" se refiere a un conjunto integrado de bases de datos con orientación temática que está diseñado para apoyar en la toma de decisiones, y donde cada unidad de datos es relevante en algún momento del tiempo. El "*Data Warehouse*" almacena información de múltiples fuentes o sistemas operacionales dispersos, la tendencia en el uso de esta tecnología en las empresas es tener un repositorio central que organice la información que se genera en la operación del negocio formalizando así las directrices de manejo de información con la tendencia de hacer oficial el origen de toda información con fines de negocio, administrativos y/o contables resolviendo así las distorsiones de información que se presentaban a partir de la toma de información de diversas fuentes.

El "*Data Warehousing*" o almacenamiento de datos es el procedimiento que facilita la creación y explotación de un "*Data Warehouse*". "*Data Warehousing*" es un proceso para organizar la información para su análisis, que hace real la gestión de conocimiento.

"*Data Mining*" es un proceso analítico para explorar información buscando patrones o relaciones entre las variables.

"*Data Mining*" es la conjunción de tres fuerzas: "*Machine Learning*" (inteligencia artificial), Estadística, y Bases de Datos. El objetivo de aplicar técnicas de "*Data Mining*" a la información corporativa es obtener una descripción de los datos y sus variables, mismas que son la causa de determinados efectos. Una vez identificado el modelo que produce una determinada situación real, el analista de exploración intentará formular predicciones.

Es importante señalar que este tipo de proyectos están en constante evolución y deben replicar la propia evolución de la empresa.

Los procesos de "*Data Warehousing*" y "*Data Mining*" están experimentando un increíble crecimiento en los últimos años.

¿Cuáles son las causas de su auge?. La primera de las causas es que la información de clientes, mercados, productos y datos externos está aumentando dada la globalización de los negocios. Así mismo se está produciendo un gran aumento en el rendimiento del "hardware"; las herramientas de análisis han progresado sustancialmente; las organizaciones están aprendiendo a sacar provecho de sus datos internos para obtener ventaja competitiva lo cual en paralelo permite la gestión del conocimiento. La información convertida en conocimiento es uno de los factores más importantes en la producción, ya que es un recurso clave para cualquier compañía que quiera sobrevivir en el competitivo y global entorno empresarial que hoy se gesta.

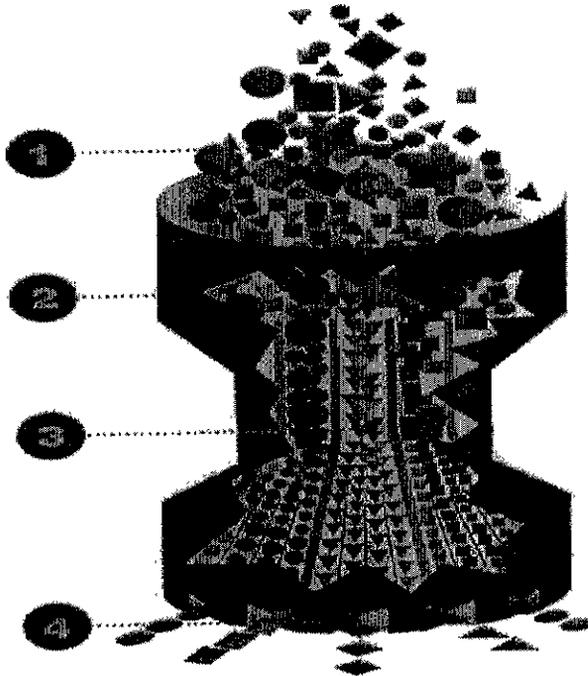
"Data Warehouse" ha empezado a proporcionar importantes ventajas competitivas que antes no podían ser visualizadas por la falta de un repositorio central de información, algunas respuestas han sido obtenidas a partir de análisis estadístico como la siguiente: La duda de quiénes son los clientes rentables y cuáles los problemáticos para las compañías aseguradoras por fin fue resuelto. El grupo de población del sexo femenino cuya edad oscila entre los 25 y los 39 años es el que registra menos reclamos por seguros de automóvil; en tanto los hombres entre los 18 y los 25 años son una verdadera amenaza porque es justamente en este grupo de población en donde se reporta el mayor número de accidentes automovilísticos.

No fue un grupo de investigadores ex-profeso quien encontró la respuesta. Simplemente se utilizó la técnica de "Data Mining" (minería de datos), en el "Data Warehouse" global de la empresa el cual está integrado a su vez con un sistema de soporte de decisiones.

Consecuencia de lo anterior la compañía aseguradora en cuestión determinó diseñar un plan especial de aseguramiento con estímulos y descuentos atractivos para retener a sus clientes rentables, es decir las mujeres de entre 25 y 39 años de edad, se decidió además la creación de productos específicos para estimular el ingreso de nuevos clientes con estas características, en cambio, para los clientes de riesgo los jóvenes de 18 a 25 años la empresa aseguradora decidió aumentar sensiblemente el costo del seguro. Una decisión de negocios basada en el uso de la tecnología.

Más allá de las definiciones formales el ejemplo descrito resume la aplicación en un caso concreto de la inteligencia de negocios, tema obligado en el ámbito de la integración de soluciones informáticas y factor de cambio en la manera de conseguir la aplicación de la tecnología a nivel empresarial, la visión del "Data Warehouse" es

transformar los datos simples y dispersos en información coherente y concretar el proceso en conocimiento del negocio y toma de decisiones, puntuales y casi en tiempo real, considérese el siguiente como el flujo de información:



1.- Los datos son recibidos de diferentes fuentes y en diferentes formatos incluyendo sistemas operacionales internos y algunas fuentes externas. Todos los datos son cargados, limpiados e integrados en el "Data Warehouse"

2.- "Data Warehouse" distingue de dónde provienen los datos y cuándo fueron generados, qué representan y cómo pueden ser usados. Las decisiones de mercado son basadas en un grupo de datos seleccionados a partir de minería de datos.

3.- "Data Warehouse" es efectivo reorganizador para soportar creativamente las consultas de decisión de mercado.

4.- El poder de “Data Warehouse” para el soporte en la toma de decisiones puede ser explotado por un sin fin de combinaciones que lo llevarán a conocer oportunidades que usted necesita saber y que existen dentro de la organización.

Algunas otras ventajas se pueden ubicar en el manejo de relaciones de mercado, esta área se concentra en los objetivos de la empresa, la gente, los procesos de negocio y la infraestructura en el confeccionado de los productos, los servicios y sobre todo, la interacción con los clientes y sus necesidades. En las empresas, ésta es un área estratégica porque el impacto del crecimiento de un 5% en la tasa de retención de clientes impactaría directamente las ganancias.

## 4-2 DATA MINING

### Data Mining.

Como se ha mencionado, el “Data Warehouse” proporciona al analista empresarial dos ingredientes esenciales. El primero es una gran cantidad de datos sobre sus clientes en forma ordenada, histórica y que expresa la estructura corporativa. El segundo es el carácter único de estos datos –ninguno de los competidores los posee-. Para extraer revelaciones fundamentales sobre el comportamiento de sus clientes, sus productos e incluso sus proveedores, la solución de “Data Warehouse” debe incorporar la minería de datos –Data Mining- a su plataforma de soporte de decisiones.

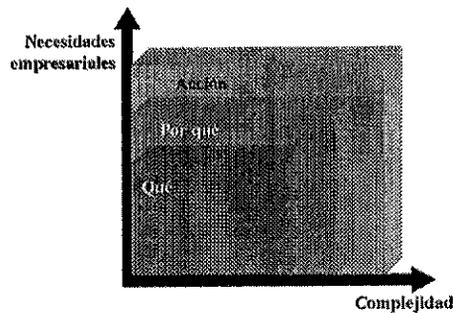
“Data Mining” apoya la modalidad de descubrimiento del soporte de decisiones. Las herramientas recorren el detalle de las transacciones para desenterrar patrones y asociaciones ocultos.

Las herramientas de procesamiento analítico permiten diversas visualizaciones de los datos, como ventas por marca, tienda, y períodos de tiempo, es decir a la medida de las necesidades. Las herramientas de minería de datos son excelentes para comprender el comportamiento de los clientes.

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 los clientes. La meta es descubrir revelaciones estratégicas competitivas para controlar la participación en el mercado y las utilidades. Los seres humanos tienen agudeza para percibir excepciones y anomalías, pero no

tienen el poder y la capacidad para encontrar relaciones en grandes volúmenes de datos. Sin embargo, una vez extraídas las relaciones y presentadas a los analistas, tienen la posibilidad de examinarlas y seleccionar las más interesantes y útiles.

Los analistas tienen un rango de necesidades de conocimiento. La primera necesidad es comprender qué está sucediendo en el negocio. La siguiente es por qué está sucediendo, ¿cuál es el comportamiento de clientes y mercados?. El conocimiento proveerá la respuesta para el ¿qué hacer?.



Entender el comportamiento de clientes y mercados y establecer una estrategia competitiva es el reto.

### **Tecnologías y Herramientas de Data Mining.**

Existe una amplia variedad de tecnología para la minería de datos, estas herramientas y tecnologías se clasifican en tres categorías:

- Análisis estadístico.
- Descubrimiento de conocimientos.
- Otros, como sistemas de visualización, análisis fractal y herramientas de propietario.

### 4-3 ANÁLISIS ESTADÍSTICO

Los sistemas de análisis estadístico y análisis de datos se usan para detectar patrones de datos. Estos patrones de datos se explican después mediante modelos estadísticos y matemáticos. Algunas de las técnicas de modelado estadístico y matemático que se emplean son el análisis de regresión lineal y no lineal, el análisis de regresión logística, el análisis univariado y multivariado y el análisis de series de tiempo.

Las herramientas de análisis estadístico se utilizan en diversas aplicaciones empresariales para:

- Incrementar la participación en el mercado y las utilidades, detectando las mejores oportunidades.
- Aumentar la satisfacción del usuario a través de la mejora en la calidad de productos y servicios
- Impulsar los márgenes a través de la modernización de la manufactura de productos y de logística (ingeniería de proceso empresarial).

Así mismo las herramientas de análisis estadístico han experimentado una gran evolución dentro del mercado de “*software*”, el cual provee herramientas diseñadas para hacer comprensiva y poderosa la exploración de datos, ofrece técnicas de análisis especializadas y poderosos gráficos.

Algunos de los temas que atiende el “*software*” estadístico son:

- Data Mining
- Data Warehouse
- On-line Analytic Processing (OLAP)
- Groupware Functionality
- Neural Networks

“*Data Mining*” se define como un proceso analítico diseñado para explorar grandes cantidades de datos (típicamente relacionados con negocios y mercados) buscando patrones consistentes de datos o relaciones sistemáticas entre variables para entonces validar el subconjunto de datos, el proceso consiste básicamente de tres etapas: exploración, construcción del modelo y validación.

El concepto de “*Data Mining*” es ahora común en la información de negocios para revelar conocimiento de las estructuras que pueden guiar en decisiones. “*Data Mining*” está actualmente basado en el concepto de Exploración y análisis de datos (EDA, Exploratory Data Analysis) y modelado por lo que comparten estas dos técnicas. En otras palabras “*Data Mining*” está relativamente menos interesado en identificar relaciones específicas entre las variables involucradas, por ejemplo revelar la naturaleza fundamental de funciones o dependencias multivariadas entre variables que no son el principal objetivo de “*Data Mining*”.

“*On-line Analytic Processing*” (OLAP), El término OLAP se refiere a la tecnología que permite a los usuarios de bases de datos multidimensionales generar en línea (tiempo real) vistas de sumarios comparativos de datos y otras consultas, OLAP puede estar integrado dentro de la corporación para permitir análisis y monitoreo en el desempeño de los negocios, el resultado final de las técnicas de OLAP pueden ser muy simples, por ejemplo tablas de frecuencias, estadística descriptiva, tabulaciones de información etc.

“*Groupware*”, Se refiere a posibilitar a un grupo de usuarios de la red de trabajo a colaborar en proyectos específicos, provee servicios de comunicaciones (como correo electrónico), documentos comunes para desarrollar, analizar y reportar; también ofrece análisis estadístico de datos, calendarización y localización, los documentos pueden contener texto, gráficas y otras formas de información como multimedia.

“*Redes Neuronales*”, son técnicas analíticas modeladas después de (hipotetizados) el proceso de aprendizaje en el sistema cognoscitivo y las funciones neurológicas del cerebro capaces de predecir nuevas observaciones (sobre variables específicas) de otras observaciones (sobre las mismas u otras variables) después de ejecutar un proceso llamado de aprendizaje a partir de los datos existentes (es una técnica de minería de datos).

La técnica de “*Peinado*”, quizás la primera técnica estadística más ampliamente utilizada históricamente identificada como análisis gráfico exploratorio de datos es el peinado, un método interactivo que permite seleccionar en una vista de un sub-conjunto de datos e identifica las características (comunes) o para examinar sus efectos entre variables relevantes. Aquellas relaciones entre variables pueden ser visualizadas por funciones ajustadas (e.g. líneas en dos dimensiones o superficies en tres dimensiones) y sus intervalos de confianza así como por ejemplo se puede examinar

cambios en aquellas funciones removiendo interactivamente (temporalmente) o añadiendo sub-conjuntos específicos de datos por ejemplo una de muchas aplicaciones de la técnica de peinado es la de seleccionar (i.e marcar) en una matriz donde cada entrada es una gráfica de datos, todos los datos que pertenecen a una cierta categoría (e.g. un nivel de ingresos medios) para examinar como aquellas observaciones específicas contribuyen en las relaciones entre otras variables en el mismo conjunto de datos (e.g. las correlaciones entre lo que se tiene y lo que se debe).

Redes Neuronales, el primer paso es para designar una arquitectura de red (que incluye un número específico de estratos, cada uno consistiendo en un cierto número de neuronas). El tamaño y estructura de red necesita concordar con la naturaleza del fenómeno estudiado (complejidad formal). Debido a que este último no es muy conocido en esta temprana etapa esta, tarea no es fácil y frecuentemente involucra múltiples ensayos y errores.

La nueva red es entonces sujeta al proceso de entrenamiento. En esta fase se aplica un proceso iterativo en el número de variables de entrada para ajustar el peso de la red para predicciones óptimas (en términos tradicionales uno podría decir, buscar un ajuste a). en los datos de muestra sobre los cuales es ejecutado el entrenamiento. Después de la fase de aprendizaje a partir de un conjunto de datos existente, la nueva red está lista y puede ser usada para generar predicciones. La red resultante desarrollada en el proceso de aprendizaje representa un patrón detectado en los datos. De este modo, en este enfoque la red es el equivalente de un modelo de relaciones entre variables en la construcción de un modelo tradicional. Sin embargo, a diferencia de los modelos tradicionales en la red las relaciones no pueden ser articuladas en los términos usuales utilizados en estadística o metodológicamente para describir relaciones entre variables (e.g. considere A positivamente relacionada con B pero sólo para observaciones donde el valor de C es bajo y D es alto). Algunas redes neuronales pueden producir predicciones precisas, éstas representan, sin embargo un enfoque de investigación típico a-teórico (se puede decir, una caja negra). Este enfoque tiene que ver sólo con consideraciones prácticas, esto es, con la validez predictiva de la solución y su relevancia aplicada y no con la naturaleza del mecanismo por sí mismo o de su relevancia para cualquier teoría del fenómeno en sí.

No obstante las técnicas de redes neuronales pueden ser usadas como un componente de análisis diseñado para construir modelos explicatorios ya que las redes neuronales pueden ayudar a explorar conjuntos de datos en busca de variables relevantes o grupos de

variables, el resultado de este tipo de exploraciones pueden entonces facilitar el proceso de construcción de modelos. Además, ahora hay software de redes neuronales que usa sofisticados algoritmos de búsqueda para las variables de entrada más relevantes, de este modo se contribuye potencialmente en el proceso de construcción de modelos.

Una de las principales ventajas de las redes neuronales es que teóricamente son capaces de aproximar cualquier función continua, de tal forma que los investigadores no necesitan tener ninguna hipótesis acerca del modelo considerado, o alguna referencia adicional. Una importante desventaja sin embargo es que al final la solución depende de las condiciones iniciales de la red y esto como se dijo antes, es virtualmente imposible interpretar la solución en términos analíticos tradicionales, tales como los utilizados para construir teorías que expliquen los fenómenos.

Algunos autores esperan usar masivamente modelos computacionales paralelos. Por ejemplo Haykin (1994) define las redes neuronales como:

“Un procesador distribuido masivamente en paralelo que tiene una tendencia natural para almacenar conocimiento experimental y hacer disponible su uso. Se parece al cerebro en dos aspectos (1) El conocimiento es adquirido por la red a través del proceso de aprendizaje y (2) Las sólidas conexiones interneuronales conocidos como pesos sinápticos son usadas para almacenar conocimiento”.

### **Uso de las Herramientas de Análisis Estadístico.**

“Data Warehouse” o “Data Mart” resultan excelentes proveedores de datos para herramientas de análisis estadístico, los usuarios empresariales seleccionan y extraen datos del negocio para probar hipótesis y obtener conocimiento de tendencias de mercado.



### **Aplicaciones del Análisis Estadístico.**

Se usa para abordar un amplio rango de aspectos empresariales, tales como:

#### Comercialización.

En la comprensión de los hábitos de compra y comportamiento del consumidor, con la identificación de segmentos del mercado. El análisis de la canasta de mercado se emplea para analizar los patrones de compra de los clientes.

#### Ventas.

Análisis de ingresos por ventas y desviaciones de lo planeado, ventas por temporada.

#### Televentas y Comercialización.

Análisis de respuestas para afinar la estrategia de televentas.

#### Administración Empresarial.

Evaluación de préstamos y análisis de riesgos de créditos, medición de pruebas de resultados y evaluación de la satisfacción del cliente.

#### Investigación Médica.

Estudio de la respuesta a diagnósticos y tratamientos.

**4-4 DESCUBRIMIENTO DE CONOCIMIENTOS.**

El análisis estadístico muestra al analista el comportamiento de las variables en términos de frecuencias, distribución, medidas de tendencia central, hace pronósticos de comportamiento en función de datos históricos, entre otras; cuando el analista plantea alguna pregunta no necesariamente sabe cuál puede ser la respuesta así que requerirá de tecnología y herramientas estadísticas para el descubrimiento de conocimientos.

El descubrimiento de conocimientos tiene su raíz en la inteligencia artificial y el aprendizaje con máquinas:

- El descubrimiento de conocimientos consiste en extraer de los datos información implícita, no trivial, que no se conocía y es potencialmente útil.
- Es el proceso de buscar en los datos sin establecer por adelantado una hipótesis e incluso así, encontrar información que muestre relaciones y patrones entre los datos.

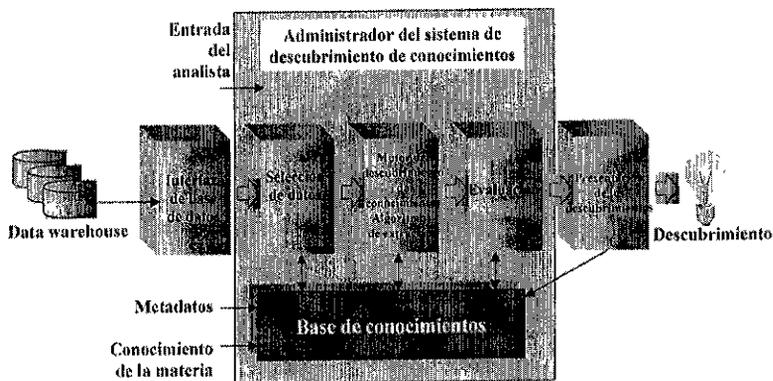
El descubrimiento de conocimientos pretende examinar una gran cantidad de datos en el "*Data Warehouse*" en busca de patrones recurrentes, detectando tendencias y desenterrando hechos.

***Estructura de Sistemas de Descubrimiento de Conocimientos.***

Una pieza de conocimientos es una relación o patrón interesante y útil entre elementos de datos con respecto a un tema y tarea específicos. El sentido de autonomía está implícito en la definición del descubrimiento de conocimientos -aprendizaje no supervisado-. La mayoría de los sistemas de descubrimiento tienen alguna entrada o guía del analista. Consideraremos al analista como parte del sistema de descubrimiento de conocimientos.

Un sistema de descubrimiento de conocimientos consta de una integración de componentes que identifican y extraen patrones y relaciones interesantes y útiles para el analista de los datos registrados en el "*Data Warehouse*".

Los componentes individuales no se encuentran tan claramente separados. Las principales entradas para el sistema son los datos del "Data Warehouse", la guía del analista y la experiencia en el tema que almacena la base de conocimientos del sistema. Los datos seleccionados del "Data Warehouse" se procesan en el motor de descubrimiento de conocimientos, en donde se aplican una serie de algoritmos de extracción para producir prospectos de patrones y relaciones. Estos prospectos se evalúan. Algunos se identifican como descubrimientos interesantes y se le presentan al analista empresarial. Algunos de estos descubrimientos pueden agregarse a la base de conocimientos para mejorar las extracciones y evaluaciones posteriores de descubrimientos.



### Interfaz de Base de Datos del "Data Warehouse".

Los sistemas de descubrimiento de conocimientos extraen datos de la base de datos del "Data Warehouse" por medio de las opciones de consulta de la base de datos. Para las bases de datos relacionales se emplea SQL. Los metadatos del "Data Warehouse" en la base de conocimientos guían la interfaz de base de datos para corregir la organización de las estructuras de datos y saber cómo se almacenan en el "Data Warehouse".

Para efectos de eficiencia y desempeño, la interfaz de base de datos del sistema de descubrimiento de conocimientos debe comunicarse directamente con el "Data Warehouse".

### Selección de Datos.

Este componente determina cuáles datos se necesitan extraer del "Data Warehouse" y cuáles son sus estructuras de datos. La base de conocimientos guía al componente de selección de datos acerca de qué extraer y cómo hacerlo. Si sólo se requiere una muestra de datos, el componente de selección de datos debe tener la capacidad para extraer y seleccionar la muestra aleatoria correcta. Además, se selecciona e incorpora en el algoritmo el tipo de datos que requiere.

### Motor de Descubrimiento de Conocimientos.

Aplica los algoritmos de extracción de la base de conocimientos a los datos que extrae el componente de selección de datos. La meta consiste en extraer patrones y relaciones entre los elementos de datos. La influencia desarrollada dentro de la base de conocimientos tiene un efecto crítico sobre los descubrimientos extraídos.

Es posible incorporar un amplio rango de algoritmos dentro del sistema, tales como dependencias de datos, reglas de clasificación, agrupamiento, resúmenes, detección de desviación, etcétera.

### Evaluación de Descubrimientos.

Ayuda al analista a examinar los patrones para seleccionar sólo los que sean de interés. Las técnicas que se emplean para analizar patrones interesantes comprenden la significia estadística, el factor de confianza o nivel de cobertura y el análisis visual.

## **Tecnología de Descubrimiento de Conocimientos.**

La tecnología fundamental son los algoritmos para patrones y relaciones. La tecnología se describe desde la perspectiva de varias categorías:

### Análisis de Dependencias.

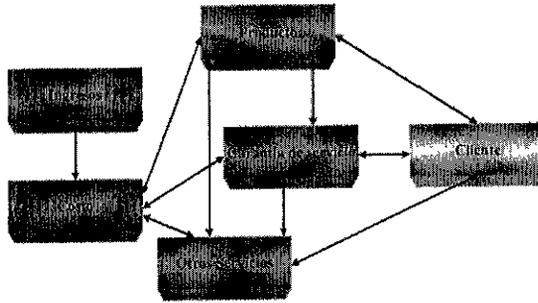
Extraen dependencias entre elementos u objetos en el "Data Warehouse".

Se utilizan para predecir el valor de un objeto de datos a partir del valor de otro. Por lo regular, las dependencias no definen una

relación exacta o segura, sino un valor de probabilidad, el factor de confianza.

Un conjunto de dependencias se muestra como una gráfica que organiza todas las dependencias en una sola estructura para comunicar un conjunto complejo de relaciones.

Gráfica de Dependencias.



Un uso del análisis de dependencias es explicar y comprender las posibles causas de los cambios descubiertos, en donde se da seguimiento a un cambio en ciertos datos por una gráfica de red de dependencias para localizar cambios en otros datos que expliquen el cambio observado.

Por ejemplo, el efecto de las garantías de servicio en las ventas de productos y los servicios de seguimiento para un conjunto de clientes.

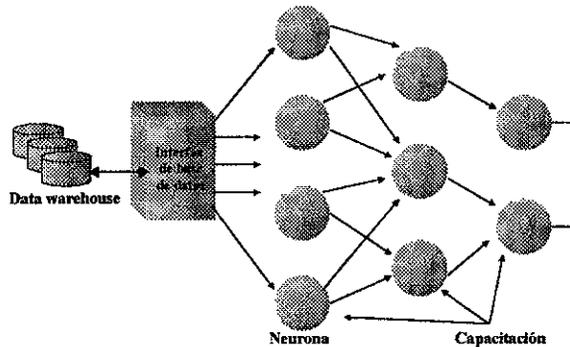
### Clasificación.

Los algoritmos de clasificación agrupan los datos en clases significativas. Se emplea este algoritmo para crear un perfil de clientes preferenciales o se usa para generar una base estándar con el fin de detectar desviaciones. Los algoritmos de agrupamiento se usan para descubrir clases en forma automática. Algunos de los algoritmos comunes de agrupamiento son el reconocimiento de patrones, la generación de perfiles, el agrupamiento lineal y el agrupamiento conceptual.

## 4-5 REDES NEURONALES.

### Redes Neuronales.

Construyen modelos aprendiendo los patrones en los datos que se analizan. Son clasificadores de tipo implícito. Se crean reglas implícitas ya que no se perciben reglas explícitas.



Esta técnica emplea "neuronas" o nodos interconectados que se organizan en capas. Por lo regular constan de tres capas: de entrada, media y de salida. Cada neurona evalúa los valores de entrada, calcula el valor total de entrada, compara el total con el mecanismo de filtrado y en seguida determina su propio valor de salida. El comportamiento se modela conectando un conjunto de neuronas. El aprendizaje ocurre modificando la "fuerza de conexión" o los parámetros que conectan las capas. Se acondicionan con muestras adecuadas de la base de datos.

Aprenden en forma supervisada o no supervisada. En la modalidad supervisada, la red neuronal intenta predecir los resultados para ejemplos conocidos. Compara sus predicciones con la respuesta objetivo y aprende de sus errores. Las redes neuronales supervisadas se emplean para predicción, clasificación y modelos de series históricas. El aprendizaje no supervisado es eficaz para la descripción de datos, pero no para la predicción de resultados. Las redes supervisadas crean sus propias descripciones y validaciones de clase y trabajan exclusivamente a partir de los patrones de datos. Se ven afectadas por tiempos prolongados de aprendizaje. Debido a que actúan como una caja negra, algunos analistas no confían en ellas.

### Detección de Desviaciones.

Extrae desviaciones o anomalías de los datos, algunas de las desviaciones de interés son:

- Anomalías que no se ajustan a las clases estándar.
- Clases muy diferentes a su padres o hijos (estructura de datos).
- Cambios de valor de un período histórico al siguiente.
- Valores extremos que están al margen de un patrón.

El análisis de desviación ayuda a filtrar y eliminar extracciones que no son del interés del motor de descubrimiento de conocimientos o datos que no coinciden, por ejemplo, el lavado de dinero. Al mismo tiempo, una desviación o anomalía puede mostrar un hecho de gran interés.

## **4-6 OTRAS TÉCNICAS Y HERRAMIENTAS DE DATA MINING.**

### ***Sistemas de Visualización.***

Permiten a los analistas hacer descubrimientos, analizando datos de manera gráfica con suficientes variables lo que permite detectar patrones y relaciones, que serían difíciles de determinar mediante algoritmos.

La representación visual de coordenadas paralelas, permite al analista empresarial exhibir al mismo tiempo las relaciones entre muchas variables, indicando coordenadas como ejes paralelos en vez de en los ejes perpendiculares de las herramientas de visualización tradicionales. Esta representación de conjuntos de datos con variables múltiples que conserva toda la información y transforma las relaciones con variables múltiples en patrones de dos dimensiones bien definidos, ayuda a manejar grandes conjuntos de datos y apoya el análisis visual y los métodos de consulta para análisis complejos. El uso de esta técnica de visualización se está difundiendo en aplicaciones de análisis de defectos de manufactura, aplicaciones de investigación médica y análisis de comercialización, para ayudar a identificar las preferencias del comprador y las tendencias de compra.

**Sistemas de Información Geográfica.**

Relacionan datos del “Data Warehouse” en diferentes ubicaciones físicas con representaciones geográficas. El analista aprecia los datos en un contexto geográfico y compara territorios para un mismo producto o diferentes productos para un mismo territorio. También es posible analizar los datos temporales en el “Data Warehouse” para ver los cambios a través del tiempo para ventas, inventarios, etc. dentro de áreas geográficas determinadas.

**Análisis Fractal.**

Las bases de datos multidimensionales proporcionan una abundante información analítica y tienen un tiempo de respuesta satisfactorio, el análisis fractal pretende identificar los patrones usando la ciencia del caos y en seguida empleando fractales para almacenarlos. La meta consiste en ofrecer respuestas estilo OLAP.

**Motores de Descubrimiento de Propietario.**

Algunos analistas han desarrollado algoritmos para construir herramientas de descubrimiento de conocimientos, estas herramientas generan reglas similares a un sistema experto a partir de datos históricos y luego, estas reglas se aplican en datos nuevos para pronóstico.

## CONCLUSIONES.

El modo de hacer las cosas ha venido cambiando en forma sustancial, el uso de tecnología en el quehacer cotidiano dentro de organizaciones de todo tipo ha marcado la pauta, el entorno global ha generado mercados potenciales nunca imaginados y han sido integradas millones de personas en organizaciones que ofrecen productos y/o servicios en diversas áreas geográficas con una misma filosofía, este hecho es objeto de análisis dada la diversidad de circunstancias que prevalecen en diferentes mercados ya que se ven involucradas diferentes idiosincrasias, circunstancias geográficas, políticas, etc.

El caso ideal sugiere integrar información de clientes y productos en un mismo formato para facilitar su manejo y explotación, aunque esto por lo general podría cumplirse, la información de operación es generada en distintas entidades regionales y en diferentes equipos de procesamiento informático (hardware). Contar con diferentes plataformas de almacenamiento de información, es una de las principales razones por las cuales el concepto "Data Warehouse" (almacén corporativo de datos), funciona adecuadamente como concentrador de información corporativa es decir, un repositorio central con información validada por las fuentes, no importa de que región provenga ésta, la fuente garantiza su integridad y unicidad ello provee una excelente fuente de información para análisis de mercado, clientes, productos etc. ya que mismos productos pueden ser ofertados en diferentes áreas geográficas con diferentes costos o ligeras modificaciones (considere el caso de los seguros para automóvil ofertados en el centro del país y los ofrecidos en el Distrito Federal), asimismo se promueve el conocimiento de clientes y productos (considere la ventaja que ofrece tener clientes de bajo riesgo), y permite el análisis temático ya que la especialización de personas en áreas concretas ayuda a analizar un mismo tema a nivel global (considere la fusión BBVA Bancomer, la cartera de la institución es analizada en Chile mientras que la Gestión en España etc. ), lo que además de promover una política de grupo permite aplicar conocimiento ya adquirido de un mercado a otro (considere el caso de las tarifas regionales tanto en España como en México, un caso que sin duda ofrece una ventaja competitiva e innovadora).

La tendencia es centralizar información de operación validada por las fuentes generadoras en un mismo formato, ello provee información para análisis e interfaces a otros sistemas, así como la creación de equipos de trabajo especializados que puedan interactuar con información global, por lo que la determinación de oportunidades de mercado en un área geográfica podría ser aplicada en algunas otras con resultados igualmente exitosos,

respecto de la operación en la organización puede observarse una sensible disminución en la cantidad de personal ya que mucho del trabajo que solía ser realizado regionalmente, ahora gracias a la disponibilidad de información global, podrá ser realizado a nivel organizacional.

Otro perfil de explotación que ofrece tener un repositorio central de datos (Data Warehouse) es la posibilidad aplicar técnicas estadísticas para determinar segmentos de mercado capaces de maximizar utilidad, se ha llegado a un punto en el desarrollo de Software y Hardware en que es posible contener información en un arreglo conveniente que exprese la trayectoria de la organización a través de sus variables, perfectamente distinguidas por la clave de la fuente generadora y etiquetadas con la fecha de origen, algunos efectos del análisis de información corporativa podrían sugerir la conveniencia de un efecto de reposicionamiento es decir, contemplar nuevas fuentes de ingresos, promover una nueva cultura de crédito y riesgo, organización basada en negocios, programas de reducción de costos, atracción de segmentos de clientes altamente productivos, aprovechamiento de la franquicia, etc. Todo ello con la garantía de que las conclusiones preliminares han sido basadas en la trayectoria de la organización es decir, están en función de su capacidad histórica.

# GLOSARIO

## A

<input type="checkbox"/> <b>Adición.</b>	Acción de combinar datos de diversas fuentes para formar una unidad de información más compleja, que con frecuencia responde consultas hechas al Data Warehouse. Casi todas las bases de datos operacionales guardan datos en la unidad más sencilla posible (normalización).
<input type="checkbox"/> <b>Agregar.</b>	Incorporar múltiples fuentes de datos o dimensiones que exprese el global de una dimensión a través de objetos medida. Ej. Dimensión: Zona Geográfica, Objeto Medida: km <sup>2</sup>
<input type="checkbox"/> <b>Análisis dinámico de datos.</b>	Comparar valores de diferentes fuentes de datos y de dimensiones distintas.
<input type="checkbox"/> <b>ANOVA.</b>	Siglas de Analysis of Variance -análisis de varianza- Prueba si un valor o conjunto de valores, en un conjunto de está dentro del rango normalmente esperado.
<input type="checkbox"/> <b>Ah-hoc</b>	Consulta diseñada a la medida, (Desarrollo a la medida)

## B

<input type="checkbox"/> <b>Base de datos operacional.</b>	Una base de datos que apoya sistemas de software que están soportando las operaciones en línea de una organización. También llamada base de datos OLTP -On Line Transaction Processing-
--	---

## C

<input type="checkbox"/> <b>Case.</b>	Siglas de Computer Aided Software Engineering -ingeniería de software asistida por computadora-. Una metodología de desarrollo de software soportada por herramientas basadas en computación para el apoyo de tareas de análisis, diseño y desarrollo.
<input type="checkbox"/> <b>CHAID.</b>	Siglas de Chi-squared Automatic Interaction Detection -detección de interacción automática de Ji cuadrada-, que es un algoritmo que iterativamente segmenta conjuntos de datos en subconjuntos mutuamente excluyentes basados en su efecto sobre variables nominales.
<input type="checkbox"/> <b>Consolidación de datos.</b>	Proceso de sintetizar porciones de información en bloques únicos de conocimiento esencial.

## D

<input type="checkbox"/> <b>Datos de herencia.</b>	Los datos de herencia están fuera de línea, en archivos históricos se cargan al Data Warehouse en referencia de fecha de generación. Proporcionan valiosa información para el
--	---

---

# GLOSARIO

---

	<b>minado de datos.</b>
<input type="checkbox"/> <b>Datos derivados.</b>	Datos producidos por cálculos o procesos que aplica el Data Warehouse a los datos provenientes de las fuentes. Los datos derivados sólo residen en el Data Warehouse y éste los usa para precalcular varios valores que se requieren para responder solicitudes. La ventaja de guardar datos derivados es evitar calcular cada vez que se ejecuta la solicitud, lo cual provoca una <b>disminución significativa en el desempeño.</b>
<input type="checkbox"/> <b>Depósito.</b>	Almacén de datos con estructura adecuada para administrar información de diferentes fuentes en un mismo formato.
<input type="checkbox"/> <b>Dilema</b>	Situación que requiere el hacer una elección entre dos o más alternativas.
<input type="checkbox"/> <b>Dimensión.</b>	La dimensión de una variable es expresada en el nivel de detalle en que puede ser referida. Ej. Considere la dimensión tiempo (año, semestre, trimestre, mes semana día, etc.)

## E

<input type="checkbox"/> <b>Esquema de bases de datos.</b>	Especificación de cómo está organizada una base de datos. (Ej. Relacional, multidimensional, etc.)
<input type="checkbox"/> <b>Estructura.</b>	Es una forma de organizar usando un conjunto de analogías simplificadas para ayudar a convertir una situación compleja en componentes entendibles.
<input type="checkbox"/> <b>Extracción de datos.</b>	Generalmente obtenidos a partir de consultas en función de metadatos definidos en la estructura de almacenamiento.
<input type="checkbox"/> <b>Extracción.</b>	Actividad de transferir datos de bases de datos operacionales (fuentes de datos) al Data Warehouse.

## G

<input type="checkbox"/> <b>Generalizar.</b>	Tendencia que garantiza un cierto comportamiento en un conjunto de datos.
<input type="checkbox"/> <b>Granulidad.</b>	Término que se usa para expresar niveles de detalle en la información. A más alto nivel de granulidad, más bajo nivel de detalle.

## H

<input type="checkbox"/> <b>Hardware.</b>	Equipo físico de computo. Ej. Monitor, impresora, CPU, etc.
---	---

---

# GLOSARIO

---

## I

- Indexado o indizado.** Llave de acceso a tablas de datos, campos con información única (no repetida para ningún otro registro. Ej. Curp) método de acceso para buscar y recobrar registros de la base de datos.

---

  - Índice de mapa de bits.** Método para indexar datos en el que casi todas las operaciones en los registros almacenados se ejecutan en los índices evitando consultar datos repetidos. Ejecutando operaciones principalmente en índices, la cantidad de lecturas a la bases de datos se reducen significativamente.

---

  - Integración de datos.** Actividad de combinar datos de diversas fuentes en el Data Warehouse, expresa tendencias por zonas o resultados a nivel grupo.
- 

## J

- "Join"** Comando de "SQL" para generar un reporte resultado de la unión de dos o más consultas a la base de datos.
- 

## L

- Limpiar y pulir.** La información es seleccionada de acuerdo a criterios establecidos para su carga.
- 

## M

- Matriz dispersa.** Matriz en la que no todos los bloques de la rejilla están llenos de datos. Condición que se aplica al contenido de bases de datos multidimensionales.

---

  - Middleware.** Término que comúnmente se aplica al software que intercambia información en forma transparente entre aplicaciones y bases de datos. Ofrece un mecanismo de conexión abstracta entre el software de aplicación y la base de datos, ocultando al programador de aplicaciones los elementos específicos que dependen de la implantación.

---

  - Modelo de datos.** Actividad de representar las categorías de datos y relaciones entre ellas como una abstracción en forma de diagrama.
- 

## N

- Navegador de información.** Opción que permite a los usuarios examinar el directorio empresarial y los datos del Data Warehouse. Apoya a los usuarios empresariales que no tienen los conocimientos para
-

---

# GLOSARIO

---

formular consultas en lenguaje de bases de datos y a cambio ofrece una herramienta para apuntar y hacer clic para viajar por el Data Warehouse.

---

## O

**OLTP.**

Siglas de On Line Transaction Processing –procesamiento de transacciones en línea-. Define operaciones a través de transacciones pre-establecida –en el momento en que ocurren- entre el operador de datos y la base de datos.

**Orientado a temas.**

Clasificado por temas de interés empresarial.

---

## R

**“Slice and Dice”**

Herramientas para explorar algún conjunto de datos, considere análisis de una dimensión a través de este tipo de herramientas se pueden hacer proyecciones y vistas parciales a través de secciones.

**Resumen.**

Actividad de incrementar la granularidad de la información en una base de datos. Un resumen reduce el nivel de detalle y es muy útil para presentar datos en soporte para la toma de decisiones.

---

## S

**SQL**

“*Structured Query Language*”, Este lenguaje es usado para definir manipular y controlar el almacenamiento de datos en una base de datos relacional.

**Software**

Se refiere a programas, sistemas operativos, paquetería, etc.

---

## W

**Data Warehouse**

Almacén corporativo de datos que representan la operación histórica del negocio, los cuales preservan la estructura organizacional.

---

## **BIBLIOGRAFÍA.**

Data Warehouse, PRIMER, By EDS, Technology Architecture, 5400 Lagacy Drive, H3-5D-66, Plano Texas 75204, USA. December 1995.

The SAS Data Warehouse, Supporting Creativity Decision Making, By SAS Institute Inc. World Headquarters, SAS Campus Drive, Cary, NC 27513 USA. 1995.

SAS Business Solutions, Making Business Decisions in a Multidimensional World, "The SAS Business Solution for On Line Analytical Processing. By SAS Institute Inc. 1997.

Information Management Perspective Series: Beat The Odds: Making Your Data Warehouse a Winner, Joseph E. Weing, General Manager and Gregory Sannik, Senior Consultant. NCR Corporation Worldwide Services. 1996.

Information Management Perspective Series: The Data That Drives The Data Warehouse Irwin Conopolsky. Partner. NCR Corporation Worldwide Services. 1996.

Matemáticas Aplicadas a la Administración y Economía. Jagdish Arya and Robin Lardner Segunda Edición Ed. Prentice Hall (PHH) ISBN 968-880-077-5. 1989.

Serie de Compedios Schaum Teoría y Problemas de Investigación de Operaciones Richard Bronson, Ph.D. Schaum-McGraw-Hill ISBN 968-451-385-5. 1982.

Data Warehousing "la integración de Información para la mejor Toma de Decisiones" Harjinder S. Gill, Prakash C. Rao. Prentice Hall Hispanoamericana, S.A. ISBN 0-7897-0714-4 1996.

Sanders, Donald H. "Informática Presente y Futuro". Mc Graw Hill 1985, ISBN 968-451567-7

Cohen, Daniel. "Sistemas de Información para la Toma de Decisiones". Mc Graw Hill 2ª ed. México, 1996.

Kendall. "Análisis y Diseño de Sistemas" Prentice Hall Hispanoamericana México, 1991.

Orfeo, G. León. "Análisis de Decisiones" Mc Graw Hill, México.

Vineder Zurbano Rafael. "Teoría de la Decisión Empresarial". Tomo 1, Biblioteca Deusto de Dirección y Organización. Deusto España.

Gallagher CharlesA. "Métodos Cuantitativos para la Toma de Decisiones en Administración". Mc Graw Hill, México

Manual, Business Objects Explorer Course  
Business Objects Americas education Centers.

Revista, Soluciones Avanzadas "Integrando la Tecnología con los Negocios" ISSN 0188-8048

Publicación Mensual de Xview S.A. de C.V.

Números: 71, Julio de 1999.  
72, Agosto de 1999.  
73, Septiembre de 1999.

Revista, Byte, "Data Warehouse Bulding Blocks". Enero de 1997

Revista, Computer World. Marzo, 1998. México.  
Computer World. Octubre, 1998. México.

Internet: [www.byte.com](http://www.byte.com)  
[www.ibm.com](http://www.ibm.com)  
[www.sas.com](http://www.sas.com)  
[www.statsoft.com](http://www.statsoft.com)