

01132  
65

**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

**FACULTAD DE INGENIERÍA**

**SISTEMA DE INFORMACIÓN  
CORPORATIVA DE LA C F E**

**T E S I S**  
QUE PARA OBTENER EL TÍTULO DE:  
**I N G E N I E R Í A E N**  
**C O M P U T A C I Ó N**  
P R E S E N T A :  
**OSCAR MIRANDA MIRANDA**

**DIRECTOR DE TESIS:**  
**M. en I. JORGE VALERIANO ASSEM**

**MEXICO, D.F.**

**2003**

A

**TESIS CON  
FALLA DE ORIGEN**





Universidad Nacional  
Autónoma de México



**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.

Autorizo a la Dirección General de Bibliotecas de la UNAM a digitalizar en formato electrónico e imprimir el contenido de mi trabajo con fines de estudio.

NOMBRE: Oscar Miranda  
Miranda

FECHA: 1 - DIC - 2007

FIRMA: [Firma]

A Dios, por que sin el nada sería posible.

Con amor, a mi esposa: Lily, por compartir su vida conmigo.

A mis hijos: Johnatan, por su amistad de tantos años y por enseñarme a ser Papá; Alex, por que siempre nos contagia con su inagotable energía; Ricky, por recordarnos que debemos decirnos: Te amo.

A la familia, por que con ellos caminamos por el mundo.

A mis amigos, de ayer, de hoy y de siempre, por que con ellos construimos el mundo.

A mi Director de tesis: M. en I. Jorge Valeriano Assem, por sus consejos, pero sobre todo por su paciencia.

A la Universidad Nacional Autónoma de México. La primera institución en la que me he sentido orgulloso de formar parte.

A la Comisión Federal de Electricidad. Por que en ella he acumulado experiencia.

## *Indice*

<b>1</b>	<b>Introducción.....</b>	<b>3</b>
<b>2</b>	<b>Antecedentes.....</b>	<b>4</b>
2.1	Estructura de la CFE.....	4
2.2	Sistema actual para el manejo de la información en la CFE.....	5
2.3	Problemática que presenta el actual sistema para el manejo de la información en la CFE.....	6
2.4	Propuesta de solución para la problemática presentada.....	6
2.5	Alcance.....	7
<b>3</b>	<b>Marco Teórico.....</b>	<b>9</b>
3.1	Bases de Datos Relacionales.....	9
3.1.1	Modelo de datos.....	10
3.1.2	Normalización.....	11
3.1.3	Lenguaje estructurado de consulta (SQL).....	12
3.2	Bases de Datos Multidimensionales.....	12
3.2.1	Introducción.....	12
3.2.2	Sistemas OLTP vs. Sistemas OLAP.....	13
3.2.3	Elementos de las bases de datos multidimensionales.....	15
3.2.4	Esquema estrella.....	16
3.2.5	Esquema Copo de nieve.....	18
3.3	Minería de Datos.....	18
3.3.1	Introducción.....	18
3.3.2	Procesamiento informático.....	19
3.3.3	Procesamiento analítico.....	19
3.3.4	Minería de datos.....	19
3.3.5	Usuarios de la minería de datos.....	21
3.3.6	Surgimiento de aplicaciones de minería de datos.....	21
3.3.7	Tecnologías y herramientas de la minería de datos.....	21
3.3.8	Análisis estadístico.....	21
3.3.9	Uso de las herramientas de análisis estadístico.....	22
3.3.10	Características de las herramientas de análisis estadístico.....	22
3.3.11	Aplicación del análisis estadístico.....	23
3.3.12	La Minería de datos y la inteligencia artificial.....	24
3.3.13	Redes Neuronales.....	26
3.4	Comunicación entre Procesos.....	29
3.4.1	Introducción.....	29
3.4.2	SOCKETS.....	30
3.4.3	Llamado de procedimientos remotos (RPC).....	30
3.4.4	CORBA.....	31
3.5	Transformación del modelo Estrella al Modelo Cubo.....	35
3.5.1	Introducción.....	35
3.5.2	Alcance.....	35
3.5.3	Conceptos Generales.....	36
3.5.4	Análisis y Representación de Almacenes de Datos.....	36
3.5.5	Modelo Conceptual de Estrella.....	37
3.5.6	Modelo Conceptual de Cubos.....	38
3.5.7	Transformación de estrellas a cubos.....	43
3.5.8	Conclusiones y Recomendaciones.....	47
<b>4</b>	<b>Arquitectura general del Almacenes de Datos.....</b>	<b>48</b>
<b>5</b>	<b>Construcción del Sistema de Información Corporativa.....</b>	<b>50</b>

<b>5.1 Análisis de información y construcción de modelos de datos.</b>	<b>50</b>
5.1.1 Alcance.....	51
5.1.2 Desarrollo de un Documento de Análisis de Sistema de Información.....	51
5.1.3 Análisis del Sistema Gerencial Comercial (GECOM).....	58
5.1.5 Conclusión de la sección de análisis.....	65
<b>5.2 Extracción/Transformación de información.</b>	<b>65</b>
5.2.1. Leer la información legada.....	65
5.2.2. Determinar información actualizada.....	66
5.2.3. Generalizar llaves para dimensiones cambiantes.....	66
5.2.5. Crear imágenes de los registros.....	66
5.2.6. Migrar la información al almacén.....	66
<b>5.3 Inserción/Carga de Información.</b>	<b>66</b>
5.3.1 Ordenar los registros cargados para crear agregados.....	67
5.3.2 Generalizar llaves para registros agregados.....	67
5.3.3 Cargar masivamente los registros checando integridad referencial.....	67
5.3.4 Manejar excepciones del proceso de carga.....	68
5.3.5 Asegurar calidad de los datos cargados.....	68
5.3.6 Publicar.....	68
<b>5.4 Diseño de la Base de Datos Multidimensional.</b>	<b>68</b>
5.4.1 Consideraciones en el diseño.....	68
5.4.2 Claves de diseño.....	69
5.4.3 Ejemplo de diseño con un cubo con información del proceso comercial en la CFE.....	69
<b>5.5 Herramientas de explotación de los datos</b>	<b>72</b>
5.5.1 Análisis en línea de procesos (OLAP).....	72
5.5.2 Sistema de Información para la Alta Dirección o Tablero de Control de la CFE.....	79
5.5.3 Cuadro de Mando Integral.....	83
5.5.4 El Portal de la CFE como cliente del Almacén de datos.....	98
<b>6 Sensibilidad al Cambio.</b>	<b>102</b>
<b>7 Aportaciones a la Ingeniería en Computación.</b>	<b>104</b>
7.1 Introducción.....	104
7.2 Estructura del temario.....	104
7.3 Asignación de tiempos para el temario.....	105
<b>9 Referencias Bibliográficas</b>	<b>110</b>

## **Sistema de Información Corporativa de la CFE**

### **1 Introducción**

El presente trabajo de tesis tiene como objetivo mostrar la aplicación de la tecnología de información, particularmente de la inteligencia de negocios, a un problema de disponibilidad de información como apoyo al proceso de toma de decisiones en las áreas directivas de la Comisión Federal de Electricidad.

Esta tesis se ha dividido en 8 capítulos. El capítulo 1 es esta introducción.

El capítulo 2: Antecedentes, presenta un panorama actual de la CFE; un análisis de su sistema para el manejo de información antes del desarrollo del proyecto objeto de esta tesis; las dificultades de este sistema y propone las alternativas de solución, así como el alcance del presente documento.

En el capítulo 3: Marco teórico, se muestra un marco teórico mínimo necesario para la comprensión del trabajo desarrollado. Debido al alcance de este trabajo, los temas se presentan a manera de resumen. Es probable que este marco teórico resulte insuficiente. Sin embargo, se decidió no ahondar más en estos temas para dejar espacio al desarrollo del proyecto. Si el lector considera necesario, puede consultar la bibliografía para adentrarse más en el estudio de esta teoría básica.

En el capítulo 4: Arquitectura general del almacén de datos, como su nombre lo indica se presenta la arquitectura general del proyecto.

En el capítulo 5: Construcción del sistema de información corporativa, se desarrolla el proyecto. Es sin duda, el capítulo más importante.

El capítulo 6: Sensibilidad al cambio, trata de la importancia del manejo del cambio, ya que el principal obstáculo de los proyectos de gran envergadura no es de origen tecnológico. El principal obstáculo se da en la gente de la organización, cuando no se ha hecho un adecuado manejo del cambio.

En el capítulo 7: Aportaciones a la ingeniería en computación, se analiza la importancia de la inteligencia de negocios en la ingeniería en computación y se propone un temario para incluir estas disciplinas en el plan de estudios de la carrera.

Por último, el capítulo 8 presenta las conclusiones.

TESIS CON  
FALLA DE ORIGEN

## 2 Antecedentes

### 2.1 Estructura de la CFE

La Comisión Federal de Electricidad es una empresa pública descentralizada del gobierno federal, la CFE genera, transmite y distribuye energía eléctrica en todo el territorio nacional para 19.31 millones de clientes, lo que representa casi a 80 millones de mexicanos.

La CFE cuenta con una capacidad instalada de 36,659 MW en 155 centrales generadoras, 36,305 km de líneas de transmisión y 624,793 km de líneas de distribución.

Para hacer que esta infraestructura se mantenga productiva, la CFE cuenta con más de 60,000 trabajadores activos, el organigrama básico de la CFE se muestra en la figura 1.1.

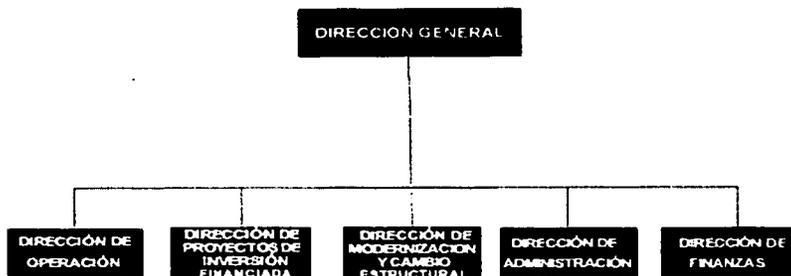


Figura 1.1 Organización de la CFE (Direcciones).

La Dirección de Operación agrupa a las siguientes subdirecciones:

- Generación
- Transmisión, Transformación y Control (TTYC)
- Distribución
- Técnica

Las subdirecciones de generación, TTYC y distribución por su cobertura nacional cuentan con aproximadamente 40 gerencias regionales.

Las principales actividades productivas de la CFE se realizan en las áreas de la Dirección de Operación, y es aquí también donde se genera casi toda la información de la empresa. A continuación se muestra la cadena de valor de la CFE. (ver figura 2.2)

TESIS CON  
FALLA DE ORIGEN

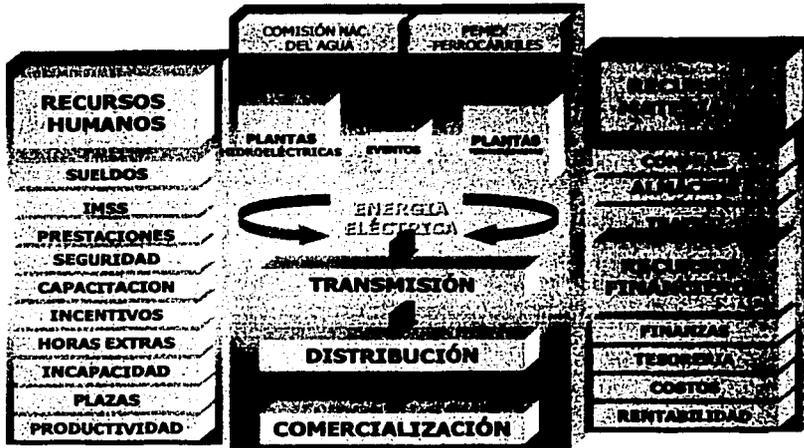


Figura 2.2 Cadena de valor de la CFE (Modelo de Negocios).

## 2.2 Sistema actual para el manejo de la información en la CFE

La CFE cuenta con diversos sistemas como apoyo a los procesos productivos (sistemas sustantivos), y sistemas que apoyan a los procesos de administración y finanzas (sistemas adjetivos).

En los primeros años de vida de la CFE, el desarrollo de los sistemas de información era una actividad centralizada por un departamento que satisfacía todas las necesidades de información de la institución apoyándose para esto con computadoras PDP-11 y más tarde VAX. Conforme la empresa fue creciendo las principales áreas de operación se avocaron a desarrollar sus propios sistemas de información.

Actualmente existe un área de desarrollo informático casi en cada una de las gerencias regionales que es coordinada por un área en su subdirección correspondiente. Las direcciones de administración y finanzas también cuentan con sus áreas de desarrollo informático. De esta forma cada área en la CFE cuenta con suficiente capacidad para resolver sus necesidades de tecnología de información.

En los últimos años las áreas de la CFE se han preocupado por desarrollar o implementar sistemas de información integrados y que aprovechen la infraestructura de cómputo y telecomunicaciones con que cuenta la institución.

Como ejemplos de estos sistemas tenemos la implantación desde 1998 del ERP R/3 de la empresa SAP y el desarrollo del Sistema Institucional de Recursos Humanos (SIRH), para las áreas de administración y finanzas.

TESIS CON  
FALLA DE ORIGEN

### **2.3 Problemática que presenta el actual sistema para el manejo de la información en la CFE.**

Aunque ha habido avances importantes en materia informática, debido a la falta de coordinación central, la comunicación entre aplicaciones de diferentes áreas para compartir información, es ineficiente y en muchas ocasiones no se ha sistematizado.

Los principales problemas que se detectan son los siguientes:

- Sistemas heterogéneos que tienen pocas posibilidades de intercomunicación.
- Problemas de consolidación y de integración de información a los niveles superiores.

Por lo anterior, aunque las áreas operativas y los mandos medios en la institución siempre tienen información en bases de datos de lo que ocurre en su ámbito, los cuadros directivos no cuentan con sistemas automáticos que proporcionen la información generada con la operación de la institución de una forma eficiente.

Es frecuente que los reportes que llegan a los directivos provengan de hojas de cálculo y que las cifras contenidas no reflejen la operación real por que hayan sido previamente manipuladas.

También es frecuente que diferentes directivos; por ejemplo: el Director General, el Director de Operación, y el Secretario General del Sindicato, en momentos diferentes de un mismo evento dan una cifra diferente de la Generación de la CFE. En realidad es probable que ninguna de las cifras sea incorrecta, el error frecuente es que la cifra que se da no se haya construido usando los mismos criterios.

Resumiendo, el problema a resolver es el siguiente: ***No se cuenta con un sistema de información que proporcione apoyo a los procesos de toma de decisiones en las áreas directivas de la institución, particularmente en la Dirección General.***

### **2.4 Propuesta de solución para la problemática presentada.**

Para solucionar estos problemas se propone la construcción de un Sistema de Información Corporativa (SICORP) basado en la tecnología de Almacenes de Datos (Data Warehouse).

El sistema deberá ser capaz de concentrar la información en un almacén de datos corporativo y proporcionar herramientas de análisis y apoyo a la toma de decisiones.

Por su naturaleza el SICORP deberá ser del tipo: Sistema de Información Ejecutiva (EIS) y Sistema para el Soporte a la Toma de Decisiones (DSS).

Los objetivos que se pretende cubrir en el presente trabajo de tesis son los siguientes:

1. El desarrollo de un Tablero de Control (Sistema EIS) que presente a los directivos Información en forma de indicadores de gestión a través de una interfase gráfica que de forma ágil y sencilla los mantenga informados de los principales aspectos en la operación de la empresa.
2. El desarrollo de un sistema de análisis en línea de proceso (OLAP) basado en cubos de información. El propósito de este sistema es proporcionar a los directivos una herramienta de apoyo al proceso de toma de decisiones.
3. El desarrollo del almacén de datos basado aplicando para esto las técnicas y metodologías requeridas en la construcción de un *Data Warehouse*.

Debido a que el SICORP esta basado en las metodologías y tecnologías de almacenes de datos (*Data Warehouse*) e Inteligencia de Negocios (*Business Intelligence*) en el desarrollo de este documento se ocuparan estos términos de forma indistinta dependiendo el contexto de la sección que se esté desarrollando.

El SICORP se convertirá en un sistema estratégico en la CFE ya que será un concentrador de información corporativa de gran valor en la institución. Su cobertura será completa en la CFE. La figura 2.3 muestra el posicionamiento del SICORP en el ambiente de procesos e información de la CFE.

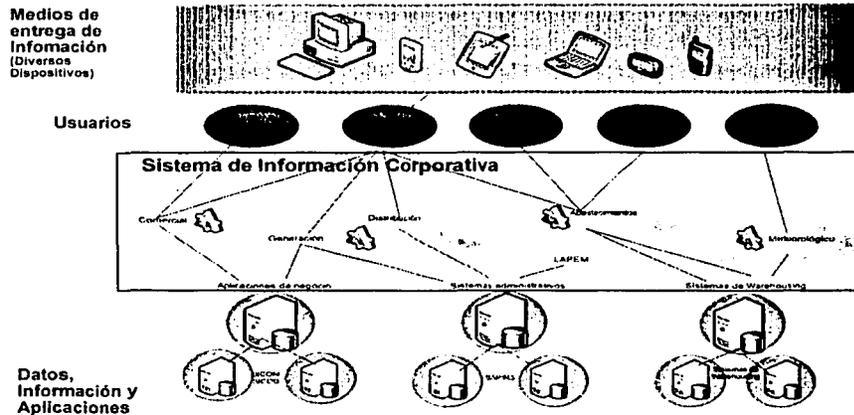


Figura 2.3. Posicionamiento del SICORP en el ambiente de procesos y sistemas de la CFE

## 2.5 Alcance.

Los sistemas de almacenes de Datos (*Data Warehouse*) comúnmente son proyectos muy extensos cuyo desarrollo y crecimiento puede ser un trabajo para varios años. Esta situación se agrava particularmente en empresas muy grandes como es el caso de la Comisión Federal de Electricidad.

Considerando lo anterior, es necesario acotar el alcance del presente proyecto de tesis en sus aspectos prácticos; a fin de que el tiempo y esfuerzo que se invierta sea razonable.

En todo caso se buscará explorar y aplicar toda la tecnología y metodologías disponibles para la construcción de un sistema de almacén de datos, seleccionando para esto un subconjunto de la información de la CFE y generando un producto completamente funcional, lo más representativo de las tecnologías que intervienen en el trabajo de construcción de un gran almacén de datos.

Aunque por motivos de tiempo es necesario acotar la parte práctica del proyecto, en lo que respecta a los aspectos teóricos en la construcción de un sistema de data warehouse se buscará no limitarse a efecto de lograr un trabajo lo más completo posible.

Por lo anterior, el presente documento de tesis no incluye todo el trabajo de planeación, análisis, diseño e implementación del Sistema de Información Corporativa de la CFE, ya que es muy extenso. Sin embargo, se buscará que se incluya todo el material práctico necesario para la clara exposición del proceso de creación de un almacén de datos corporativo en cualquier empresa de mediana a grande.

Un proyecto de almacén de datos corporativo incluye el software de Extracción, Transformación, Transporte y Carga de datos al Almacén; así como las herramientas de acceso a los datos: Análisis en línea de procesos OLAP, tableros de control y minería de datos.

### 3 Marco Teórico

El presente capítulo tiene como propósito presentar un marco teórico mínimo necesario para la comprensión de la teoría general de sistemas de almacenes de datos que se desarrollará en los siguientes capítulos.

Aún cuando el desarrollo completo de cada uno de los temas tratados en este capítulo puede ocupar volúmenes completos, el objetivo que se busca es recordar las bases teóricas necesarias para comprender los aspectos teóricos y metodológicos tratados en el presente trabajo de tesis.

#### 3.1. Bases de Datos Relacionales

Una base de datos es un sistema de mantenimiento de registros basados en computadoras, es decir, un sistema cuyo propósito general es registrar y mantener información. Tal información puede estar relacionada con cualquier cosa que sea significativa para la organización donde el sistema opera; en otras palabras, cualquier dato necesario para los procesos de toma de decisiones inherentes de la administración de esa organización.

El software, el hardware y los procedimientos para manejar la base de datos conforman un sistema de administración de bases de datos (DBMS), este sistema hace posible acceder datos integrados que cruzan límites operacionales u organizativos dentro de una empresa.

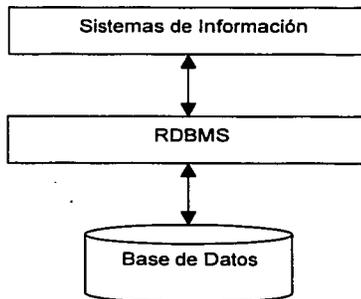


Figura 3.1 Esquema básico de un sistema de base de datos.

En una empresa que utilice un sistema de bases de datos debe existir una persona específica cuya responsabilidad central sea controlar los datos de operación. Esta persona es el administrador de bases de datos (DBA).

Algunas de las ventajas de tener un control centralizado de datos son:

- Puede reducirse la redundancia.
- Puede evitarse la inconsistencia.
- Los datos pueden compartirse.
- Pueden hacerse cumplir las normas establecidas.
- Pueden aplicarse restricciones de seguridad.
- Puede conservarse la integridad.
- Pueden equilibrarse los requerimientos contradictorios.

### 3.1.1. Modelo de datos

Una de las principales responsabilidades del administrador de la base de datos es desarrollar un modelo conceptual también llamado modelo de la empresa, el modelo empresa, las entidades y sus relaciones es la herramienta utilizada para representar a la reorganización conceptual de los datos. Este modelo conceptual se utiliza para organizar, visualizar, planear y comunicar ideas de manera abstracta.

#### Modelo Entidad-Relación (MER)

El modelo entidad relación permite crear el diseño de un fenómeno real para un sistema de base de datos obteniéndose un diagrama basado en tres elementos: entidades, atributos y relaciones.

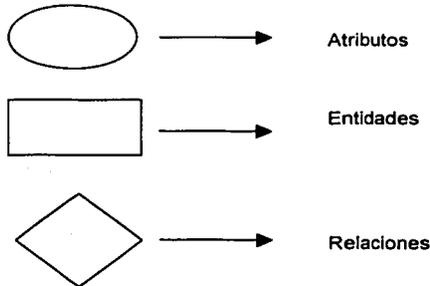


Figura 3.2 Símbolos de los diagramas Entidad-Relación

**Atributos:** Es un valor descrito o propiedad asociada a una entidad; sus características son:

- Valor asociado
- Rango de validación
- Estructura de datos asociada

Los atributos pueden ser: nombre, calle, ciudad, etc.

**Entidad:** Es un objeto que contiene ciertos atributos que lo caracterizan como un elemento básico que se desea administrar, guardándose información del mismo; sus características son:

- Debe tener existencia propia
- Debe ser un concepto genérico
- Debe contar con propiedades de interés

Las entidades pueden ser:

- a) Concreta, como una persona, un automóvil o un libro.
- b) Abstracta, como una fecha, edad o un concepto.

**Relación:** Asociación entre dos o más entidades y corresponde al número de ocurrencias que pueden enlazarse de una entidad a otra.

La mayoría de las relaciones son BINARIAS; no obstante, pueden existir relaciones que incluyan a más de dos conjuntos de entidades. Ejemplos:

- cuenta-cliente: conjunto de relaciones binarias.
- cuenta-cliente-sucursal: conjunto de relaciones ternarias.

Una relación también puede tener atributos descriptivos. Por ejemplo, la relación Cuenta-Cliente entre las entidades cliente: Juan y cuenta: 1000 puede tener el par [(fecha, Oct 12 1997)], indicando la última vez que interactuaron las entidades, es decir, la fecha del último acceso del cliente a su cuenta.

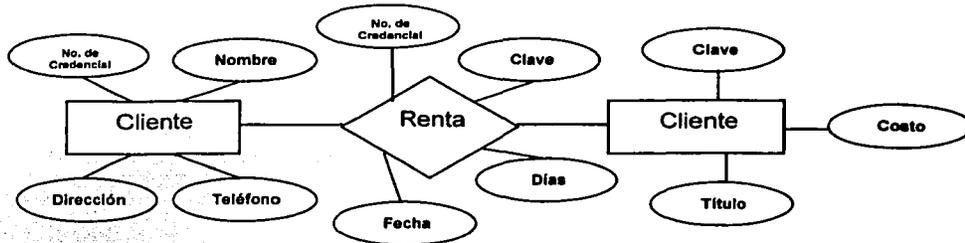


Figura 3.3 Ejemplo de diagrama Entidad-Relación

Por otro lado existe la cardinalidad del mapeo, con la cual se puede especificar la cantidad de entidades que podrán asociarse mediante una relación.

Las cardinalidades existentes para dos conjuntos de entidades A y B y conjunto de relaciones R; estas son:

1. UNA A UNA: Una entidad de A puede asociarse únicamente con una entidad de B.
2. UNA A MUCHAS: Una entidad de a puede asociarse con cualquier cantidad de entidades de B.
3. MUCHAS A UNA: Cualquier cantidad de entidades de A puede asociarse con una entidad de B.
4. MUCHAS A MUCHAS: Cualquier cantidad de entidades de a puede asociarse con cualquier cantidad de entidades en B.

### 3.1.2. Normalización.

Este proceso sirve para agrupar a los campos de datos en tablas que representan a las entidades y sus relaciones. La teoría de la normalización está basada en la observación de que un cierto conjunto de relaciones tiene mejores propiedades en un medio de inserción, actualización y supresión, que las que tendrían otros conjuntos de relaciones conteniendo los mismos datos.

La razón de usar el procedimiento de normalización es asegurar que el modelo conceptual de la base de datos funcionará. Esto no significa que una estructura no normalizada no funcionará, sino que puede causar algún problema cuando los programadores de aplicación traten de modificar la base de datos.

Las formas normales son:

- Primera forma normal
- Segunda forma normal
- Tercera forma normal

**Primera forma normal:** El primer paso de la normalización consiste en transformar los campos de datos a una tabla de dos dimensiones. Lo que se requiere usualmente en este paso es la

eliminación de ocurrencias repetidas de campos de datos, de tal manera que se obtenga un archivo fijo.

**Segunda forma normal:** Es establecer las claves y relacionarlas con los campos de datos. En la primera forma normalizada, el renglón entero de la tabla (cadena) depende de todos los campos de claves. En esta forma se hace un intento de establecer los campos de datos que están relacionados con alguna parte de la clave completa. Si los campos de datos sólo dependen de una parte de la clave, la clave y los campos conectados a la clave parcial son susceptibles de separarse en registros independientes. En conclusión esta forma pretende la división de la primera tabla normalizada, en una serie de tablas en las que cada campo sólo depende de la clave completa.

**Tercera forma normal:** Esta normalización consiste en separar los campos de las segundas relaciones normales que, aunque dependan sólo de una clave, deben tener una existencia independiente en la base de datos. Esto se hace de forma tal que la información sobre estos campos pueda introducirse separadamente a partir de las relaciones en las que se encuentra implicada.

### 3.1.3. Lenguaje estructurado de consulta (SQL)

SQL (Structured Query Lenguaje) es un lenguaje flexible y eficiente, con características que han sido diseñadas para manipulación y examen de datos relacionales. Este lenguaje es para todos los tipos de usuarios y de ambientes, por ejemplo usuarios finales, programadores o administradores de base de datos.

SQL es algo más que un lenguaje de consulta a pesar de la palabra consulta que figura en su nombre. Primero, no solo ofrece funciones de recuperación, sino también una diversidad de operaciones de actualización. Segundo, puede utilizarse tanto desde una terminal en línea como desde un programa de aplicación en línea.

La operación fundamental en SQL es la transformación que se representa sintácticamente como un bloque SELECT- FROM- WHERE.

Su sintaxis general es:

```
SELECT  {<atributo1><alias>, <atributo2><alias>,...}
FROM    {<tabla1><alias>, <tabla2><alias>,...}
WHERE   <condición>
```

## 3.2. Bases de Datos Multidimensionales

### 3.2.1 Introducción.

Las necesidades de información en los cuadros directivos de las empresas siempre han existido, es bien sabido que mientras más y mejor información se tenga de la operación de la institución que se dirige, mejores resultados podemos esperar del proceso de toma de decisiones. Esta situación se ha acentuado ahora que las empresas enfrentan una feroz competencia global, de esta forma es indispensable estudiar cuidadosamente todas las implicaciones de una decisión que afectará positiva o negativamente el rumbo de la empresa.

Con el paso del tiempo se le ha dado un valor primordial a la información, convirtiéndola en uno de los activos más valiosos de las empresas. Esto sin duda también hace evidente un proceso de cambio cultural en el que el uso de las computadoras es cada vez más intenso en cada actividad que desempeñamos, y la actividad directiva, no puede ser la excepción.

Para que se cuente con suficiente información como herramienta en el proceso de toma de decisiones en los escritorios de los altos ejecutivos adicionalmente al cambio de cultura

Informática que se ha dado en los últimos años, también ha sido necesario el desarrollo y maduración de algunas tecnologías y metodologías que anteriormente no era posible tener.

Una de estas tecnologías son las llamadas bases de datos multidimensionales, estas bases de datos incorporan mecanismos para optimizar las consultas de información analítica.

### **3.2.2. Sistemas OLTP vs. Sistemas OLAP.**

Desde el punto de vista de propósito del almacenamiento y consulta de la información almacenada en las bases de datos, actualmente existen dos tipos de sistemas: OLTP (OnLine Transaction Process) y OLAP (OnLine Analytical Process), Sistemas transaccionales y Sistemas de análisis en línea de procesos.

La mayoría de los sistemas que apoyan en la operación de la empresa, son del tipo OLTP, en este tipo de sistemas de información básicamente se busca eficiencia en los mecanismos de almacenamiento y recuperación de información y están dirigidos a procesos transaccionales. Los sistemas basados en la metodología y tecnología OLAP son básicamente sistemas del tipo DSS (Sistemas para el soporte a la toma de decisiones) y EIS (Sistemas de información Ejecutiva) que buscan reducciones notables en los tiempos de recuperación de grandes volúmenes de información y presentan características interesantes en el análisis de los datos recuperados.

La fuente de alimentación de los sistemas OLTP es la operación misma de la empresa, mientras que la fuente de alimentación de los sistemas OLAP son los sistemas OLTP ya que son estos los que mantienen cada registro generado en las múltiples transacciones que ocurren cotidianamente.

Las características de un sistema OLTP son las siguientes:

- Procesamiento de las transacciones de un negocio en tiempo real.

Los sistemas OLTP conducen esencialmente los procesos de los negocios por la ruta de las transacciones en tiempo real. Los sistemas OLTP cambian continuamente para representar el estado actual del negocio. De esta manera los sistemas OLTP procesan nuevas transacciones, los datos son actualizados o insertados dentro del sistema OLTP inmediatamente.

- Contener estructuras de datos optimizadas para entradas y ediciones.

Debido a que el rendimiento de esos sistemas es crítico para mantener la ruta de los procesos esenciales del negocio, las estructuras de los datos son optimizadas para la entrada y edición de datos.

- Limitada capacidad de soporte a decisiones.

El soporte a la toma de decisiones no es la meta prioritaria de un sistema OLTP. Los reportes en los sistemas operacionales pueden suministrar el dato más actual. Sin embargo, entrar directamente al sistema fuente puede tener un impacto negativo sobre el rendimiento del sistema y producir reportes inconsistentes debido a la volatilidad del sistema OLTP.

Las características de los sistemas OLAP son las siguientes:

La tecnología OLAP da una alternativa a la tecnología de bases de datos relacionales, ofreciendo rapidez, vista de datos flexible, análisis y navegación. Las siguientes son características de las tecnologías OLAP.

- Las bases de datos OLAP tienen un optimizado esquema para las consultas rápidas por parte de los usuarios.
- Las consultas OLAP son muy rápidas y permiten mas interactividad de uso para los usuarios.
- Las bases de datos OLAP tiene un motor de calculo robusto para análisis numérico.
- OLAP es un modelo de datos conceptual e intuitivo.
- OLAP provee una vista multidimensional de los datos.
  - Los usuarios pueden hacer drill down o drill up a través de varios niveles de sumarización de datos.  
El Drill down es una exploración hacia adentro en los niveles de detalle de los datos y el drill up es una exploración hacia fuera.
  - Los usuarios pueden rotar vistas de datos. Es decir, pueden intercambiar columnas por renglones.

Los diferentes tipos de esquemas de almacenamiento y recuperación de información OLAP son los siguientes:

- ROLAP
- MOLAP
- HOLAP

Los tres esquemas cumplen con las características funcionales descritas anteriormente. Sus diferencias se encuentran en la forma de almacenamiento y recuperación de la información y se describen a continuación:

**ROLAP:** El origen de los datos es una base de datos relacional, el diseño de la base de datos tiene la forma especial que lo habilita para las consultas a través de herramientas OLAP. La implementación de las operaciones multidimensionales se encuentra incorporada no en la base de datos sino en las herramientas de acceso a datos. La eficiencia en el cálculo de agregaciones es baja, esta situación impacta en el tiempo de respuesta. Algunas de las ventajas de este esquema es que es más económico y generalmente permite manejar información a mayor detalle, es decir, más desagregada. Como ejemplo de herramientas de tipo ROLAP tenemos a Discoverer de Oracle y MetaCube de Informix.

**MOLAP:** El origen de los datos es una base de datos multidimensional, las estructuras lógicas de los datos son cubos, en los que se destacan los hechos, las dimensiones y sus niveles jerárquicos. La implementación de las operaciones multidimensionales se encuentra incorporada en el manejador de la base de datos por lo que tienen una mayor eficiencia en general que las soluciones basadas en el esquema ROLAP. Sus principales desventajas son que son caras y cuando se trata de manejar información detallada su eficiencia baja.

**HOLAP:** La solución HOLAP es una combinación del esquema ROLAP y del MOLAP. Esto es muy usado en la implementación de grandes almacenes de datos. Su aplicación busca resolver el siguiente problema:

Trabajando el esquema ROLAP, cuando un cubo es grande, es decir, cuando tiene muchos hechos o variables y muchas dimensiones, y el cálculo de las agregaciones (Roll-up) es complejo; el tiempo requerido para procesarlo es muy alto y el espacio requerido para su almacenamiento es considerable. Esto es por que el número de posibles cruces de información y agregaciones crece de forma importante.

Para resolver este problema se toma la decisión de partir el cubo de forma que el cálculo de las agregaciones se efectúa en la modalidad de MOLAP y el manejo de información detallada de realiza en modo ROLAP, directamente de la base de datos relacional. Cuando el usuario hace una consulta, si lo hace en los niveles y subniveles más agregados no nota la diferencia entre la solución MOLAP y la HOLAP, pero si quiere accesar cierto nivel de detalle entonces el sistema pierde rendimiento respecto a una solución MOLAP pura.

Es frecuente que se decida ganar en eficiencia de almacenamiento aún cuando se pierda un poco de eficiencia en el manejo de los datos de mayor detalle.

### 3.2.3. Elementos de las bases de datos multidimensionales.

#### Tabla de Hechos.

La tabla de hechos es el centro de la base de datos multidimensional, es la tabla donde se encuentran los hechos o variables de interés en la empresa. Como su nombre lo indica, almacena hechos, datos sobre cosas concretas que suceden en la empresa. Los hechos son siempre valores numéricos, ejemplo de hechos típicos en las bases de datos multidimensionales: Ventas, Empleados, Productos, Utilidades, Pérdidas, etc.

#### Granularidad.

La granularidad de la tabla de hechos es la medida de un registro en la tabla de hechos que comunica el contexto de los datos numéricos almacenados en la tabla de hechos. La granularidad de la tabla de hechos es típicamente definida por el menor nivel del detalle almacenado en cada una de las dimensiones asociadas con la tabla de hechos.

Tomando estas dependencias sobre la dimensión de datos, la integridad de los datos en la tabla de hechos dependen sobre el mantenimiento de la granularidad de la tabla de hechos y haciendo cumplir la llave única en las tablas de dimensiones.

A la granularidad de la tabla de hechos deben adherirse los principales diseños:

- Cada tabla de hechos puede tener solamente una granularidad definida.
- Todas las medidas de una tabla de hechos deben conformar la granularidad definida.
- Cuando tienes datos en dos diferentes niveles de detalles, necesitaras implementar mas de una tabla de hechos que contengan los datos.

#### Dimensiones.

Las tablas de dimensiones mantienen información que le proporciona el contexto a los datos almacenados en la tabla de hechos.

Para ilustrar mejor el concepto dimensiones nos valdremos del siguiente ejemplo: considérese el hecho "ventas" con valor de 4000, el valor numérico almacenado tiene significado únicamente hasta que se "dimensiona" con información procedente de las tablas de dimensiones de la siguiente manera (Venta: 4000, Fecha: 02/02/2002, Región: Norte, Producto: Refrescos).

Como se puede ver en el ejemplo, las dimensiones permiten ubicar claramente un hecho en el contexto adecuado.

Una tabla de dimensión:

- Describe entidades de negocios.

En un esquema de estrella, una tabla de dimensión representa una simple entidad de negocio como puede ser un producto o un cliente.

- Contiene atributos que proporcionan el contexto para los datos numéricos.

Los atributos de las columnas son altamente correlacionados y descriptivos. Estos son usados para proveer los datos numéricos almacenados en la tabla de hechos. Por ejemplo un usuario puede solicitar las ventas mensuales medidas por cliente y región.

- Representa la organización de los datos en jerarquías.

En cada dimensión, podemos organizar los datos dentro de una o más jerarquías. Construyendo jerarquías de datos permiten a los usuarios ver los detalles y los datos sumados. Físicamente cada tabla de dimensión contiene una o muchas jerarquías. Una jerarquía contiene niveles que rompen las dimensiones dentro de categorías.

PK	Customer Id
	account num
	lname
	fname
	address1
	address2
	address3
	address4
	city
	state_province
	country

PK	State Id
	State_name
	region
	Country

PK	time Id
	the_date
	the_day
	the_month
	the_year
	day_of_month
	week_of_year
	quarter
	fiscal_period

Figura 3.4 Ejemplo de dimensiones.

#### Dimensión Padre-hijo.

Una tabla padre-hijo permite una dimensión con una jerarquía desbalanceada donde algunas de las ramas de las jerarquías contienen mas niveles que otras ramas. Una dimensión padre-hijo es similar a la dimensión copo de nieve en la que cada unión crea un nuevo nivel. La diferencia de una dimensión copo de nieve es que los niveles mas altos en la jerarquía padre-hijo derivan de la misma tabla como del menor nivel.

#### Dimensiones de tiempo.

Cerca de cada base de datos OLAP se tiene una dimensión de tiempo. Frecuentemente una dimensión de tiempo contiene meses como el menor nivel del detalle-agregado dentro de cuartos y años. Algunas veces una dimensión de tiempo contiene días como el menor nivel del detalle. En ocasiones al monitorear operaciones de manufactura la dimensión de tiempo tiene minutos o hasta segundos como el menor nivel del detalle.

Una ventaja de crear una tabla de dimensión de tiempo es que puede contener propiedades adicionales para una fecha como son la temporada para un mes y la bandera de fiestas para un día.

Una dimensión de tiempo separada reduce el total de espacio almacenado en un data warehouse. En suma diseñando y procesando una dimensión basada sobre una tabla con solo unos pocos registros es mucho más rápida que extraer los valores de la dimensión de una enorme tabla de hechos.

Una tabla de dimensión de tiempo separada puede ser utilizada con mas de una tabla de hechos. Esto permite crear y compartir dimensiones de tiempo para múltiples cubos. También permite compartir dimensiones para crear cubos virtuales y comparar medidas de múltiples cubos.

#### 3.2.4. Esquema estrella.

Para facilitar la recuperación de datos y el análisis, un data warehouse organiza los datos en estructuras fáciles de usar llamadas esquemas de estrellas.

Un esquema de estrella se caracteriza por tener una tabla central de hechos que es rodeada por tablas de muchas dimensiones que contienen descripciones desnormalizadas de la tabla de hechos.

Las características de un esquema de estrella son los siguientes:

1. El centro de la estrella es la tabla de hechos: esta tabla contiene métricas del negocio – medidas que describen los eventos específicos en un negocio como son las transacciones bancarias o el producto de las ventas. Estos pueden tener también datos agregados como son las ventas sumadas del mes.
2. Los puntos de la estrella son las tablas de dimensiones (Dim\_empleados, Dim\_productos, Dim\_clientes, Dim\_Exportación y la Dim\_tiempo). Esos puntos de la estrella son utilizados para describir la información que existe en un proceso de negocio específico y provee el contexto numérico de datos.
3. La granularidad de la tabla de hechos es una sentencia de definición de lo que representa la tabla de hechos. Esta tabla de hechos contiene los datos de las ventas por producto, empleado, exportaciones y clientes.

La tabla de hechos es la tabla central en un esquema de estrella que representa datos numéricos dentro del contexto de las unidades de negocio.

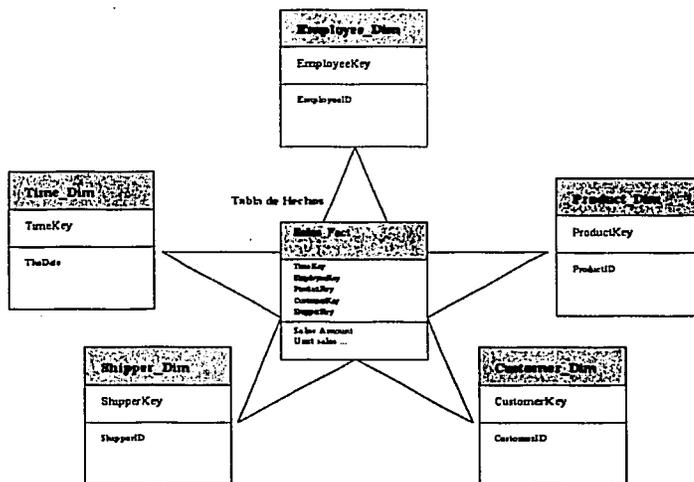


Figura 3.5. Ejemplo de Diagrama Estrella.

#### Métricas.

Una métrica es una columna numérica cuantitativa dentro de la tabla de hechos. Las métricas generalmente representan los valores que son analizados, como una unidad de ventas o número de empleados. Las medidas son generalmente numéricas y sumadas debido a que esos valores son la base desde donde se pueden realizar los cálculos.

### Llaves Foráneas.

Una llave foránea es la representación de una dimensión en la llave primaria dentro de tabla de hechos. Las llaves primarias son tomadas de las llaves primarias de cada tabla de dimensiones. La combinación de esas llaves es típicamente el identificador único para cada registro de la tabla de hechos.

#### 3.2.5. Esquema Copo de nieve.

Un esquema copo de nieve es una variación del esquema de estrella en el cual las dimensiones de jerarquías son almacenadas en múltiples tablas de dimensiones. Por ejemplo la dimensión del producto es almacenada en tres tablas: la categoría de productos, subcategoría del producto y el producto. En el ejemplo cada tabla de dimensión corresponden a diferentes niveles en la dimensión –Categoría, subcategoría y producto.

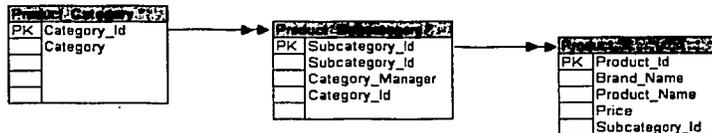


Figura 3.6 Dimensión Copo de Nieve

Porque se almacenan las dimensiones de datos a través de múltiples tablas, el esquema copo de nieve es más normalizado que un estricto esquema de estrella conteniendo simples tablas para cada dimensión. Solamente una tabla de dimensión se une a la tabla de hechos, y entonces las otras tablas de dimensiones son unidas con la relación a la llave foránea.

### 3.3. Minería de Datos

#### 3.3.1. Introducción

Herramientas de soporte de decisiones es el término genérico para referirse a las aplicaciones y herramientas del data warehouse que se emplean para recuperar, manipular y analizar los datos, y para presentar después los resultados.

Estas herramientas se utilizan en dos modalidades: verificación y descubrimiento. En la modalidad de verificación el usuario empresarial crea una hipótesis e intenta confirmarla accediendo a los datos en el data warehouse. Las herramientas que implementan la modalidad de verificación son de consulta, de sistemas de reporte y de análisis multidimensional. En la modalidad de descubrimiento, las herramientas intentan descubrir características de los datos como patrones de compra o de asociación entre la adquisición de artículos diferentes. En la modalidad de descubrimiento, el usuario empresarial no conoce ni sospecha los patrones y asociaciones descubiertos. La herramienta de minería de datos [13] es un ejemplo de la modalidad de descubrimiento.

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

### 3.3.2 Procesamiento informático.

El procesamiento informático apoya a la modalidad de verificación del soporte de decisiones. Comprende técnicas como análisis estadísticos básicos y de datos, consultas y reportes. Los datos que se acceden y procesan pudieran ser históricos o bastante recientes, y pudieran estar un poco o muy resumidos. Los resultados se presentan en forma de reportes o gráficas.

### 3.3.3. Procesamiento analítico.

También el procesamiento analítico apoya a la modalidad de verificación del soporte de decisiones. Su meta consiste en hacer que los datos estén disponibles para el usuario de la empresa en su perspectiva de las dimensiones empresariales. Este procesamiento maneja capacidades de análisis de subconjuntos, profundización (drill-down) y condensación y adición (roll-up). Los datos que se emplean en el procesamiento analítico por lo general históricos tanto al nivel resumen como al de detalle.

### 3.3.4. Minería de datos.

La minería de datos apoya la modalidad de descubrimiento del soporte de decisiones. Las herramientas de minería de datos corren los datos detallados de transacciones para desenterrar patrones y asociaciones ocultos. Por lo regular los resultados generan extensos reportes o se les analiza con herramientas de visualización de datos descubiertos.

El procesamiento informático es excelente y rentable para el despliegue masivo de consultas, análisis y reportes de datos de dos o tres dimensiones. Las herramientas de procesamiento analítico permiten diversas visualizaciones de los datos, como ventas por marca, tienda, temporada y periodo de tiempo, las cuales se pueden definir, consultar y analizar. Las herramientas de minería de datos son esenciales para comprender el comportamiento de los clientes.

El data warehouse proporciona al gerente empresarial dos ingredientes esenciales, el primero es una gran cantidad de datos sobre sus clientes así como la historia entre el cliente y la organización, el segundo es el carácter único de estos datos.

La minería de datos es una arma esencial en el arsenal del soporte de decisiones del analista. La minería de datos auxilia a los usuarios empresariales en el proceso 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 situaciones ocultas en los datos que revelan comportamientos interesantes en los procesos de negocio. Estos descubrimientos deberán proporcionar ventajas competitivas para controlar la participación en el mercado y las utilidades.

Los seres humanos tienen la agudeza para percibir excepciones y anomalías, pero no tienen la potencia y la capacidad para inferir relaciones en grandes volúmenes de datos. Sin embargo una vez extraídas las relaciones y presentadas a los analistas empresariales, estos las examinan y seleccionan las más interesantes y útiles.

Los analistas empresariales tienen un rango de necesidades. La primera necesidad es saber qué está sucediendo en el negocio. La siguiente es saber por qué está sucediendo, ¿cuál es el comportamiento de los clientes y mercados?. La última necesidad es ¿qué puede hacerse?, ¿Cuáles acciones se pueden tomar?. El valor de un analista para los gerentes es más alto cuando genera una recomendación factible.

Comprender el comportamiento y los pronósticos de clientes y mercados, y lo que puede hacerse, son los retos para las técnicas tradicionales de análisis. Las consultas, reportes y

análisis multidimensional tradicionales se concentran en lo que está sucediendo y, en menor medida, en el por qué.

La minería de datos se concentra en llenar la necesidad de descubrir el por qué para luego predecir y pronosticar las posibles acciones con cierto factor de confianza para cada predicción.

Las herramientas de minería de datos son un componente importante del sub-bloque de Análisis y reportes del bloque de Acceso. Las herramientas de minería de datos utilizan el sub-bloque de Análisis y recuperación para tener una interfaz con el data warehouse y con el mercado de datos. Muchas de las herramientas de minería de datos también emplean el componente de depósito local del bloque de Acceso y recuperación, a fin de almacenar los datos en estructuras de datos de propietario para análisis subsecuentes y presentaciones de los resultados.

La mayoría de las herramientas de minería de datos puede con facilidad saltarse al data warehouse o el mercado de datos y acceder de manera directa la fuente de datos. Tradicionalmente, las herramientas de minería de datos acceden a los datos de la fuente. Sin embargo, los datos del data warehouse o del mercado de datos están refinados, integrados y estandarizados. La estandarización eliminó aspectos como las convenciones de nombres múltiples, las estructuras ocultas de codificación y los campos faltantes. Los datos operacionales en las fuentes son por lo general inconsistentes y están dispersos en muchas aplicaciones. Además se requieren datos históricos para descubrir patrones temporales de interés.

La minería de datos difiere en varias formas del procesamiento informático y analítico. Ver tabla 3.1.

	<b>El procesamiento informático/analítico</b>	<b>minería de datos</b>
<b>Enfoque</b>	Datos de resumen	Datos de transacción o de detalle
<b>Dimensiones</b>	Limitadas	Muchas
<b>Cantidad de atributos</b>	Total de decenas	Cientos para cada dimensión
<b>Tamaño del dataset</b>	De reducido a mediano para cada dimensión	Millones para cada dimensión
<b>Enfoque del análisis</b>	¿Qué está sucediendo en el negocio?	¿Por qué está sucediendo? Acciones de predicción y pronóstico
<b>Técnica de análisis</b>	Rebanar y picar	Descubrir automáticamente
<b>Proceso de análisis</b>	Analista empresarial iniciado y controlado	Datos y sistema iniciado orientación mínima al analista empresarial
<b>Factor de confianza</b>	Derivado por el analista empresarial	Derivado de los datos
<b>Estado de la tecnología</b>	Desarrollada	Desarrollo en análisis estadístico incipiente en descubrimiento de conocimiento

Tabla 3.1 Procesamiento informático/analítico vs minería de datos.

Los datos en el data warehouse deben estar al nivel del detalle correcto. Debe de existir una estrecha cooperación entre los analistas empresariales y los profesionales de la tecnología de la información.

Para formar la mezcla correcta de actividades de minería de datos, son cruciales tres ingredientes: usuarios, aplicaciones empresariales y tecnología y herramientas.

### 3.3.5. Usuarios de la minería de datos.

Los usuarios clave en perspectiva de la minería de datos son los analistas empresariales, los peritos en estadística y los profesionales de la tecnología de la información que auxilian a los usuarios empresariales. Quienes obtienen beneficios de los resultados de minería de datos son los gerentes empresariales y los ejecutivos, que desean que desean entender los factores de éxito del negocio con base en datos completos del cliente, y utilizar luego este conocimiento para afinar las estrategias de producción, precios y comercialización; mejorar el nivel de éxito de las estrategias; e impulsar el balance.

Hasta la fecha las empresas han dependido del procesamiento informático y analítico para medir y comprender la estabilidad de un negocio. El procesamiento informático – consultas y reportes – es más sencillo de usar, pero requiere de una estrecha dirección del analista. Los analistas preguntan cuestiones específicas y verifican las cuestiones o hipótesis con los datos. Para este fin, los datos deben estar bien organizados. El procesamiento analítico (OLAP) requiere de menos dirección del analista, aunque los datos deben estar organizados en una forma especial (base de datos multidimensional), o accederse bien de manera especial (Visión multidimensional). En ocasiones se utiliza una combinación de técnicas de consulta y OLAP para comprender el comportamiento del cliente o para construir perfiles de segmentos de mercado; pero el proceso de aplicar estas técnicas es conducido esencialmente por el analista empresarial. En estos casos, este proceso también se conoce como minería de datos.

### 3.3.6 Surgimiento de aplicaciones de minería de datos.

La tecnología de minería de datos se ha utilizado principalmente en aplicaciones de comercialización, ventas y análisis de crédito; y se ha aplicado con éxito en áreas empresariales con el más alto potencial, tales como la segmentación de clientes y del mercado y el análisis de comportamiento del cliente, en particular de los sectores de menudeo, bancario y financiero. Hoy en día, una confluencia de tres fuerzas importantes conduce al crecimiento de la minería de datos:

- o La tecnología del data warehouse para proporcionar un gran banco de datos bien organizados e históricos.
- o Hardware en paralelo, producto de base de datos y herramientas a precios razonables.
- o Tecnología y herramientas para minería de datos cada vez mas desarrolladas.

### 3.3.7. Tecnologías y herramientas de la minería de datos.

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

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

### 3.3.8. Análisis estadístico.

Los sistemas de análisis estadísticos (también conocidos como análisis de datos), se usan para detectar patrones no usuales de datos. Estos patrones estadísticos se explican después, mediante modelos estadísticos y matemáticos. Algunas de las técnicas de modelado estadístico y matemático que se emplean son el análisis lineal y no lineal, el análisis de regresión continua y logística, el análisis de univariación y multivariación, y el análisis de series históricas.

Las herramientas de análisis estadístico se utilizan en diversas aplicaciones empresariales: incrementan la participación en el mercado y las utilidades, detectando las mejores oportunidades, aumentando la satisfacción del usuario por medio del mejoramiento de la calidad en productos y servicios (programas de administración de la calidad total), e impulsar los márgenes a través de la modernización de la manufactura de productos y de la logística (ingeniería de proceso empresarial). Las herramientas de análisis estadístico existen desde hace tiempo y son las herramientas más desarrolladas que se tienen para la minería de datos. Han servido para reducir el tiempo de análisis, lo que a su vez conduce a una mejor toma de decisiones.

Anteriormente, las herramientas de análisis estadístico estaban dirigidas principalmente a especialistas en estadística y profesionales para aplicaciones técnicas y de ingeniería. Sin embargo muchas empresas están experimentando las presiones paralelas de reducir costos – eliminar personal de apoyo- y un tiempo para comercialización cada vez más corto. Por lo tanto las herramientas de análisis estadístico deben evolucionar hasta el punto en donde puedan ser adoptadas con éxito y utilizadas por los analistas empresariales. Estos analistas son expertos en su campo, pero no son programadores ni peritos en estadística. Necesitan seleccionar los campos correctos del data warehouse, extraerlos y luego analizarlos. Los analistas empresariales no pueden dedicar sus escasos recursos a aprender la forma correcta de escribir programas de cómputo, manipular bases de datos y desarrollar enfoques y estrategias formales de análisis estadístico. Las habilidades clave que se esperan en los analistas empresariales son las siguientes:

- Experiencia en su campo – conocimientos del tema
- Capacidad para resolver problemas – saber como estructurar el análisis para abordar los aspectos diarios de la empresa.

Las herramientas de análisis estadístico se encuentran en un punto medio entre las herramientas para el descubrimiento de conocimientos por una parte y las herramientas para procesamiento informático por la otra.

### **3.3.9 Uso de las herramientas de análisis estadístico.**

Para utilizar una herramienta de análisis estadístico, los usuarios empresariales deben seleccionar y extraer datos adecuados del data warehouse o del mercado de datos.

A continuación los usuarios empresariales deben invocar las funciones de visualización y analítica, disponibles en las herramientas de análisis estadístico, para descubrir relaciones entre los datos y construir modelos estadísticos y matemáticos a fin de interpretar los datos. Se usa un proceso interactivo e iterativo para definir el modelo; la meta consiste en desarrollar el modelo de mejor ajuste para convertir los datos en información. Los analistas empresariales con capacidad para resolver problemas y experiencia en su campo son fundamentales para seleccionar el modelo que se ajusta mejor. El analista empresarial tiene una confianza natural en el modelado para producir resultados factibles.

### **3.3.10 Características de las herramientas de análisis estadístico.**

Dada la complejidad de muchas de las tareas del análisis estadístico, las herramientas para el efecto deben de ofrecer lo siguiente:

- Funciones de visualización – estas funciones ayudan a descubrir relaciones entre grandes cantidades de datos
- Funciones exploratorias – estas funciones ayudan a elegir la función estadística y el modelo correcto que se ajustan a los datos. Algunas de estas funciones son tablas multidimensionales de pivoteo, ayuda orientada al análisis, e identificación de valores extremos y distantes. La herramienta debe producir y presentar cuadros, gráficas y

tablas para el análisis empresarial, en forma dinámica y automática, como parte del proceso de exploración.

- Funciones y operaciones estadísticas – estas funciones y operaciones ofrecen un rico conjunto de herramientas, tales como el análisis de regresión, tanto continuo como logístico; el análisis de series históricas, incluyendo auto correlación; transformaciones rápidas de Fourier y pronósticos; análisis de variación múltiple; ANOVA; CHAID; pruebas no paramétricas; y análisis de respuesta múltiple.
- Funciones de administración de datos – estas funciones profundizan al detalle, examinan subconjuntos de datos, discriminan valores extremos, comparan subconjuntos, etcétera.
- Funciones de grabación y reproducción – estas funciones graban los pasos del análisis, transfieren los registros a otro analista empresarial, y reproducen luego tarea completa del análisis. Las funciones de grabación deben incluir los pasos del análisis, el proceso de selección del conjunto de datos, una paleta o carrusel de cuadros, gráficas seleccionadas y cualquier otra información que se vaya a comunicar. Esta es la clave para comunicar y compartir, tanto los resultados de la tarea de análisis estadístico, como las técnicas analíticas y el proceso aplicados.
- Herramientas de presentación – estas herramientas comunican los datos complejos y el análisis en cuadros, tablas, gráficas y tablas sencillas. La herramienta debe convertir con rapidez los datos de un tipo de cuadro y, cuando se necesite, exhibirlo en un tipo de cuadro diferente. La herramienta debe también mostrar la variedad de tipo de cuadros, gráficas y tablas a los que se ajustan los datos, de modo que se seleccionen con facilidad las mejores opciones de presentación. Un conjunto básico de los cuadros y gráficos requeridos consiste en gráficas lineales x-y y de dispersión, gráficas de cuadro, histogramas, gráficas de barra (de pastel y de área), gráficas de intervalos, gráficas de superficie tridimensional y contorno, cuadros estadísticos (como de Pareto y de barra X) y reportes tales como tabulaciones cruzadas.
- Juego de herramientas del desarrollador – utilice este juego para enlazarse con facilidad a aplicaciones de escritorio y complementar componentes para el análisis estadístico y la elaboración de cuadros, gráficas y reportes.
- Tiempo de respuesta responsable – este período que se mide en minutos o incluso en horas, es aceptable para algunas decisiones empresariales pero existen sus excepciones.

### 3.3.11 Aplicación del análisis estadístico.

El análisis estadístico de los datos del data warehouse se usa para abordar un amplio rango de aspectos empresariales, incluyendo los siguientes:

- Comercialización – esta área de análisis tiene que ver con la comprensión de los hábitos de compra y el comportamiento del consumidor y con la identificación de segmentos del mercado. El análisis de la canasta de mercado es una de las técnicas que se emplean para analizar los patrones de compra de los clientes.
- Ventas – esta área de análisis se refiere a los análisis de ingresos por ventas y las desviaciones de lo planeado.
- Tele ventas y comercialización – esta área de análisis corresponde a la conducción y análisis de respuestas para afinar la estrategia de tele ventas.
- Administración empresarial – esta área de análisis se relaciona con las aplicaciones para evaluar prestamos y analizar riesgos de crédito, medir pruebas de resultados y evaluar la satisfacción del cliente.
- Investigación médica – esta área de análisis tiene que ver con el estudio de la respuesta a los diagnósticos y tratamientos.
- Garantía del producto – esta área de análisis se concentra en vigilar la calidad del producto y analizar los datos de reparaciones con garantía.

### 3.3.12 La Minería de datos y la inteligencia artificial

#### Introducción

Algunos sistemas que son sólo parcialmente conocidos, producen una cantidad inmensa de datos; estos datos con frecuencia contienen valiosa información que puede resultar muy útil y ser vista como vetas de oro por los ojos de un ejecutivo de una corporación.

Las dimensiones de las bases de datos grandes (montañas) y sus velocidades de crecimiento, hacen muy difícil para un humano su análisis y la extracción de alguna información importante (oro). Aún con el uso de herramientas estadísticas clásicas esta tarea es casi imposible.

El descubrimiento de conocimiento en base de datos (KDD) combina las técnicas tradicionales con numerosos recursos desarrollados en el área de la inteligencia artificial. En estas aplicaciones el término "Minería de Datos" (Data mining) ha tenido más aceptación.

En algunos casos las herramientas provenientes de la inteligencia artificial son nuevas, no del todo comprendidas y carentes de un soporte teórico formal. Pero en este caso el objetivo es tan valioso, que los resultados prácticos han rebasado a la elegancia académica.

**Concepto:** Veamos la definición más común.

Es esta: "Una actividad de extracción cuyo objetivo es el de descubrir hechos contenidos en las bases de datos".

En la mayoría de los casos se refiere a un trabajo automatizado. Si hay alguna intervención humana a lo largo del proceso, este no es considerado como minería de datos por parte de algunas personas.

La palabra "descubrimiento" está relacionada con el hecho de que mucha de la información valiosa es desconocida con anterioridad. En todo caso, estas técnicas pueden ayudar a confirmar cualquier sospecha sobre el comportamiento del sistema en un particular contexto.

En cuanto a los hechos escondidos, Estos estarán principalmente bajo la forma de reglas; las cuales nos ayudarán a entender el modelo del sistema relacionado con los datos observados. Por otra parte, las reglas también pueden ser usadas en las predicción de ciertos estado del sistema.

**Aplicaciones:** Algunas de las áreas donde la minería de datos ha sido exitosa.

**Detección de fraudes:** Esta es una aplicación que puede ser considerada como una técnica de clasificación. En efecto, cuando el algoritmo analiza una gran cantidad de transacciones, el mismo tratará de categorizar aquellas que sean ilegítimas mediante la identificación de ciertas características que estas últimas tengan en común.

Esto puede ser usado en las corporaciones para prevenir que se culmine un proceso que muestre pertenecer a una "clase" peligrosa

**Análisis de riesgos en créditos:** Esta es una aplicación similar a la anterior, pero con la ventaja de la existencia de maneras tradicionales para realizarlo. El clásico procedimiento de asignación de puntos puede ser complementado y mejorado con la ayuda de la minería de datos.

**Clasificación de cuerpos celestes:** Debido a la gran contribución a estas tareas por parte del reconocimiento de imágenes y los pre-procesamientos involucrados, esta aplicación también puede considerarse como perteneciente al área del reconocimiento de patrones de imágenes.

**Minería de texto:** Con billones de páginas en la red, se requieren de nuevas tecnologías para encontrar, clasificar y detectar patrones particulares en la información disponible. La esencia de los métodos de la minería de datos aplicados a los datos numéricos, puede también ser aplicada a datos de texto.

**Cuando son útiles:** Algunas restricciones iniciales a las expectativas.

**Sistemas parcialmente desconocidos:** Si el modelo del sistema que produce los datos es bien conocido, entonces no necesitamos de la minería de datos ya que todas las variables son de alguna manera predecibles. Este no es el caso del comercio electrónico, debido a los efectos del comportamiento humano, el clima y de decisiones políticas entre otros.

En estos casos habrá una parte del sistema que es conocida y habrá una parte aparentemente de naturaleza aleatoria. Bajo ciertas circunstancias, a partir de una gran cantidad de datos asociada con el sistema, existe la posibilidad de encontrar nuevos aspectos previamente desconocidos del modelo.

**Enorme cantidad de datos:** Bases de datos muy grandes compensan las limitaciones de un modelo incompleto. Esto es particularmente cierto cuando las redes neuronales y otras técnicas adaptativas son utilizadas. En estos casos, se requiere suficiente cantidad de datos para el entrenamiento y la verificación.

**Potente hardware y software:** Muchas de las herramientas presentes en la minería de datos están basadas en el uso intensivo de la computación, en consecuencia convenientes equipos y software eficientes aumentarán el desempeño del proceso, el cual a veces debe vérselas con producciones de datos del orden de los Gbytes/hora.

**Dos tipos de descubrimientos:** Se pueden perseguir diferentes objetivos.

**Descripción:** El principal producto del proceso de la minería de datos es el descubrimiento de reglas. Estas mostrarán nuevas relaciones entre las variables o excepciones a las ya establecidas. Ello enriquecerá el análisis y la descripción del modelo y ayudará en la planificación y en el diseño de futuros cambios.

Es posible que algunas de las reglas descubiertas no puedan ser cambiadas, pero si resulta posible realizar modificaciones apropiadas en la organización con el propósito de mejorar su desempeño.

**Predicción:** Una vez descubiertas reglas importantes que gobierne el sistema, estas pueden ser utilizadas para estimar algunas variables de salida. Puede ser en el caso de secuencias en el tiempo, o bien en la identificación e interrupción a tiempo, de una futura mala experiencia de crédito.

En esta tarea, se complementan las técnicas estadísticas tradicionales con aquellas provenientes de la inteligencia artificial. Conceptos adaptativos como los algoritmos genéticos y las redes neuronales, permiten realizar predicciones más acertadas, especialmente en casos de gran complejidad y con relaciones internas no-lineales.

**Algunas herramientas utilizadas:** De la Inteligencia artificial.

**Redes Neuronales:** Grupo de unidades no-lineales interconectadas y organizadas por capas. Estas pueden ser funciones matemáticas y números almacenados en computadoras digitales, pero pueden ser elaboradas también mediante dispositivos analógicos como los transistores de efecto de campo (FET).

A pesar del incremento en velocidad y de la escala de integración en los semiconductores, la mejor contribución de las redes neuronales tendrá que esperar por computadoras más rápidas, masivas y paralelas.

**Mapas característicos de Kohonen:** Es una red neuronal del tipo de entrenamiento no-supervisado. Los datos son mostrados a la estructura y esta se sensibiliza a los patrones presentes. Una vez entrenada es capaz de identificar tales patrones en nuevos datos.

**Reconocimiento de patrones:** Se trata de un grupo de técnicas orientadas a evaluar la similitud y las diferencias entre señales. Se involucran en esto a varios tipos de procesamiento tales como la transformada de Fourier.

**k-nearest neighbor:** Un procedimiento para clasificar a los "records" de un archivo mediante la identificación de grupos (clusters) y decidiendo a cual grupo pertenece cada uno de los "records".

**Algoritmo Genético:** Imitando la evolución de las especies mediante la mutación, reproducción y selección, estos algoritmos proporcionan programas y optimizaciones que pueden ser utilizados en la construcción y entrenamiento de otras estructuras como las redes neuronales.

**El ejemplo de los pañales y las cervezas:** Un caso famoso acerca del comportamiento de los consumidores.

Una situación muy popular sucedió en una cadena de víveres en E.U. Utilizando un software de minería de datos para estudiar el comportamiento de sus clientes, encontraron relaciones interesantes entre pañales, cervezas, hombres, y día de la semana.

Encontraron que los días jueves y sábado, los hombres que compraban pañales también compraban cerveza. Información como esa, que no siempre es evidente a primera vista, puede ser utilizada para reubicar la mercancía en lugares más estratégicos, en este ejemplo, manteniendo a los pañales y a las cervezas cercanos unos de otros.

Este resultado suministrado por un proceso de minería de datos, puede ser analizado en profundidad por expertos humanos. Si ellos encuentran una explicación razonable, esta seguramente será de mucha ayuda para que los ejecutivos de la empresa alcancen sus objetivos de una manera más eficiente.

### 3.3.13 Redes Neuronales.

Esta tecnología puede ser desarrollada en software o en hardware y es capaz de aprender y predecir.

Es adecuada para problemas que hasta ahora eran resueltos sólo por el cerebro humano y difíciles para las máquinas lógicas secuenciales.

Las Redes Neuronales son utilizadas para la predicción, la minería de datos, el reconocimiento de patrones y los sistemas de control adaptativo.

Pueden ser combinadas con otras herramientas de la Inteligencia Artificial (IA), tal como la lógica difusa, los algoritmos genéticos y los sistemas expertos.

#### Concepto.

Es un sistema compuesto por un gran número de elementos básicos, agrupados en capas y que se encuentran altamente interconectados; Esta estructura posee varias entradas y salidas, las cuales serán entrenadas para reaccionar (valores 0), de una manera deseada, a los estímulos de entrada (valores 1).

Estos sistemas emulan, de una cierta manera, al cerebro humano. Requieren aprender a comportarse y alguien debe encargarse de enseñarles o entrenarles, en base a un conocimiento previo del entorno del problema



Figura 3.7 Redes neuronales, un sistema con entradas y salidas

### Aplicación

Esta tecnología es muy útil en unos cuantos problemas de características muy especiales. A grandes rasgos, estas aplicaciones son aquellas en las cuales se dispone de un registro de datos y nadie sabe exactamente la estructura y los parámetros que pudieran modelar el problema. En otras palabras, grandes cantidades de datos y mucha incertidumbre en cuanto a la manera de como estos son producidos.

Como ejemplos de las aplicaciones de las redes neuronales (Neural Networks) se pueden citar: las variaciones en la bolsa de valores, los riesgos en préstamos, el clima local, el reconocimiento de patrones (rostros) y la minería de datos (data mining).

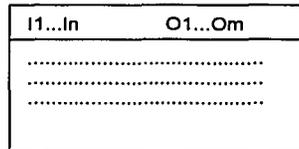


Figura 3.8 Datos en entrenamiento

### Diseño

Se puede realizar de varias maneras. Por ejemplo en hardware utilizando transistores de efecto de campo (FET) o amplificadores operacionales, pero la mayoría de las RN se construyen en software, esto es en programas de computación.

Existen en internet herramientas muy buenas y flexibles que pueden simular muchos tipos de neuronas y estructuras.

- **Elemento básico. Neurona artificial.** Pueden ser con salidas binarias, análogas o con codificación de pulsos (PCM). Es la unidad básica de procesamiento que se conecta a otras unidades a través de conexiones sinápticas.
- **La estructura de la red.** La interconexión de los elementos básicos. Es la manera como las unidades básicas se interconectan.

### Elemento básico. Neurona artificial.

Una neurona artificial es un elemento con entradas, salida y memoria que puede ser realizada mediante software o hardware. Posee entradas ( $I$ ) que son ponderadas ( $w$ ), sumadas y comparadas con un umbral ( $t$ ).

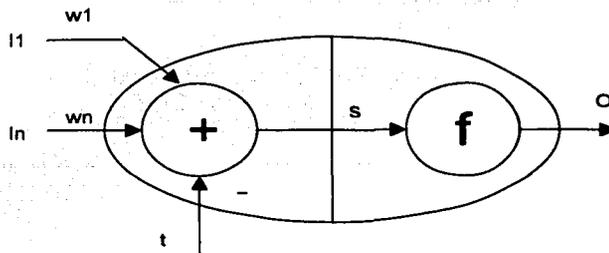


Figura 3.9 Elemento básico. Neurona artificial

$$s = I1 * w1 + I2 * w2 + \dots + In * wn - t$$

$$o = f(s)$$

La señal computada de esa manera, es tomada como argumento para una función no lineal (f), la cual determinará el tipo de neurona y en parte el tipo de red neuronal.

#### La estructura de la red (Neural Network). La interconexión de los elementos básicos.

Es la manera como las unidades (artificial neurons) comunican sus salidas a las entradas de otras unidades. Por lo general estas están agrupadas en capas, de manera tal, que las salidas de una capa están completamente conectadas a las entradas de la capa siguiente; en este caso decimos que tenemos una red completamente conectada. Es posible tener redes en las cuales sólo algunas de las unidades están conectadas, también podemos tener conexiones de realimentación, conectando algunas salidas hacia entradas en capas anteriores (no se confunda esto con el "back propagation").

Para obtener un resultado aceptable, el número de capas debe ser por lo menos tres. No existen evidencias, de que una red con cinco capas resuelva un problema que una red de cuatro capas no pueda. Usualmente se emplean tres o cuatro capas.

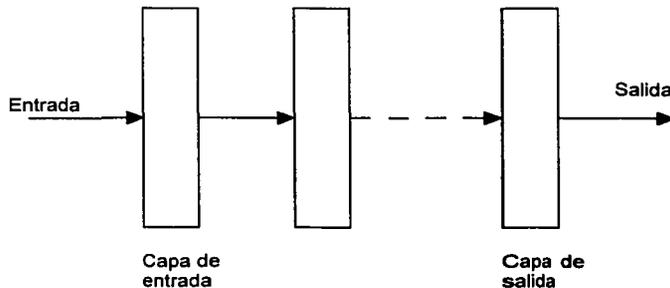


Figura 3.11 La interconexión de elementos básicos

#### Aprendizaje. El proceso de entrenamiento.

Este proceso consiste en una adaptación progresiva de los valores de las conexiones sinápticas, para permitir a la Red Neuronal el aprendizaje de un comportamiento deseado. Para lograr esto, alimentamos a la red con una entrada de los datos de entrenamiento, comparamos

la salida de la red con la salida de los datos de entrenamiento; la diferencia se usa para computar el error (cuadrático medio) de la respuesta de la red. Con un algoritmo apropiado es posible retocar los valores de los pesos sinápticos con el fin de reducir el error. Estas correcciones deben realizarse varias veces o ciclos, para todo el conjunto de entradas-salidas de los datos de entrenamiento.

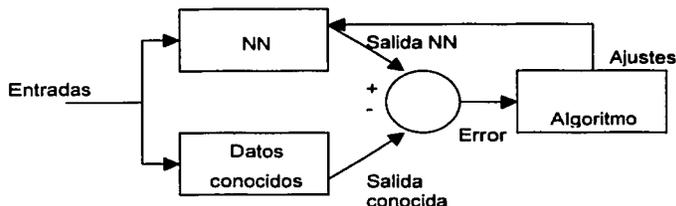


Figura 312 El proceso de aprendizaje

### Ejecución. Comportamiento final de la Red Neuronal.

Para este trabajo debemos disponer de una red entrenada. Es posible alimentar a este sistema con una nueva entrada (nunca antes vista), una situación nueva, y nuestra Red Neuronal producirá una respuesta razonable ó inteligente en sus salidas. Puede tratarse de la predicción de un valor en la bolsa en ciertas circunstancias, el riesgo de un nuevo préstamo, una advertencia sobre el clima local o la identificación de una persona en una nueva imagen.

El futuro de las Redes Neuronales estará determinado en parte por el desarrollo de chips ad hoc, avances en la computación óptica/paralela y tal vez en un nuevo tipo de unidad química de procesamiento.

## 3.4. Comunicación entre Procesos.

### 3.4.1 Introducción.

Como todo el desarrollo de la computación, la comunicación entre procesos es un campo que ha tenido un gran desarrollo en los años recientes. En este capítulo exploraremos los principales métodos que existen en la actualidad para comunicar procesos. Pero antes repasemos algunos conceptos.

Nos interesa comunicar procesos que no se ejecutan en la misma computadora, procesos que incluso pudieran haberse desarrollado para plataformas diferentes y por lenguajes de programación diferentes. Es por esto que es muy importante utilizar métodos estandarizados de comunicación, estos métodos pueden aplicarse para comunicar procesos que pueden correr, por ejemplo, en un equipo UNIX con otros desarrollados en equipos NT.

En un principio, con el *mainframe*, el sistema se controlaba de forma centralizada. Con la introducción de los PC's se pudo descargar gran parte de la carga de trabajo que tenían que soportar los *mainframes*, apareciendo por lo tanto la arquitectura cliente/servidor. Así se sustituyeron estos "monstruos" por servidores UNIX mucho más económicos. Esto trajo consigo una mayor independencia de las secciones dentro una empresa para crear aplicaciones que se adaptasen mejor a sus necesidades y no depender tanto del control central.

En este tipo de sistema, los datos se hallan almacenados en el servidor, el interfaz de usuario en el cliente y la lógica de control en cualquiera de ellos. Tenía la desventaja de que al cambiar alguna parte del cliente, el nuevo programa debía ser distribuido a todos los usuarios.

Esta arquitectura evolucionó hacia la multicapa cliente/servidor. Divide al servidor en varias capas, para poder aislar al cliente de los cambios producidos en la aplicación. De esta forma se podrán hacer cambios en la aplicación con poca probabilidad de que afecten al cliente. Además permite más flexibilidad en la aplicación al poder colocar los módulos del servidor en máquinas diferentes.

A continuación se pueden encuadrar los sistemas distribuidos. En estos sistemas todas las funciones de la aplicación se exponen como objetos, cada uno de los cuales puede usar los servicios proporcionados por otros objetos del mismo o de diferente sistema. Por lo tanto se oscurece la diferencia entre cliente y servidor: los clientes pueden crear objetos que actúen como servidores. Proporcionan una gran flexibilidad. Ésta se basa en que la interfaz (protocolo de comunicación entre dos módulos) de un módulo especifica a los demás los servicios que ofrece y cómo se usan. Mientras la interfaz no varíe, se podrán hacer todos los cambios que se deseen dentro del módulo. Se deben incluir servicios que permitan localizar a los diferentes módulos.

Pasemos a la descripción de los métodos:

### 3.4.2 SOCKETS

Los Sockets [12] son un punto de comunicación por medio del cual los procesos pueden emitir o recibir información, funcionan como una interfaz entre la capa de la aplicación y la de transporte.

En el interior de un proceso se identificará por un descriptor, parecido al usado para la identificación de archivos:

- o permite redirección de la entrada y salida estándar
- o permite utilización de aplicaciones estándar sobre la red
- o todo nuevo proceso, (*fork()*) hereda los descriptors de socket de su padre.
- o canal físico de comunicación, (capa física de la ISO-OSI).

Permite, dado dos procesos que se comunican a través de ellos, despreocuparse de:

- o forma de codificación de señales para disminuir probabilidad error en la transmisión.
- o (capa enlace).
- o nodos red por los cuales tiene que pasar, (capa red).
- o formar paquetes de la información a transmitir y buscar la ruta de la computadora.
- o El origen y el destino, (capa transporte).

De una forma muy resumida un servidor abre su puerto y espera peticiones del cliente, el cliente abre su puerto y escribe una petición al servidor, el servidor proporciona el servicio al cliente.

### 3.4.3 Llamado de procedimientos remotos (RPC)

No es nuevo para cualquier programador de lenguajes como C o Pascal el usos de funciones y procedimientos para empaquetar código que se usará frecuentemente en la ejecución del programa. Cuando hacemos esto estamos haciendo cálculo centralizado o computación centralizada. Para lograr esto pero de forma distribuida se utiliza el llamado de procedimientos remotos.

La especificación RPC [12] fue creada por Bireel & Nelson en 1984, permite a los programadores llamar procedimientos localizados en otras máquinas, un proceso *x* en una máquina *A*, puede llamar a un procedimiento y en una máquina *B*, es posible enviar información entre los procesos.

El modelo RPC establece su mecanismo de comunicación a través de petición-respuesta. El procedimiento cliente envía un mensaje de petición al procedimiento servidor, el cual regresa

un mensaje de respuesta. Los procesos cliente y servidor se comunican por medio de dos stub; un stub para el cliente y otro para el servidor. Un stub es una interfaz de comunicación que implanta el protocolo RPC y especifica como los mensajes se construyen e intercambian.

Un stub contiene funciones que asocian una llamada a un procedimiento local en una serie de llamadas de función RPC en la red. El stub del cliente interviene y envía un proceso a ejecutarse a otra máquina a través de la red. El stub servidor efectúa las funciones que llaman a las rutinas que despachan el servicio, regresando por la red el resultado al stub cliente.

Cuando se construye una aplicación distribuida se tiene dos opciones para su diseño orientado a la comunicación y diseño orientado a la aplicación. El diseño orientado a la comunicación inicia con el protocolo de comunicación. Diseñando el formato y sintaxis de los mensajes. Diseñando los componentes del cliente y servidor; especificando como cada uno reacciona al llegar los mensajes y los mensajes de salida que cada uno genera.

El diseño orientado a la aplicación, inicia elaborando un programa de aplicación convencional, construyendo y probando la versión que opera en una sola máquina. Divide el programa en dos o más piezas y adiciona protocolos de comunicación, los cuales permiten a cada pieza ejecutarse en una computadora por separado.

El modelo RPC utiliza el diseño orientado a la aplicación. Primeramente se diseña un programa convencional que solucione el problema, y entonces se divide el programa en dos piezas, las cuales se compilarán y trabajarán en dos computadoras por separado. Antes de que cada programa se compile, se deben agregar los protocolos de comunicación.

#### **3.4.4 CORBA.**

En los sistemas distribuidos es muy importante la definición de la interfaz y ciertos servicios como la búsqueda de módulos, lo que hace CORBA [11] es proporcionar un estándar para poder definir estas interfaces entre módulos, así como algunas herramientas para facilitar la implementación de dichas interfaces en el lenguaje de programación escogido. Adicionalmente se han incluido algunos servicios estándar accesibles a todas las aplicaciones en CORBA. Además CORBA proporciona todo el mecanismo que permite a los distintos módulos de una aplicación comunicarse entre sí.

Las dos principales características de CORBA es que es independiente tanto de la plataforma como del lenguaje de la aplicación. La independencia de plataforma significa que los objetos de CORBA se pueden utilizar en cualquier plataforma que tenga una aplicación CORBA ORB. La independencia de lenguaje se refiere a que los objetos CORBA y los clientes se pueden implementar en cualquier lenguaje de programación. Así, a un objeto CORBA no le hará falta saber el lenguaje en que ha sido escrito otro objeto con el que se esté comunicando.

La especificación CORBA fue desarrollada por el Grupo de Administración de Objetos (OMG). Su misión es proporcionar un entorno de trabajo arquitectural común para aplicaciones orientadas a objetos basadas en especificaciones de interfaces ampliamente disponibles. Esto se logra con el establecimiento de la Arquitectura para el Manejo de Objetos (OMA) de la cual CORBA es sólo una parte. La OMA consta de la función ORB, servicios de objetos (servicios CORBA), facilidades comunes (facilidades CORBA), interfaces de dominio y objetos de aplicación. El papel de CORBA en el OMA es implementar la función ORB.

La primer versión surgió en Diciembre de 1990 y fue seguida rápidamente por CORBA 1.1 que definía el lenguaje de definición de interfaces (IDL) así como los interfaces de programación de aplicaciones (API) para comunicar las aplicaciones con los ORBs. La versión de CORBA 2.0 (1994) ya especifica un protocolo estándar para la comunicación entre ORBs. Así surgió el protocolo inter-ORBs de Internet (IIOP) aplicable sólo a redes basadas en TCP/IP.

## Arquitectura de CORBA

CORBA es una arquitectura orientada a objetos. Por eso exhibe muchas de las características comunes a otros sistemas orientados a objetos, incluyendo herencia de interfaces y polimorfismo. Lo que diferencia a CORBA y lo hace más interesante es que proporciona esta capacidad incluso cuando se utiliza con lenguajes no orientados a objetos como C y COBOL.

Los dos pilares fundamentales de la arquitectura CORBA son el ORB (*Object Request Broker*), que dirige la comunicación entre objetos CORBA, y el IDL (*Interface Definition Language*), que define las interfaces de los componentes de la aplicación sobre los que se construyen las aplicaciones CORBA.

### El Object Request Broker (ORB).

El *Object Request Broker* es uno de los pilares fundamentales de CORBA. Un ORB es un componente software cuyo propósito es facilitar la comunicación entre objetos. Esto lo hace proporcionando una serie de facilidades, una de las cuales es localizar un objeto remoto dada una referencia a ese objeto y otra es la ordenación de los parámetros y valores de retorno y de invocaciones a métodos remotos.

Su funcionamiento es el siguiente: cuando un módulo de una aplicación quiere usar un servicio proporcionado por otro módulo obtiene una referencia del objeto que provee ese servicio. Después de obtenerla, el módulo puede invocar métodos en ese objeto. La primera responsabilidad del ORB es resolver peticiones de referencias de objetos, permitiendo a los módulos de la aplicación establecer conexión entre ellos.

Otra de las responsabilidades del ORB es el *marshaling* y el *unmarshaling*. Después de que un módulo de la aplicación haya obtenido una referencia del objeto cuyos servicios quiere usar, ese módulo ya puede invocar métodos en ese objeto. Generalmente estos métodos necesitan parámetros como entrada y devuelven otros parámetros como salida. El ORB debe recibir los parámetros de entrada del módulo que llama al método y *marshal* estos parámetros. Esto quiere decir que el ORB traduce los parámetros a un formato (*on-the-wire format*) que puede ser transmitido por la red hasta el objeto remoto. El ORB también *unmarshals* los parámetros devueltos, convirtiéndolos a un formato que el módulo llamante entienda.

Todo esto se hace de forma transparente a la intervención del programador. Una aplicación cliente invocará un método remoto y recibirá los resultados como si el método fuese local.

Gracias a este proceso se consigue la **independencia de plataforma**, debido a que los parámetros se traducen en la transmisión a un formato independiente de la plataforma (el *on-the-wire format* forma parte de las especificaciones CORBA) y en recepción se convierten al formato específico de la plataforma. Un cliente ejecutándose en un sistema Macintosh podrá invocar métodos de un servidor que se ejecute en un sistema UNIX. Además de la independencia del SO usado, las diferencias de hardware (como el ordenamiento de los bytes, *endianness*, la longitud de las palabras, etc.) son irrelevantes puesto que el ORB hace las conversiones necesarias automáticamente.

En resumen, las responsabilidades del ORB son:

1. Dada una referencia a un objeto por un cliente, el ORB localiza la correspondiente implementación del objeto (el servidor).
1. Cuando el servidor está localizado, el ORB asegura que el servidor está preparado para recibir la petición.
2. El ORB del lado del cliente acepta los parámetros del método que se está invocando y *marshals* los parámetros a la red.

3. El ORB del lado del servidor *unmarshals* los parámetros de la red y se los entrega al servidor.
4. Los parámetros de retorno, si existen, se *marshal/unmarshal* del mismo modo.

La mayor ventaja que ofrece el ORB es el tratamiento de los datos independientemente de la plataforma.

#### Lenguaje de definición de interfaces (IDL).

La otra pieza fundamental de la arquitectura CORBA es el uso del lenguaje de definición de interfaz (IDL) que especifica interfaces entre objetos CORBA. Es un lenguaje estándar.

Los interfaces descritos en IDL se pueden ajustar a cualquier lenguaje de programación lo que hace que las aplicaciones y los componentes CORBA sean **independientes del lenguaje** utilizado para implementarlos. El pliego de condiciones IDL es el responsable de asegurar que los datos son intercambiados correctamente entre lenguajes diferentes. Por ejemplo, el tipo *long* en IDL es un entero de 32 bits con signo, que corresponde a un *long* de C++ o a un *int* de Java.

Esta independencia del lenguaje se consigue gracias al *language mapping*, que es un pliego de condiciones que empareja las construcciones en IDL con las de un lenguaje de programación particular. Por ejemplo en el C++ *mapping* el "interface" de IDL corresponde a una "class" de C++. La OMG ha definido un número de *language mapping* estándar para muchos lenguajes, tales como C, C++, COBOL, Java, Smalltalk, etc. *Mappings* para otros lenguajes existen, pero o no son estándar o están en proceso de serlo.

Otra característica importante de IDL es que no se trata de un lenguaje de implementación (no se pueden escribir aplicaciones en IDL), su único propósito es definir interfaces. Comparando con C++, se podría decir que las definiciones IDL son análogas a los ficheros de cabecera para las clases (que no contienen la implementación de la clase, sino que más bien definen su interfaz). Con respecto a Java sería parecido a las definiciones de los interfaces Java.

CORBA no dicta el uso de un lenguaje en particular, por lo que deja a los diseñadores de la aplicación la elección del lenguaje que más se ajuste a sus necesidades. Incluso se pueden utilizar varios lenguajes para implementar cada parte de la aplicación. Por ejemplo, el cliente puede estar escrito en Java que asegura que pueda funcionar en cualquier máquina, mientras que el servidor se puede implementar en C++ para obtener unas buenas prestaciones.

#### El modelo de comunicaciones de CORBA.

Vamos a intentar entender el papel de CORBA en una red de ordenadores. Típicamente ésta consiste en sistemas que están físicamente conectados. Esta capa física proporciona el medio a través del cual tiene lugar la comunicación. Más allá de la capa física se encuentra la capa de transporte, que incluye los protocolos responsables de mover los paquetes de datos desde un punto hacia otro. CORBA es neutral respecto estos protocolos de red, es independiente del protocolo utilizado y podría funcionar con cualquiera.

El estándar CORBA especifica lo que se conoce como el Protocolo General Inter-ORB (GIOP) que, en un nivel alto, establece un estándar para la comunicación entre varios ORB's de CORBA. GIOP es tan sólo un protocolo general, por lo que el estándar CORBA también especifica otros protocolos que detallan GIOP para usar un protocolo de transporte particular (como TCP/IP o DCE). El utilizado para redes TCP/IP se denomina Internet Inter-ORB Protocol (IIOP). Los fabricantes deben implementar este protocolo para ser considerados conformes a CORBA. Este requerimiento ayuda a asegurar la **interoperabilidad** entre productos CORBA de diferentes fabricantes, aunque cada uno puede tener además sus propios protocolos.

Un ORB puede soportar cualquier protocolo, pero siempre debe que incluir el IIOF (se negocia el protocolo a usar al establecerse la conexión entre los ORB's). De todas formas, los ORB's de CORBA normalmente se comunican usando el IIOF debido, en parte, a que este protocolo es el correspondiente al protocolo TCP/IP, que es el usado en Internet. En el argot CORBA/IIOF las referencias a objeto se pasan a denominar Referencias a Objetos Interoperables o IOR's.

Resumiendo, las aplicaciones CORBA se construyen encima de los protocolos derivados de GIOP (como el IIOF). Estos protocolos están, a su vez, encima de los protocolos de transporte (TCP/IP, DCE). Las aplicaciones CORBA no están limitadas a usar sólo uno de estos protocolos, sino que se puede usar un puente para interconectar aplicaciones situadas en redes que trabajen con diferentes protocolos. Se puede ver que, más que suplantar los protocolos de transporte, la arquitectura CORBA crea otra capa (la capa del protocolo Inter-ORB) que utiliza estos protocolos como soporte. Esta es otra de las claves de la interoperabilidad entre las aplicaciones CORBA ya que no se dicta el uso de un protocolo de transporte particular.

### **El modelo de objetos de CORBA.**

En CORBA, todas las comunicaciones entre objetos se hacen a través de referencias de objeto (explicadas posteriormente). Además, la visibilidad de los objetos se obtiene únicamente pasando referencias a esos objetos. Los objetos no pueden ser pasados por valor (al menos en la especificación actual de CORBA). En otras palabras, los objetos remotos en CORBA permanecen remotos, no hay actualmente ninguna forma de mover o copiar un objeto de un sitio a otro.

Todas las arquitecturas orientadas a objetos ofrecen un modelo de objetos, que describe cómo se representan los objetos en el sistema. Tres de las principales diferencias entre el modelo de objetos de CORBA y los modelos tradicionales radican en la forma semi-transparente de distribuir los objetos en CORBA, el tratamiento de las referencias a objetos y el uso de los llamados adaptadores de objetos (como el BOA -Basic Object Adapter-).

Para un cliente CORBA, una llamada a un método remoto es exactamente igual a una llamada a un método local. La naturaleza distribuida de los objetos en CORBA es transparente a los usuarios de dichos objetos. Debido a que los objetos se encuentran distribuidos hay más posibilidades de que se produzca un fallo, por lo que CORBA debe ofrecer la forma de manejar estas situaciones. Lo resuelve ofreciendo un conjunto de excepciones de sistema, que pueden ser lanzadas por cualquier método remoto.

En una aplicación distribuida, hay dos métodos posibles de obtener acceso a un objeto de otro proceso. Uno se conoce como "paso por referencia". Se explica muy bien mediante un ejemplo. Supongamos dos procesos, A y B. El primer proceso (A) pasa una referencia a un objeto suyo al segundo proceso (B). Cuando el proceso B invoca un método en ese objeto, el método lo ejecuta el proceso A puesto que él posee el objeto. El proceso B sólo tiene visibilidad del objeto (a través de la referencia al objeto) y por lo tanto sólo puede pedir al proceso A que ejecute los métodos en su lugar. Es decir, pasar un objeto por referencia significa que un proceso concede visibilidad de uno de sus objetos a otro proceso mientras retiene la propiedad del objeto. Las operaciones en el objeto a través de la referencia al objeto las procesa el propio objeto.

El segundo método de pasar un objeto se denomina "paso por valor". El estado actual del objeto se pasa al módulo que lo pide donde se creará una nueva copia del objeto. Siguiendo con el ejemplo de antes, cuando el proceso B invoque métodos en el objeto se ejecutarán en dicho proceso (en la copia) en vez de en el proceso A, donde el objeto original reside. Por lo tanto el estado del objeto original no cambia.

Como se dijo antes, en el modelo de objetos de CORBA todos los objetos se pasan por referencia. Al pasar un objeto por valor es necesario asegurarse de que el módulo que recibe el

objeto tiene implementaciones para los métodos soportados por dicho objeto. En cambio esto no es necesario cuando los objetos son pasados por referencia.

Sin embargo, hay algunos problemas asociados a pasar los objetos por referencia. Todas las llamadas a los métodos serán remotas, por lo que si un componente invoca repetidas veces métodos en un mismo objeto remoto se producirá una gran cantidad de sobrecarga en la comunicación entre los dos componentes (sería más eficiente pasar el objeto por valor, de forma que el componente pueda manipular el objeto localmente).

El estándar CORBA define una serie de adaptadores de objetos (*object adapters*) cuyo primer propósito es conectar la implementación de un objeto con su ORB. La OMG proporciona tres adaptadores de objetos como muestra: el BOA (*Basic Object Adapter*), el *Library Object Adapter* y el *Object-Oriented Database Adapter*, estos dos últimos útiles para acceder a objetos almacenados de forma persistente. El BOA suministra a los objetos CORBA un conjunto común de métodos para acceder a las funciones del ORB. Estas funciones van desde la autenticación del usuario hasta la activación de objetos. El BOA es, en definitiva, el interfaz entre los objetos CORBA y el ORB.

Una característica importante del BOA es su capacidad de activar y desactivar objetos. Provee cuatro tipos de políticas de activación, que indican cómo se deben inicializar los componentes de la aplicación (cómo se va a acceder al servidor):

1. *Shared server* (servidor compartido), en la que un único servidor es compartido entre múltiples objetos.
1. *Unshared server* (servidor no compartido), un servidor contiene sólo un objeto.
2. *Server-per-method* (servidor por método), que automáticamente arranca un servidor cuando se invoca un método en un objeto, y termina el servidor cuando el método regresa.
3. *Persistent server* (servidor persistente), en la que el servidor es arrancado manualmente (por un usuario, un demonio del sistema, ...).

Esta variedad de políticas de activación permite elegir el tipo de comportamiento que más se adapte a un determinado servidor. Por ejemplo, un servidor que tarde mucho en arrancar trabajaría mejor como *persistent server*; por el contrario un servidor que arranque rápidamente bajo demanda podría trabajar bien como *server-per-method*.

## 3.5. Transformación del modelo Estrella al Modelo Cubo.

### 3.5.1. Introducción.

El objetivo de este documento es proporcionar una guía para la transformación de la representación del Modelo Estrella de las Bases de Datos Relacionales al modelo de Cubos que utilizan las Bases de Datos Multidimensionales.

Como consecuencia del seguimiento de esta guía se espera contar con una documentación homogénea en estructura y contenido que sea utilizable como plataforma para el diseño lógico y la implementación, tanto de las estructuras de Almacén en un Modelo de Cubos requeridos por las herramientas de explotación de tipo OLAP [1] (*On Line Analytical Processing*).

### 3.5.2. Alcance.

Esta es una guía para la transformación del modelo conceptual de Estrellas de las estructuras de los datos de un proceso de CFE (Distribución) al Modelo conceptual de Cubos.

Como resultado del trabajo de diseño conceptual de las estructuras del Almacén de Datos de los procesos mencionados por medio de las Estrellas, se debe contar con un documento que

para hacer la transformación de dichas representaciones conceptuales de Estrellas a Cubos y así proceder al diseño lógico y físico de las estructuras de datos del propio almacén en la Base de Datos Multidimensional haciendo uso de sus herramientas de tipo OLAP para la explotación del mismo.

El resto de este documento se subdivide en 5 secciones principales. La primera es un resumen de algunos conceptos generales de Almacén de Datos, la segunda y tercera son el modelo de estrella y de cubos y la cuarta, es la guía de cómo transformar del modelo conceptual de estrellas a cubos para el modelo de datos del Almacén y su implementación en la Base de Datos Multidimensional. La quinta son conclusiones y recomendaciones.

### 3.5.3. Conceptos Generales.

De manera sencilla el Almacén de Datos (AD) es una colección de datos e información corporativa derivada directamente de los sistemas operacionales internos a la empresa u organización y de algunos orígenes de datos e información externa, con los siguientes atributos:

- Una colección de datos integrada*
- Orientada a entidades*
- Variante en el tiempo y*
- No volátil*

*Cuyo objetivo primordial es el soporte en toma de decisiones directivas.*

Esto es, el propósito específico es soportar la toma de decisiones en un negocio, no las operaciones de éste. Por lo anterior, un Almacén de Datos es diferente, tanto en intención, como en extensión, a una base de datos operacional, entendiéndose por bases de datos operacionales, aquellas cuyo objetivo es proporcionar información puntual y actual en las empresas y organizaciones. Por lo tanto, los usuarios, el contenido y la estructura de los datos son diferentes; las computadoras, la administración y los ritmos de trabajo son diferentes; las técnicas y los estilos de diseño son diferentes.

La tendencia actual es moverse hacia configuraciones de software y hardware más potentes. Con dichas configuraciones, se tiene la posibilidad de procesar grandes volúmenes de datos e información de manera analítica. Las empresas deberán de hacer uso de estas tecnologías emergentes para evitar el tener un déficit de información o una sobrecarga de información, entendiéndose sobrecarga al tener una excesiva cantidad de datos e información imposible de ser usada, analizada para el descubrimiento de más información de ella.

También es importante hacer notar que para el desarrollo de un proyecto que involucre un AD no sólo es necesario el cambio de tecnología de software y hardware, sino que es necesario cambiar la metodología de análisis, especificación de requerimientos, diseño, implementación y pruebas.

En la siguiente sección, se presenta una descripción breve una metodología para el análisis y representación de la especificación de requerimientos de un AD.

### 3.5.4. Análisis y Representación de Almacenes de Datos.

Para el análisis y representación de los Almacenes de Datos, se utilizan métodos y modelos de representación de información que ayuden al grupo de desarrollo a dos tareas fundamentales: conceptualizar y representar dicha estructura y facilitar la comunicación e interacción con el usuario y cliente del AD.

El modelo más utilizado para la representación de la estructura de un AD es el conocido como Modelo de Estrella. Este modelo será explicado en la siguiente sección.

### 3.5.5 Modelo Conceptual de Estrella

En esta sección se describirá brevemente el modelo conceptual Modelo de Estrella que se utiliza para el análisis y representación de Bases de Datos Multidimensionales (BDM) que forman parte de los AD.

El componente fundamental de una BDM es un arreglo o matriz de N Dimensiones.

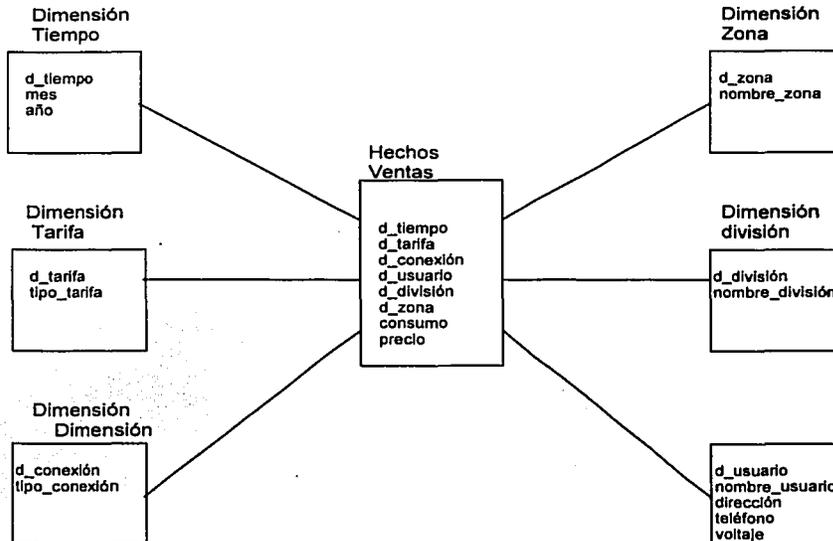


Figura 3.13 Modelo Conceptual de Estrella.

Esta componente se representa gráficamente por medio de una Estrella. En la figura se muestra un ejemplo de una BDM representada gráficamente por medio de una estrella.

Los extremos de la Estrella son las Dimensiones, como se indica en la figura y el Centro de la Estrella son los Hechos. Los Hechos son medidas, como en el caso de **consumo y precio**, registradas en la intersección de las Dimensiones. Para la representación gráfica de estos conceptos se usa rectángulos para las Tablas, tanto la de Hechos como las de Dimensiones y líneas que unen a dichos rectángulos. Dentro de los rectángulos de Dimensiones se escriben aquel Atributo o aquellos Atributos que serán las Dimensiones de la BDM. Dentro del rectángulo de Hechos se escriben los atributos de dichos Hechos, como se mostró en la figura anterior.

Como se mencionó, el modelo Estrella representa los conceptos de Dimensión y Hechos de las BDM. La manera de representarse es por medio de Tablas, por lo que tendremos N Tablas de Dimensiones y una Tabla de Hechos.

Las N dimensiones presentan las posibilidades siguientes:

- La dimensión puede tener un solo atributo.
- La Dimensión puede estar compuesta de dos o más atributos.

En el modelo, las Dimensiones pueden presentar una mezcla de las posibilidades mencionadas. Es decir, la Estrella puede estar constituida de N Dimensiones de un solo atributo o puede ser que de las N Dimensiones, m sean de un solo atributo y las N-m sean de dos o más atributos o finalmente, las N Dimensiones están constituidas de dos o más atributos.

La Tabla de Hechos presenta un "renglón" constituido de uno o m Hechos. Dichos Hechos están Dimensionados sobre las Dimensiones que constituyen la Estrella. Para el ejemplo mostrado en la figura, vemos los Hechos **consumo** y **precio**. Ambos hechos están Dimensionados sobre las Dimensiones mostradas en la figura.

La Tabla de Hechos muestra explícitamente además de los Hechos en sí, las Dimensiones sobre las cuales se analizarán los datos.

Para mostrar conceptualmente, como se comporta el modelo al momento de presentar datos e información, se planteará un ejemplo hipotético. Por ejemplo, una pregunta que se puede hacer a dicha BDM es el **consumo durante todo el tiempo de un tipo de Usuario, con un cierto tipo de tarifa**, dejando las otras Dimensiones tomar cualquiera de sus valores posibles. El resultado será una Tabla donde aparecerán los registros de todo el tiempo existente registrado en la BDM de los datos donde la combinación de **Tipo de Usuario, Tarifa** y los otros valores de las otras Dimensiones sean posibles, en tanto que donde la combinación no sea posible aparecerá en blanco.

Basándose en el resultado presentado por el sistema, el usuario podrá concluir teniendo una visión global de la empresa y no sólo puntual como lo proveería un sistema de transacciones.

En esta sección, se presentó de manera breve y general lo que es la representación de BDM por medio de un Modelo Estrella. También se ejemplificó de manera sencilla el uso que se puede dar a esta representación y un resultado sencillo que no sería posible con otro tipo de esquema o sistema. En la siguiente sección, se discutirá acerca de la representación de las mismas BDM usando el enfoque de Modelo de Cubos.

### 3.5.6. Modelo Conceptual de Cubos

A continuación se explica el concepto de Modelo de Cubo.

Existen diferencias entre el concepto de Cubo y Estrella, aún cuando ambos son representaciones conceptuales de BDM e intentan exactamente lo mismo. Las diferencias son:

- La representación visual
- El cubo no hace uso del concepto de Tabla para representar las Dimensiones y los Hechos
- Los Hechos de las Estrellas en los Cubos se consideran Variables
- Un renglón de N Hechos en la Tabla de Hechos de la Estrella se representa en N Cubos diferentes, con las mismas dimensiones que en la Estrella.
- Las Dimensiones de más de un atributo en la Estrella se representa como varias Dimensiones en el Cubo, tantas Dimensiones como atributos existan en la Estrella.
- Los Cubos manejan el concepto de Jerarquía, que permite representar varias Dimensiones de la Estrella en una sola Dimensión del Cubo pero jerarquizada o permite representar una Dimensión de varios atributos en la Estrella, también de manera jerarquizada.

Gráficamente un cubo se representa como se muestra en la siguiente figura. Para fines prácticos se considera un Cubo de sólo tres dimensiones, el límite de dimensiones lo determina la BDM que se utilice; Sin embargo, no se recomiendan mas de 12.

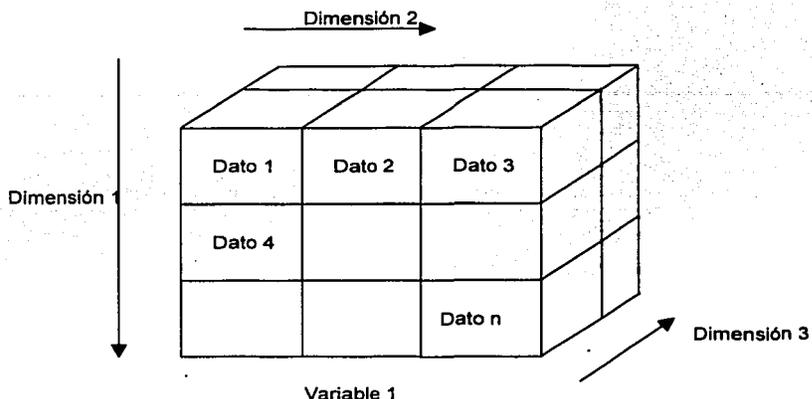


Figura 3.14 Modelo Conceptual de Cubos

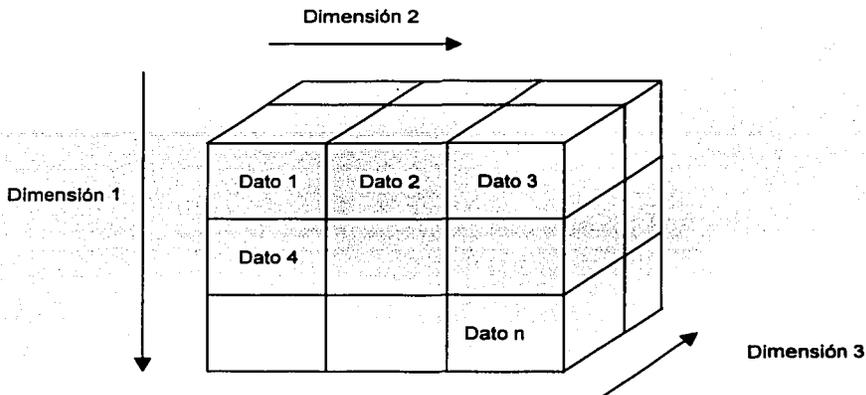
Una base de datos en la BDM es un solo archivo que contiene objetos que organizan, almacenan y manejan datos en un formato optimizado para el procesamiento analítico en línea.

Existen tres objetos fundamentales en la base de datos:

Objeto	Descripción
VARIABLES	Las variables tienen datos. Los datos pueden ser numéricos o tipo texto.
DIMENSIONES	Organiza e indexa los datos almacenados en una variable. Esto permite seleccionar y trabajar con un subconjunto de los datos.
RELACIONES	Las relaciones ligan los valores de una dimensión con los valores de otra dimensión. Esto permite crear jerarquías a través de ligar juntos valores de las dimensiones en un formato padre-hijo lo cual facilita la agregación. Las relaciones permiten seleccionar un subconjunto de los datos, basados en alguna característica común compartida.

Con respecto a la figura y a la tabla, lo que tenemos en la BDM es: la Variable que contiene los datos Dato 1, Dato 2, ..., Dato n está dimensionada por las Dimensiones: Dimensión 1, Dimensión 2 y Dimensión 3.

La mayoría de BDM dimensionan las variables hasta en 32 dimensiones, pero el número de dimensiones que puede existir es una base de datos es ilimitada. Lo anterior, permite que distintas variables compartan las mismas dimensiones. En el caso de la figura del Cubo, la variable 1 está dimensionada por las dimensiones 1, 2 y 3, pero otra variable, por ejemplo la 2, puede estar dimensionada por esas mismas Dimensiones, como se muestra en la siguiente figura. Para este caso, los datos 1, 2, 3, ..., n son valores de la Variable 2.



Para explicar el concepto de relación, se hará referencia a la siguiente figura, en la cual se muestra una relación de tipo jerárquica que permite la BDM.

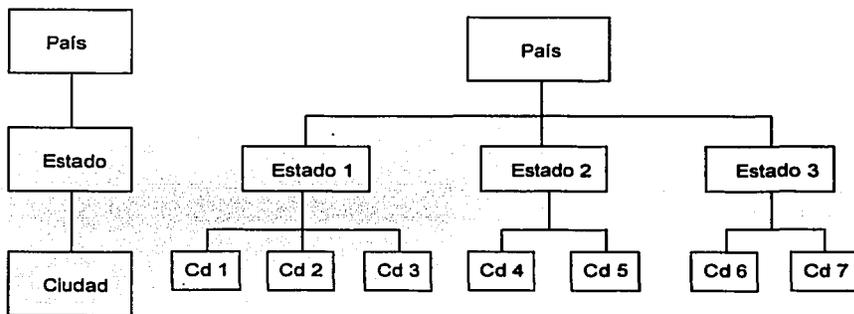


Figura 3..15 Liga entre valores de la misma dimensión.

En la figura anterior, se presenta la liga entre valores de la misma Dimensión y se le conoce como relación consigo misma (*self-relation*) y es una jerarquía de una Dimensión geográfica. Se tiene una variable que está dimensionada por geografía, donde cada ciudad pertenece a un estado y los estados pertenecen al país. La jerarquía facilita la agregación de datos.

La jerarquía puede ser múltiple sobre la misma Dimensión. Por ejemplo, las ciudades fueron agrupadas por estado, pero estas mismas ciudades pueden ser jerarquizadas o agrupadas como se muestra en la siguiente figura.

TESIS CON FALLA DE ORIGEN

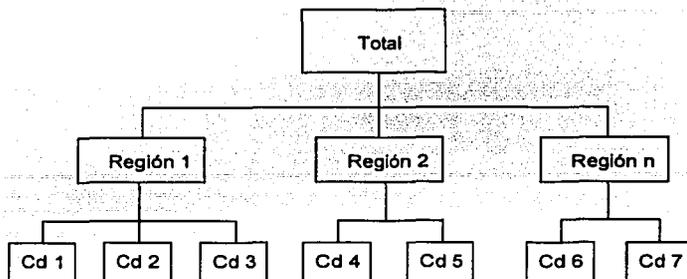


Figura 3.16 Jerarquía múltiple en la misma dimensión.

De lo anterior es posible por medio de la BDM analizar las variables dimensionadas desde dos jerarquías diferentes. Si la variable es producción de energía eléctrica por ciudad en una fecha determinada, entonces es posible conocer la producción de energía eléctrica por región o por estado. Esto se puede identificar durante el proceso de análisis, al identificar que las ciudades, como en este ejemplo, pueden ser "agrupadas" por estado o por región. La variable tendrá por dimensiones ciudad y fecha de registro y el tipo de agrupamiento será por región o estado para la dimensión ciudad.

A continuación se explica la correspondencia que existe entre los conceptos del Modelo Estrella y los conceptos que maneja el Modelo de Cubos.

En las Estrellas se manejan Hechos, en el Cubo se manejan Variables. En las Estrellas se manejan Dimensiones, en el Cubo también. De lo anterior, un renglón de N Hechos de la Tabla de Hechos corresponde a N Variables en el Modelo de Cubos. Con respecto a las Dimensiones, lo ideal es que cada Dimensión en la Estrella sea de un solo atributo y corresponda a una única Dimensión en el Cubo. Sin embargo, como se explicó en la sección del Modelo Estrella, una Dimensión en dicho modelo puede estar compuesta de más de un atributo. Para hacer la representación de dicho tipo de Estrellas, a continuación, se presentan las posibles situaciones que se dan entre las dimensiones de las estrellas y su correspondencia con las del cubo.

Al modelar Cubos las estrellas definidas durante el proceso de análisis, se encuentran las siguientes posibilidades:

- 1 dimensión de una Estrella corresponde a 1 Dimensión en el Cubo.
- 1 Dimensión en una Estrella corresponde a N Dimensiones en el Cubo.
- n Dimensiones en una Estrella corresponden a una Dimensión Jerarquizada en el Cubo.
- Un conjunto de N Hechos en la Estrella corresponde a N Variables distintas en el Cubo, una Variable por cada Hecho.

En la siguiente figura se presenta una Estrella con 3 Dimensiones y un Hecho. Es el caso más sencillo y se usará para fines demostrativos. En la siguiente sección se usará de ejemplo la Estrella del sistema de Combustibles para mostrar las ideas que presentan en esta sección.

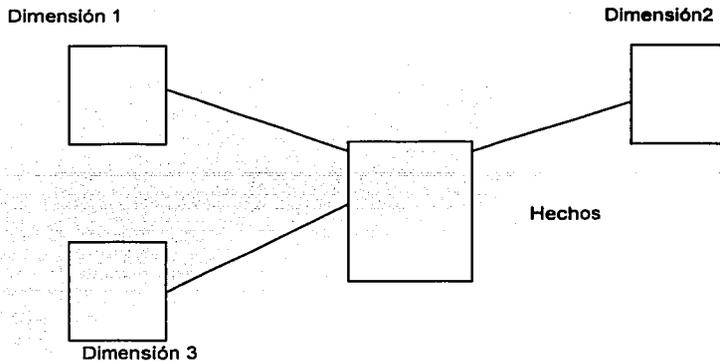


Figura 3.17. Estrella con N dimensiones.

En la figura anterior, tenemos una Estrella con N Dimensiones y su correspondiente Tabla de Hechos. Suponiendo que cada Dimensión es un solo atributo, dicha estrella tiene su Cubo correspondiente y se representa en la siguiente figura.

Cada Dimensión de la Estrella es una Dimensión en el Cubo y suponiendo que la Tabla de Hechos contiene un solo Hecho, entonces su correspondiente en el Cubo es la Variable 1. Si las Dimensiones en la Estrella sólo tienen un atributo, entonces cada Dimensión en la Estrella será representada por una Dimensión en el Cubo.

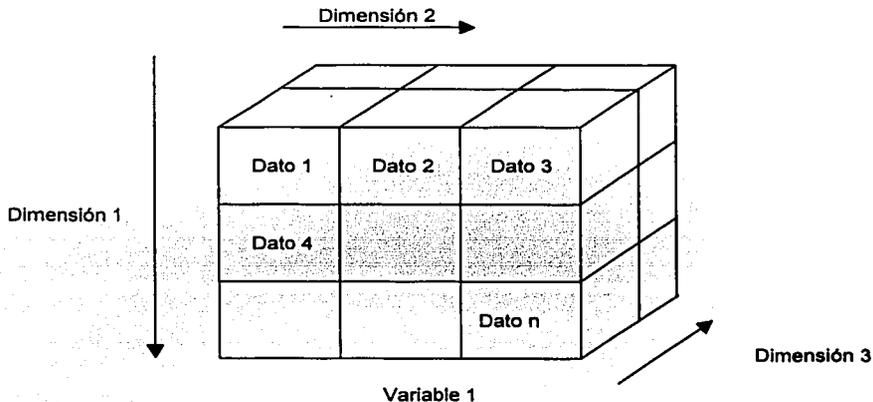


Figura 3.18. Representación de dimensiones en un cubo.

TESIS CON  
FALLA DE ORIGEN

A continuación se expondrán los casos generales para la transformación de estrellas a Cubos.

<b>Estrellas</b>	<b>Cubos</b>
N Dimensiones de un solo atributo cada Dimensión	Se representan como N Dimensiones en el Cubo, una Dimensión de la estrella corresponde a 1 Dimensión en el Cubo. Esta relación es 1 - 1
N Dimensiones de un solo atributo cada Dimensión	Estas Dimensiones se representan como 1 Dimensión con una sola Jerarquía en el Cubo. Esta relación es N - 1 con jerarquía sencilla.
N Dimensiones de un solo atributo cada Dimensión	Se representan como una Dimensión de Múltiples Jerarquías en el Cubo. Esta relación es N-1 con Jerarquía múltiple.
La Dimensión tiempo entre las N Dimensiones	La Dimensión tiempo se representa como una Dimensión con Jerarquía sencilla con el objetivo de ayudar al usuario final del AD a analizar de manera sumariada las Variables que tengan al tiempo como Dimensión. Esta relación es 1 - 1 con Jerarquía sencilla.
Una Dimensión de más de un atributo	Esta dimensión se representa en el Cubo con N Dimensiones. Esta relación es 1 - N
Una Dimensión de más de un atributo	Esta dimensión se representa en el Cubo con 1 Dimensión Jerarquizada. Esta relación es 1 - 1 Jerarquizada.
Una Dimensión de más de un atributo	Esta dimensión se representa en el Cubo con 1 Dimensión y con Jerarquía Múltiple. Esta relación es 1 - 1 con Jerarquía múltiple.
Más de 32 Dimensiones	Este caso se puede resolver de dos maneras: 1) Jerarquizar las 32 Dimensiones en menos de 32 Dimensiones en el Cubo ó 2) Crear Variables con subconjuntos de Dimensiones, esto es, de las 32 Dimensiones en la estrella en las que los Hechos correspondientes están Dimensionadas, identificar subconjuntos de dimensiones y crear Variables con menos de 32 Dimensiones.
Un Hecho de un conjunto de Hechos	Una variable
Un conjunto de Hechos	Un conjunto de Variables

Tabla 3.2 Estrellas vs Cubos

Las opciones anteriores cubren todas las posibilidades y el proceso de representar de Estrellas a Cubos puede ser directo y obvio o puede ser motivo de un análisis adicional en conjunto con el "propietario" de la Estrella correspondiente y/o además con el usuario final.

### 3.5.7 Transformación de estrellas a cubos.

A continuación se explica la manera de transformar de una representación a otra.

La manera de transformar de Estrellas a Cubos no es muy complicada, es un proceso que se basa en la experiencia del analista y sobre todo en el conocimiento que se tienen del dominio del problema y del sistema en cuestión, y en el cual es la siguiente:

- Se toma la Estrella y se analizan las diferentes Dimensiones.
- No dibujar de nuevo la Estrella, sino sólo hacer referencia al documento donde se presentó el modelo.
- Se identifican si las Dimensiones son de un solo atributo o de varios.
- Algunas transformaciones de Dimensiones son obvias por lo que hace de manera directa. Lo obvio se da en que la Jerarquía es natural al problema o que las Dimensiones son de un solo atributo y no existe manera de Jerarquizar por condiciones del problema.
- Se identifican los Hechos y se transforman en variables sobre las mismas Dimensiones de la Estrella.

□ Si existen dudas sobre como representar las diferentes Dimensiones y/o Hechos, entonces se recomienda hablar con quién desarrolló la Estrella o con el usuario o cliente final del sistema para resolver las dudas. No se recomienda tomar decisiones sin consultar al analista o cliente/usuario pues podría ocasionar una gran cantidad de trabajo inútil con la consabida pérdida de tiempo y objetividad del sistema.

□ En el caso de que la Estrella original pose más de 32 Dimensiones, se recomienda lo mismo que en el caso anterior.

□ Se construye una tabla con 4 columnas:

1. En la primera se hace referencia a las Dimensiones de las Estrellas
2. En la segunda las Dimensiones del Cubo
3. En la tercera se escribe el tipo de relación, esto es, 1 - 1, 1 - N, N - 1 con Jerarquía sencilla (JS) o múltiple (JM) y
4. En la cuarta, las relaciones y Jerarquías correspondientes, de las Dimensiones del Cubo yendo de la más elemental a la de mayor agregación.

### Ejemplo:

A continuación se presenta un ejemplo de dicha transformación tomando como referencia la estrella del sistema de Combustible (CB).

En la siguiente figura se presenta de manera conceptual el Diagrama estrella del Sistema CB. En este caso, dibujamos de nuevo la estrella para fines de demostración, sin embargo, como se mencionó anteriormente, sólo se hará referencia al documento donde está el análisis del sistema y la figura correspondiente.

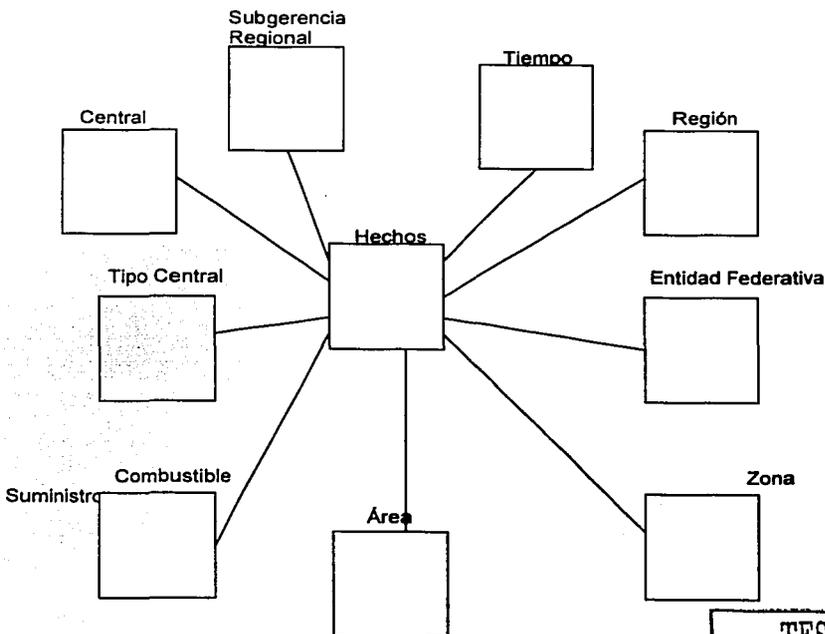


Figura 3.19 Diagrama de Estrella del Sistema de Energéticos.

TESIS CON  
FALLA DE ORIGEN

Del documento se tiene la tabla de Dimensiones y la de Hechos. En este caso, las Dimensiones son:

1. Tiempo
2. Tipo Central
3. Combustible
4. Área
5. Zona de Suministro
6. Entidad Federativa
7. Región
8. Subgerencia Regional

Los Hechos que se identificaron fueron:

1. ca\_poca2
2. ca\_poca3
3. ca\_poca4
4. nu\_cocon
5. nu\_cocol
6. nu\_cogan
7. nu\_codin
8. nu\_cocan
9. nu\_excon
10. nu\_exdin
11. nu\_excacn
12. nu\_recon
13. nu\_recoi
14. nu\_redin
15. nu\_redii
16. nu\_recan
17. nu\_recai
18. nu\_traco
19. nu\_tradi
20. nu\_traca
21. Nu\_derco
22. Nu\_derdi
23. Nu\_derca
24. Nu\_gence

Cada Dimensión es de un solo atributo y los 24 Hechos están Dimensionados sobre las 8 Dimensiones mencionadas.

Se hace un análisis de las Dimensiones y se encuentran para la Dimensión Central las siguientes jerarquías:

- Pemex
- Geográfica
- CENACE
- Generación

Entonces, la representación en Cubos y de acuerdo al procedimiento establecido, quedaría de la siguiente forma:

N-1J CENTRAL

AREA-CENACE  
Entidad Federativa – PAIS  
Zona de Suministro – PEMEX  
Subgerencia-Región-MEXICO

ESTRELLA  
ORIGINAL  
DE OCHO  
DIMENSIONES



1-1 Tipo  
Central

1-1  
Combustible  
1-1j Tiempo

MES  
BIMESTR  
E  
SEMESTR  
E  
AÑO

Y la tabla quedaría de la siguiente forma.

DIM.ESTRELLA	DIM.CUBO	TIPO DE MAPEO	RELACIONES (JERARQUÍA)
Central Area Entidad Federativa Zona de Suministro Subgerencia Región	Central	JM	Area - CENACE Entidad Federativa - País Zona de Suministro - Pemex Región-Subgerencia- Distribución
Tiempo Diario	Tiempo Diario	JM	Día-mes-año Mes-bim-sem-año Mes-trim-sem-año
Tipo Central	Tipo Central	1 - 1	Total
Tipo Comb	Comb	1 - 1	Total

Tabla 3.3 Representación de un cubo en forma de tabla

**JM = Jerarquía Múltiple**

**JS = Jerarquía Sencilla**

Las variables serán para este caso exactamente los 24 Hechos de la Estrella y estarán Dimensionadas sobre las 4 Dimensiones presentadas en la tabla.

Habrà que aclarar que Centrales pertenecen a:

- Qué Áreas
- Qué Entidad Federativa
- Qué Zona de Suministro
- Qué Región
- Qué tipo de central
- Qué tipo de combustible

Esto se presenta haciendo referencia al documento o construyendo una tabla como la siguiente.

Central	Tipo central	Combustible	Región	Zona de Suministro	Área	Sub-Gerencia	Estado
C01	Turbo gas	Gas	R01	ZS04	A03	SG01	ESTADO 1
C02	Vapor	Gas	R01	ZS02	A03	SG02	ESTADO 1
C03	Vapor	Combustóleo	R01	ZS01	A04	SG01	ESTADO 1
C04	Carbón	Carbón	R02	ZS04	A03	SG03	ESTADO 2
C05	Combustión Interna	Diesel	R02	ZS03	A01	SG02	ESTADO 2

C06	Carbón	Carbón	R03	ZS03	A02	SG01	ESTADO 2
C07	Vapor	Combustóleo	R03	ZS03	A01	SG01	ESTADO 3
C09	Turbo gas	Gas	R04	ZS05	A04	SG02	ESTADO 3
C10	Vapor	Combustóleo	R04	ZS05	A05	SG03	ESTADO 3

Tabla 3.4 Representación de un cubo en forma de tabla

### 3.5.8. Conclusiones y Recomendaciones

En el desarrollo de esta guía, se presentaron conceptos generales de modelado de BDM para AD. La transformación de Estrellas a Cubos es generalmente de manera sencilla y directa. Sin embargo, como se mencionó, es importante reconocer en el problema cuando es de esa manera y cuando se requiere de un análisis adicional.

La BDM permite de manera sencilla y natural la implementación del modelo de Cubos, sin embargo, no así las modificaciones de haberse dado algún error ya sea de análisis o de representación.

El Hardware que se ha utilizado es una PC con procesador PENTIUM a 120 MHZ y con 16 MB de memoria principal y disco duro de 1.2 GB. El desempeño ha sido bajo, tanto en la construcción a través de Administrator como la manipulación con la herramienta OLAP Analyzer. Por lo anterior, se recomienda utilizar tanto para el desarrollo e implementación, como para la explotación del AD, Hardware más poderoso.

Ante dudas de jerarquizar o de construir Cubos con subconjuntos de Dimensiones y/o Hechos, es necesario que se consulte tanto al analista original de la Estrella como posiblemente al usuario/cliente, para evitar tiempo y esfuerzo perdido. Por lo anterior, se recomienda que en las actividades de diseño e implementación de la BDM se añada la actividad de consulta. Claro, es verdad que los documentos serán revisados y finalmente autorizados, pero se considera mejor hacer revisiones y sus posibles modificaciones preventivas que correctivas.

## 4 Arquitectura general del Almacenes de Datos

Las partes que integran un sistema de Almacenes de Datos son tres:

### A. El acopio de la información.

#### a. Extracción.

Se refiere a los programas que se acoplan a los sistemas donde residen los datos de las áreas para extraerlos, este proceso deberá ser automático y periódico.

#### b. Transformación

Una vez extraído el dato es posible que requiera de los siguientes procesos antes de insertarlo al almacén.

##### i. Limpieza

##### ii. Filtrado

##### iii. Extrapolación

Los procesos de transformación tienen como objetivo garantizar que cada dato ingresado al almacén esté libre de errores.

#### c. Transporte.

Procesos que transportan los datos extraídos y transformados de la computadora del área a la computadora central para el data warehouse.

#### d. Inserción.

Programas a través de los cuales se transforma la información de su estructura original, muy probablemente relacional a una estructura multidimensional. Típicamente se mapean bases de datos relacionales a multidimensionales.

### B. El almacenamiento.

El almacenamiento puede ser relacional o multidimensional, lo más usado en sistemas de data warehouse es bases de datos multidimensionales aunque en ocasiones para abatir costos se utilizan bases de datos relacionales con diseños en forma de estrella.

### C. La explotación.

Existen varias formas de explotar la información contenida en el almacén de datos, las más usadas son las siguientes.

#### a. Análisis en línea de Procesos (OLAP). Este tipo de herramientas proporcionan importantes características para el análisis de datos en varias dimensiones (criterios de clasificación y agregación de datos).

#### b. Tableros de control. Presentan la información en forma de índices de operación, generalmente incluyen semáforos y alarmas.

#### c. Minería de datos. Proporciona algoritmos para efectuar hallazgos en grandes volúmenes de información.

#### d. Los elementos presentados en esta arquitectura general de un sistema de almacenes de datos serán explicados más adelante conforme se desarrolla el proyecto.

TESIS CON  
FALLA DE ORIGEN

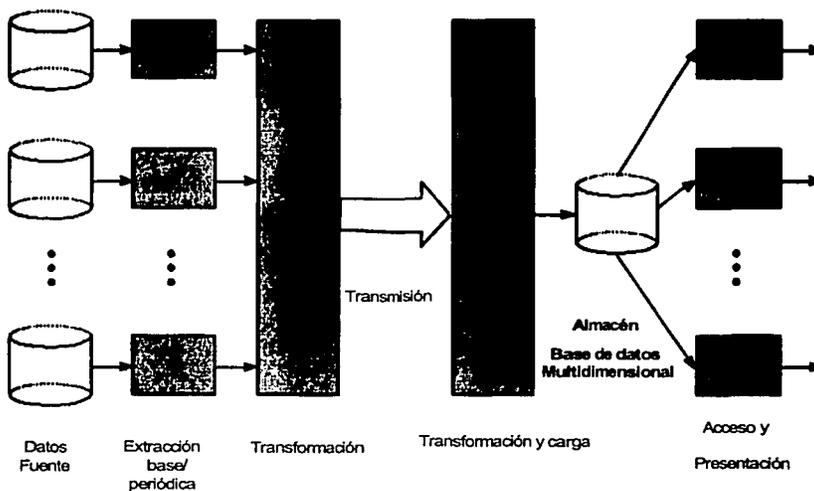


Figura 4.1 Arquitectura general del Almacén de Datos

TESIS CON FALLA DE ORIGEN

## 5 Construcción del Sistema de Información Corporativa

Las etapas en el desarrollo del sistema son las siguientes:

### 5.1 Análisis de información y construcción de modelos de datos.

El primer paso en la construcción del Sistema de Información Corporativa de la CFE es analizar los sistemas de información para determinar cual es la información interesante para el proceso de toma de decisiones. A este tipo de información le llamaremos información directiva.

La información directiva se encuentra en los sistemas que sirven a las áreas para realizar su operación cotidiana, en ocasiones y con el propósito de apoyar a las áreas gerenciales para la concentración de información y elaboración de reportes estadísticos se han desarrollado sistemas concentradores de información en su mayoría directiva. Tal es el caso del sistema GECOM de la Gerencia Comercial.

Para exponer adecuadamente la fase de análisis de información, estudiaremos el caso de los sistemas fuentes de información del proceso de Distribución en la CFE.

Cada agencia de la CFE (aproximadamente 840) cuenta con un pequeño servidor UNIX que corre un sistema con el que se realizan las actividades de facturación y cobranza, el nombre de este sistema es SICOM (Sistema Comercial), se desarrolló en el lenguaje cobol a finales de los 80s. Todas las agencias están enlazadas a la red de la CFE por lo que es posible mantener comunicación TCP/IP con cada instalación del SICOM.

El SICOM cada mes genera un grupo de archivos (archivos gerenciales) que contienen la información que deberá concentrarse en la Gerencia Comercial. Adicionalmente al SICOM se cuenta con el GECOM (Sistema Gerencial Comercial) este tiene como objetivo concentrar la información que genera en SICOM y generar reportes estadísticos. Aunque el propósito del GECOM es eminentemente de apoyo a la toma de decisiones, no se desarrolló siguiendo la tecnología de almacenes de datos, por lo que sus posibilidades de reportes y análisis están limitadas. Sin embargo, el GECOM mantiene una base de datos que en estructuras de almacenamiento claras y sencillas centraliza toda la información relevante del proceso comercial de la CFE. Por lo anterior, el sistema del que se tomará la información del área comercial para incluirla al Data Mart de distribución, será el GECOM.

Además de las actividades comerciales, en las áreas de la Subdirección de Distribución de la CFE también se realizan todas las actividades relacionadas con la distribución de la energía eléctrica desde las grandes subestaciones que reciben la electricidad de las líneas de alta tensión hasta el usuario final. El sistema que maneja la información relativa a los procesos de distribución es el SIAD (Sistema Institucional de Administración de Distribución), este sistema fue desarrollado con la herramienta Borland Delphi y opera en dos capas (Cliente/Servidor) manteniendo una base de datos centralizada desarrollada en informix. Por lo anterior, la información directiva del proceso de distribución deberá tomarse del sistema SIAD.

Después de identificar los sistemas de donde se tomará la información es necesario conocer detalladamente la información que incluyen y las estructuras de almacenamiento que usan para mantenerla. Por lo que se realiza un análisis de datos detallado.

Si se cuenta con buena documentación técnica de los sistemas fuente, es posible que el trabajo de análisis se abrevie. Sin embargo, se deberá ser muy cuidadoso en asegurarse que se cuente con una excelente documentación de modelado de datos.

### 5.1.1 Alcance

Esta es una guía de documentación del análisis de sistemas de información operacionales en el proceso de Distribución de la CFE.

Como resultado del trabajo de análisis de los sistemas operacionales contemplados en este proyecto se debe dar respuesta a los siguientes planteamientos.

- ¿Qué documentación existe disponible del sistema?
- ¿Qué función o funciones realiza el sistema?
- ¿Cuáles son las estructuras y los datos que maneja el sistema?
- ¿Cuáles son los datos que maneja el sistema que son de interés corporativo?
- ¿Cuál es el flujo de datos en la aplicación entre aplicaciones del proceso y con aplicaciones de otros procesos?
- ¿Quiénes son los usuarios del sistema?
- ¿Cuál es la periodicidad en la captura de los datos?
- ¿Cuál es la periodicidad en la generación de reportes?
- ¿Cuál es la disponibilidad y la calidad de datos históricos y cuál es su medio de almacenamiento?
- ¿Cuál es la ubicación física del sistema, contemplando el caso en que el sistema se encuentre en diversos niveles jerárquicos organizacionales?
- ¿Cuáles son las plataformas de hardware, software y comunicaciones (marcas y versiones) sobre las que se encuentra el sistema?
- ¿Cuál es el estado de la(s) versión(es) actual(es) y su estabilidad?
- ¿Cuál es la disponibilidad de un sistema con datos reales o representativos para pruebas?
- ¿Quién es la persona para contacto y medio para el mismo (teléfono, correo electrónico, etc.)?

### 5.1.2 Desarrollo de un Documento de Análisis de Sistema de Información

El documento de análisis deberá incluir las secciones que se especifican a continuación y para las que se especifica su contenido y alcance.

#### Sección de Introducción

En esta sección se introduce en el estudio del sistema objeto del análisis.

#### Sección de Generalidades

En esta sección se documentan las generalidades del sistema analizado, presentando un panorama sobre los objetivos y/o las funciones del sistema, para qué se utiliza, quién lo utiliza, en qué plataforma se encuentra y su relevancia para el Sistema de Información Corporativa.

#### Sección de Objetivos y Alcance

En esta sección se establece el objetivo y el alcance del documento de análisis. Una oración típica para el establecimiento del objetivo sería:

*El objetivo de este documento es presentar el resultado del análisis del sistema X.*

Asimismo en esta sección se desglosa el contenido del documento en sus secciones principales.

#### Sección de Introducción al Sistema

En esta sección se hace una descripción general del sistema objeto del análisis que se documenta.

En términos generales, se amplía la descripción general hecha en la Introducción del documento, consistente en los objetivos y/o las funciones del sistema, para qué se utiliza, quién lo utiliza, en qué plataforma se encuentra y su relevancia para el Sistema de Información Corporativa.

#### **Sección de Estructuras de Datos**

En esta sección se documentan las relaciones existentes entre los archivos o las tablas de base de datos del sistema objeto del análisis.

La documentación se hace a través de uno o varios diagramas Entidad-Relación y de prosa que describe la semántica de dichos diagramas.

#### **Sección de Diccionario de Datos**

En esta sección se documentan las tablas o los archivos de base de datos del sistema objeto de análisis.

Para cada una de las tablas o archivos identificados en la sección anterior del documento se presenta información sobre su nombre y propósito, así como de los datos que almacena: nombre de campo, descripción de campo, tipo de dato, dominio o restricciones en los valores que puede tomar, pertenencia a llave primaria o alterna, pertenencia a llave foránea.

Las llaves se documentan con la siguiente sintaxis:

Si un atributo es la llave primaria o es parte de ésta, se marca con un número "1". Si un atributo es llave alterna o parte de una llave alterna, se marca con un "2" para la primera llave alterna, un "3" para la segunda y así sucesivamente.

Si un atributo es llave foránea o parte de una llave foránea, se marca con una letra "f" seguida del nombre de la tabla de la que se hereda el atributo entre paréntesis.

Si un atributo satisface varios criterios de llave, como por ejemplo es llave primaria y es llave foránea, cada uno de estos hechos se marca y se separa por comas.

En la siguiente página se incluye el formato de la matriz que será empleada para esta documentación.

#### **Sección de Especificaciones de Datos de Interés Corporativo o Información Directiva.**

En esta sección se identifican cuáles de los datos que se manejan en el sistema objeto del análisis son de interés corporativo.

Los datos de interés corporativo deben identificarse como *hechos*, en términos del Almacén de datos.

Los datos de interés corporativo pueden documentarse de diferentes maneras, según sea más conveniente.

- Haciendo una o varias matrices con columnas "Nombre de Tabla" y "Nombre de Atributo".
- Haciendo para cada tabla con atributos de interés una lista de éstos.
- Haciendo una o varias matrices con columnas "Nombre de Tabla" y "Nombre de Atributo Excluido".
- Haciendo para cada tabla con atributos de interés una lista de los atributos que se excluyen.

### **Sección de Flujo de Datos**

En esta sección se documenta el flujo de datos entre diferentes bloques funcionales del sistema objeto de análisis, con otros sistemas del proceso de CFE (Generación, Control o Transmisión) y con sistemas de otros procesos.

La documentación se hace a través de uno o varios diagramas de flujo y de prosa que describe la semántica de dichos diagramas.

Los diagramas de flujo de datos se harán con la notación propuesta por Yourdon [10].

También se deberá documentar el diccionario de datos del sistema origen, para esto se proponen las tablas 5.1 y 5.2:



<b>Tabla:</b> dd_ar					
<b>Descripción:</b> Esta tabla contiene el catálogo de áreas o centros de trabajo de generación en CFE.					
<b>Cardinalidad:</b> Un registro por cada área de trabajo.					
Atributo	Descripción	Unidad	Tipo de Dato	Dominio	Llave
área	Clave del área		char(5)		1
área_c	Clave del área consolidadora		char(5)		
cenace	Clave alterna de acceso al área		char(3)		
nombre	Nombre del área		char(50)		
password	clave de acceso al área		char(13)		
nivel	Nivel del área dentro de la organización (Central, Subgerencia, Región o Nacional)		char(1)	C, S, R ó N	
fua	Fecha de última actualización		date		
hua	Hora de última actualización		char(8)		

Tabla 5.2 Ejemplo de matriz de documentación de una tabla.

**FALTA**

**PAGINA**

**56 DE 56**

**Sección de Usuarios y Uso del Sistema**

En esta sección se documenta quiénes son los usuarios del sistema, en términos de su localización en la estructura jerárquica de CFE y qué uso hacen del sistema en términos de cuál es la periodicidad en la captura de datos y cuál es la periodicidad en la generación de reportes.

**Sección de Ubicación del Sistema**

En esta sección se documenta la localización física del sistema. Para los sistemas que se encuentran descentralizados, se documenta en qué niveles de la estructura jerárquica de CFE se encuentran, como por ejemplo en Planta, Región de Generación y Subdirección de Generación, o por ejemplo en Agencia, Zona, División y Gerencia de Distribución.

**Sección de Plataformas del Sistema**

En esta sección se documentan las plataformas de hardware y software sobre las que se encuentra implementado el sistema objeto del análisis, considerando los casos en los que haya múltiples versiones instaladas y en operación.

Se deberá documentar con el mayor detalle posible los nombres y las versiones de computadoras, sistemas operativos, manejadores de archivos, manejadores de base de datos y lenguajes de implementación.

Asimismo, se documentará si el sistema hace uso de telecomunicaciones, por qué medio y para qué propósito.

**Sección de Disponibilidad, Instalación y Pruebas**

En esta sección se identificará la existencia o no y, en su caso, la disponibilidad de un sistema para pruebas, ya sea real, o con datos realistas.

**Sección de responsables del Sistema**

En esta sección se presentan los datos de los responsables del sistema objeto del análisis. Por responsable se entiende ya sea la o las personas que lo desarrollaron o que lo mantienen y la o las personas que se encargan de su explotación.

Se deberá identificar uno de los responsables, el tipo de responsabilidad que tiene y los datos de localización, como ubicación física, teléfono y correo electrónico.

**Sección de Conclusiones y Recomendaciones**

En esta sección se presentarán las conclusiones y recomendaciones resultado del análisis llevado a cabo.

Como conclusión se puede reiterar sobre las generalidades del sistema analizado, presentando información sobre los objetivos y/o las funciones del sistema, para qué se utiliza, quién lo utiliza, en qué plataforma se encuentra y su relevancia para el Sistema de Información Corporativa.

Como recomendaciones se puede argumentar más a fondo sobre la inclusión o no del sistema como fuente de datos para el Sistema de Información Corporativa, así como lo que podrían ser consecuencias o puntos importantes a considerar a raíz de tal decisión, desde el punto de vista del Almacén de Datos, en relación por ejemplo con redundancia de datos, necesidades de unificación de criterios, de sincronización de actividades, etc.

## Sección de Referencias

En esta sección se lista la documentación disponible del sistema analizado, tal como manuales de usuario, manuales técnicos, ejemplos de reportes impresos. También se pueden incluir referencias a artículos u otras publicaciones técnicas relacionadas con el sistema o con la actividad para la que el sistema se encuentra en operación.

Cada referencia se numera de la 1 a la N y ese número se emplea entre paréntesis cuadrados en los lugares apropiados del texto de las secciones del documento de análisis.

A continuación se presenta un ejemplo de referencia:

[1] Centro de Capacitación Occidente. Manual Técnico GECOM 1.0, Tomo 1. 27 de abril de 1998.

### 5.1.3 Análisis del Sistema Gerencial Comercial (GECOM).

Siguiendo la guía que se desarrolló en la sección anterior, a continuación se presenta como un ejemplo, el documento *Análisis del GECOM*. A continuación se presenta el desarrollo de las secciones del documento.

#### I. Generalidades

Con el propósito de brindar información oportuna sobre el proceso de comercialización de la CFE a los ejecutivos de la Subdirección de Distribución, se ha desarrollado el Sistema Gerencial de la Base de Datos Comercial (GECOM), que permite la elaboración de reportes con información estratégica para la toma de decisiones.

#### II. Objetivos y Alcance

El objetivo de este documento es presentar el resultado del análisis del sistema GECOM.

El alcance de este documento es presentar el análisis del sistema a partir de las conversiones sostenidas con el responsable asignado para esta tarea.

El contenido de este documento comprende las siguientes 12 secciones:

La sección 1 corresponde a la introducción.

La sección 2 presenta una descripción general del sistema.

La sección 3 presenta los diagramas Entidad-Relación de las estructuras de datos del sistema.

La sección 4 incluye la documentación del diccionario de datos.

La sección 5 contiene los datos del sistema que son de interés corporativo.

La sección 6 describe el flujo de datos del sistema.

La sección 7 identifica los diferentes tipos de usuarios del sistema.

Las secciones de la 8 a la 11 incluyen información relativa a la plataforma del sistema, su ubicación física, así como su disponibilidad para pruebas y responsables del sistema.

La sección 12 incluye algunas observaciones o conclusiones obtenidas después del análisis del sistema.

Por último se incluye una sección de referencias.

### III Introducción al Sistema

El GECOM tiene como función principal la generación de reportes de la Base de Datos Comercial.

La elaboración de reportes se basa en la selección de datos de interés para el usuario tales como: División, Zona, Agencia, Entidad Federativa, Municipios, Poblaciones y Tarifas, permitiendo la creación de filtros que el sistema procesa, par ser visualizados e impresos.

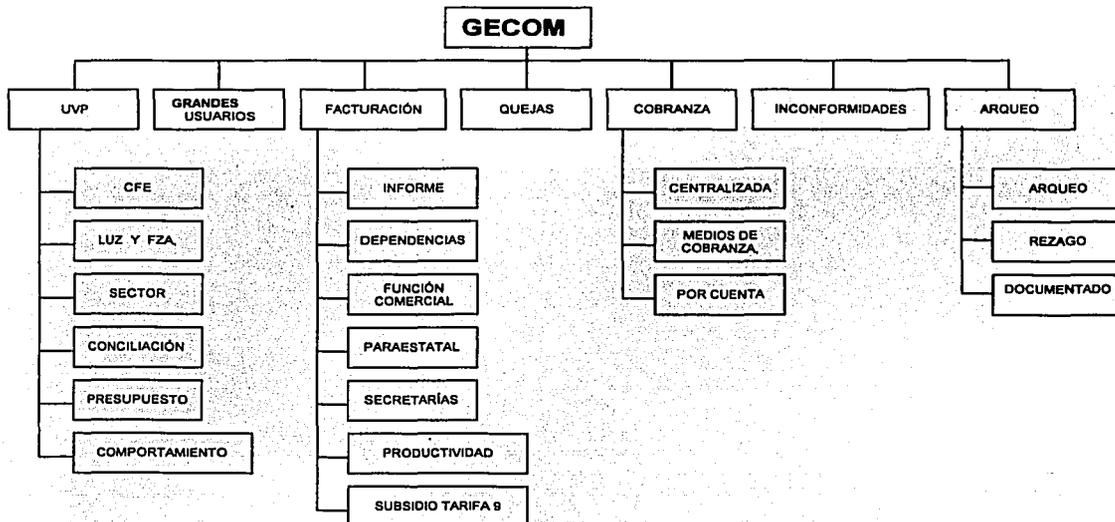


Figura 5.1. Diagrama general del GECOM

### IV Estructuras de Datos

Esta sección incluye los diagramas entidad-relación del sistema. Los diagramas presentados corresponden a cada uno de los módulos mencionados en la sección anterior.

En la figura 7.2 se presentan ejemplo de diagramas entidad-relación tomados del documento completo de análisis del GECOM.

TESIS CON  
FALLA DE ORIGEN

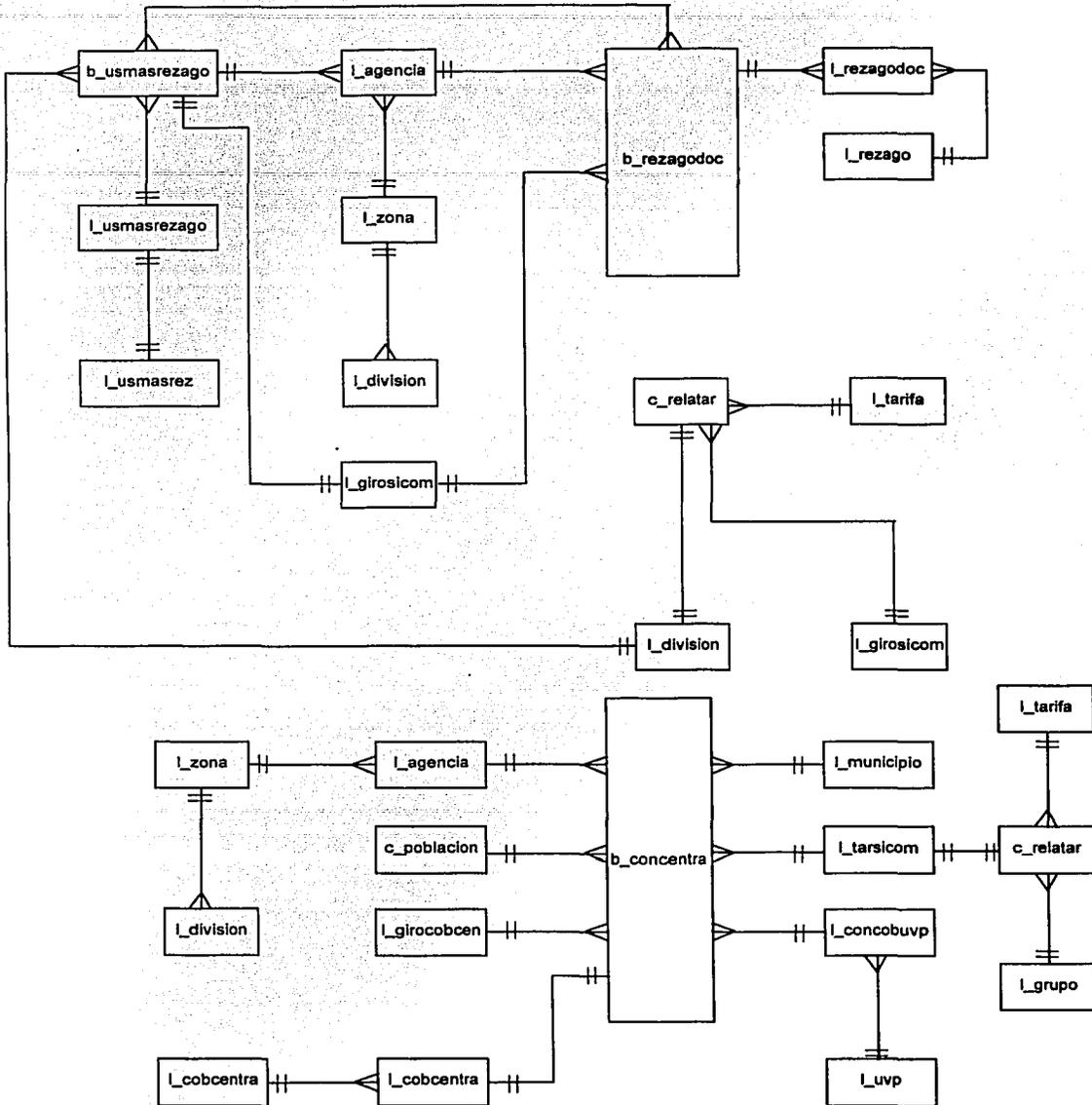


Figura 5.2. Diagramas para el módulo de Cobranza y Arqueo.

## **V Diccionario de Datos**

A continuación se presentan las descripciones de las tablas que componen la base de datos del Sistema.

La base de datos está compuesta por cuatro diferentes tipos de tablas.

Las tablas base, las cuales contienen los datos relacionados con las entidades que componen el sistema como arqueos, cobranza, uvp, rezago, quejas e inconformidades.

Las tablas tipo catálogo incluyen información sobre poblaciones, tarifas, cuentas, etc.

Las tablas tipo lista incluyen información sobre conceptos tales como empresas, universidades, delegaciones, divisiones, municipios, entidades federativas, tipo de inconformidades y tipo de quejas, etc.

Las tablas temporales almacenan la información que proviene del SICOM.

<b>Tabla: b_arqueo</b>				
<b>Descripción:</b> Información de arqueo y de rezago				
Atributo	Descripción	Tipo de dato	Dominio	Llave
cl_area	Clave del área de responsabilidad	Char	No nulos	1
cl_municipio	Clave del Municipio	Char	No nulos	1
cl_población	Clave de la Población	Char	No nulos	1
cl_tarifa	Clave de la tarifa eléctrica	Char	No nulos	1
cl_giro	Clave del giro del usuario	Char		1
cl_registro	Fecha de registro	Char	No nulos	1
cl_ciclo	Clave del ciclo de facturación	Char	No nulos	1
cl_tipanti	Clave del tipo de antigüedad	Char	No nulos	1
cl_conarq	Clave del concepto del arqueo	Char	No nulos	1
Un_conarq	Importe del concepto del arqueo	Char	No nulos	1
<b>Tabla: b_arqueoproceso</b>				
<b>Descripción:</b> Información de arqueo y rezago procesada				
Atributo	Descripción	Tipo de dato	Dominio	Llave
cl_area	Clave del área de responsabilidad	Char	No nulos	1
cl_municipio	Clave del Municipio	Char	No nulos	1
cl_población	Clave de la Población	Char	No nulos	1
cl_tarifa	Clave de la tarifa eléctrica	Char	No nulos	1
cl_giro	Clave del giro del usuario	Char		1
cl_registro	Fecha de registro	Char	No nulos	1
cl_tipanti	Clave del tipo de antigüedad	Char	No nulos	1
cl_conarq	Clave del concepto del arqueo	Char	No nulos	1
Un_conarq	Importe del concepto del arqueo	Decimal	No nulos	
<b>Tabla: b_baleen</b>				
<b>Descripción:</b> Información del balance de energía				
Atributo	Descripción	Tipo de dato	Dominio	Llave
cl_area	Clave del área de responsabilidad	Char	No nulos	1
cl_registro	Fecha de registro del balance de energía	Date	No nulos	1
cl_conbalene	Clave del concepto del balance de energía	Char	No nulos	1
Un_conbalene	Valor del concepto del balance de energía	Decimal	No nulos	

Tabla 5.3 Ejemplo de diccionario de datos

## **VI Especificaciones de Datos de Interés Corporativo**

La función principal del GECOM es la generación de reportes orientados para la toma de decisiones, a partir de la información proveniente de otros sistemas.

Los reportes incluyen información sobre:

- Usuarios, ventas y productos
- Facturación
- Cobranza
- Arqueo

Un reporte puede generarse a través del uso de los siguientes criterios de selección:

- Giro
- División
- Zona
- Agencia
- Entidad Federativa
- Municipios
- Poblaciones
- Tarifas
- Fecha

Las tablas involucradas para llevar a cabo este proceso se presentan en el diagrama 6.7 de la siguiente sección.

## **VII Usuarios y Uso del Sistema**

El sistema actualmente está en proceso de instalación. Se instalará en cada una de las divisiones de la CFE,

La información mensual que requiere el GECOM debe estar en los servidores de cada división al 5° día hábil de cada mes siguiente.

Los archivos generados por división son recuperados de manera automática mediante el uso de un extractor y se concentran en Oficinas Nacionales.

Los administradores del sistema en la Gerencia Comercial, localizada en Oficinas Nacionales.

La información histórica almacenada en el sistema ha sido recuperada y concentrada desde la GIT y data desde 1988.

La emisión de reportes es mensual.

## **VIII Ubicación del Sistema**

El sistema está en proceso de instalación en las divisiones de la CFE como se explicó en la sección anterior.

## **IX Plataformas del Sistema**

Los requerimientos del GECOM son:

---

**EQUIPO DE COMPUTO****CAPACIDAD DE ALMACENAMIENTO****SOFTWARE**

Servidor. Sistema operativo SCO UNIX versión 3.0 en adelante  
Informix On-line  
Informix Star

Terminal Windows 95 o posterior  
Archivo ejecutable del proyecto  
Borland Databases Engine (BDE)  
SQL Link  
ReportSmith Run-Time

Adicionalmente,

INFORMIX NET  
Emulador (TCP/IP o VISUAL APEAL)

**X Disponibilidad, Instalación y Pruebas.**

Actualmente se encuentra disponible la versión 1.0 del GECOM, en cuanto a su documentación existen dos manuales:

Manual Técnico actualizado al 27 de abril de 1998.

**XI Responsables****XII Conclusiones y recomendaciones**

El GECOM es una herramienta primordial para la Gerencia Comercial para la obtención de información oportuna relacionada con el proceso de comercialización de la energía eléctrica. Su objetivo es proporcionar a los ejecutivos información estratégica para mejorar el proceso de toma de decisiones, por lo cual se considera indispensable la inclusión de la información del GECOM en el almacén del SICORP.

Es conveniente mencionar que para la elaboración del presente documento se utilizó un esquema de la base de datos que ha sido actualizado recientemente.

En la siguiente etapa del proyecto se consideraran dichos cambios.

**XIII Referencias**

Manual Técnico GECOM 1.0, Tomo I y II

### **5.1.5 Conclusión de la sección de análisis.**

Una vez concluida la sección de análisis debemos estar en condiciones para responder a las preguntas planteadas en la sección 6.1.1 *Alcance*. Este análisis detallado del entorno de la información de la empresa, en este caso del área de distribución de la CFE; es el punto de partida para iniciar realmente con el diseño y desarrollo del Sistema de Información Corporativa. Si no se alcanza un adecuado entendimiento de las fuentes de información al almacén de datos no se recomienda pasar a las siguientes etapas del proyecto.

Para que el trabajo de análisis arroje los resultados esperados, lo deberá realizar un especialista con conocimientos en las áreas de sistemas operativos; bases de datos; ingeniería de software y Redes de computadoras.

### **5.2 Extracción/Transformación de información.**

Una vez identificados los datos de interés corporativo el proceso siguiente es extraerlos desde el sistema fuente, para éste fin es necesario desarrollar programas de extracción de datos a los que llamaremos "extractores".

Los extractores tienen bajo su responsabilidad extraer del sistema fuente el conjunto de datos previamente seleccionado, sus principales características son que deberán ejecutarse automáticamente y periódicamente. Automáticamente por que no deberá ser necesaria la intervención de un operador para su ejecución y periódicamente por que el programa deberá ser capaz de ejecutarse cada que los datos a mover se encuentren debidamente validados, esto se da en periodos determinados, particularmente los datos en la Subdirección de Distribución de la CFE tienen una periodicidad mensual.

Aún cuando se recomienda que los programas extractores se ejecuten en el equipo de cómputo donde opera el programa fuente de los datos y generen un archivo plano que posteriormente será transportado, transformado e insertado a las estructuras de datos multidimensionales del almacén; es necesario ser flexible y adaptarse a las condiciones de la arquitectura particular de cada sistema fuente. Estas condiciones deberán estar contenidas en el documento de análisis del sistema.

En el caso del programa GECOM que contiene los datos del proceso comercial de la CFE, la base de datos de la que habrá que extraer la información se encuentra en el RDBMS Informix, en un servidor UNIX enlazado en la misma red de datos al servidor que contiene el almacén de datos en el RDBMS Oracle. Por lo anterior el proceso de extracción estará a cargo de un programa escrito en SQL que en periodos mensuales genere un archivo plano y lo deposite en el servidor destino.

Los pasos generales para migrar información de los sistemas operacionales al Almacén de Datos son los siguientes:

#### **5.2.1. Leer la información legada.**

Es muy común que la información fuente se encuentre en sistemas legados, en la mayoría de los casos estos sistemas no cuentan con documentación actualizada de la operación o si la tienen esta documentación no presenta las características requeridas en el punto 6.1. Cuando se presenta este caso es necesario desarrollar por completo la fase de análisis como se vio anteriormente.

En casos extremos, por falta de documentación y principalmente por que algunos sistemas legados corren en plataformas obsoletas no es posible leer la información por que no se conoce el formato en que está almacenada.

### **5.2.2. Determinar información actualizada.**

Debido a que la información se deberá extraer en paquetes periódicos, es importante identificar en el sistema legado la información a través de estampas de tiempo. En algunas, particularmente cuando los datos guardados no contienen la estampa de tiempo, este proceso puede resultar difícil. Es el caso de sistemas que manejan el tiempo a través de almacenar diferentes archivos de datos ubicados en el tiempo únicamente por el nombre del archivo.

### **5.2.3. Generalizar llaves para dimensiones cambiantes.**

Si se quiere, por ejemplo, rastrear descripciones de clientes que han cambiado, las llaves respectivas deben ser generalizadas para añadir nuevos registros de dimensión. Generalizar estas llaves es tan fácil como aumentar 2 o 3 dígitos de versión o tan complicado como crear llaves anónimas.

### **5.2.4. Combinar fuentes separadas bajo un sólo registro.**

En el proceso de extracción, también es importante tomar en cuenta que en la mayoría de los casos las tablas de hechos y dimensiones se obtienen en una relación muchos a uno; por lo que es frecuente que el código SQL que se genere para extraer los datos tome datos de diferentes archivos o tablas fuente.

### **5.2.5. Crear imágenes de los registros.**

Otro aspecto recomendable que se deberá tomar en cuenta en el proceso de extracción es crear una imagen de los registros obtenidos. No es recomendable que se inserten directamente los registros al almacén sin que se genere un archivo con cada uno de los registros que se cargarán. Los registros contenidos en este archivo no deberán ser agregados y cada dato deberá estar en formato estándar (fechas, números, cantidades)

### **5.2.6. Migrar la información al almacén.**

Se puede hacer en cualquier momento entre los pasos 7.2.2 al 7.2.8, pero se recomienda hacerla en este punto.

Es aquí donde se deberán aplicar las rutinas de transformación que aplique:

- Las rutinas de transformación (comparar, combinar, transformar, etc.) se implementan más naturalmente en la máquina fuente.
- Es el proceso formal de transformar datos de la fuente de información a una arquitectura común.
- La información no se debe pasar de una BD a otra sin que ésta sea transformada formalmente, para prevenir y resolver las desigualdades.
- La transformación de datos no es una tarea trivial.

## **5.3 Inserción/Carga de Información.**

Los insertores son los programas encargados de cargar los datos tomados de los sistemas fuentes a las estructuras multidimensionales del almacén de datos corporativo.

El proceso de extracción/transformación generalmente construye información basada en el modelo estrella. Sin embargo, para que los almacenes de datos tengan mayor rendimiento generalmente los podemos encontrar en el mercado, basados en tecnología de base de datos multidimensional.

Los siguientes aspectos se deberán tomar en cuenta en el diseño y desarrollo de los programas de carga de información al almacén de datos:

### 5.3.1 Ordenar los registros cargados para crear agregados.

Crear agregados es ordenar los registros y luego crear renglones con resúmenes.

En la medida de lo posible, se deben crear los agregados fuera del DBMS, por dos razones:

□ Existen herramientas para hacer esto fuera del DBMS.

□ Agiliza la carga masiva de los registros, ya que es más fácil insertar cada agregado; además quedan como respaldo en el snapshot.

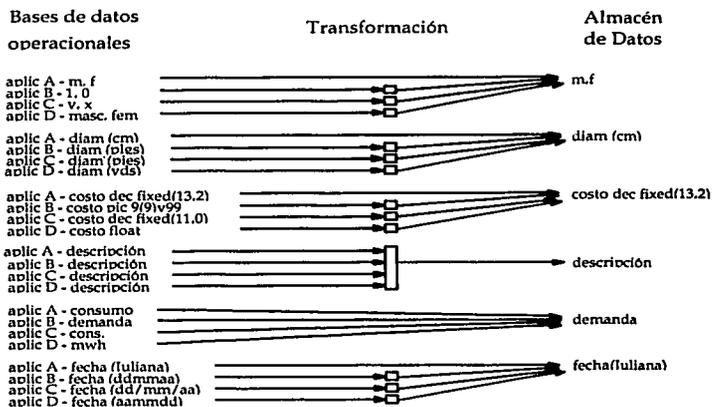


Figura 5.3 Migrar la información al almacén (Transformación)

### 5.3.2 Generalizar llaves para registros agregados.

Los registros agregados necesitan llaves completamente artificiales y no deben tener conflictos con las llaves de los registros base.

### 5.3.3 Cargar masivamente los registros checando integridad referencial.

Una vez que todos los registros han sido ensamblados deben pasarse al almacén, es deseable eliminar los índices en esta etapa y es importante reforzar la integridad referencial.

En un esquema de estrella, la integridad referencial significa que los componentes de la tabla de hechos siempre tienen llaves foráneas con respecto a cada una de las dimensiones. En almacenes con millones de registros es muy costoso checar la integridad referencial.

### 5.3.4 Manejar excepciones del proceso de carga.

TESIS CON  
FALLA DE ORIGEN

Seguramente habrá registros que no pudieron ser cargados, por ejemplo por violación de la integridad referencial. Se deben concentrar esos registros, componerlos (ya sea que el problema está en ellos o en la tabla de hechos) y cargarlos nuevamente.

Si por falta de criterios sistematizados en la solución problemas de carga no es posible cargar algunos registros, se deberán almacenar en un archivo de log que posteriormente será analizado por personal especializado a efecto de que se corrijan los problemas presentados.

### **5.3.5 Asegurar calidad de los datos cargados.**

La calidad es el estado de totalidad, validez, congruencia, oportunidad y exactitud que hace a la información apropiada para un uso específico.

Totalidad: Grado en el que están disponibles todos los datos necesarios para cumplir con las demandas de información.

Validez: Grado en el que los datos tienen valores que caen dentro de sus respectivos dominios.

Congruencia: Grado de coherencia lógica entre los datos que los libera de variación o contradicción.

Oportunidad: Grado en el que los datos pueden ser obtenidos al tiempo requerido y/o especificado.

Exactitud: Grado de acuerdo entre un conjunto de valores de datos y su correspondiente conjunto de valores correctos.

Antes de que empiece a funcionar el Almacén de Datos, se debe asegurar la calidad de la información cargada para su consumo. Los usuarios esperan información confiable.

Para este efecto se recomienda elaborar procedimientos a través de los cuales sea posible sistematizar y controlar todos los procesos operativos requeridos en la alimentación del almacén de datos. De ser posible se deberá incluir esto en el manual de aseguramiento de calidad de la institución a efecto de que quede certificado el proceso bajo la norma ISO 9000.

Con lo anterior se contará con un conjunto de índices de calidad del almacén de datos que podrá ser monitoreado a través de un tablero de control.

### **5.3.6 Publicar**

Publicar a los usuarios que la información esta lista para ser leída. Puede ser a través de llamadas telefónicas, e-mails o incluso avisarles personalmente.

Esto debe incluir un resumen del status de carga del día anterior y el índice de calidad correspondiente.

## **5.4 Diseño de la Base de Datos Multidimensional**

Una vez que se han analizado los sistemas fuente, y se han diseñado y desarrollado los programas de extracción, transformación, transporte y carga; el siguiente paso es diseñar las estructuras de datos del almacén, para los aspectos teóricos del modelado de bases de datos multidimensionales.

### **5.4.1 Consideraciones en el diseño.**

El diseño debe ser independiente

- del cambio.
- de los datos.
- de los esquemas.
- del lenguaje.
- de los sistemas operativos.
- de los métodos de acceso.
- de los manejadores de bases de datos.

#### **5.4.2 Claves de diseño.**

Estas claves se recomiendan para que se haga un buen diseño, de su cumplimiento depende el futuro éxito del proyecto.

##### **Entender los esquemas fuente**

Del buen entendimiento de los esquemas fuente depende en buena medida el éxito del diseño. Es necesario documentar y comprender adecuadamente, los tipos de datos y atributos de la información que se integrará así como condiciones en la administración y evolución propia de cada sistema fuente como es el caso del control de cambios.

##### **Mapeo de datos fuente**

El proceso de transformación se deberá realizar cuidadosamente, de manera que no se cometan errores en la interpretación y conversión de datos.

##### **Desnormalización**

Es una práctica frecuente y correcta la desnormalización de estructuras de datos relacionales cuando se modela esa misma información en estrellas o cubos.

##### **Carga inicial y carga periódica.**

Se deberán considerar los criterios de carga de datos dependiendo la periodicidad y la disponibilidad de recursos de cómputo. Lo que se hace frecuentemente es una carga inicial completa y después reprocesar los cubos con cargas incrementales con los registros faltantes.

##### **Herramientas de explotación**

Al diseñar las estructuras del almacén también es importante considerar el tipo de aplicación (OLAP, Minería, Integraciones ad hoc, etc.) que se utilizará para explotar los datos. Algunas herramientas OLAP requieren llaves adicionales para su operación.

##### **Metadatos**

Se deberán considerar los metadatos y su administración. Recuérdese que los metadatos son información relativa a los propios datos de la Base de Datos.

#### **5.4.3 Ejemplo de diseño con un cubo con información del proceso comercial en la CFE.**

En la figura 5.4 se muestra el diseño del cubo de usuarios, ventas y producto. A continuación se presentan los diccionarios de datos de la tabla de hechos y las dimensiones así como sus metadatos

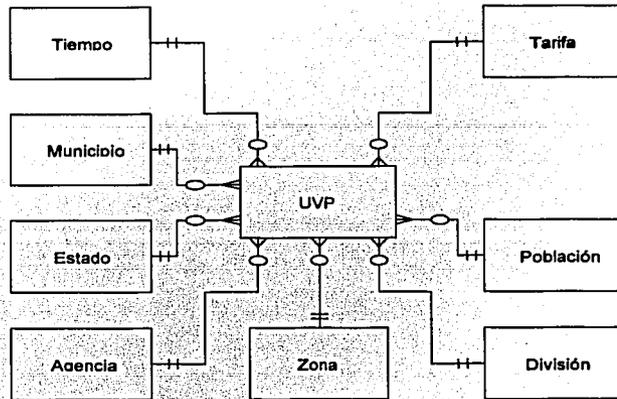


Figura 5.4 Estrella de Usuarios, Ventas y Productos

Tabla: hUVP						
Descripción: Información de usuarios, ventas y productos						
Cardinalidad: mensual						
Atributo	Descripción	Fuente	Unidad	Tipo de Dato	Dominio	Llave
id_fecha	Identificador de la fecha			Int	Secuencial 1,2...	1
id_agencia	Identificador de la agencia			int	Secuencial 1,2...	1
id_zona	Identificador de la zona			int	Secuencial 1,2...	1
id_división	Identificador de la división			int	Secuencial 1,2...	1
id_tarifa	Identificador de la tarifa			int	Secuencial 1,2...	1
ImporteV	Importe por concepto de ventas	nu_uvp	pesos	Decimal(16,2)		
ImporteU	Importe por concepto de usuarios	nu_uvp		Decimal(16,2)		
ImporteP	Importe por concepto de productos	nu_uvp		Decimal(16,2)		

Tabla 5.4 Diccionario de datos de un cubo

TESIS CON  
 FALLA DE ORIGEN

## 5.5 Herramientas de explotación de los datos

Hasta el momento, el desarrollo del Sistema de Información Corporativa tiene como principal producto una gran base de datos. Gracias al trabajo de análisis de las fuentes de información y diseño y desarrollo del software de extracción, transformación, transporte y carga de información; contamos con una base de datos que se actualiza automática y periódicamente desde donde se originan los datos.

En esta sección se presentarán algunos programas de explotación a un almacén de datos. Como toda base de datos, es posible consultarla desde cualquier sistema capaz de conectarse vía ODBC, OLE DB, o cualquier otro medio similar. Sin embargo, para que este tipo de bases de datos pueda servir como base para las aplicaciones de soporte a la toma de decisiones; típicamente las aplicaciones que la accesan son OLAP *Análisis en línea de procesos*, Tablero de Control y minería de datos.

### 5.5.1 Análisis en línea de procesos (OLAP)

Para realizar el análisis de procesos en línea se desarrollo una aplicación que cuenta con una interfase de usuario Web para ser usada en la intranet de la CFE.

Esta aplicación esta basada en la metodología OLAP (Online Analytical Process) [1] para analizar procesos en línea. Las aplicaciones OLAP están orientadas al análisis de datos. Para mayor referencia de las características de un sistema basado en tecnología OLAP,

La interfase de usuario del SICORP esta completamente orientada a la Web para ser accedada desde la Intranet de la CFE. En el acceso se solicita una clave de usuario y un password, de esta forma es posible asignar diferentes niveles de acceso a la información. Ver figura 5.5.

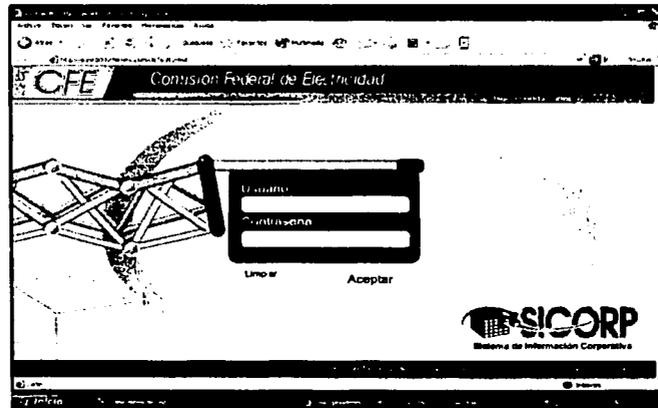


Figura 5.5 Acceso al sistema OLAP del SICORP.

Una vez que el usuario proporciona su clave y password, el sistema presenta un menú organizado por los procesos de la cadena de valor de la CFE. Ver figura 5.6.

Al seleccionar las opciones del menú, el sistema nos lleva a un cubo de información. Visto esquemáticamente, el cubo se representa (para tres dimensiones) como se muestra en la figura 5.7. Los datos se almacenan y se recuperan haciendo referencia a sus dimensiones, también podemos ver a las dimensiones como dominios y criterios de clasificación y/o selección de información en una base de datos.

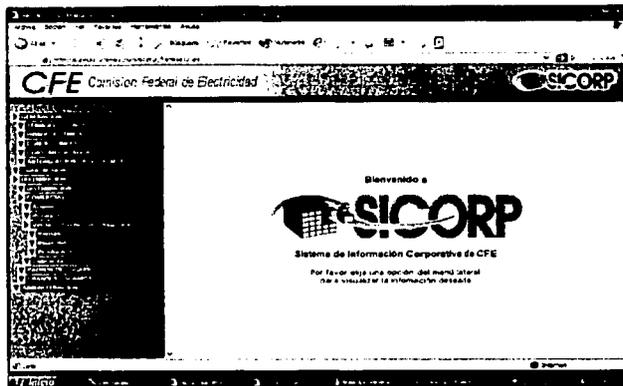


Figura 5.6 Menú del SICORP.

En el ejemplo de la figura 5.7, la métrica 5; una venta por un valor de 5. La podemos ubicar en el contexto de la empresa al identificar sus dimensiones: Tarifa, Regiones y Tiempo. Siguiendo el ejemplo, encontramos que la venta 147, se realizó con tarifa 5; en la región norte y en el año. Adicionalmente los cubos de las bases de datos multidimensionales almacenan datos agregados y desagregados, de manera que si queremos desagregar el dato en una o más de sus dimensiones los podemos hacer fácilmente, a este proceso se le llama "Drill Down".

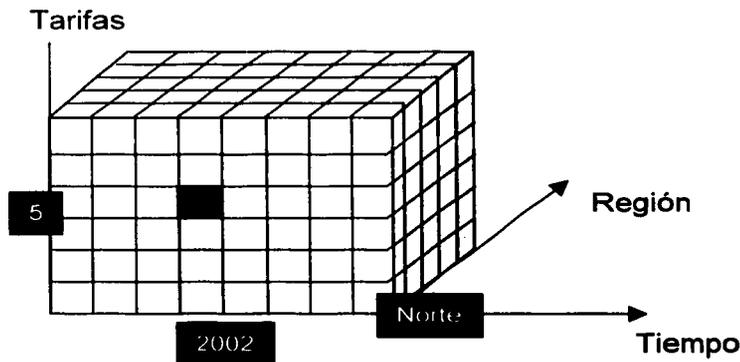


Figura 5.7 Cubo multidimensional.

#### Rotación de Cubos.

Para entender mejor las características de los cubos y las ventajas que estas nos dan en las actividades de análisis de información; veamos un ejemplo completo con la herramienta OLAP del SICORP.

Se trabajará con un cubo que contiene información de personal, este cubo es generalmente usado por el área de recursos humanos para diversas actividades de análisis de información. Exploraremos las funciones de Rotación, Drill Down, Drill Up; Análisis en línea y construcción de cubos virtuales.

Obsérvese que para todas las láminas el cubo mostrado contiene la misma información del área de recursos humanos, pero vista en diferentes perspectivas. Esta es la característica principal de las herramientas OLAP, el poder manipular estos cubos de información nos da fabulosas oportunidades de análisis.

La siguiente figura muestra el resultado de aplicar la función de rotación en el cubo. Nótese que este cubo cuenta con seis dimensiones; en las columnas el Género; en los renglones el estado Civil y en las listas superiores: Dependientes económicos; Tipo de jornada; Áreas de responsabilidad y Tipo de contrato. Después de la operación de rotación, las columnas se han intercambiado por los renglones. Generalizando, en la rotación es posible intercambiar cualquier dimensión por otra.

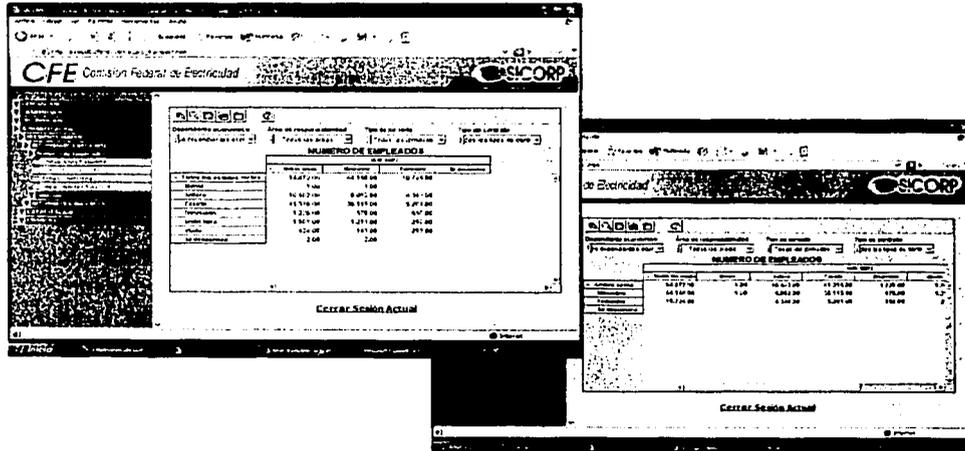


Figura 5.8 Rotación de Cubos.

### Drill Down y Drill Up

Para presentar el concepto de Drilling, es necesario recordar el concepto de jerarquías en las dimensiones.

El concepto de jerarquía se aplica a las dimensiones; en una dimensión existen jerarquías cuando es posible agregar los elementos de su dominio en subconjuntos. Por ejemplo, la dimensión tiempo se puede agrupar en años, semestres, cuartos, trimestres, bimestres, meses, semanas y días; las definiciones y cantidad de niveles corresponden a los diseños particulares de cada cubo.

Dicho lo anterior, la función de Drill Down en un cubo, permite explorar en un mayor nivel de detalle la información de una dimensión. Por ejemplo, si se estaba observando información anual, será posible seleccionar un año y explorar el comportamiento de los datos mensualmente.

La función de Drill Up hace lo contrario al Drill Down, si se tiene un cubo desplegando datos de algún nivel de detalle de alguna dimensión; será posible agregar nuevamente esa información al nivel anterior. En la figura 5.9 y 5.10 se presentan ejemplos de Drill Down y Drill Up respectivamente.

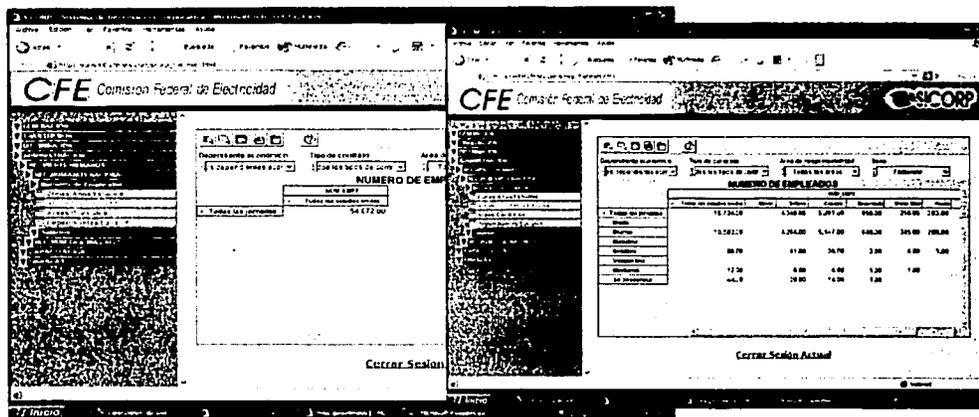


Figura 5.9 Drill-Down.

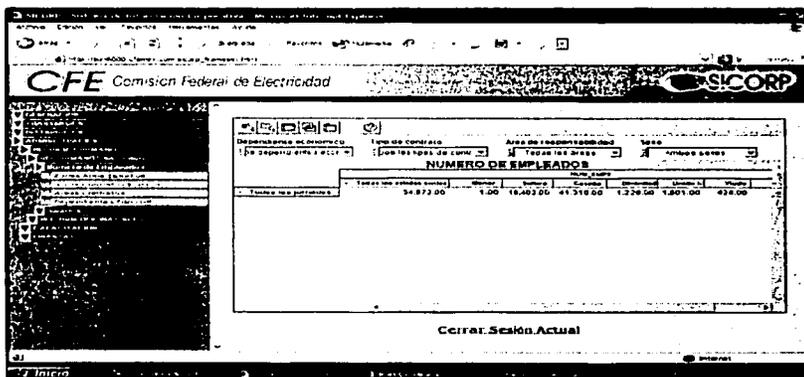


Figura 5.10 Drill-Up

### Análisis Multidimensional

Cuando el trabajo de construcción de un almacén de datos es exitoso. Es decir, cuando se cuida que el análisis de las fuentes de información; la alimentación a la base de datos (extracción, transformación, transporte y carga); el diseño de las estructuras de datos y las herramientas de explotación se realizaran adecuadamente. Entonces, estaremos en condiciones de proporcionar poderosas herramientas de consulta en línea a las áreas de análisis, planeación y control de gestión de la CFE.

Para ilustrar lo anterior se presenta el siguiente ejemplo. Supóngase que es interés de algún funcionario de la CFE saber la distribución de personal por género y tipo de contrato con el propósito de determinar si lanzará una campaña para promover que las mujeres participen más en puestos ejecutivos en la CFE.

Como hemos mencionado anteriormente estas herramientas deberán apoyar al proceso de toma de decisiones, veamos como funciona:

Como hemos visto con los ejemplos anteriores, contamos con un cubo de personal que incluye además de algunas otras dimensiones, la dimensión de género y tipo de contrato. Para contestar esta pregunta lo primero que haremos, rotando el cubo, ubicar la dimensión de sexo (género) en las columnas y la de tipo de contrato en los renglones; después mediante el proceso de Drilling nos aseguraremos de que las otras dimensiones presenten los datos más agregados a efecto de que la selección incluya a todos los trabajadores de la CFE. El resultado de lo anterior se ve de forma tabular en la figura 5.11. y de forma gráfica en la figura 5.12.

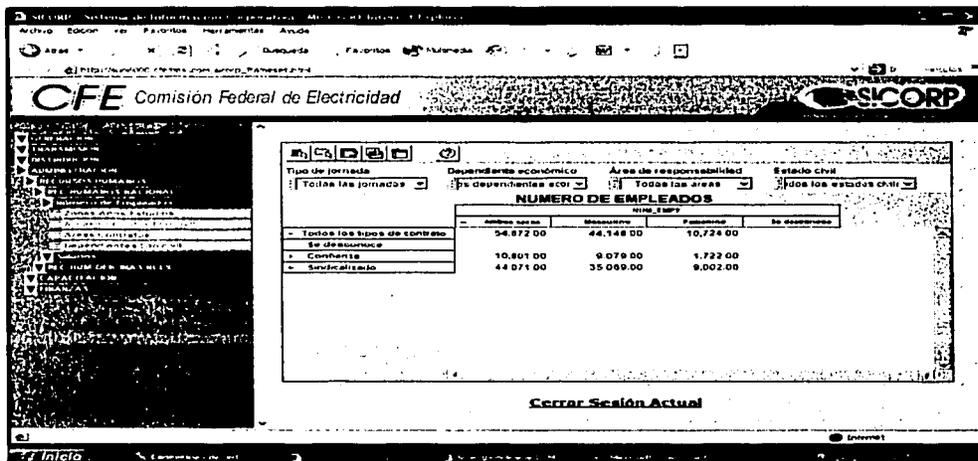


Figura 5.11 Análisis multidimensional. Distribución por género en los tipos de contrato.

La selección realizada se puede observar en forma tabular y gráfica directamente en la herramienta OALP.

Aún cuando las herramientas de análisis OLAP proporcionan muchas posibilidades de exploración de datos; en ocasiones es necesario exportar los datos obtenidos a una hoja de cálculo para

mejorarles el formato de presentación o incluir algunos datos calculados que el sistema OLAP no contemplaba en su diseño inicial. En este caso la consulta obtenida se envió a excel obteniendo lo que se presenta en la figura 5.13.

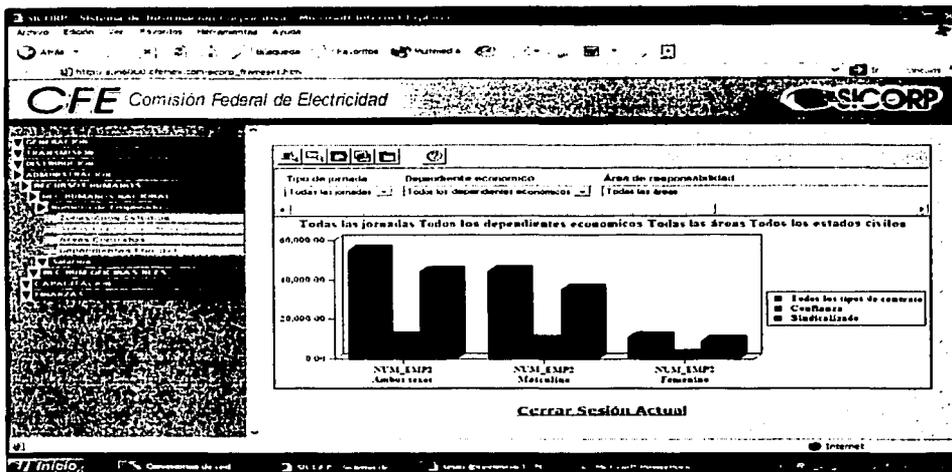


Figura 5.12. Análisis Multidimensional. Vista Gráfica

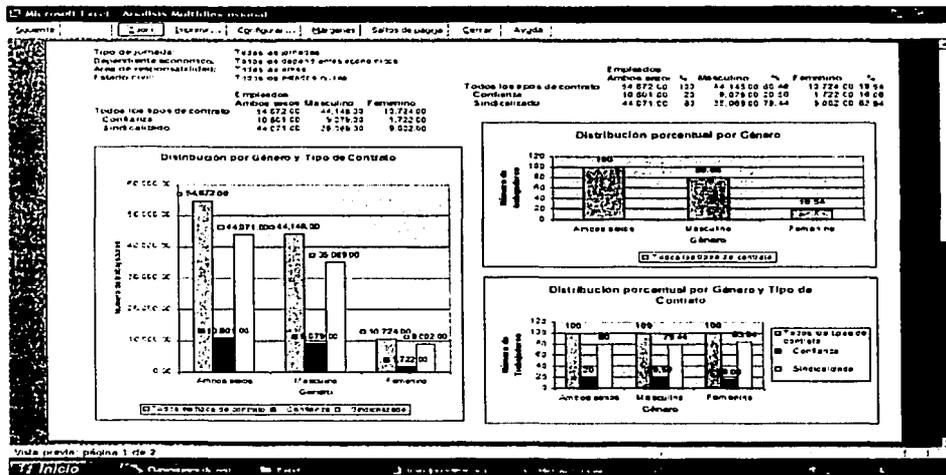


Figura 5.13. Análisis Multidimensional. Apoyo con Excel.

Al analizar los datos de la lámina anterior obtenemos los siguientes hechos:

1. De los 54,872 trabajadores activos en la CFE; 44,148 son hombres, lo que representa el 80.46% del total; y 10,724 son mujeres, lo que representa el 19.54%.
2. De los 54,872 trabajadores activos en la CFE; 10,801 son de confianza, lo que representa el 20% de la muestra; y 44,071 son sindicalizados, lo que representa el 20%.
3. De los 44,148 hombres; 9,079 son de confianza, lo que representa el 20.56%; y 35,069 son sindicalizados, lo que representa el 79.44%.
4. De las 10,724 mujeres; 1,722 son de confianza, lo que representa el 16.06%; y 9,002 son sindicalizados, lo que representa el 83.94%.

Para comprender mejor el análisis numérico es necesario explicar la siguiente regla de negocio:

Regla de negocio incluida en el análisis: Trabajadores sindicalizados y de confianza.

Los trabajadores sindicalizados o de base comprenden la base trabajadora de la CFE y están protegidos por un contrato colectivo de trabajo que detalla la relación entre la empresa, los trabajadores y su sindicato. El contrato colectivo de trabajo especifica las obligaciones y derechos de los trabajadores.

Los trabajadores de confianza ocupan los cargos de dirección, jefaturas de diferentes niveles y actividades de análisis, investigación y desarrollo. Por lo anterior, los trabajadores de confianza perciben mejores salarios que los sindicalizados y aunque no son regidos completamente por el contrato colectivo de trabajo, tienen mejores opciones de crecimiento y desarrollo personal y laboral.

Adicionalmente a la regla de negocio anterior también será necesario considerar la siguiente condición social:

Aunque las condiciones de desarrollo profesional son mejores para los trabajadores de confianza; su nivel de participación en actividades de alta responsabilidad es mayor. Por lo anterior los trabajadores de confianza tienen menos tiempo para atender sus asuntos personales que los sindicalizados.

Por esta condición es fácil explicarse que hay más mujeres en puestos sindicalizados que de confianza. Esto les da mayor oportunidad de estar al pendiente de sus familias, actividad que tradicionalmente ha desarrollado la mujer.

Análisis de los hechos anteriores y la regla de negocios presentada:

1. De todos los trabajadores en la CFE 80.46 son hombres y 19.54 mujeres, si consideramos también que del 100% , 20% son de confianza y 80% sindicalizados. Estos porcentajes se ven muy parecidos, por lo que de este primer análisis se pudiera pensar que la distribución de género y tipo de contrato es muy homogénea en la CFE.
2. Continuando con el análisis encontramos que; el 20.56% del total de los hombres son de confianza y el 79.44 sindicalizados; y del total de las mujeres el 16.06% son de confianza y el 83.94% sindicalizadas. Al analizar este segundo punto, encontramos que existe una diferencia de 4.5 puntos porcentuales entre la distribución de género y tipo de contrato que favorece a los hombres.
3. Analizando la regla de negocio y condición social anteriormente presentada es muy fácil explicar estos 4.5 puntos porcentuales.

Conclusiones del análisis anterior:

1. El ejecutivo decidió que no hace falta instrumentar una campaña que fomente mayor participación de las mujeres en puestos de confianza en la CFE.
2. Sin embargo, después del análisis de las cifras anteriores el ejecutivo decidió instrumentar una campaña para fomentar las contrataciones de mujeres en la CFE, ya que la proporción de 80% hombres y 20% mujeres le pareció muy desigual en un México que busca la igualdad entre hombres y mujeres.

### 5.5.2 Sistema de Información para la Alta Dirección o Tablero de Control de la CFE.

Como se puede observar en la sección anterior, las herramientas del tipo OLAP proporcionan grandes posibilidades de análisis de información. Son parte fundamental en el soporte a la toma de decisiones de las organizaciones en la actualidad. Sin embargo, la mayoría de los ejecutivos de alto nivel tienen muy poco tiempo como para que ellos personalmente analicen la información contenida en el almacén de datos, esta tarea la realizan generalmente sus asesores o sus áreas de análisis.

Los directivos requieren de herramientas informáticas que de una forma rápida y sencilla los informen de los aspectos más importantes en la operación de su empresa, esto se hace a través de índices o indicadores agrupados por las perspectivas fundamentales de la institución.

Los índices que se le presentan al directivo deberán ser los mínimos suficientes para que en pocos minutos se informe de las condiciones reales en la operación de la empresa.

Estos índices se presentan en un Tablero de Control cuya presentación es tabular y gráfica.

A este tipo de sistemas se les conoce generalmente como EIS *Ejecutive Information System* o SIE (en español).

Para resolver esta necesidad se desarrolló para la CFE un Sistema Ejecutivo al que se llama Sistema de Información para la Alta Dirección (SIADIR).

A Continuación se describe la operación y las partes fundamentales del sistema.

La estructura de almacenamiento y presentación de la información en el SIADIR corresponde a la estructura organizacional de los procesos en la CFE. Las áreas de la Dirección de Operación a nivel de subdirecciones también corresponden a los procesos, que son: Generación, Transmisión, Control, Distribución y Comercial. Esta estructura es jerárquica y se modeló en el sistema en una estructura arbolada, en los nodos terminales se tienen los indicadores. Es decir, a cada área se le asignan un conjunto de indicadores, estos indicadores son tomados del almacén de datos.

El menú principal del sistema presenta la estructura orgánica de la CFE (los procesos) y otra estructura arbolada que contiene los productos de información. Los productos de información son los reportes o formatos, agrupaciones especiales de índices que corresponden a necesidades específicas de la dirección.

En la figura 5.14 se observa el menú del sistema. En la figura 5.15 se observa la estructura de procesos y productos, respectivamente, parcialmente desagregada.

Al seleccionar alguna de las opciones del menú; ya sea un proceso o un producto de información, el sistema presenta el tablero de control. La siguiente figura presenta una pantalla tabular correspondiente al producto *Junta de Gobierno*.

Los indicadores tienen las siguientes características: Descripción, unidad de medida, valor programado, valor real, polaridad, temperatura y semáforo.

La temperatura determina el nivel de validación del dato, en la mayoría de los casos se consideran dos valores; validado y no validado, para este caso el validado se representa con el renglón en negritas y el no validado con tonalidad más baja.

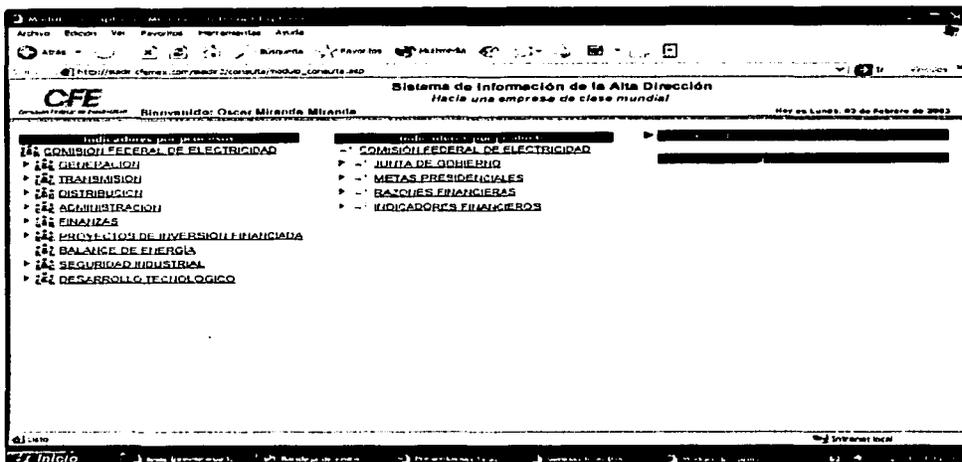


Figura 5.14. Menú del SIADIR.

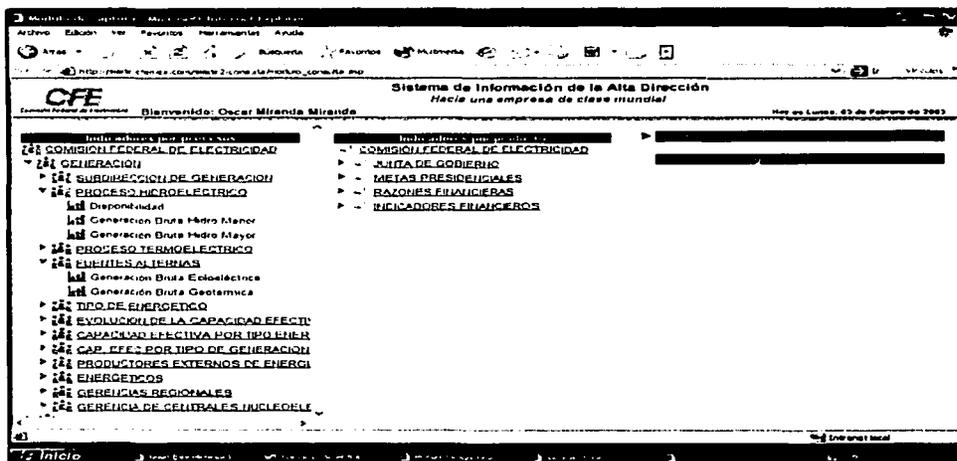


Figura 5.15 Estructura Arbolada. Los indicadores en los nodos terminales.

TESIS CON  
FALLA DE ORIGEN

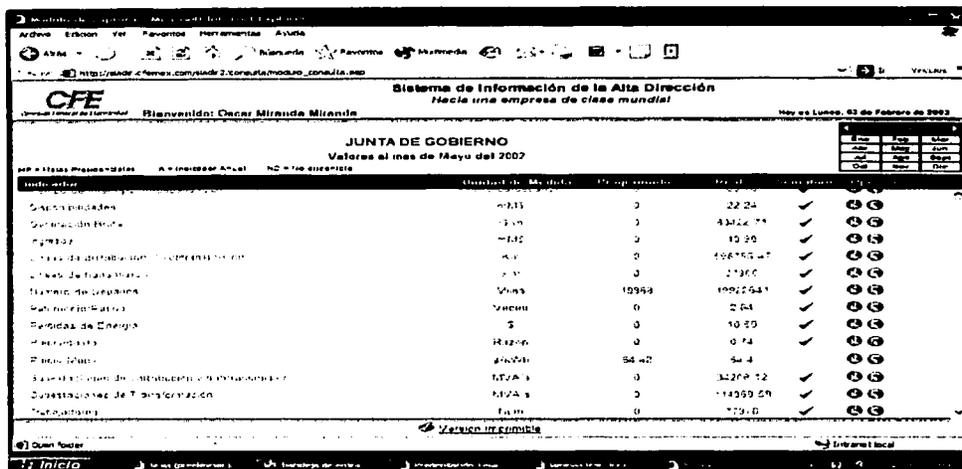


Figura 5.16 Tablero de Control del SIADIR.

El semáforo indica que tanto se acerca el valor real del programado, generalmente un semáforo verde indica que se excedió satisfactoriamente la meta especificada para ese índice en un periodo de tiempo dado; el rojo denota incumplimiento. Es posible además asignar umbrales. La polaridad puede ser positiva o negativa: es positiva cuando se calcula el semáforo evaluando como satisfactorios valores mayores a la meta; es negativa en caso contrario. Por ejemplo, la generación de energía eléctrica es un Índice con polaridad positiva; la accidentalidad es un Índice con polaridad negativa. La siguiente figura 5.17 muestra indicadores del producto Metas Presidenciales vistos a través de una representación gráfica.

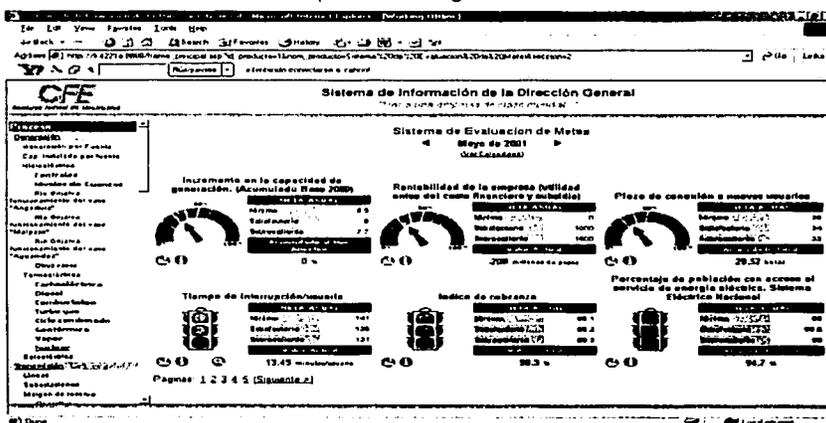


Figura 5.17. Tablero de Control. Representación Gráfica de Indicadores.

Un sistema ejecutivo como el SIADIR debe de presentar información gráfica para que con pocas pantallas el usuario se informe rápidamente de la operación de la empresa. Sin embargo, cuando se encuentra algo que requiera mayor atención es importante que el sistema sea capaz de dar más ayuda en el proceso de toma de decisiones. Para resolver esto, se le incluyó a los tableros de control la posibilidad de presentar cubos de información y gráficas de datos históricos de los indicadores.

La figura 5.18 muestra la funcionalidad OLAP del SIADIR, aun cuando esta funcionalidad es inferior al sistema basado en cubos mostrado anteriormente, proporciona suficientes capacidades de análisis y exploración de información para las áreas directivas.

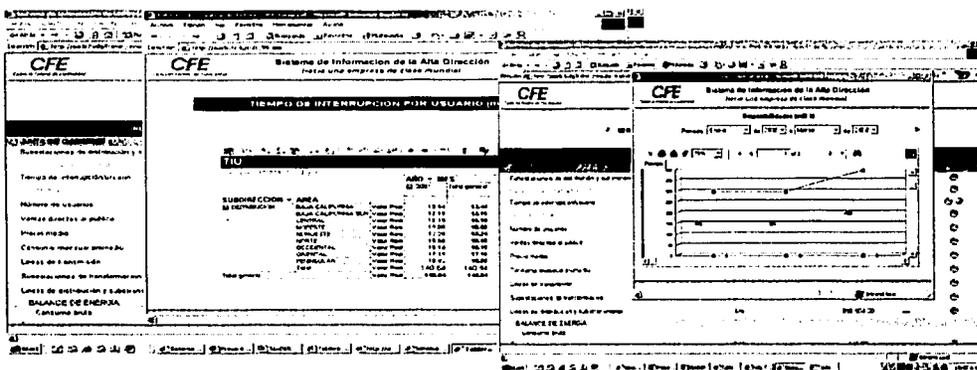


Figura 5.18 Opciones de análisis gráfico en el Tablero de Control del SIADIR

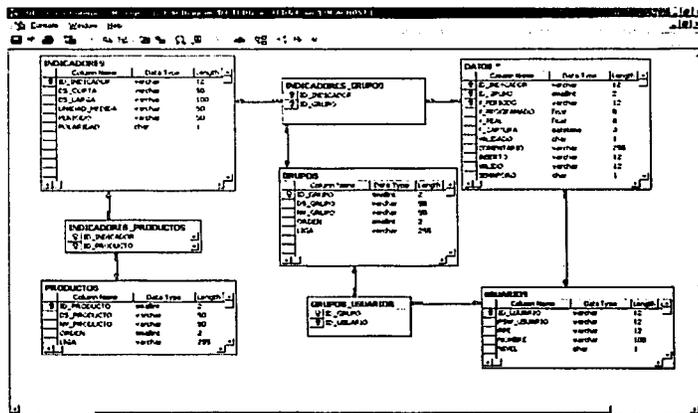


Figura 5.19 Esquema de la Base de Datos del SIADIR.

Al ser el SIADIR un sistema orientado a resolver las necesidades de información del directorado en la CFE, con el propósito de proporcionarles mayores elementos de toma de decisiones se les desarrollo un módulo de información geográfica. Este módulo presenta capas que contienen información de ubicación de líneas, subestaciones y centrales. Ver figura 5.20.

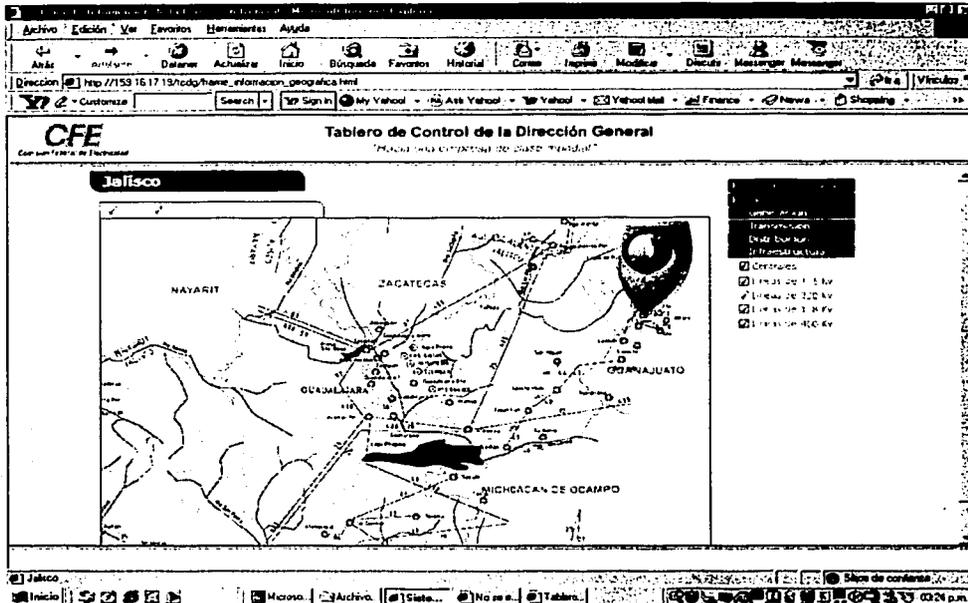


Figura 5.20 Sistema de Información Geográfica del SIADIR

### 5.5.3 Cuadro de Mando Integral.

#### Introducción.

El Cuadro de Mando Integral (CMI) proporciona a los directivos el equipo de instrumentos que necesitan para navegar hacia un éxito competitivo futuro. Hoy en día, las organizaciones están compitiendo en entornos complejos y, por lo tanto, es vital que tengan una exacta comprensión de sus objetivos y de los métodos que han de utilizar para alcanzarlos.

El CMI traduce la estrategia y la misión de una organización en un amplio conjunto de medidas de la actuación, que proporcionan la estructura necesaria para un sistema de gestión y medición estratégica.

El entorno de la era de la información, tanto para las organizaciones fabricantes como para las de servicios, exige tener nuevas capacidades para tener éxito competitivo. La habilidad de una empresa para movilizar y explotar sus activos intangibles o invisibles, se han convertido en algo mucho más decisivo que invertir y gestionar sus activos tangibles y físicos. El CMI proporciona el balance necesario entre las diferentes perspectivas que las empresas deben manejar eficientemente.

#### **La medición de la estrategia empresarial.**

Una de las actividades más importantes en las áreas de dirección debe de ser la medición de la actuación, contrastándola con lo que anteriormente se planeo. El CMI, a través de la vinculación de los objetivos estratégicos con los indicadores de actuación; permite completar una traza de control que va desde el establecimiento de la visión hasta el seguimiento de las actividades estratégicas.

#### **Perspectivas.**

El cuadro de mando integral sigue poniendo énfasis en la consecución de los objetivos financieros, pero también incluye los inductores de actuación de esos objetivos financieros. El cuadro de mando mide la actuación de las organizaciones desde cuatro perspectivas equilibradas: las financieras, los clientes, los procesos internos, y la formación y el crecimiento. El CMI permite que las empresas puedan seguir la pista de los resultados financieros, al mismo tiempo que observan los progresos de la formación de aptitudes y la adquisición de los bienes intangibles que necesitan para un crecimiento futuro.

#### **Balancear las perspectivas.**

Tradicionalmente se ha dado más importancia a la medición y control de los indicadores financieros ya que son los que mantienen informados a los accionistas del valor de sus activos. En esta nueva era, en la que cada vez tiene mayor importancia el capital intelectual. Se ha hecho más evidente la necesidad de fortalecer todas las perspectivas para garantizar el éxito financiero a mediano y largo plazo.

De esta forma para lograr el éxito financiero, se deberá fortalecer la relación con los clientes, lo que se logrará con el desarrollo eficiente de los procesos internos que sólo empleados motivados pueden llevar a cabo.

A continuación se describen brevemente las características de las perspectivas.

#### **La perspectiva financiera**

El Cuadro de Mando Integral puede hacer que los objetivos financieros sean explícitos, y que se adapten a la medida de las unidades de negocio, en las diferentes fases de su crecimiento y ciclo de vida. Aquí se utilizan los objetivos financieros para aumentar e intensificar la rentabilidad, los rendimientos de los activos y los ingresos. Y permite a la alta dirección especificar como se evaluará el éxito de la empresa a largo plazo.

Ejemplo de indicadores:

- Total de activo
- Margen de beneficio
- Valor de mercado
- Flujo de caja
- Costos totales

**La perspectiva del cliente**

La perspectiva del cliente mide la forma en que el cliente nos ve: demoras mínima; una sensación placentera sobre productos agradables o una buena línea de productos. Pero también es la forma de como nosotros lo vemos a él: ¿Qué penetración de mercado tienen nuestros productos en grupos de edades diferentes o en sectores?; ¿Aumenta el número de nuestros clientes?; ¿Somos vulnerables porque vendemos a una cantidad limitada de clientes importantes?. El objetivo de estas perspectivas es buscar la satisfacción total del cliente.

Ejemplo de indicadores:

- Número de clientes
- Cuota de mercado
- Clientes perdidos
- Número de quejas

**La perspectiva del proceso interno**

En la perspectiva del proceso interno, los ejecutivos identifican los procesos críticos internos en los que la organización debe ser excelente. Estos procesos permiten a la unidad de negocio:

- Entregar las propuestas de valor que atraerán y retendrán a los clientes.
- Satisfacer las expectativas de excelentes rendimientos financieros de los accionistas.

Las medidas de los procesos internos se centran en los procesos internos que tendrán el mayor impacto en la satisfacción del cliente y en la consecución de los objetivos financieros de una organización.

Ejemplo de indicadores:

- Productividad
- Calidad
- Nivel de tecnología
- Utilización de la capacidad
- Plazos de entrega
- Colas y tiempos de espera

**La perspectiva de aprendizaje y crecimiento**

La cuarta perspectiva del CMI, la formación o aprendizaje y el crecimiento, identifica la infraestructura que la empresa debe construir para crear una mejora y crecimiento a largo plazo. Las perspectivas del cliente y del proceso interno identifican los factores más críticos para el éxito actual y futuro.

La formación y el crecimiento de una organización proceden de tres fuentes principales: las personas, los sistemas y los procedimientos de la organización. Los objetivos financieros, de clientes y de procesos internos del CMI revelarán grandes procedimientos; al mismo tiempo, mostrarán qué será necesario para alcanzar una actuación que represente un gran adelanto. Algunos indicadores que se pueden ver en esta perspectiva son:

- Motivación y dedicación
- Liderazgo
- Eficiencia en el trabajo

- Orientación

### Vinculación de las perspectivas con la planeación estratégica.

Una vez establecida la planeación estratégica así como los índices por perspectiva, el siguiente paso en el proceso de creación en el CMI es la alineación de los índices con los objetivos estratégicos. En el proceso de alineación se establecen los niveles de contribución de cada índice hacia los objetivos, de esta forma se establecen las bases para los mecanismos de medición que informarán el grado en que la actuación de las diferentes perspectivas de la empresa incide en la consecución de los objetivos estratégicos.

### Creación de un Cuadro de Mando Integral.

En la siguiente tabla se muestran los pasos que se deben de llevar a cabo en los proyectos para la implantación de un CMI en una empresa. El tiempo sugerido de implantación se calculo en tres meses considerando una empresa mediana (menos de 1000 trabajadores), para empresas grandes este tiempo puede ser mayor

#### Los Pasos del Proceso de Creación

Paso	Descripción	Procedimiento	Tiempo sugerido
1	Definir el sector, describir su desarrollo y el papel de la empresa	Entrevistas con el mayor número posible de personas, preferiblemente realizadas por alguien externo a la empresa para obtener una visión objetiva. Investigación sobre la situación y las tendencias del sector.	1-2 meses
2	Establecer/confirmar la visión de la empresa	Seminario conjunto con la asistencia de altos directivos, y líderes de opinión.	1-2 reuniones de un día y medio cada una
3	Establecer las perspectivas	Seminario con asistencia de altos directivos, el grupo encargado del proyecto y alguien con experiencia previa en proyectos de cuadro de mando integral.	1-2 días
4	Desglosar la visión según cada una de las perspectivas y formular metas estratégicas generales	Seminario conjunto con el mismo grupo del segundo paso.	Ver más abajo.
5	Identificar los factores críticos para tener éxito	En el seminario anterior	Total incluyendo el cuarto paso: 2-3 días ya incluido más arriba; si no, 1-2 días
6	Desarrollar indicadores, identificar causas y efectos y establecer un equilibrio.	En el seminario anterior, si es posible. A veces un cierto intervalo beneficioso.	Ya incluido más arriba; si no, 1-2 días
7	Establecer el cuadro de mando al más alto nivel.	Determinación final de la alta dirección y el grupo para el proyecto. Preferentemente, con la participación	1-2 días

8	Desglose del cuadro de mando e indicadores por unidad organizativa	de alguien con experiencia previa en proyectos de cuadro de mando integral. Adecuado para un proyecto dividido en unidades organizativas apropiadas bajo liderazgo del grupo encargado del proyecto. Preferentemente, todo el personal involucrado debería participar en el trabajo que el proyecto adjudica a cada unidad; una forma adecuada de trabajar sería un seminario. Informe sobre avances y coordinación con la alta dirección. La ayuda de un experto constructor de cuadros de mando resulta especialmente importante para alinear los indicadores y los factores de éxito.	Total de 2 hasta X meses. Para cada seminario local, entre medio día y un día como mínimo.
9	Formular metas	Propuestas de los líderes de cada unidad. Aprobación final de metas por la alta dirección.	
10	Desarrollar un plan de acción	Preparación a cargo de cada grupo para el proyecto.	
11	Implementación del cuadro de mando.	Asegurada por control activo bajo la responsabilidad general de la alta dirección.	

**Primer paso: Definir el sector, describir su desarrollo y el papel de la empresa.**

En esta etapa se deberá llegar a un consenso sobre las características y requisitos del sector así como llegar a una definición clara de la posición actual de la empresa y su papel, además vislumbrar hacia el futuro para la elaboración de estrategias. La forma adecuada de hacer este trabajo es mediante entrevistas individuales, principalmente con los máximos directivos y los líderes de opinión más influyentes de la empresa. El siguiente paso es un seminario con el propósito de alcanzar un consenso sobre los aspectos que tendrán importancia en el futuro. Este procedimiento también requiere una combinación de investigación y entrevistas con personas que tienen intereses en la empresa y con gente de los distintos niveles de la empresa.

La imagen global que dan los participantes se presentará de forma reducida. Se trata de una presentación que generalmente servirá como una excelente base de discusión entre ellos y en los pasos siguientes del proceso.

**Segundo paso: Establecer / confirmar la visión de la empresa.**

Es esencial confirmar desde el principio la existencia real de la visión de la empresa, es decir como quiero que sea la empresa.

- **Visión:** Descripción imaginativa y alentadora del papel y objetivos futuros de una organización, que significativamente va más allá de su entorno actual y posición competitiva.
- **Estrategias:** Principios que muestran como se han de alcanzar los objetivos principales de una organización en un periodo concreto. Describe las reglas, los hechos y las decisiones fundamentales que la empresa necesita tomar.

- **Objetivos o metas:** Indican de una manera mas precisa que una declaración de misión se debe alcanzar y el momento en el que se deben obtener los resultados.

Para tener éxito, la empresa debe centrarse en un camino común de desarrollo basado en una visión común, compartida por toda la organización, de la situación interna y externa y de los principales factores de éxito. Con este enfoque, la empresa tendrá la flexibilidad y el dinamismo que necesita para su continua adaptación al cambiante entorno empresarial.

#### **Tercer paso: Establecer las perspectivas.**

Una vez que se han establecido la visión y el concepto del negocio, es hora de considerar las diferentes perspectivas del cuadro de mando. Como ya ha quedado apuntado, hay cuatro perspectivas: **la financiera, la del cliente, la del proceso interno y la de formación y crecimiento**. Sin embargo algunas empresas han preferido añadir otra perspectiva, como la del empleado o humana. La elección de perspectivas debe hacerse principalmente por lógica empresarial, con una clara interrelación entre ellas. La perspectiva de desarrollo, entonces, debe indicar de qué manera propone la dirección desarrollar la organización, así como los productos y servicios ofrecidos con el propósito de simplificar los procesos y/o añadir valor para los clientes. Estos efectos se deben observar a continuación desde la perspectiva financiera. Cualquier cambio de perspectiva tiene que basarse en razones estratégicas más que en algún modelo impuesto por las personas con intereses en la empresa.

#### **Cuarto paso: Desglosar la visión según cada una de las perspectivas y formular metas estratégicas generales.**

El modelo de CMI es principalmente una herramienta para la formulación e implementación de la estrategia de la empresa. El modelo se debe considerar como un instrumento para transformar una visión y estrategia abstractas en indicadores y metas concretas. En otras palabras, un cuadro de mando integral bien formulado es la presentación de la estrategia de la empresa. El propósito de este paso es, entonces, transformar la visión en términos tangibles a partir de las perspectivas establecidas, y del modelo y del método. Este proceso es una parte muy importante del verdadero trabajo de desarrollo empresarial, y otra es formular la estrategia global en términos más generales.

Una manera de poner esta fase del proceso en marcha es pedirle a los participantes que describan las reglas de procedimiento generales que más fácil y eficazmente llevarían a la empresa a la visión buscada. La descripción se debe basar en las distintas perspectivas y, en cierto modo, en cada una de ellas. Un punto de partida adecuado podría ser la descripción de cómo será la empresa en el futuro. A partir de ella se pueden identificar reglas de procedimiento y estrategias, adecuadas en unos cuantos aspectos: rentabilidad a corto y largo plazo; formas de competir de la empresa, por ejemplo en la fijación de precios y en los plazos de entrega; organización de la empresa y por tanto del tipo de formación que se debe cultivar y del que se debe disponer dentro de ella. Otras estrategias a fijar están relacionadas con las áreas en las que la empresa desarrollara sus productos y servicios y con la persona que será responsable del desarrollo. Una vez completada esta fase, el grupo contará con una declaración para cada perspectiva que indicará las estrategias principales, por orden de prioridad, para alcanzar la visión buscada.

#### **Quinto paso: Identificar los factores críticos para obtener éxito.**

Este paso significa dejar atrás las descripciones y estrategias presentadas anteriormente y pasar a discutir y juzgar qué es lo que hace falta para que la visión tenga éxito y cuales son los elementos que afectarán más a los resultados. En otras palabras, la empresa tiene que decidir ahora cuáles son los factores críticos para tener éxito y clasificarlos por orden de prioridad. Una forma adecuada de empezar este proceso sería formar grupos de discusión para determinar, por ejemplo, cuáles

son los cinco factores más importantes para alcanzar las metas estratégicas establecidas previamente. En la sesión posterior, que resume y continúa la discusión, generalmente se alcanza un acuerdo amplio sobre un conjunto de factores reales que son indispensables para tener éxito. La lista de estos será la base para proceder a desarrollar los indicadores clave.

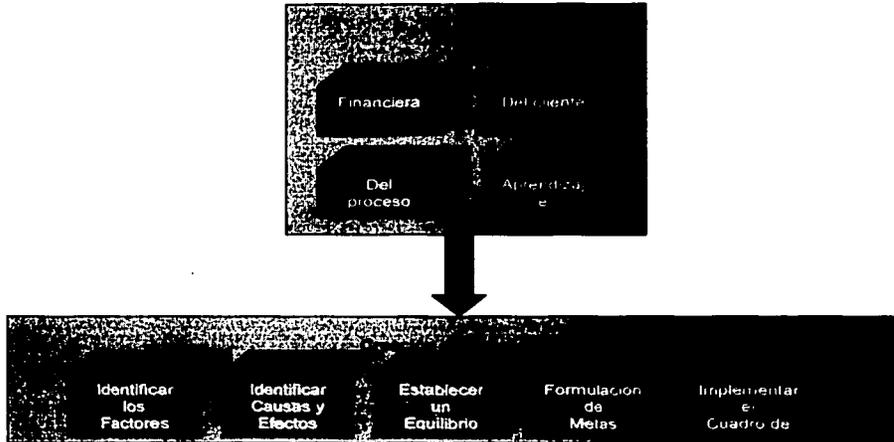


Figura 5.21 Proceso de creación de indicadores

#### **Sexto paso: Desarrollar indicadores, identificar causas y efectos y establecer un equilibrio:**

En esta etapa desarrollamos relevantes indicadores clave para su posterior uso en nuestro trabajo. Al igual que en otros pasos aquí no se rechaza ninguna idea y todos los pensamientos se usan en el proceso. Sólo en la fase final especificamos y ordenamos por orden de prioridad los indicadores que parecen más relevantes, que se pueden supervisar y que, de hecho permiten que se realice la medición.

El gran reto es encontrar claras relaciones causa-efecto y crear un equilibrio entre los diferentes indicadores de las perspectivas seleccionadas. Por tanto, es importante discutir si se puede alcanzar un equilibrio entre los diferentes indicadores de modo que las mejoras a corto plazo no entren en conflicto con las metas a largo plazo. Los indicadores de las diferentes perspectivas no deben llevar a la suboptimización, sino que deben ajustarse a la visión globalizada y servir de apoyo a la estrategia general.

TESIS CON  
FALLA DE ORIGEN



Figura 5.22 Estableciendo Causas.efectos para indicadores

### Séptimo paso: Establecer el cuadro de mando global.

Una vez completados los pasos anteriores, se prepara el cuadro de mando al más alto nivel para su aprobación y presentación a las personas involucradas. A efectos de facilitar su implementación, es importante que todos los miembros de la organización reciban información sobre el trabajo y el proceso de ideas que han sido necesarios para crear el cuadro de mando. También es ventajoso que los participantes reciban documentación suplementaria con textos explicativos, posibles enfoques y sugerencias para el trabajo en grupo que faciliten el proceso continuo de desglosar el cuadro de mando.

### Octavo paso: Desglose del cuadro de mando e indicadores por unidad organizativa.

Dependiendo del tamaño de la empresa y la organización, el cuadro de mando y los indicadores al más alto nivel generalmente se descomponen y aplican a unidades organizativas de niveles inferiores. Dado que uno de los propósitos del cuadro de mando es hacer que los empleados vean claramente de qué forma la visión de la empresa y sus metas generales afectan las operaciones de todos los días, es necesario desglosarlo para que sea tangible y comprensible.

El éxito que cada empresa tiene en un indicador concreto suele verse influido por distintas actividades realizadas por diferentes unidades pertenecientes a varios niveles de la organización. Si se supone que la mayor cantidad posible de empleados tiene que ver cómo su trabajo ayuda a que la empresa alcance buenos resultados con sus indicadores al más alto nivel, éstos tienen que detallarse al máximo.

Se recomienda que se desarrollen guías que faciliten las fases subsiguientes del trabajo. Dichas guías deben cubrir todas las preguntas, desde cómo interpretar el cuadro de mando al más alto

nivel hasta qué se espera de cada individuo en particular. Para evitar malas interpretaciones, una sencilla lista de palabras y conceptos puede ser de gran ayuda.

**Noveno paso: Formular metas.**

Se deben formular metas para cada indicador que se use. Una empresa necesita metas a corto y a largo plazo para poder medir su marcha de forma continua y llevar a cabo las acciones correctivas necesarias a tiempo. Como ya hemos indicado, es esencial que estas metas sean coherentes con la visión global y la estrategia general, y que no entren en conflicto entre sí. Por esta razón, deben alinearse tanto horizontalmente como verticalmente. También es importante crear un proceso para especificar las responsabilidades de fijar metas y para medir resultados.

**Décimo paso: Desarrollar un plan de acción.**

Finalmente, para completar el cuadro de mando, también debemos especificar los pasos que se han de dar para alcanzar las metas y la visión establecidas. Este plan de acción debe incluir a todas aquellas personas responsables y un calendario para los informes provisionales y definitivos. Dado que estos planes tienden a ser masivos y muy ambiciosos, se recomienda que el grupo acuerde una lista de prioridades y un calendario, porque así se evitarán muchas expectativas no manifestadas que más adelante pueden ser fuente de frustración e irritación. En combinación con los indicadores que se han desarrollado, esta lista es uno de los principales documentos de la dirección para ejercer su función de control.

**Undécimo paso: Implementación del cuadro de mando.**

Para mantener el interés en el cuadro de mando hace falta seguirlo continuamente para que cumpla con la función encomendada de herramienta dinámica de gestión. Con este fin, serán necesarias soluciones de Tecnología de Información que faciliten los informes y la recolección de datos.

También es importante que el cuadro de mando se use a través de toda la organización en los aspectos diarios de la gestión. Si de este modo proporciona las bases para la agenda diaria de cada unidad, tendrá una función natural en los informes y el control mediante su impacto en las operaciones de cada día. El plan de implementación, por lo tanto, debe incluir reglas y sugerir formas que aseguren el seguimiento de un cuadro de mando integral que forme parte del trabajo diario de la empresa. Finalmente, los indicadores seleccionados deben poderse cuestionar continuamente, especialmente los de control a corto plazo, y en los casos apropiados, deben reemplazarse por otros más actuales.

**El Enterprise Smart Dashboard (ESD).**

Aplicando la metodología de CMI presentada anteriormente en esta sección, se desarrolló un producto informático que permita el control y seguimiento de la estrategia empresarial a través de la aplicación de las metodologías para desarrollo de almacenes de datos e inteligencia de negocios mostradas a lo largo del proyecto de tesis.

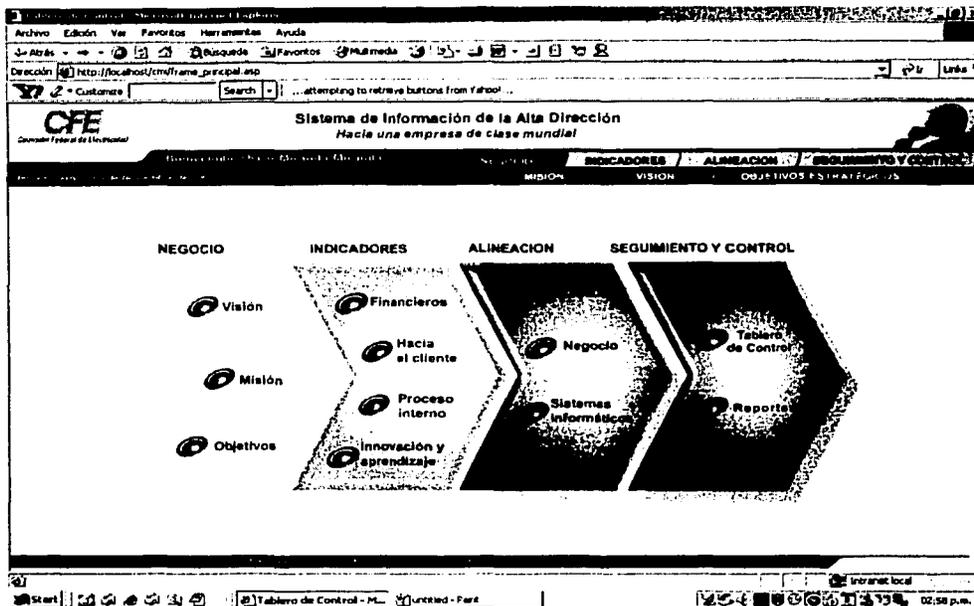


Figura 5.23 Pantalla principal de navegación del ESD.

El Enterprise Smart Dashboard (ESD) es una herramienta digital que permite la implementación de un cuadro de Mando Integral (Balanced Scorecard) en cualquier institución pública o privada. Aún cuando el ESD es una poderosa herramienta; es importante notar, que previamente a su aplicación es necesaria la creación del Cuadro de Mando Integral siguiendo la metodología propuesta por Kaplan y que se resume en los pasos de la tabla anterior.

El ESD de manera fácil y ordenada proporciona los elementos para registrar y explotar la información requerida en el proceso de creación y gestión de la estrategia empresarial, apoyándose para esto de las más modernas herramientas de Inteligencia de Negocios.

El ESD es una eficiente herramienta que sirve como un excelente apoyo en la definición, comunicación y gestión de la estrategia empresarial. Entre sus funciones más importantes se encuentra: El registro y formulación de la misión/visión y objetivos estratégicos; la administración de una base de datos de indicadores estratégicos; la alineación de estos indicadores con el negocio (los objetivos estratégicos) y con las tecnologías de información existentes en la institución; el seguimiento y control de la estrategia a través de un tablero de control y reportes periódicos de contribución.

Para lograr esto, el ESD cuenta con los módulos que se describen a continuación:

- Configuración

- Administración de usuarios
- Negocio
- Indicadores
- Alineación
- Seguimiento y control

La pantalla principal de navegación del ESD se puede ver en la figura 7.23.

### Configuración

En este módulo se encontrarán las opciones que se requieren para adaptar el aspecto visual del ESD a las necesidades de la empresa. Es posible cambiar el juego de colores; especificar el logo y los títulos de las pantallas que se despliegan, así como la presentación gráfica de los indicadores.

### Administración de usuarios

The screenshot shows the 'Módulo de Usuarios' interface. At the top, there is a navigation bar with 'NEGOCIO', 'INDICADORES', and 'ALINEACION'. Below this, a menu includes 'REPORTES', 'TABLERO DE CONTROL', 'Consulta', 'Captura', and 'Administración'. The main content area is titled 'MÓDULO DE USUARIOS' and contains a table of users.

Nombre completo	Nivel de acceso
<input type="checkbox"/> Oscar Miranda Miranda	1 - Administración
<input type="checkbox"/> Manuel Lavalle	2 - Validación
<input type="checkbox"/> Israel Galvan	3 - Carga
<input type="checkbox"/> Bruno Arce	4 - Consulta
<input type="checkbox"/> Celia Gómez Jiménez	4 - Consulta

Below the table, there is a text input field: 'Eliminar usuario(s) seleccionado(s)'. The interface also includes a search bar and a 'Agregar nuevo usuario' link.

Figura 5.24 Administración de Usuarios

Con el propósito de establecer niveles de seguridad en los accesos al ESD, el sistema cuenta con un módulo administrador de usuarios a través del cual es posible limitar el uso de algunos módulos a usuarios específicos.

## Negocio

El módulo de negocio está dirigido al registro de la misión/visión y a los objetivos estratégicos para cada una de las perspectivas. Para la operación del ESD esta información es el insumo más importante. Ver figura 5.25.

Es importante tomar en cuenta que para que el ESD funcione adecuadamente, es necesario que se desarrolle previamente el análisis del negocio a través de la metodología de CMI. De esta forma, si el planteamiento de los objetivos estratégicos es adecuado; la operación del cuadro de mando a través de los siguientes módulos del sistema también será satisfactoria.

Figura 5.25 Visión, Misión y Objetivos Estratégicos.

## Indicadores

Este módulo permite la administración de una base de datos de indicadores clasificados por perspectiva. Adicionalmente a esta clasificación, los indicadores también se organizan por grupos, estos grupos pueden ser, áreas, regiones, unidades de negocio o cualquier otra forma en que una empresa acostumbre organizarse o verse.

TESIS CON  
FALLA DE ORIGEN

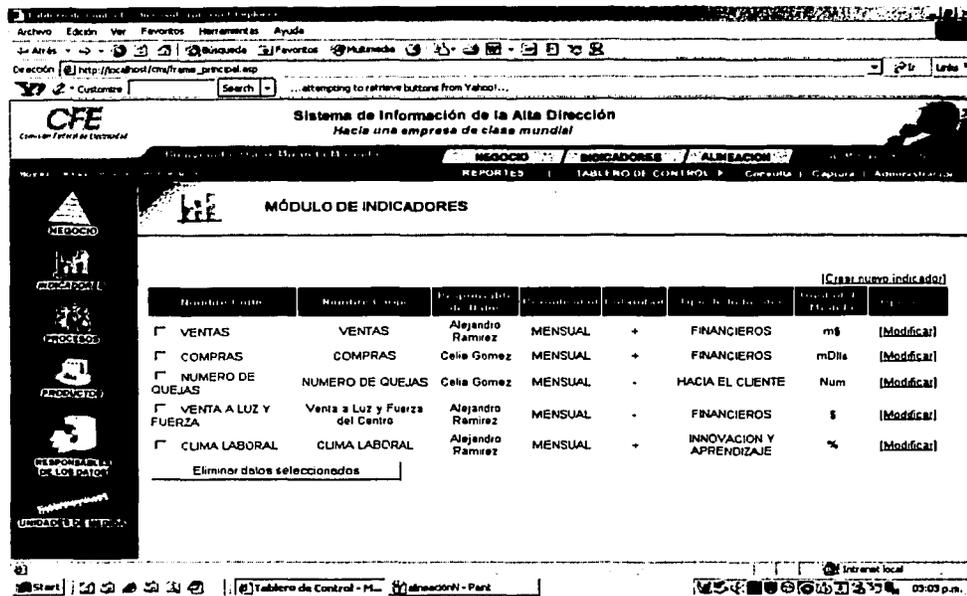


Figura 5.26 Indicadores.

### Alineación

El ESD permite la alineación de los indicadores con los objetivos estratégicos asignándoles niveles de contribución (ver figura X), de esta forma cada objetivo estratégico deberá estar cubierto al 100% con los indicadores que miden la actuación de la empresa que incide en el cumplimiento del objetivo; de igual forma, cada indicador deberá estar alineado a algún objetivo estratégico, de no ser así, es posible que la actividad que este indicador mide, no contribuya a la consecución de la estrategia empresarial.

El ESD también alinea los indicadores con las tecnologías de información, de manera que puede construir la información de cada indicador, conectándose al ambiente de bases de datos de la institución.

En las figuras 5.27 y 5.28 se muestra esquemáticamente el proceso de alineación al negocio y a las tecnologías de información. En las figuras 5.29 y 5.30 se muestra la implementación de estos procesos en el ESD.

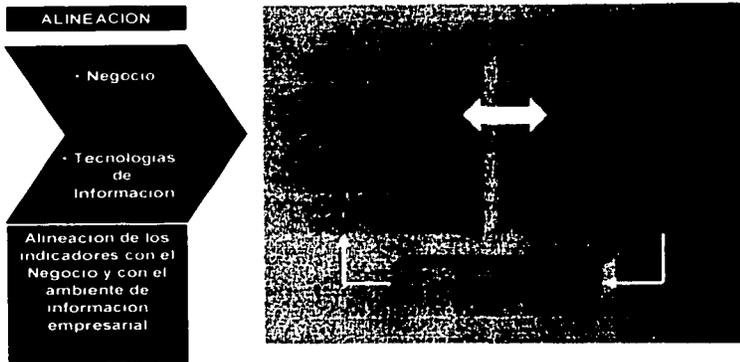


Figura 5.27 Alineación al negocio.

## Enterprise Smart Dashboard

Etapa 3

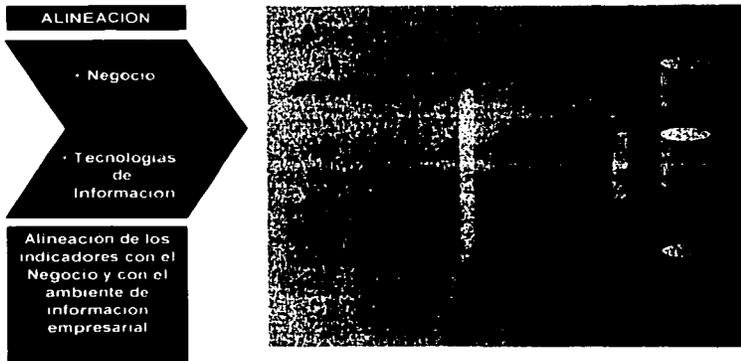


Figura 5.28 Alineación a las tecnologías de información.

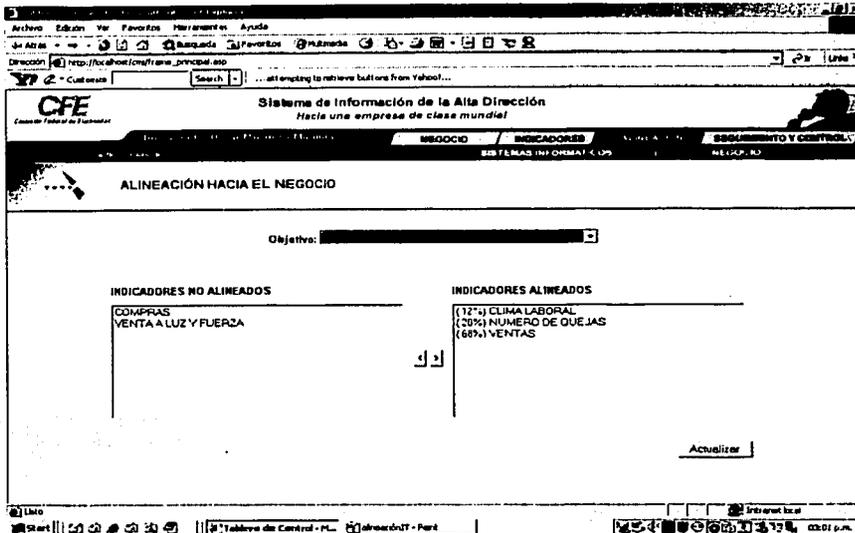


Figura 5.29 Alineación al negocio.

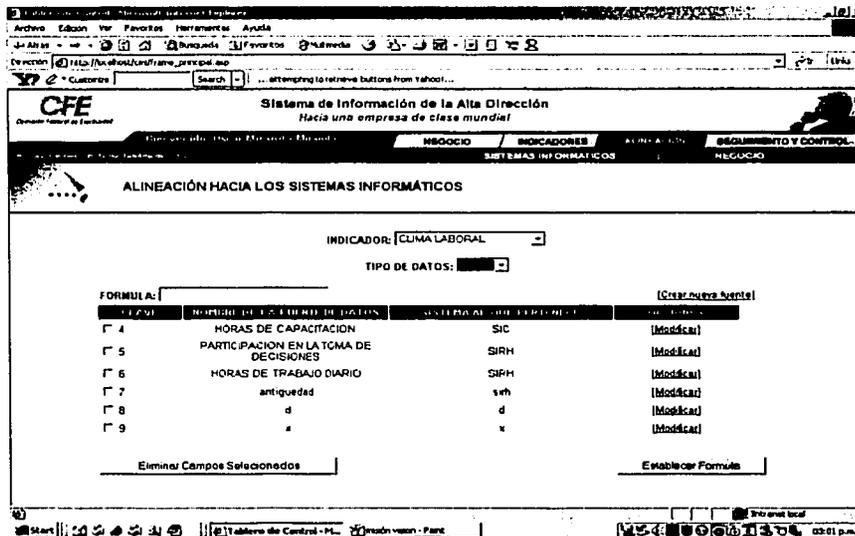


Figura 5.30 Alineación a los Sistemas Informáticos.

### Seguimiento y control

El módulo de seguimiento y control del ESD considera dos aspectos fundamentales para la medición de la estrategia empresarial:

**El Tablero de Control:** La presentación gráfica de los valores reales y programados de los indicadores por periodo.

**Matriz de contribución por periodo:** Un informe de la contribución (en porcentaje) de la actuación en un periodo determinado a la estrategia empresarial.

		objetivos			
		Ob 1	Ob 2	Ob n.	Semaforo
Indices.	<b>Financieros</b>				
	Ahorros que influyen en las compras .	20%	10%	10%	●
	Penas por morosidad en los pagos.	10%	15%	10%	●
	<b>Cliente</b>				
	La entrega puntual como la define el cliente.	13%	25%	23%	●
	La calidad del prod como la define el cliente	12%	15%	10%	●
	<b>Proceso interno</b>				
	Recopilación de datos fiable y oportuna	10%	8%	8%	●
	Evaluación de sistemas internos de calidad	28%	7%	10%	●
	<b>Innovación y mejora.</b>				
	La calidad del entorno como la definen los empleados.				No alineado
	<b>total</b>	<b>93%</b>	<b>80%</b>	<b>85%</b>	<b>91.66%</b>

Figura 5.31 Matriz de Contribución en un periodo.

Este módulo proporciona las herramientas necesarias para medir la actuación de la organización. Esta medición se puede hacer hacia procesos muy concretos a través de comparar indicadores con sus valores programados o meta, o en lo general al comparar la actuación de un periodo (todos los indicadores) con la estrategia empresarial.

#### 5.5.4 El Portal de la CFE como cliente del Almacén de datos.

La CFE, como una empresa de clase mundial, cuenta con un portal como medio de comunicación en la internet. El portal de la CFE cubre los siguientes objetivos:

1. *Difusión de la imagen corporativa.* Para lograr este objetivo, el portal explica que es la CFE, su historia y objetivos, así como los aspectos más importantes en su operación y el impacto de ésta en el desarrollo del país. Al fomentar el conocimiento de la empresa y la contribución de ésta al país se busca elevar la imagen que la población en general tiene de la CFE.

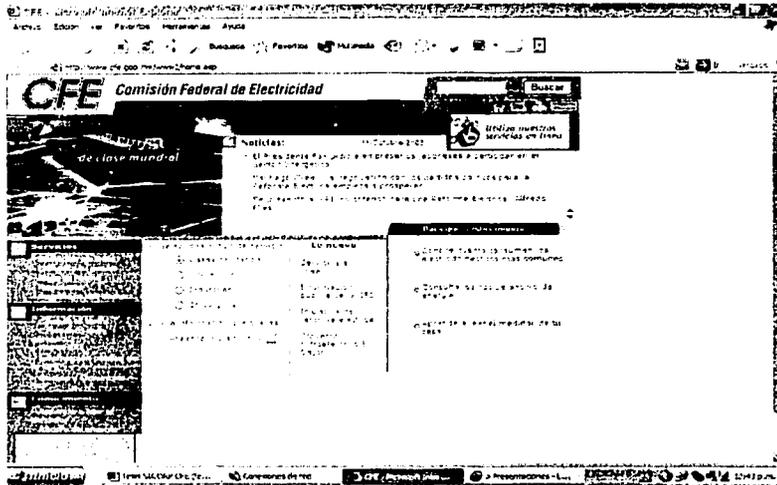


Figura 5.32 Portal corporativo de la CFE

En cumplimiento con la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental, recientemente se incluyó al portal una sección con las 17 fracciones del artículo 7 de dicha ley.

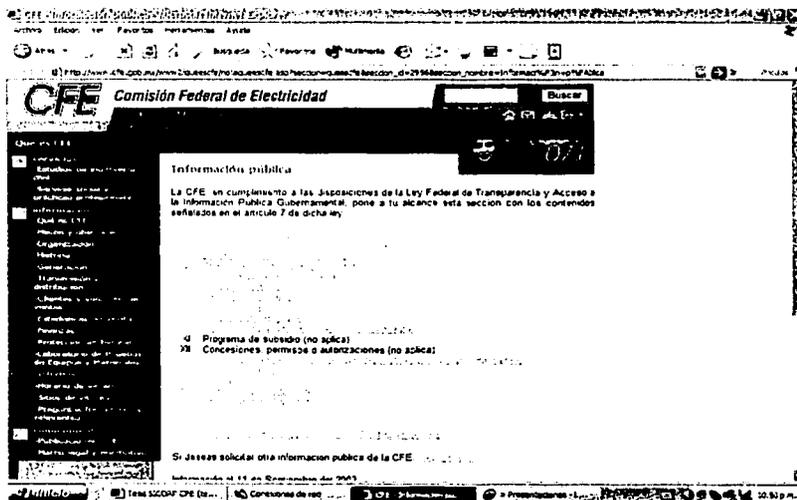


Figura 5.33 Información pública de la CFE.

2. *Difusión y promoción de la venta de diversos bienes y servicios.* El portal también se constituye como una ventanilla a través de la cuál diversas áreas de la CFE promueven la venta de bienes y servicios, como es el caso del laboratorio de pruebas, estudios de ingeniería civil e información de licitaciones y contratos. Esta sección está dirigida principalmente a clientes y proveedores.
3. **Servicios Electrónicos a clientes de la CFE.** Aprovechando las ventajas que la internet representa en el desarrollo de negocios en línea y con el propósito de abatir costos de operación, manteniendo estándares de atención de clase mundial; la CFE ofrece una serie de servicios en línea a sus clientes.

Los servicios electrónicos permiten que los clientes de la CFE en la comodidad de sus hogares o centros de trabajo puedan solicitar, desde la contratación de la luz, la consulta de sus consumos, hasta efectuar pagos en línea.

Esta tecnología además de facilitar la interacción de nuestros clientes con la empresa, fomentando el mejoramiento de la imagen, permite una significativa reducción de costos de operación en la atención tradicional en las agencias de la CFE.

Como ejemplo de las ventajas que los Servicios Electrónicos traen a la institución considérese el caso de la consulta de recibo que en promedio atiende 60,000 peticiones mensualmente.

A continuación se relacionan los servicios electrónicos que actualmente se proporcionan:

- Contáctanos
- Consulta de recibo
- Pago de recibo
- Contratación de luz
- Aviso de fallas en la luz
- Aclaración de recibo
- Solicitud de libranza
- Cambios en instalaciones en la CFE
- Revisión de medidor
- Consulta del estado de una solicitud

Para que la CFE este en posibilidades de proporcionar eficientemente esta serie de servicios a través del portal, ha sido necesario el fortalecimiento de su infraestructura tecnológica en materia de informática y telecomunicaciones. Este desarrollo tecnológico ha sido paulatino y en el han participado todas las áreas de la empresa.

Una de las piezas claves en el desarrollo tecnológico de la CFE que han permitido que se cumplan los objetivos que se plantearon para el portal ha sido el almacén de datos corporativo. El portal es uno de los clientes más importantes de la información contenida en el almacén de datos.

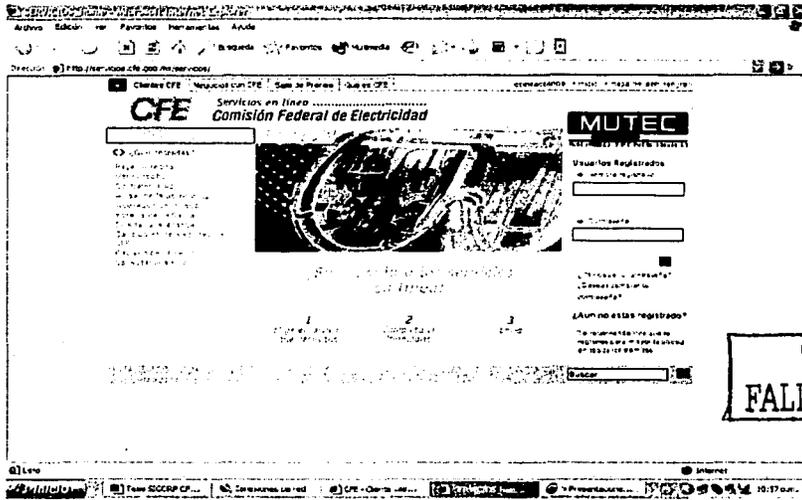


Figura 5.34 Servicios en línea.

El canal “Qué es CFE” es un buen ejemplo del tipo de información que el portal lee del almacén de datos corporativo. En la figura 5.35 se presenta un ejemplo de una consulta a la estadística de venta de energía.

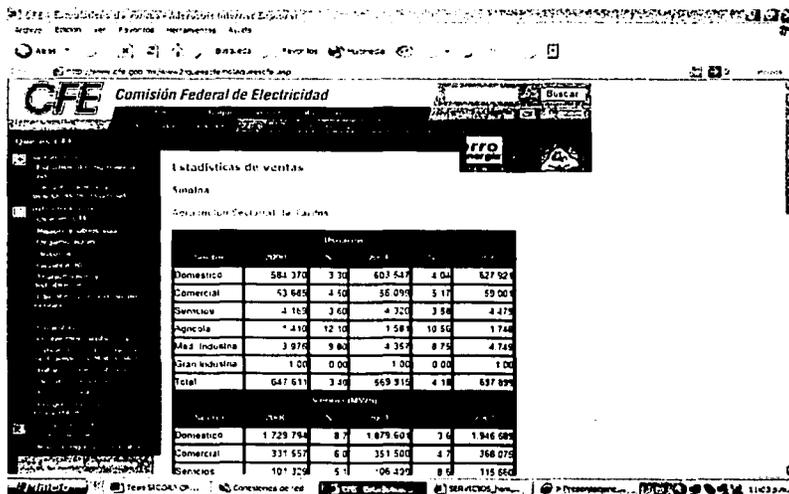


Figura 5.35 Estadística de ventas proveniente del almacén de datos vista desde el portal de la CFE.

## 6 Sensibilidad al Cambio.

En todo proyecto en el que se pretenda hacer algún cambio en la forma de trabajo se deberá tener mucho cuidado de manejar adecuadamente la reacción que estos cambios puede causar en las personas. Grandes proyectos, que pudieron haber dado muchos beneficios a algunas organizaciones han fracasado por un pésimo manejo del cambio.

El cambio es algo que siempre nos acompaña, desde que nacemos hasta que morimos experimentamos una larga serie de cambios, esto podría hacernos pensar que estamos acostumbrados a los cambios. Sin embargo, aun cuando el cambio esta presente en cada parte de nuestras vida, es alguna de las cosas que más tememos los seres humanos. En realidad las personas nos afianzamos a nuestros hábitos, los hábitos y la rutina nos dan seguridad. Esto no quiere decir que pensemos que los cambios sean malos, pero no queremos arriesgarnos a averiguarlo.

Esta condición natural en la naturaleza humana hace que proponer mejores prácticas en las organizaciones siempre encuentre resistencia de un gran número de participantes debido al miedo que se tiene al cambio.

Adicionalmente a las reacciones personales respecto al cambio, la organización como entidad también presenta resistencia al cambio; las estructuras organizacionales jerárquicas; los sindicatos; las políticas y normas vigentes, todos estos son factores que también pueden retrasar o impedir el establecimiento de los proyectos.

Por lo anterior, es muy importante tomar en cuenta esto antes de iniciar cualquier proyecto que afecte la forma de trabajar de las personas.

Estrategia para la administración del cambio en un proyecto de almacenes de Datos:

- A. Lo primero que se debe de garantizar es que todo el personal de la organización que tendrá alguna relación con el proyecto lo conozca y se sensibilice de sus bondades. Para lograr esto el proyecto deberá pertenecer a toda la organización y se deberá involucrar en la toma de decisiones, administración y control a cada área estratégica, empezando con la dirección general. Por lo que se propone la siguiente organización:

**Patrocinador**

Es muy importante contar con un patrocinador para el proyecto. El patrocinador deberá ser el director de la organización o alguna persona cercana a la dirección.

**Comité de información**

Se deberá establecer un comité de información con un representante (con autoridad) de cada área estratégica. Este comité tendrá bajo su responsabilidad el establecimiento de las políticas y procedimientos para la conformación del proyecto, el seguimiento y supervisión de los planes de trabajo y la asignación de los recursos necesarios.

**Comité técnico**

El comité técnico estará conformado por un especialista en informática de cada área estratégica y tendrá bajo su responsabilidad el establecimiento de las políticas en materia de informática y telecomunicaciones; proporcionar los elementos informáticos requeridos por cada área y participar activamente en la difusión del sistema en sus respectivas áreas.

**Grupo técnico**

Se deberá conformar un grupo de trabajo de especialistas en diferentes áreas de la informática para desarrollar el proyecto. El grupo se deberá organizar de la siguiente

manera: líder del proyecto; responsable del acopio de información; responsable de la base de datos y responsable de las aplicaciones de análisis.

- B. La segunda actividad importante en el manejo del cambio es la promoción y difusión del proyecto. Se deberán hacer uso de todos los medios disponibles para realizar estas actividades, esto dependerá de los recursos disponibles y se definirá en el comité de información. Entre las cosas que se puede hacer para promover el proyecto es publicar en la Intranet información general del proyecto, avances y logros; mantener un foro de discusión; publicar panfletos; etc.

En el desarrollo del proyecto seguramente se presentarán problemas, estos problemas pueden ser de diferentes niveles, desde la resistencia de participantes de niveles bajos en la organización hasta poca participación o incluso bloqueos graves de gerentes o altos directivos. Por lo anterior es muy importante establecer los mecanismos para dar solución a cada problema, en esto es muy importante establecer las normas de comunicación y ejecución de acciones de autoridad. Todo esto deberá establecerse por escrito, deberá ser autorizado por el comité de información y comunicado a todos los participantes del proyecto.

## **7 Aportaciones a la Ingeniería en Computación.**

### **7.1 Introducción**

En el presente trabajo de tesis se han aplicado algunos de los conocimientos que se adquieren en los programas actuales de la carrera de ingeniería en computación. Esta formación ha sido muy importante para el desarrollo del proyecto, particularmente lo relacionado a técnicas y metodologías de programación, bases de datos, sistemas operativos y redes de computadoras.

Adicionalmente a los conocimientos que la carrera aportó, para el desarrollo del proyecto, fue necesaria la aplicación de algunas otras disciplinas del área de la tecnología de información (TI). Estas áreas de estudio se pueden agrupar e identificar con un término que actualmente ya es común: Inteligencia de Negocios (Business Intelligence).

La inteligencia de negocios trata de la aplicación de la tecnología de información como apoyo a los procesos de toma de decisiones. En este sentido, la inteligencia de negocios proporciona todas las herramientas computacionales que los altos directivos de las organizaciones requieren para hacer de la información un activo que los ponga en una posición ventajosa con sus competidores.

La velocidad con la que las empresas deben reaccionar en este mundo globalizado, exige que las tecnologías de información aplicadas, proporcionen velozmente información adecuada para que se logre un eficiente proceso de toma de decisiones. Esto implica que los medios de acopio, almacenamiento y entrega de la información se orquesten de manera que fluya con oportunidad la información necesaria para la toma de decisiones.

Para lograr lo anterior, existen una serie de disciplinas de estudio que van desde las tradicionales bases de datos relacionales hasta la aplicación de técnicas de inteligencia artificial. Aún cuando diversas carreras relacionadas con las tecnologías de información encuentran un campo fértil de aplicación en la inteligencia de negocios; la Ingeniería en Computación, por su carácter de ingeniería, por su capacidad de aplicar modelos matemáticos a la solución de problemas reales, debe de aplicarse fuertemente a la solución de problemas relacionados con la inteligencia de negocios, particularmente donde el diseño y las técnicas de inteligencia artificial son requeridas.

Por lo anterior, como parte de las aportaciones del presente trabajo de tesis, se propone una nueva materia para la carrera; en ésta se sintetizarán los temas que se consideran más convenientes para preparar al alumno en el desarrollo de sistemas de información como soporte a los procesos de toma de decisiones. El nombre de la materia puede ser: Inteligencia de Negocios.

Debido a que la aplicación de la inteligencia de negocios es extensa; y por la falta de tiempo, en el temario propuesto no se entra a demasiado detalle en el estudio de algunos temas. Se exponen los temas de forma que el alumno tenga una idea general de su aplicación y se despierta su interés para que posteriormente puede profundizar conforme a las necesidades de un proyecto específico.

### **7.2 Estructura del temario**

Para el desarrollo del temario se dividió la asignatura: inteligencia de negocios; en los siguientes capítulos: Comunicación entre procesos; bases de datos multidimensionales; cuadro de mando integral y minería de datos. La estructura propuesta se muestra en la figura 7.1.

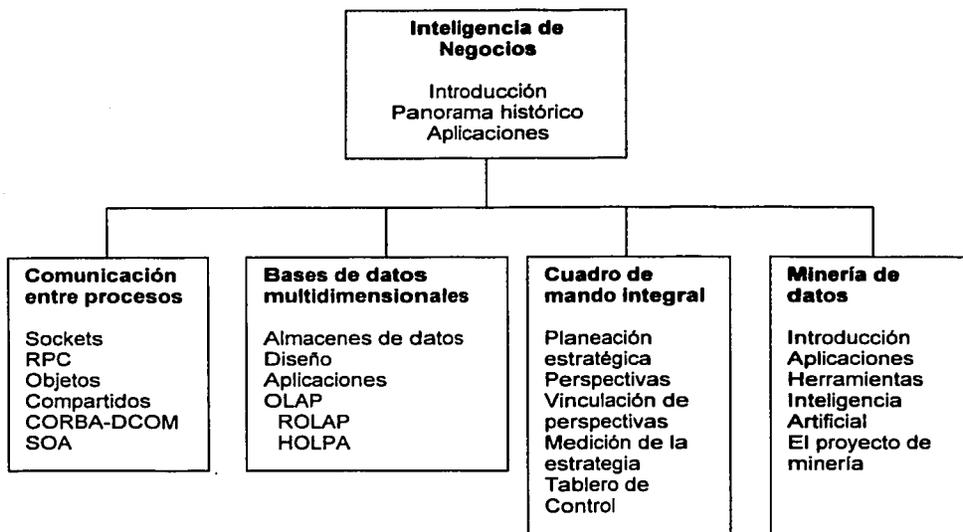


Figura 7.1 Estructura del temario

### 7.3 Asignación de tiempos para el temario

En esta sección se define el tiempo que se considera necesario para impartir cada uno de los temas propuestos para la asignatura.

- Para la asignación de tiempos se consideraron los siguientes criterios:
- Se tomó en cuenta el grado de complejidad de cada tema.
- Se incluyeron temas que no se desarrollaron en el presente trabajo de tesis por considerarlos necesarios para la comprensión total de la asignatura.
- Dado que se propone que la asignatura se imparta en el décimo semestre de la carrera se tomó en cuenta que algunos temas se debieron tratar en materias predecesoras.
- Con el propósito de mantener la secuencia y evitar en lo posible los cortes en la exposición de cada tema, se ajustaron los tiempos de los temas tomando en cuenta el tiempo que dura cada clase en la Facultad de Ingeniería.

El método que se siguió para la asignación del número de horas fue el siguiente:

- Se asignó el tiempo necesario para la asignación de los temas conforme a los criterios anteriores.
- Se sumaron el número de horas totales para el desarrollo del curso

- Para el tiempo total de horas se tomo en cuenta una materia de 8 créditos en la Facultad de Ingeniería.
- Por último se hizo una corrección de los tiempos propuestos para sintetizar un temario que se ajuste a las características establecidas por la Facultad de Ingeniería.

A continuación se presenta una matriz que detalla el número de horas propuesto inicialmente y la suma total de estas.

<b>Capítulo 1. Introducción a la Inteligencia de Negocios.</b>	<b>Tiempo 1.5 h.</b>	<b>Observaciones</b>
1.1. Introducción.	30 min.	En este capítulo se presenta una introducción y conceptos básicos de la inteligencia de negocios. Se considera hora y media suficiente para crear interés en el alumno.
1.2 Panorama histórico.	30 min.	
1.3 Aplicaciones.	30 min.	
<b>Capítulo 2. Comunicación entre procesos.</b>	<b>Tiempo 12.5 h.</b>	<b>Observaciones</b>
2.1 Sockets	2.0 h.	Se consideran 12.5 horas para tratar todos los temas relacionados con la comunicación entre procesos. Desde los sockets hasta la actuales plataformas de integración basadas en web services.
2.2 RPC	1.5 h	
2.3 CORBA	2.0 h	
2.4 COM	2.0 h	
2.5 Web services	2.0 h	
2.6 Modelo SOA	1.5 h	
2.7 Plataformas de integración	1.5 h	
<b>Capítulo 3. Bases de datos multidimensionales.</b>	<b>Tiempo 19.0 h.</b>	<b>Observaciones</b>
3.1 Almacenes de datos	30 min.	Este es el capítulo más importante de la asignatura, en el se introduce al concepto de bases de datos multidimensionales y sistemas de almacenes de datos. Se consideraron 19 horas para cubrir todos los temas del capítulo.
3.1.1 Arquitectura general de los almacenes de datos	30 min.	
3.1.2 Herramientas ETL	1.5 h.	
3.1.3 Modelo de la base de datos	1.5 h.	
3.1.4 Herramientas de acceso a datos	1.5 h.	
3.2 Diseño del almacén de datos	1.0 h	
3.2.1 Tabla de hechos	15 min.	
3.2.2 Granularidad	10 min.	
3.2.3 Dimensiones	15 min.	
3.2.4 Dimensión padre-hijo	10 min.	
3.2.5 Dimensiones de tiempo	10 min.	
3.2.6 Esquema estrella	1.0 h.	
3.2.7 Esquema copo de nieve	1.0 h.	
3.2 Transformación del modelo estrella al modelo cubo	2.0 h.	
3.3 Ejemplo	2.5 h.	
3.4 OLAP	2.5 h	
3.5 Sistemas EIS	1.5 h.	
3.6 Portales.	1.0 h	
<b>Capítulo 4. Cuadro de mando integral.</b>	<b>Tiempo 14.0 h.</b>	<b>Observaciones</b>
4.1 Planeación estratégica	2.0 h.	Aún cuando en el capítulo anterior se trataron temas relacionados con sistemas de acceso a los datos del almacén; por su relevancia y
4.2 Perspectivas de la empresa.	1.0 h.	
4.3 Vinculación de perspectivas.	30 min.	
4.4 Medición de la estrategia.	30 min.	

4.5 Tablero de Control. 4.6 Proyecto de CMI. 4.6.1 Definir el Sector. 4.6.2 Establecer la visión. 4.6.3 Establecer las perspectivas. 4.6.4 Desglosar la visión en perspectivas. 4.6.5 Identificar factores críticos 4.6.6. Desarrollar indicadores 4.6.7 Establecer el cuadro de mando al más alto nivel. 4.6.8 Desglose del cuadro de mando. 4.6.9 Formulación de metas. 4.6.10 Desarrollo del plan de acción. 4.6.11 Implementación del cuadro de mando.	2.0 h. 1.0 h. 1.0 h. 30 min. 20 min. 20 min. 30 min. 1.0 h. 30 min. 1.0 h. 30 min. 20 min. 1.0 h.	actualidad, se presenta un capítulo con el desarrollo completo de un cuadro de mando integral. Se consideraron 14 horas para el desarrollo de este capítulo.
<b>Capítulo 5. Minería de datos</b>	<b>Tiempo 17.0 h.</b>	<b>Observaciones</b>
5.1. Introducción 5.2 Aplicaciones 5.3 El procesamiento informático/analítico y la minería de datos. 5.4 Herramientas. 5.5 La Inteligencia Artificial en la minería de datos 5.5.1 Redes neuronales 5.5.2. Mapas característicos de kohonen 5.5.3 Reconocimiento de patrones 5.5.4 K-nearest neighbor 5.5.5 Algoritmos genéticos 5.6 El proyecto de minería	1.0 h. 30 min. 1.0 h. 1.0 h. 1.0 h. 2.5 h. 2.0 h. 2.5 h 2.0 h. 2.5 h 1.0 h.	Por su relevancia en la ingeniería de software, se consideraron 17 horas para tratar temas relacionados con la minería de datos y la inteligencia artificial.

**Tiempo total****64.0 horas.**

Resumiendo: Se propone una nueva materia optativa para la carrera de Ingeniería en Computación.

Nombre de la asignatura: Inteligencia de Negocios.

Créditos: 8

Duración del curso: 16 semanas, 64 horas.

Semestre: 10°

## 8 Conclusiones.

El desarrollo de los almacenes de datos y de sistemas de inteligencia de negocios para aprovechar estas grandes bases de datos, hoy en día es una imperiosa necesidad para las organizaciones medianas y grandes. A través de estos sistemas de información es como los directivos pueden tener información oportuna para aprovechar en los procesos de toma de decisiones. Es bien sabido que en la medida que una organización reacciona, es capaz de enfrentar la creciente competencia y dinamismo de los mercados actuales. La capacidad de reacción está determinada por la capacidad de contar con información oportuna y confiable.

El desarrollo de los sistemas de información nunca ha sido una tarea fácil. Sin embargo, cuando se aplican las metodologías y tecnologías adecuadas, ésta tarea puede hacerse de forma ordenada y arrojar muy buenos resultados. Este es el caso de la aplicación de la inteligencia de negocios en las organizaciones particularmente grades como es la CFE.

Adicionalmente a los aspectos tecnológicos y metodológicos inherentes al desarrollo de cualquier sistema de información, en la construcción de un almacén de datos para una empresa grande como la CFE es muy importante considerar también un problema fundamental. El crecimiento de los sistemas de información de los que se abastecerá el almacén de datos se ha dado de forma muy desorganizada. Es posible encontrar sistemas heredados hechos con tecnologías obsoletas y que difícilmente se pueden integrar a los modernos sistemas de red y por ende, la comunicación con ellos es limitada. Otra situación problemática que se deberá enfrentar es la diversidad de plataformas en las que operan los sistemas. Por ejemplo, encontramos sistemas que corren en ambientes Windows (NT, 2000 o 2003) desarrollados con diversos modelos de objetos (ASP o ASPX) que deberán convivir con diversas plataformas UNIX (Solaris, AIX, etc.) desarrollados también con una gran variedad de lenguajes de programación.

Tomando en cuenta las consideraciones anteriores, en la CFE la construcción de un sistema de almacén de datos con todas las aplicaciones de inteligencia de negocios que se presentaron en este trabajo de tesis, fue una importante tarea en la que participaron muchas personas. Recordemos que para desarrollar esta tesis se decidió considerar para algunas secciones el ejemplo de la Subdirección de Distribución y éste no se desarrollo completamente para cuidar el espacio y claridad del material que se presenta.

Son muchos los beneficios que un proyecto de este tipo proporcionan a las organizaciones. En la CFE, además de proporcionar diferentes herramientas como apoyo a los procesos de análisis, control de gestión y toma de decisiones; éste proyecto impulsó el desarrollo de la informática en las diferentes áreas de la empresa. Aún cuando las etapas de extracción, transporte y carga de información deben alimentar a la base de datos corporativa tomando la información de los sistemas fuente tal y como están desarrollados, al hacer el análisis de los sistemas y desarrollar el software para conectarse a ellos en muchos de los casos se aprovechó para migrar a plataformas modernas o en el peor de los casos se encontraron problemas de diseño que también fueron corregidos.

Hoy en día, cuando las empresas aspiran a ser clasificadas como *empresas de clase mundial*, es de gran importancia el valor que la información tiene en las áreas directivas, ya no es posible tomar decisiones sin este valioso activo intangible.

El presente trabajo de tesis ilustra el desarrollo del proyecto de software más importante de la Comisión Federal de Electricidad en los últimos años. El Sistema de Información Corporativa se ha convertido en "El Sistema de Información de la CFE" y se encuentra presente desde los equipos de cómputo de las agencias donde se paga el recibo de energía o en las plantas generadoras; pasando por las computadoras de los analistas, hasta en la computadora portátil que utiliza el Director General cuando se encuentra de gira.

El desarrollo de un proyecto de almacenes de datos que cumpla con el objetivo de informar a las áreas directivas con la operación de la empresa a través de herramientas informáticas es sin duda un reto en el campo de la ingeniería de software. Se involucran varias disciplinas de la ingeniería en computación que van desde las de infraestructura como es el caso de las redes o sistemas operativos hasta las de sistemas de soporte a la toma de decisiones basados en software de inteligencia de negocios. Sin embargo, los factores críticos de éxito en este tipo de proyectos no son necesariamente de origen técnico.

En la mayoría de los casos un proyecto de este tipo fracasa por situaciones de administrativas de planeación y administración del proyecto. El mayor de los factores de fracaso es la falta de patrocinador. Por su cobertura, el patrocinador para este proyecto deberá ser el Director General de la empresa, de no ser así es muy probable que no se llegue a buen término y la mejor decisión debería ser posponer el inicio del proyecto hasta tener al patrocinador adecuado.

Como se pudo ver en el desarrollo del presente trabajo de tesis, la aplicación de la tecnología por parte de los ingenieros, se realiza también para satisfacer necesidades del tipo empresarial y de negocios. Por lo anterior es importante mantener una adecuada sensibilidad de las necesidades de nuestros usuarios, y para esto en ocasiones un ingeniero deberá involucrarse en disciplinas administrativas y de negocios a efecto de desarrollar proyectos que aporten gran valor a las instituciones.

En la actualidad, la Ingeniería en Computación tiene una amplia aplicación en todos los temas relacionados con la Inteligencia de Negocios. El Ingeniero, por su formación orientada al desarrollo de modelos para dar solución a diversos problemas y su capacidad de abstracción y análisis, es el profesional adecuado para el diseño de grandes bases de datos, mecanismos de acoplamiento entre sistemas y desarrollo de sistemas donde se apliquen modelos matemáticos para maximizar las ventajas competitivas de las organizaciones. Ejemplo de esto lo encontramos en la minería de datos y en el análisis multidimensional de grandes bases de datos.

## 9 Referencias Bibliográficas

1. Inmon H., W . "What is a Data Warehouse". Ed: Prism Solution, Inc., 1:1 (1995).
2. Kenan Systems Corporation . "An introduction to multidimensional database technology".(1993-1995), pp. 1-29.
3. Hackathorn, Richard."Data Warehousing: Energizes your Enterprise". Ed: Datamation , (February 1, 1995).
4. Raden, Neil. "Modeling the Data Warehouse". Ed: Archer Decision Sciences, Inc , (January 29, 1996).
5. McElreath, Jack."Data Warehouses: An Architectural Perspective". Ed: Perspectives , (October 1995).
6. Hufford., Duane. "Data Warehouse quality". Ed: Data Management Review . (February-March 1996).
7. Kimball, Ralph."Dealing with Dirty Data". Ed: DBMS Magazine (September 1996).
9. Robert S. Kaplan, David P. Norton: "Cuadro de Mando Integral". Ed. HBS Press (1996).
10. Edward - Yourdon "Análisis Estructurado Moderno". Ed: Prentice Hall Hispanoamericana, S. A., pp. 136-147, 500-511. (1993).
11. Robert Orfali. "Client/Server Programming with JAVA and CORBA". Ed: WILEY (1997)
12. W. Richard Stevens, "UNIX. Network Programming", Ed. Prentice-Hall, Volume 1, 2nd Edition
13. Jlawei Han, Micheljne Kamber "Data Mining: Concepts and Techniques" (2000)